Efficient Continual Learning in Vision

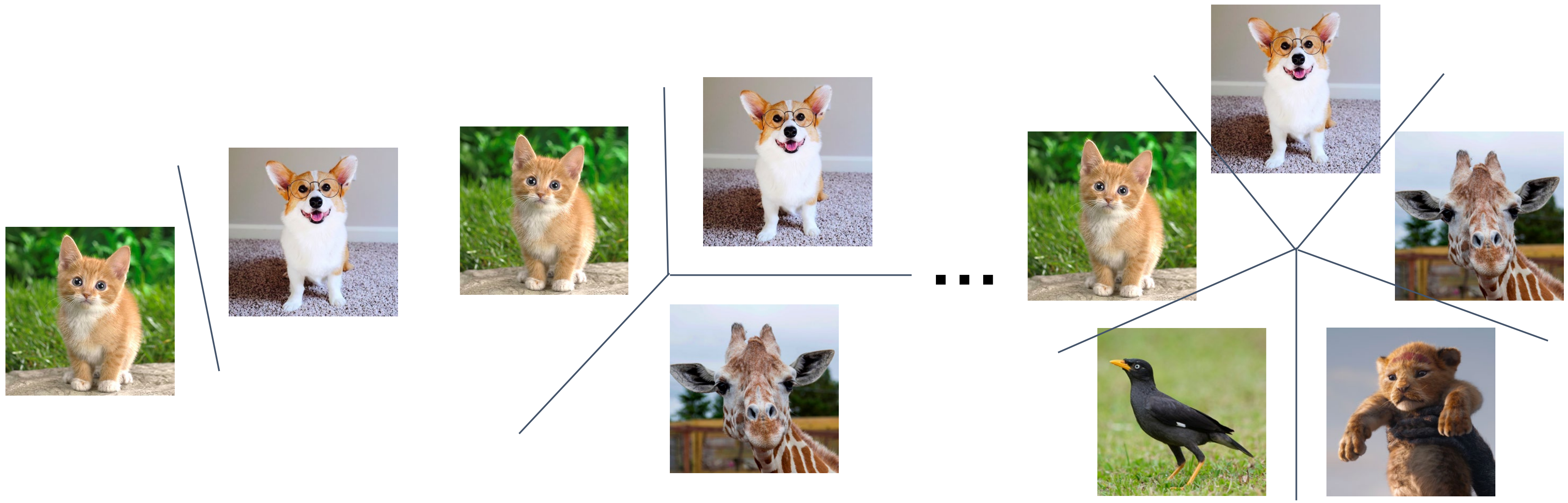
Jay Z. Wu

Show Lab, National U. of Singapore

https://zhangjiewu.github.io

Continual Learning (CL)

A common class-incremental setting

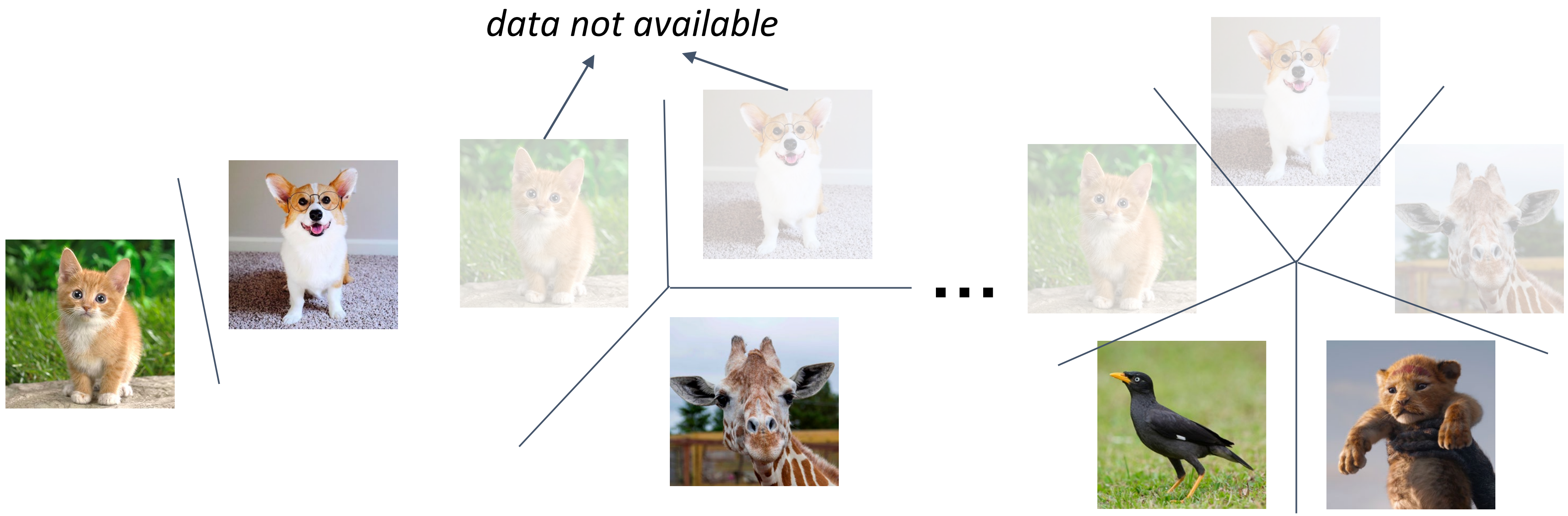


New classes:	*{cat, dog}*	*{giraffe}*	*{bird, lion, …}*
Learned classes:	{ }	*{cat, dog}*	*{cat, dog, giraffe, …}*

time

Continual Learning (CL)

A common class-incremental setting

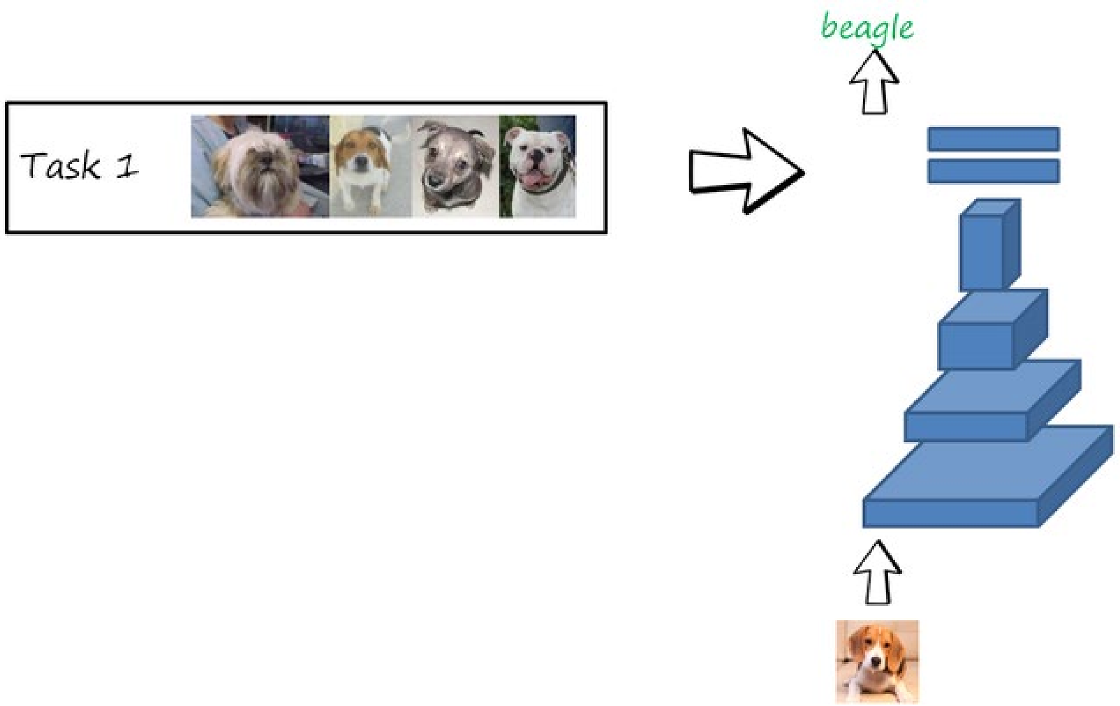


New classes:	*{cat, dog}*	*{giraffe}*	*{bird, lion, …}*
Learned classes:	{ }	*{cat, dog}*	*{cat, dog, giraffe, …}*

time

Catastrophic Forgetting

The primary challenge in CL



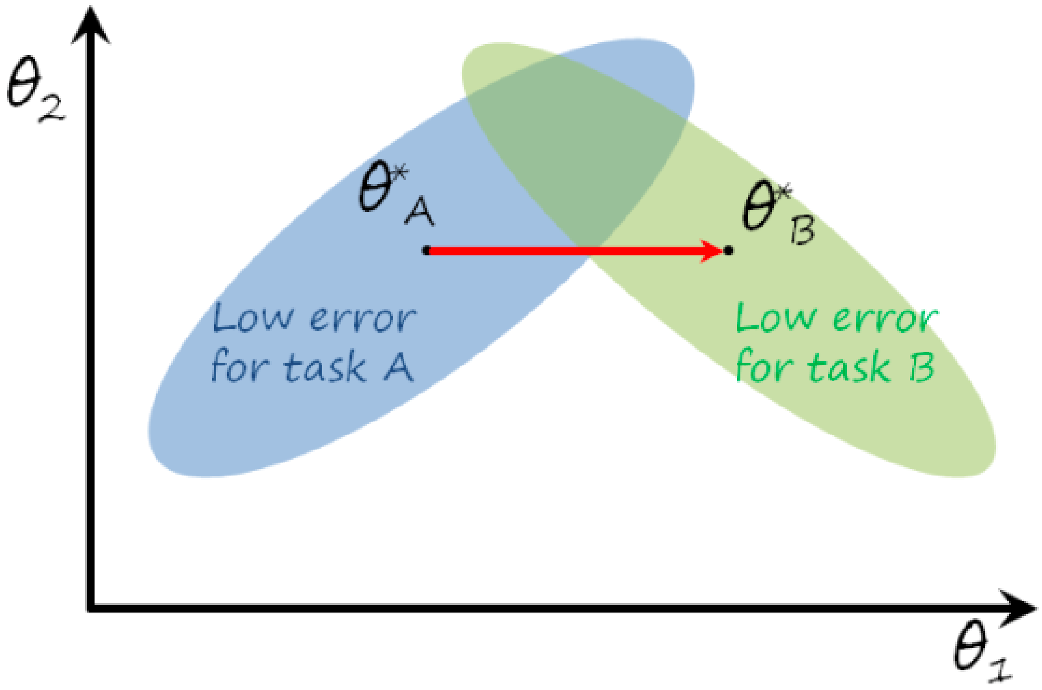


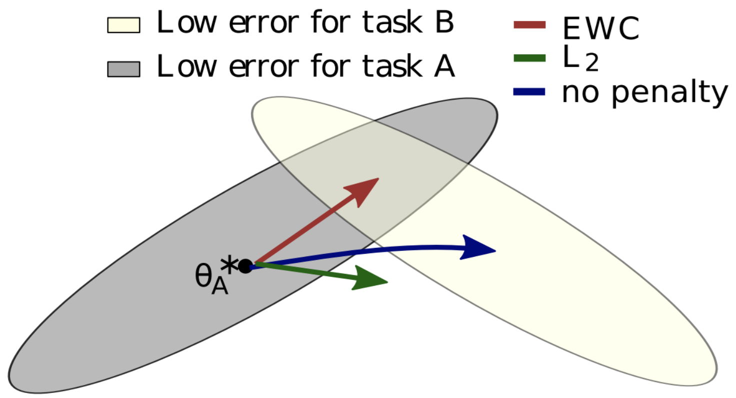
Image Credit to [Luis Herranz's blog]()

Catastrophic Forgetting

Standard solutions

Regularization-based method

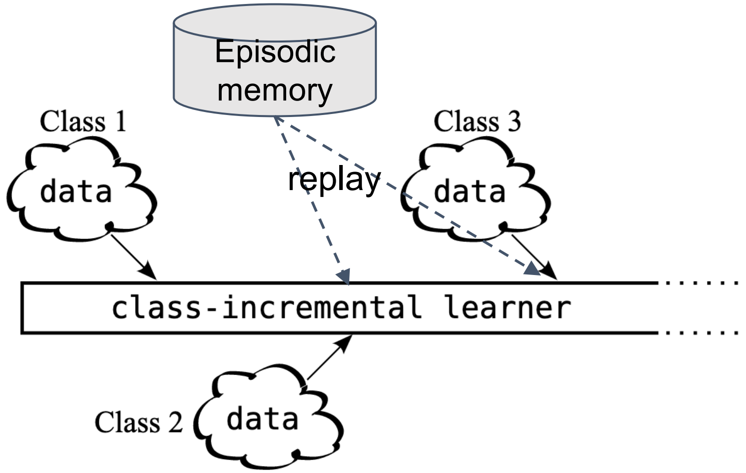
- consolidate prior knowledge when learning on new data using an extra regularization term
- would fail when task boundary is blur



[Kirkpatrick et al., 2017]

Replay-based method

- explicitly retrain on a limited subset of stored samples while training on new data
- effective in complex real-world tasks



adapted from [Rebuffi et al., 2017]

Kirkpatrick et al. "Overcoming catastrophic forgetting in neural networks." PNAS 2017.
Rebuffi et al. "icarl: Incremental classifier and representation learning." CVPR 2017.

Complex data & tasks

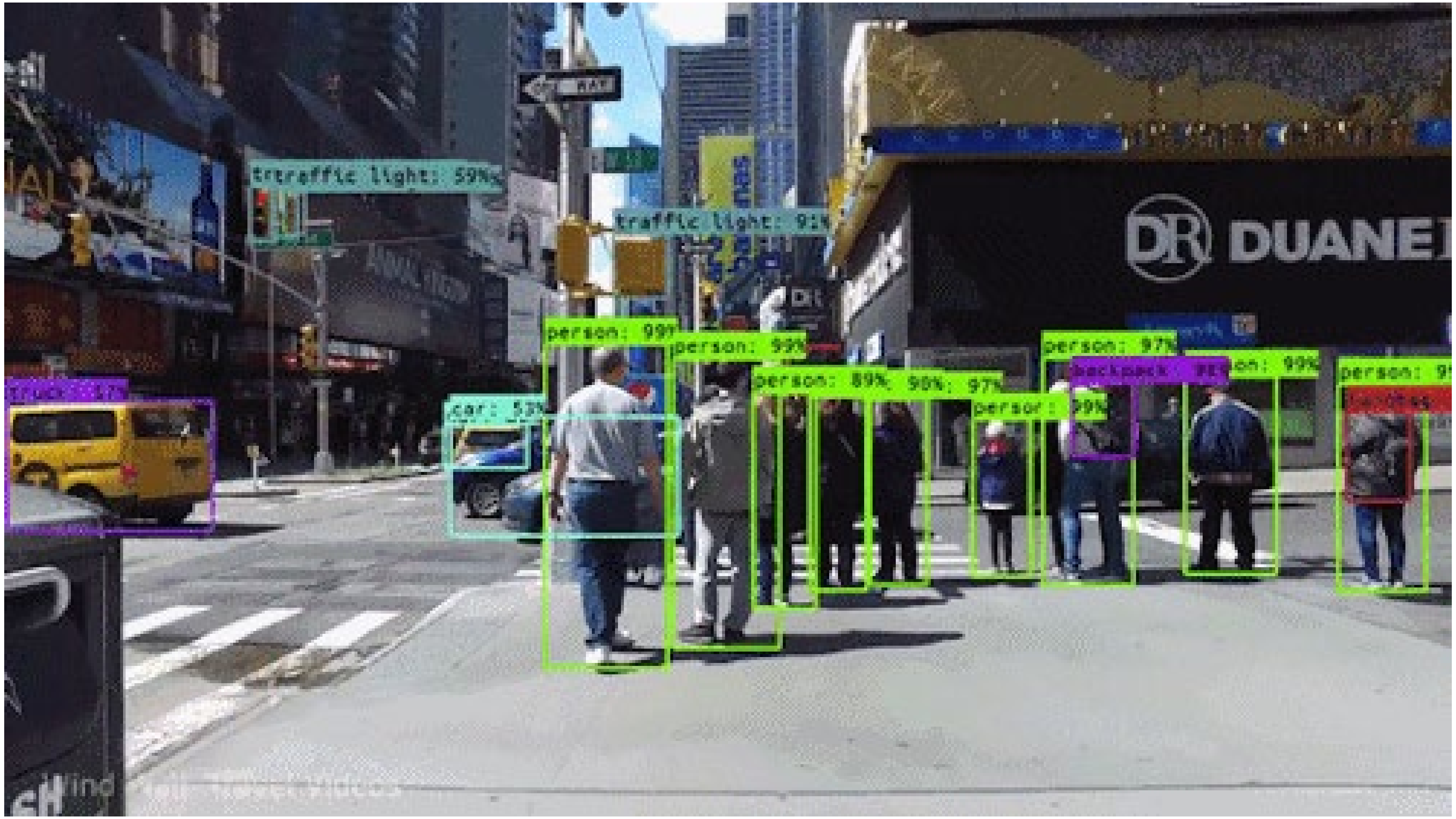
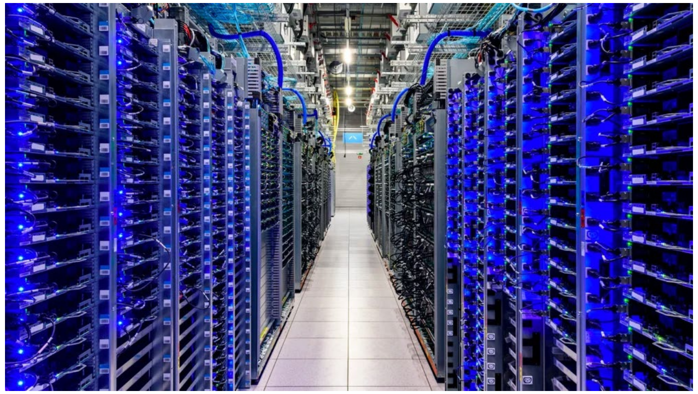


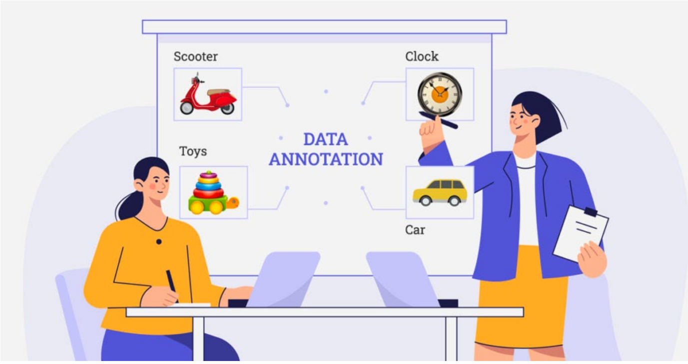
Image source: Towards Data Science

Real-world Continual Learning

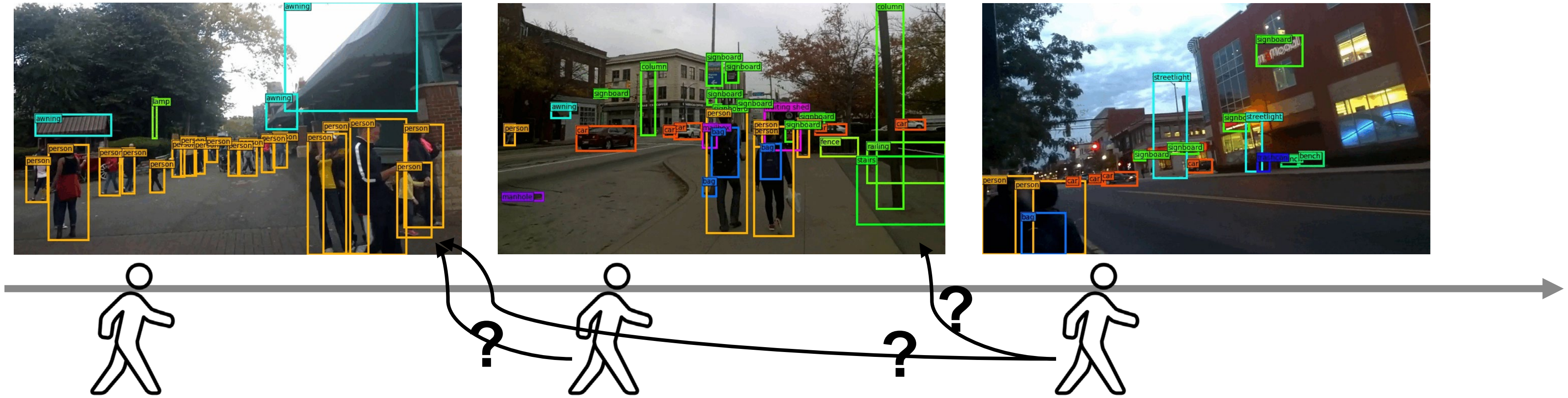
Constrained computation & annotation

Offline Learning





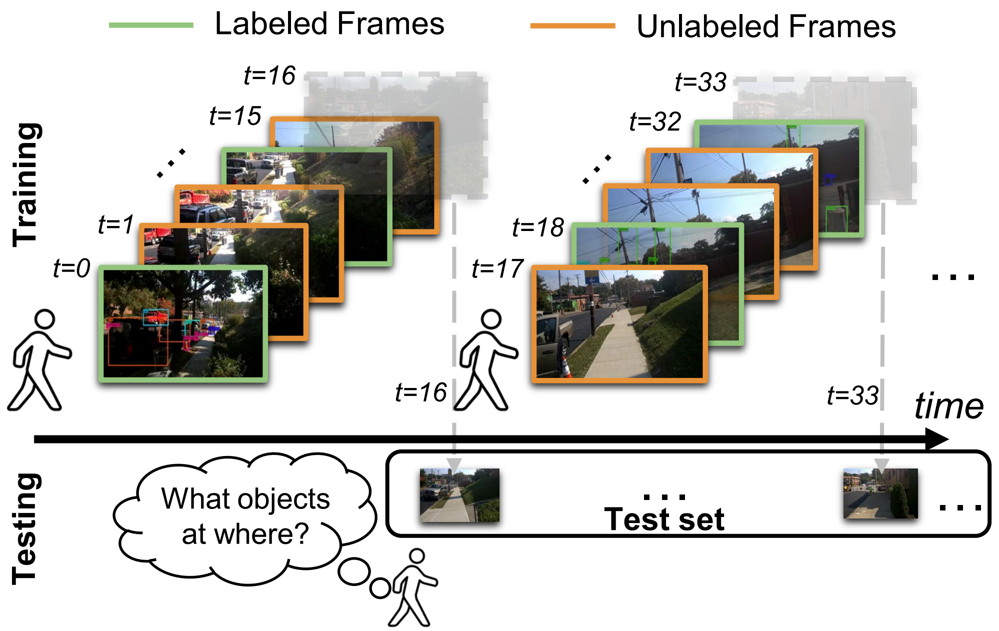
Online Learning



Prior setting **[Wang et al., 2021]**

Label-Efficient Online Continual Object Detection

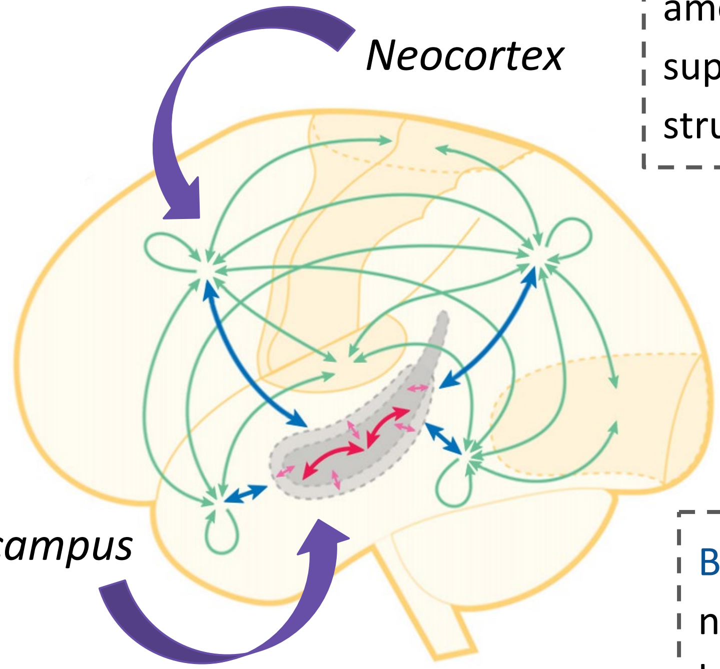
Prior setting **[Wang et al., 2021]**



Complementary Learning System (CLS)

How does human brain learn?

Rapid learning in the hippocampus supports initial learning of arbitrary new information



Complementary Learning System
(adapted from [Kumaran et al., 2016])

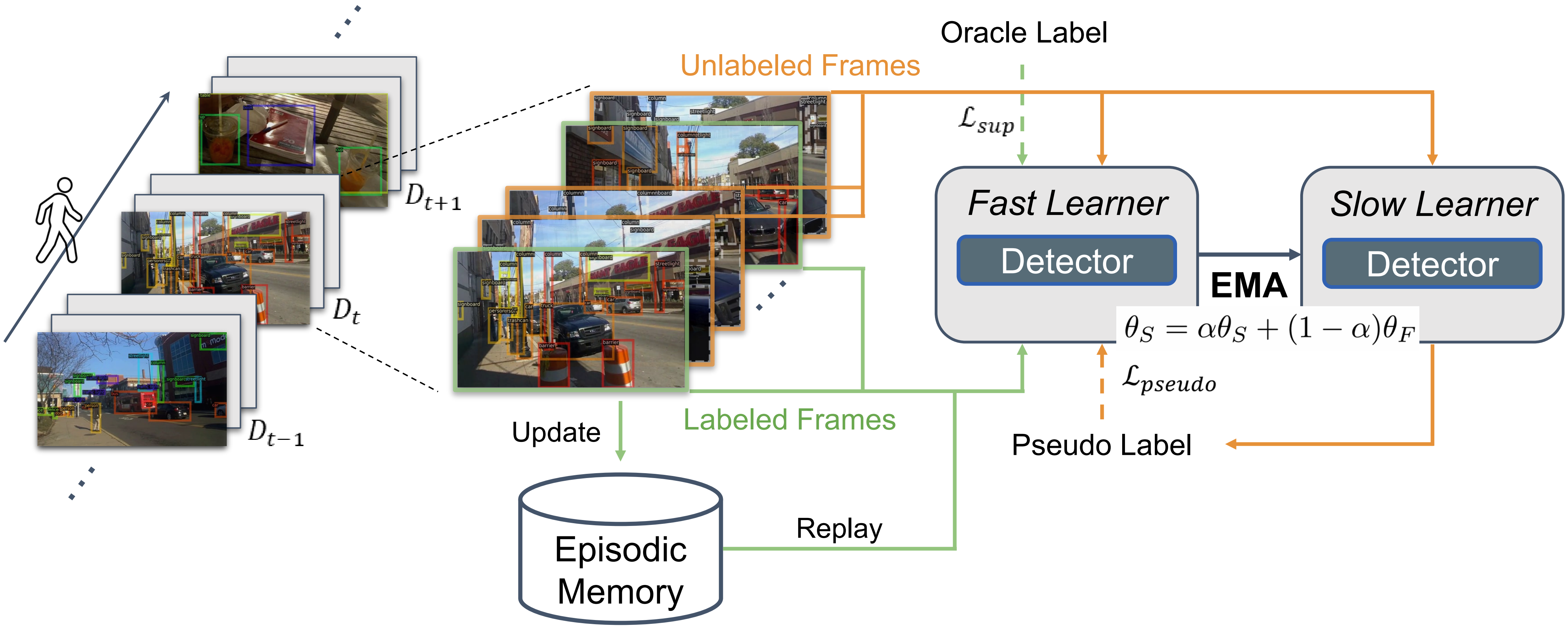
Connections within and among neocortical areas support gradual acquisition of structured knowledge

Bidirectional connections link neocortical representations to hippocampus for storage, retrieval, and replay

Kumaran et al. "What learning systems do intelligent agents need? Complementary learning systems theory updated." Trends in cognitive sciences 2016.

Efficient-CLS

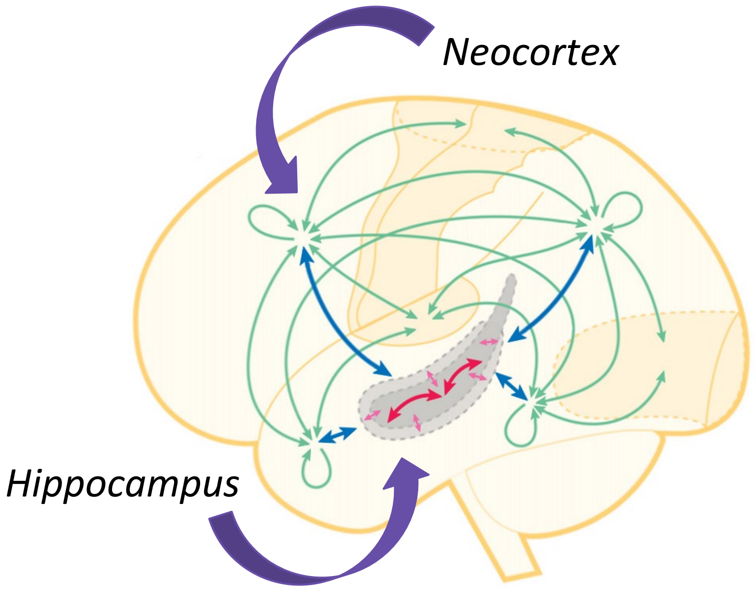
A plug-and-play module inspired by CLS

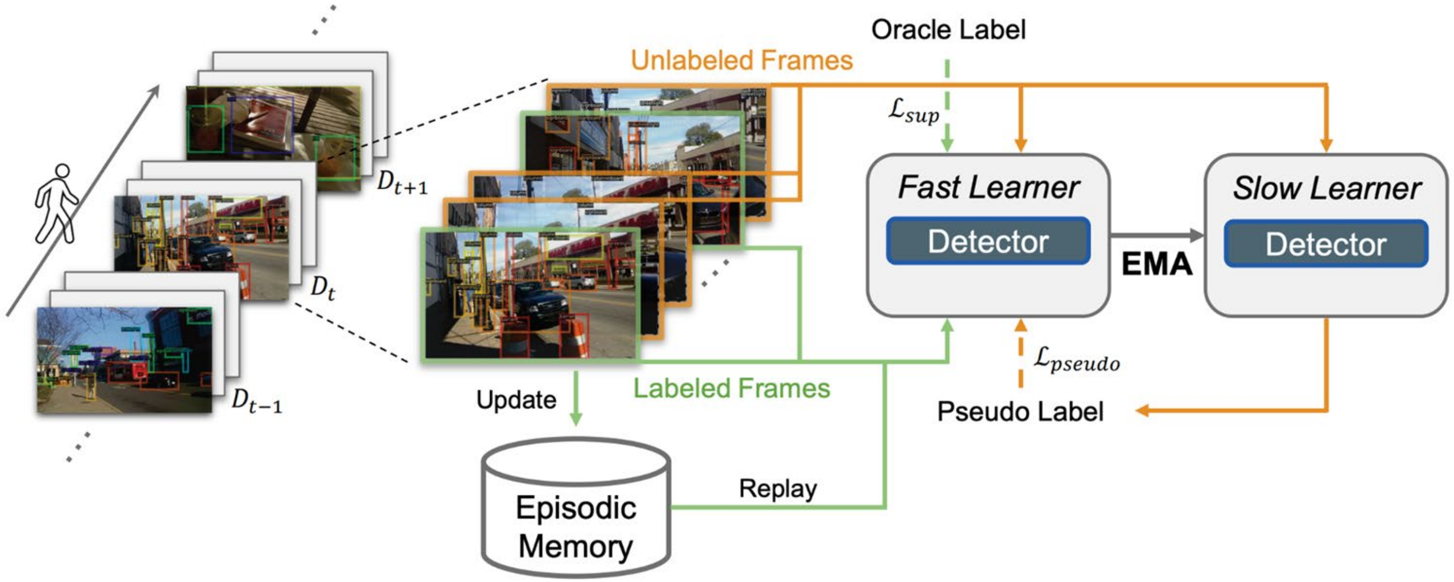


Wu et al. "Label-efficient online continual object detection in streaming video." ICCV 2023.

Efficient-CLS

A plug-and-play module inspired by CLS





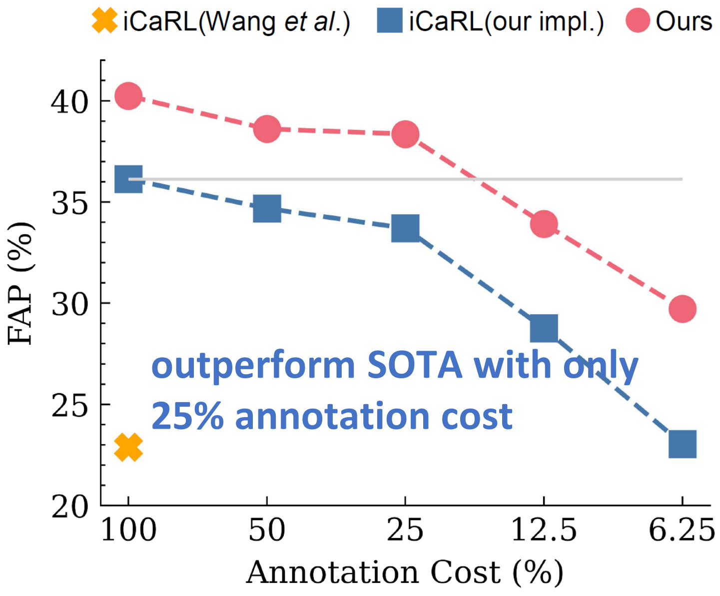
Fast Learner: quickly encodes new knowledge from current data stream and then consolidate it to the slow learner

Slow Learner: accumulates the acquired knowledge from fast learner over time and guides the fast learner with meaningful pseudo labels

Wu et al. "Label-efficient online continual object detection in streaming video." ICCV 2023.

Efficient-CLS

SOTA performance with minimal annotation cost and forgetting



outperform SOTA with only 25% annotation cost

	Annotation Cost	OAK FAP (↑)	OAK CAP (↑)	OAK F (↓)	EgoObjects FAP (↑)	EgoObjects CAP (↑)	EgoObjects F (↓)
Incremental	100%	8.38	7.72	0.03	10.21	3.55	1.48
Offline Training	100%	48.28	35.23	-	86.18	59.81	-
EWC	100%	7.73	7.02	-0.12	5.15	1.60	0.57
iOD	100%	7.92	7.14	0.98	8.80	2.64	0.00
iCaRL(Wang *et al.*)	100%	22.89	16.60	-2.95	37.61	21.71	2.79
iCaRL(our impl.)	100%	36.14	26.26	-4.89	60.80	36.41	-0.60
w/ Efficient-CLS	25%	38.36(+2.22)	26.64(+0.38)	-8.20(-3.31)	61.26(+0.46)	39.58(+3.17)	-3.48(-2.88)
	100%	40.24(+4.10)	28.18(+1.92)	-8.10(-3.21)	67.05(+6.25)	40.36(+3.95)	-3.67(-3.07)
A-GEM	100%	36.94	26.19	-5.54	58.79	35.88	-8.38
w/ Efficient-CLS	25%	37.06(+0.12)	26.36(+0.17)	-7.76(-2.22)	63.06(+4.27)	39.46(+3.58)	-7.49(+0.89)
	100%	39.87(+2.93)	27.97(+1.78)	-7.17(-1.63)	66.94(+8.15)	39.57(+3.69)	-11.68(-3.30)
GDumb	100%	35.27	25.29	-6.59	58.85	36.38	-5.21
w/ Efficient-CLS	25%	37.67(+2.40)	25.59(+0.30)	-9.30(-2.71)	62.70(+3.85)	38.78(+2.40)	-8.86(-3.65)
	100%	38.61(+3.34)	26.04(+0.75)	-9.14(-2.55)	63.55(+4.70)	38.98(+2.60)	-7.50(-2.29)
DER++	100%	37.79	25.24	-2.87	55.82	30.84	-6.08
w/ Efficient-CLS	25%	37.93(+0.14)	25.64(+0.4)	-8.90(-6.03)	59.70(+3.88)	34.15(+3.31)	-11.21(-5.13)
	100%	39.61(+1.82)	26.73(+1.49)	-8.30(-5.43)	62.01(+6.19)	33.09(+2.25)	-11.05(-4.97)

compatible with existing CL methods

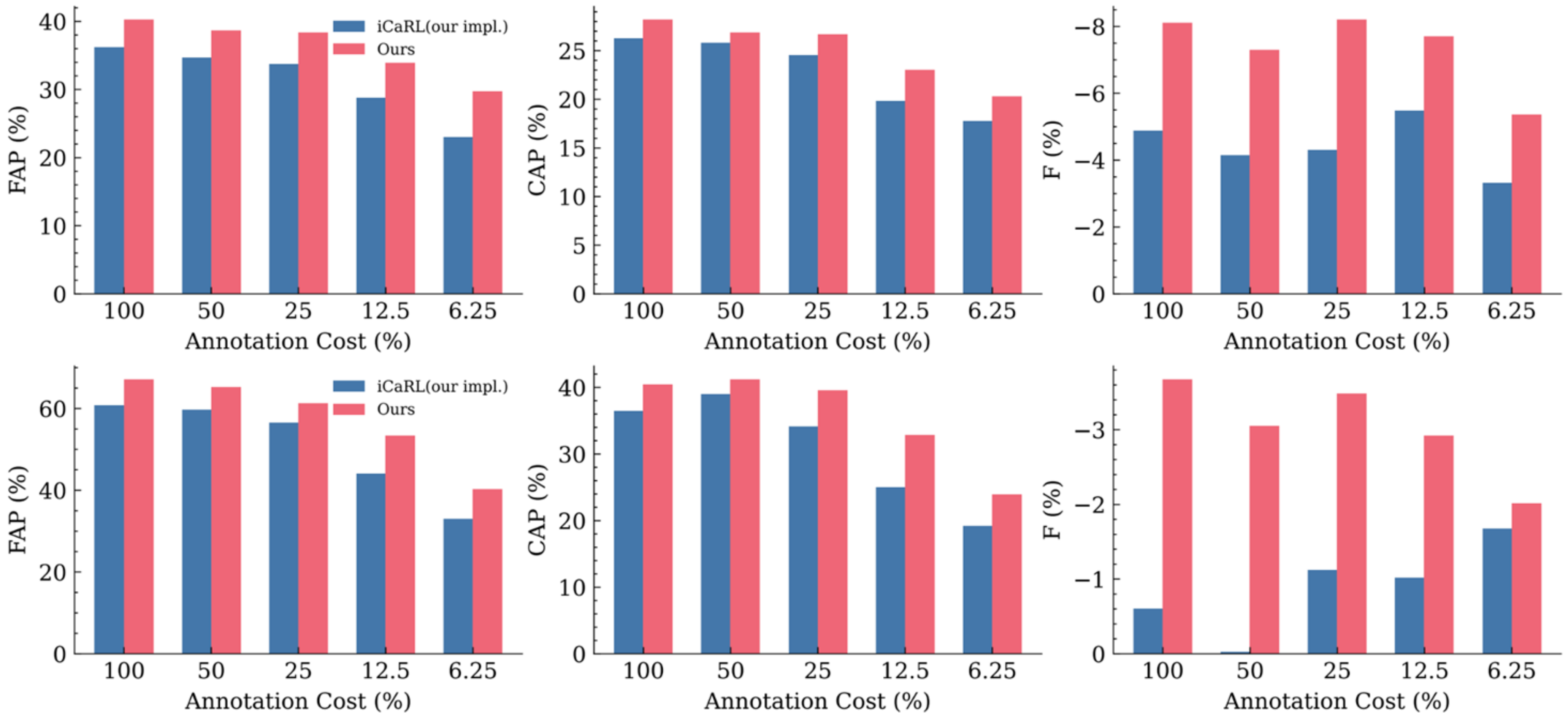
Wu et al. "Label-efficient online continual object detection in streaming video." ICCV 2023.

Efficient-CLS

Consistent improvement over all annotation costs

OAK

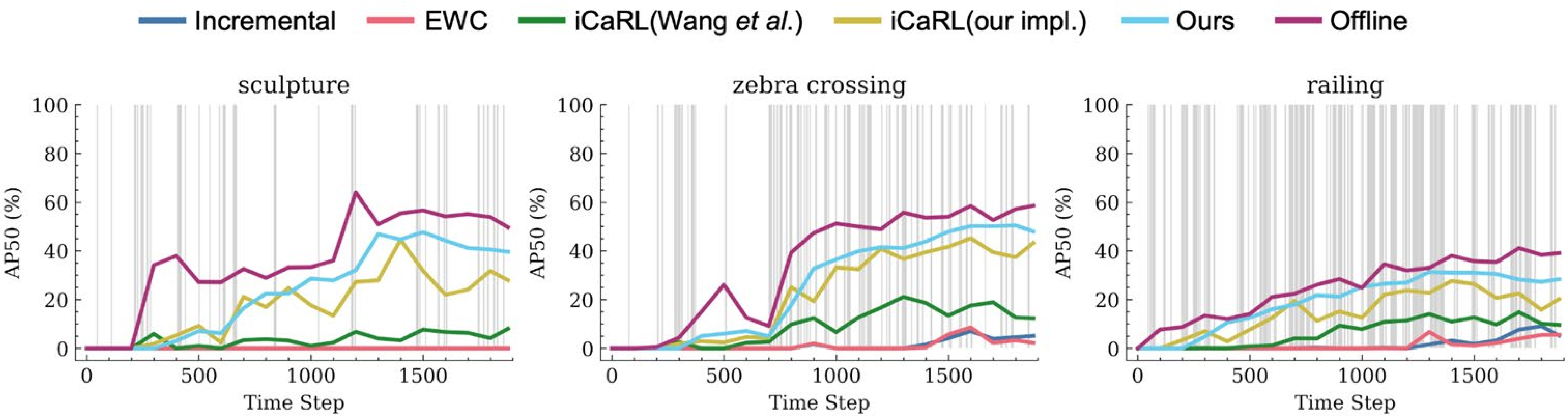
EgoObjects



Wu et al. "Label-efficient online continual object detection in streaming video." ICCV 2023.

Efficient-CLS

Reduced forgetting even when class appears infrequently



Wu et al. "Label-efficient online continual object detection in streaming video." ICCV 2023.

Efficient-CLS

Ablation on proposed components

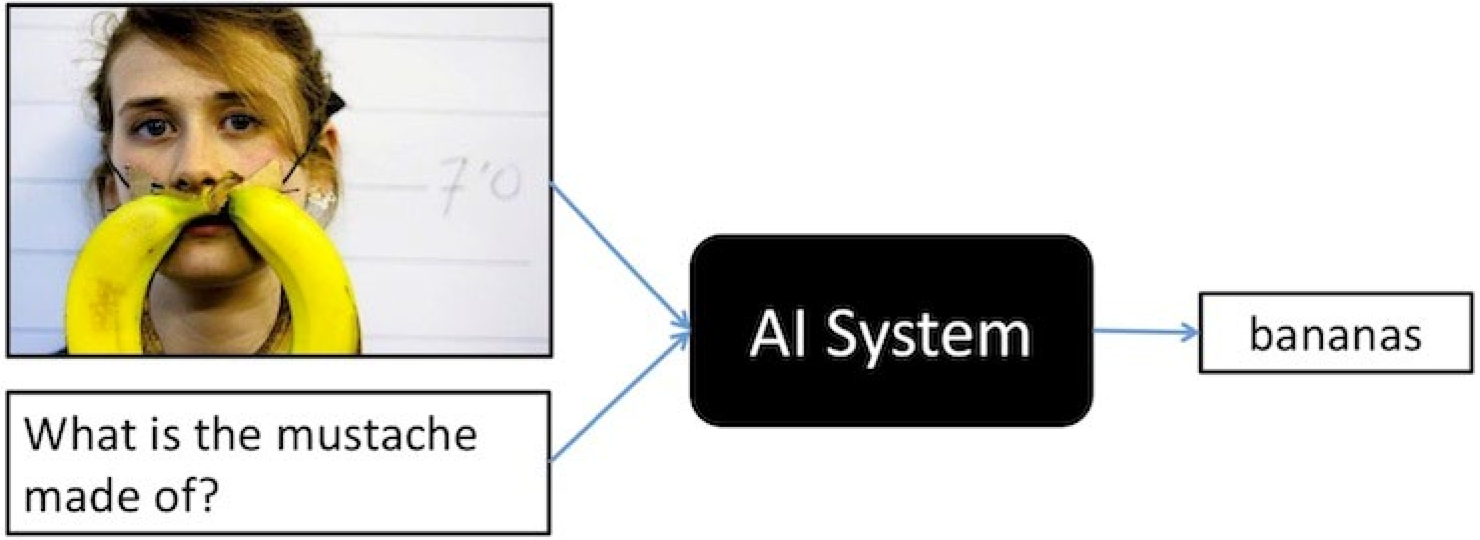
EMA	PL	50% FAP (↑)	50% CAP (↑)	50% F (↓)	25% FAP (↑)	25% CAP (↑)	25% F (↓)	12.5% FAP (↑)	12.5% CAP (↑)	12.5% F (↓)	6.25% FAP (↑)	6.25% CAP (↑)	6.25% F (↓)
✗	✗	34.68	25.78	-4.15	33.70	24.57	-4.30	28.76	19.80	-5.48	23.04	17.75	-3.31
✓	✗	35.74	25.77	-4.82	34.79	25.62	-4.35	31.72	21.16	-7.24	27.84	20.03	-3.96
✗	✓	35.61	25.56	-3.76	34.95	25.65	-3.65	31.60	22.44	-4.83	26.39	19.50	-1.99
✓	✓	**38.61**	**26.90**	**-7.29**	**38.36**	**26.64**	**-8.20**	**33.92**	**23.04**	**-7.71**	**29.72**	**20.31**	**-5.36**

- **EMA** effectively consolidates knowledge and avoid forgetting.
- Naive pseudo-labeling can improve AP, but fails to prevent forgetting.
- **Pseudo-labeling + EMA** achieves best results with minimal forgetting.

Wu et al. "Label-efficient online continual object detection in streaming video." ICCV 2023.

Real-world Continual Learning

Visual question answering (VQA)

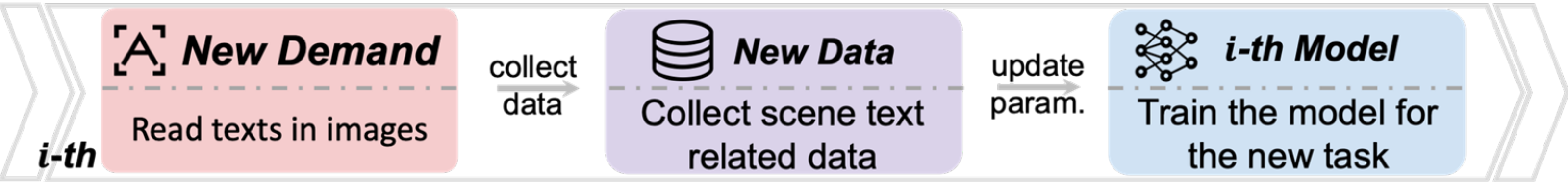


Source: [Antol et al., 2015]

Antol et al. "Vqa: Visual question answering." ICCV 2015.

Continual Learning for VQA

Scene-incremental scenario



Shop

Q: Where is the elevator in this picture? **A:** On the left.

Sports

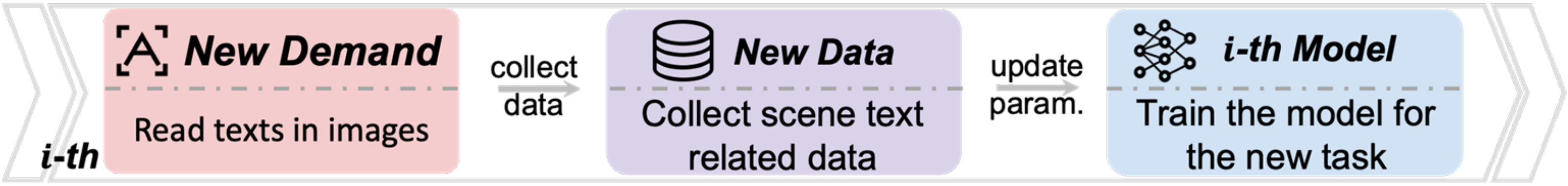
Q: What are the men holding? **A:** Ski poles.

Office

Q: Is there a laptop in this office? **A:** No.

Continual Learning for VQA

Function-incremental scenario



Attribute Recognition

Q: What <u>color</u> is the snow board on the right? **A:** Yellow.

Knowledge Reasoning

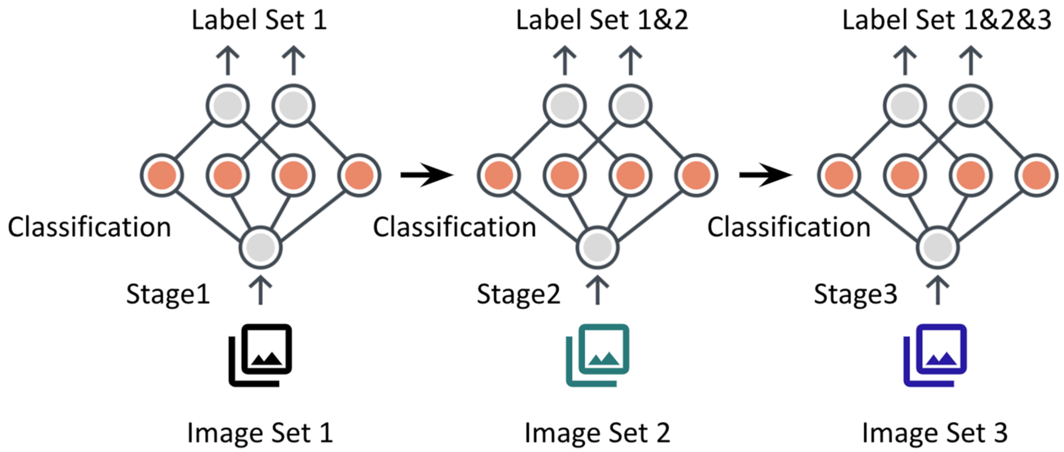
Q: What object <u>can be used to transport people</u>? **A:** Bus.

Scene Text Recognition

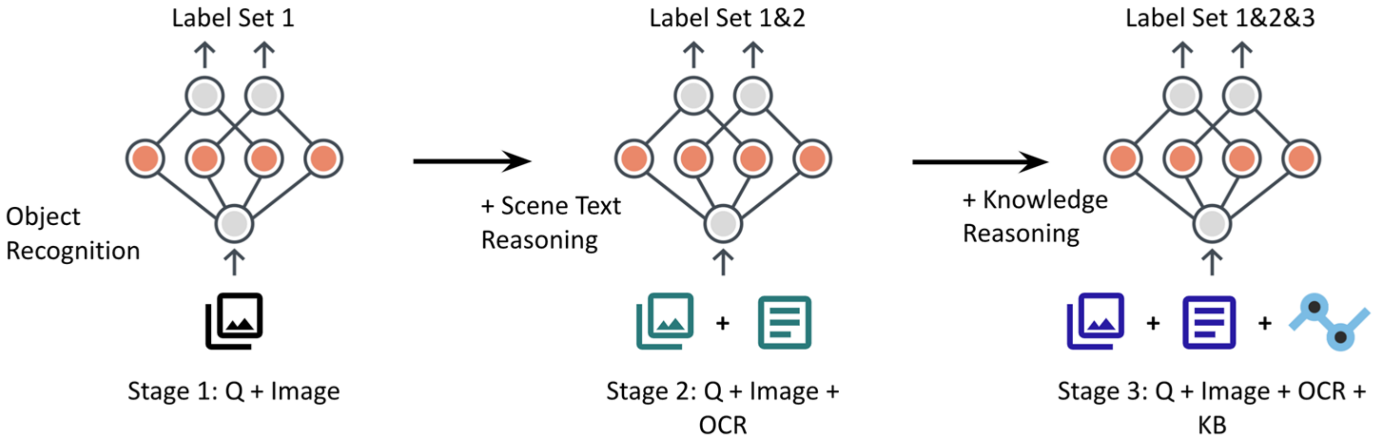
Q: What is the brand of this phone? **A:** <u>Nokia</u>.

Continual Learning for VQA

CL for classification *vs.* CL for VQA



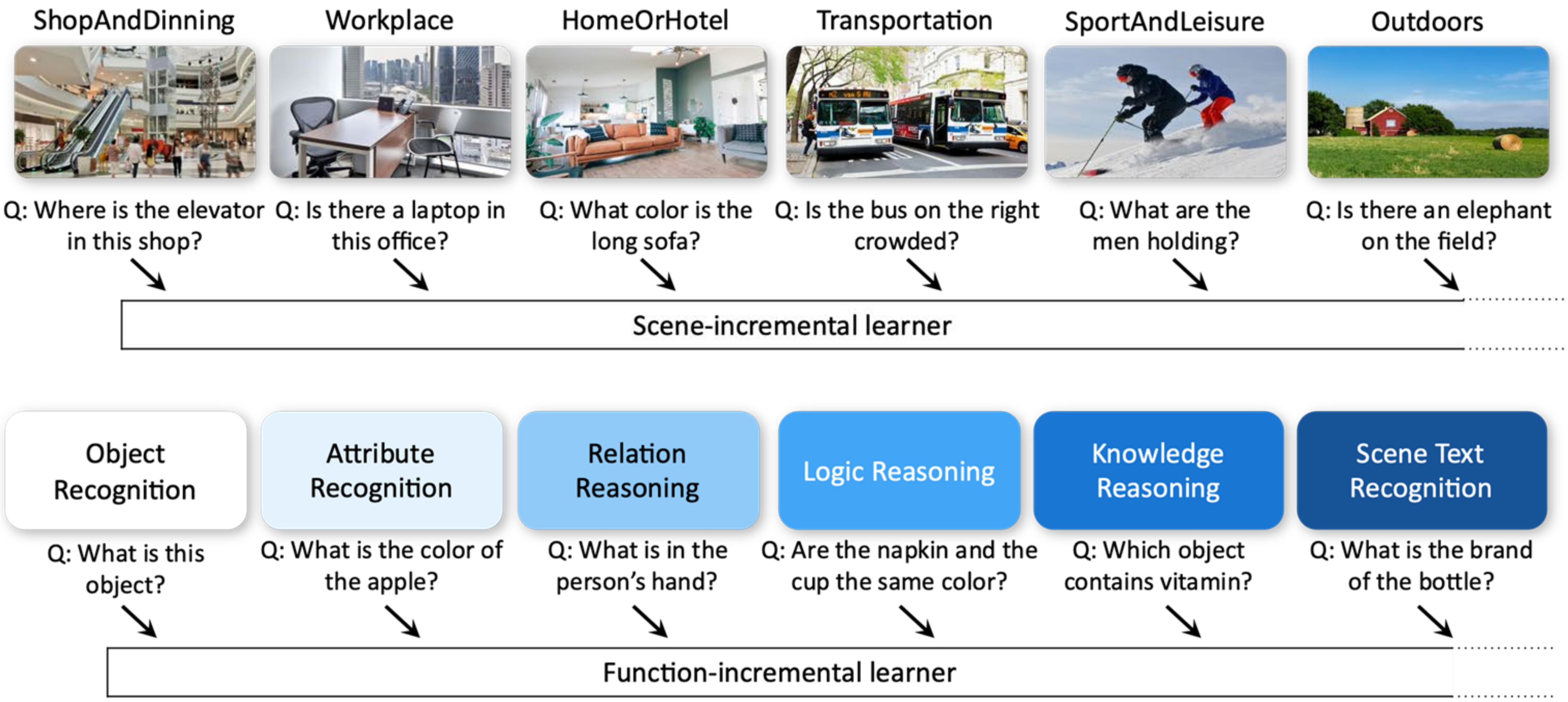
- one modality (vision)
- one function (classification)
- focus on catastrophic forgetting and interference in representation



- multi-modality (V + L)
- multiple functions (object recognition, attribute recognition, logic reasoning)
- focus on catastrophic forgetting in representation & reasoning

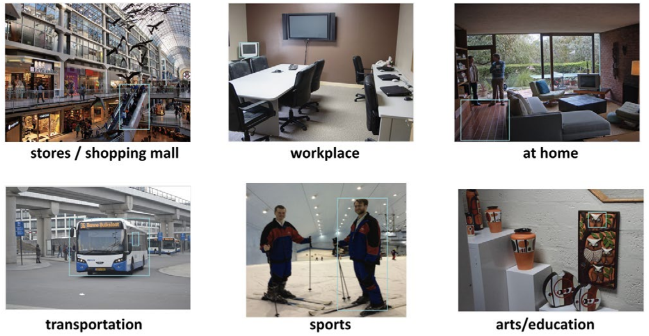
CLOVE

A benchmark for **C**ontinual **L**earning **O**n **V**isual qu**E**stion answering

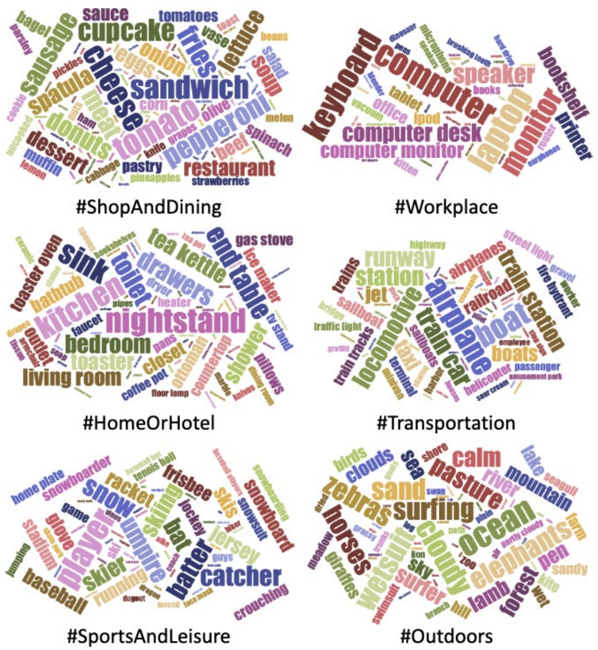


Lei et al. "Symbolic replay: Scene graph as prompt for continual learning on vqa task." AAAI 2023.

Data construction for CLOVE-Scene



scene classification



scene-specific QA selection

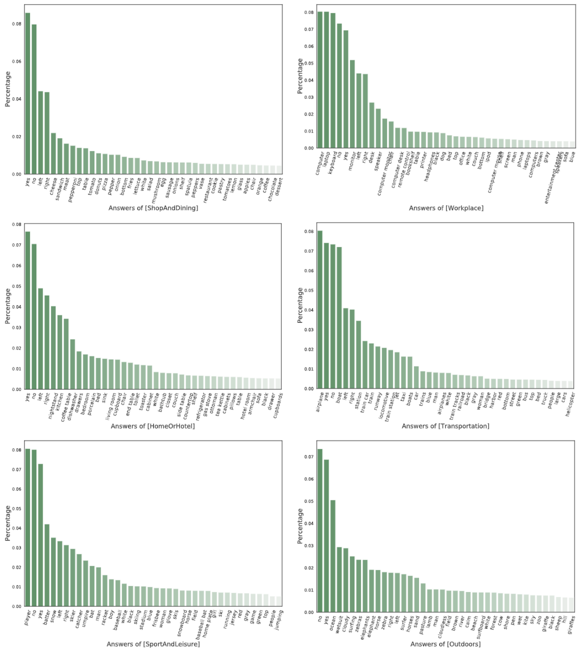


Figure S3: Answer distribution of each task in CLOVE-scene. We show the top-40 frequent answers.

smooth answer distribution

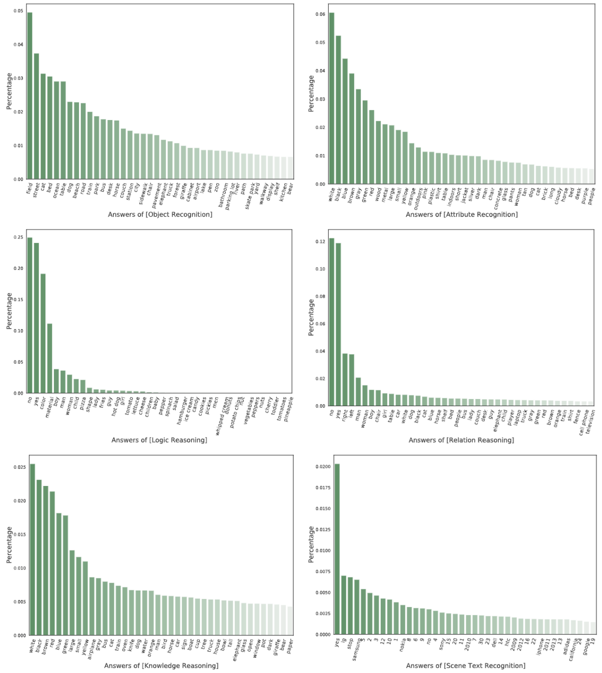
Lei et al. "Symbolic replay: Scene graph as prompt for continual learning on vqa task." AAAI 2023.

CLOVE

Data construction for CLOVE-Function

Stage	Operation	Argument
Object	Select, Query, Choose	name
Attribute	Query, Verify, Choose, Filter	color, material, weather….
Relation	Relate, Verify, Choose	rel
Logic	Different, Same, Common, Choose	same color, choose healthier,….
Knowledge Reasoning	Find w/ KG	
Scene Text Recognition	Scene text recognition	

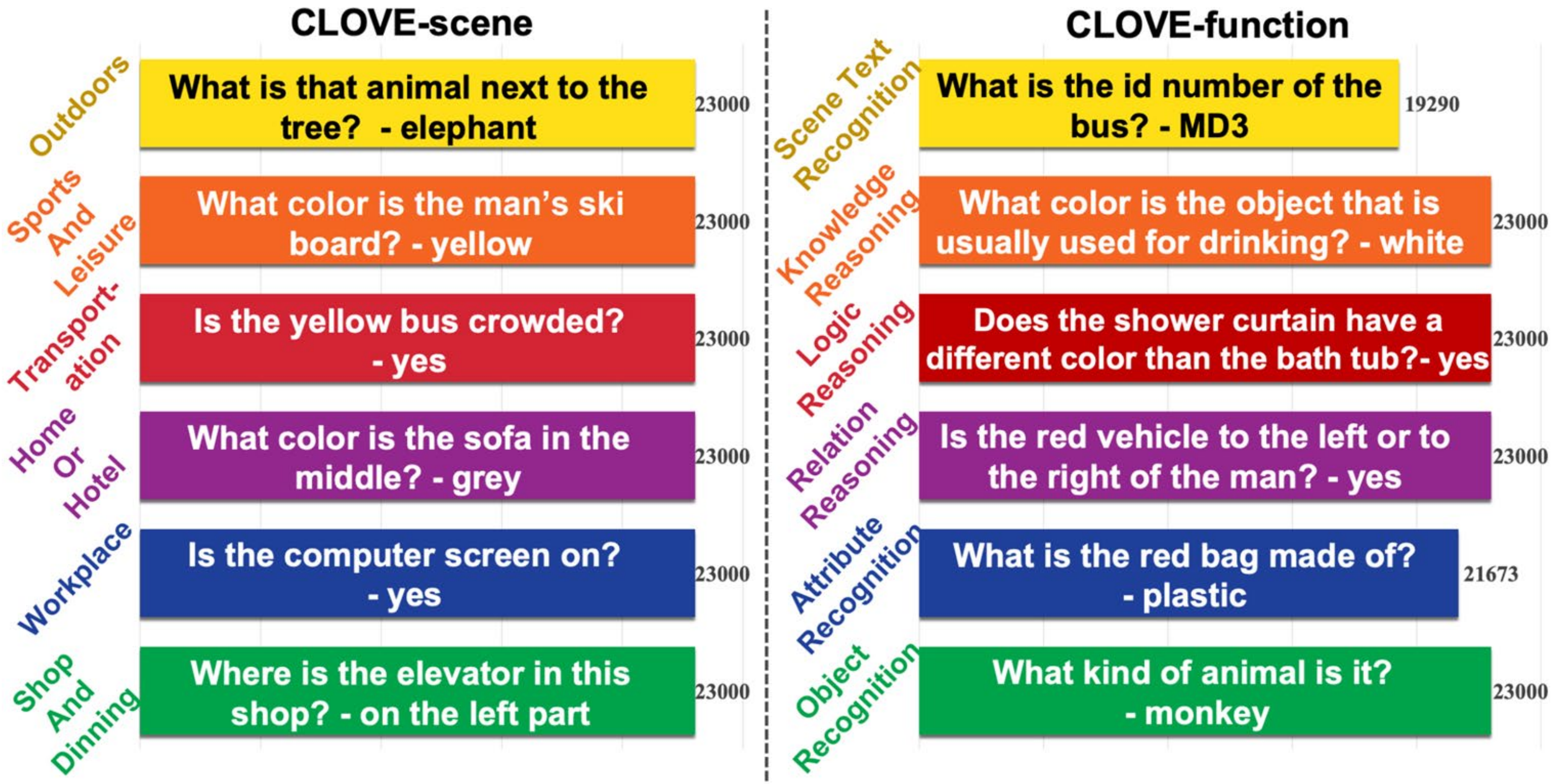
Function assignment given the rules



Smooth answer distribution

Lei et al. "Symbolic replay: Scene graph as prompt for continual learning on vqa task." AAAI 2023.

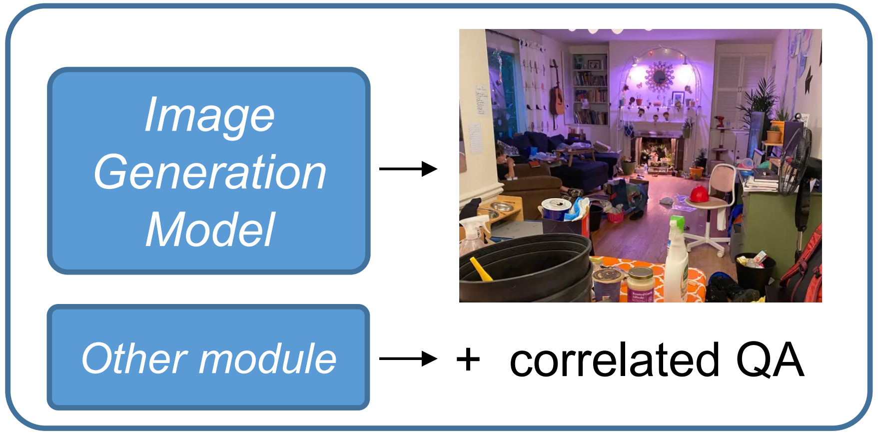
QA examples



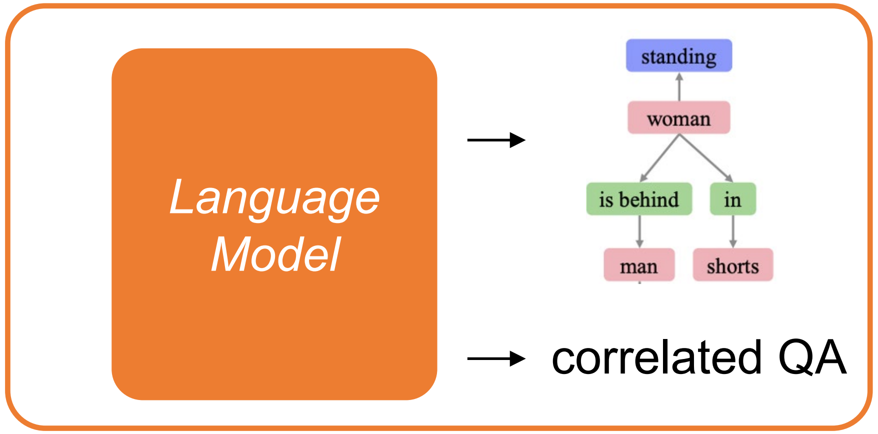
Lei et al. "Symbolic replay: Scene graph as prompt for continual learning on vqa task." AAAI 2023.

Scene Graph as Prompt for symbolic replay

Image replay *vs.* scene graph replay



Replay Image + Q + A

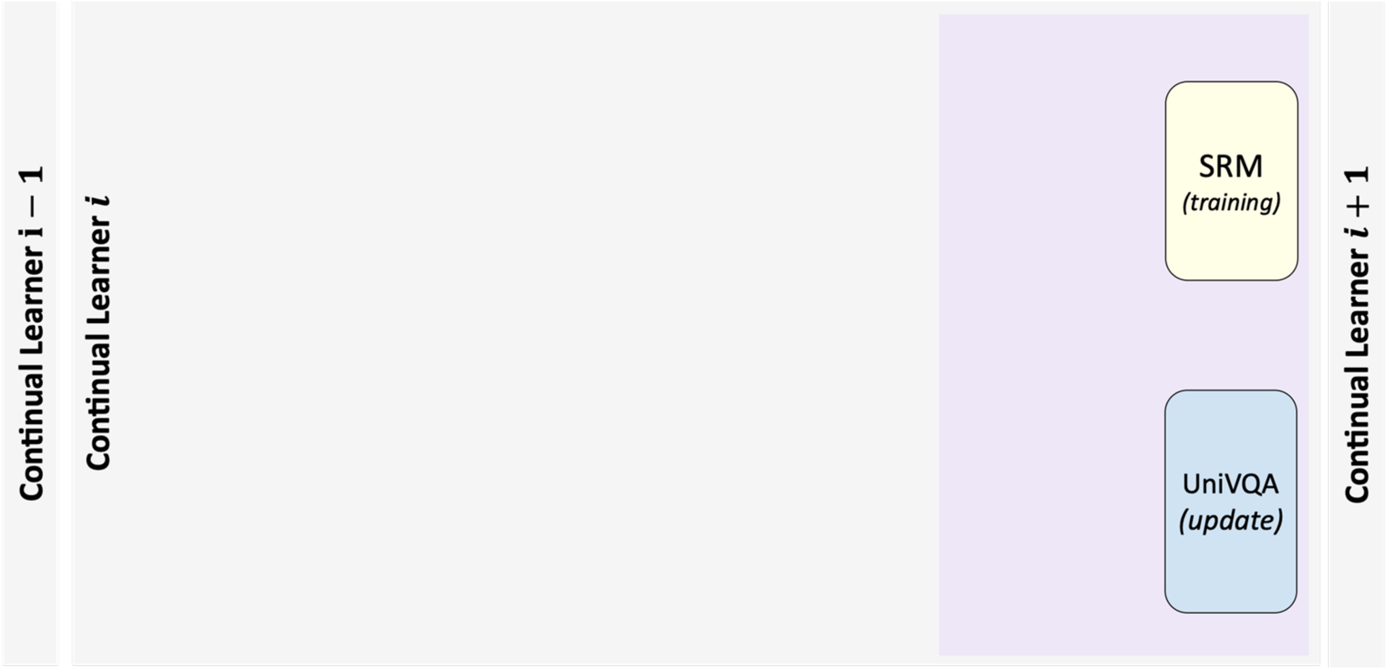


Replay Scene Graph + Q + A

Lei et al. "Symbolic replay: Scene graph as prompt for continual learning on vqa task." AAAI 2023.

Scene Graph as Prompt for symbolic replay

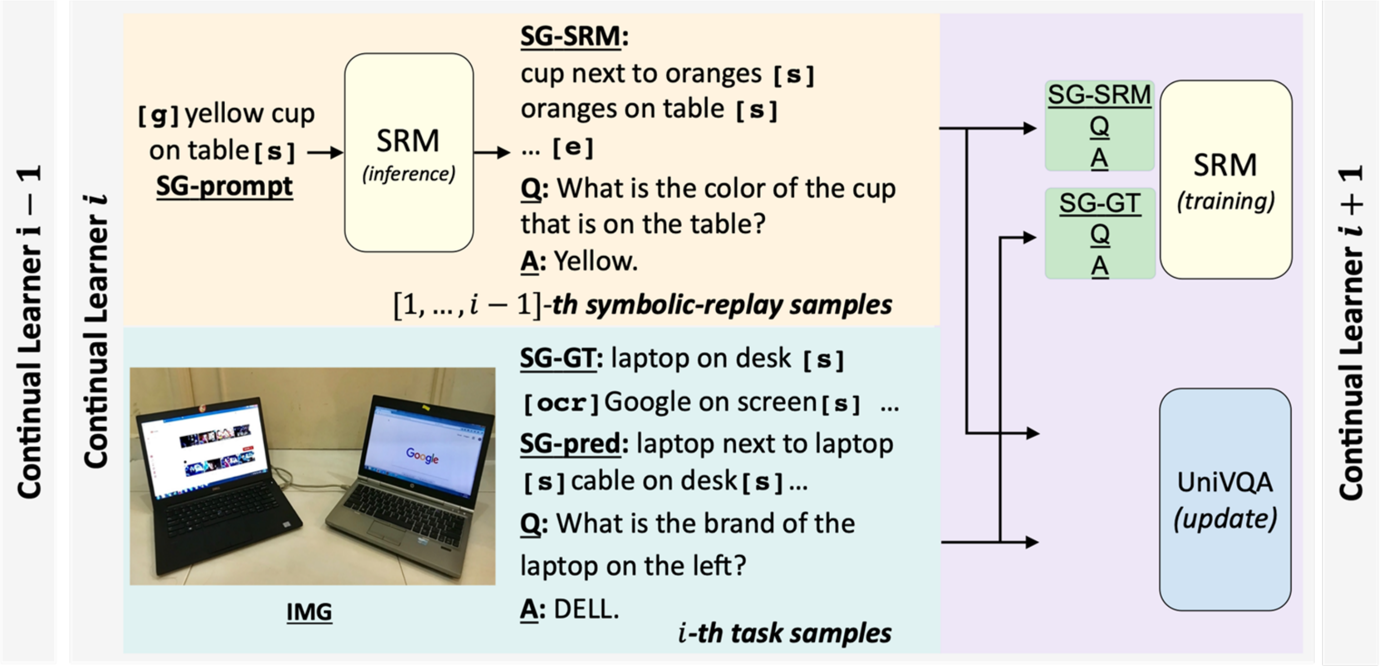
Overall framework



Lei et al. "Symbolic replay: Scene graph as prompt for continual learning on vqa task." AAAI 2023.

Scene Graph as Prompt for symbolic replay

Overall framework

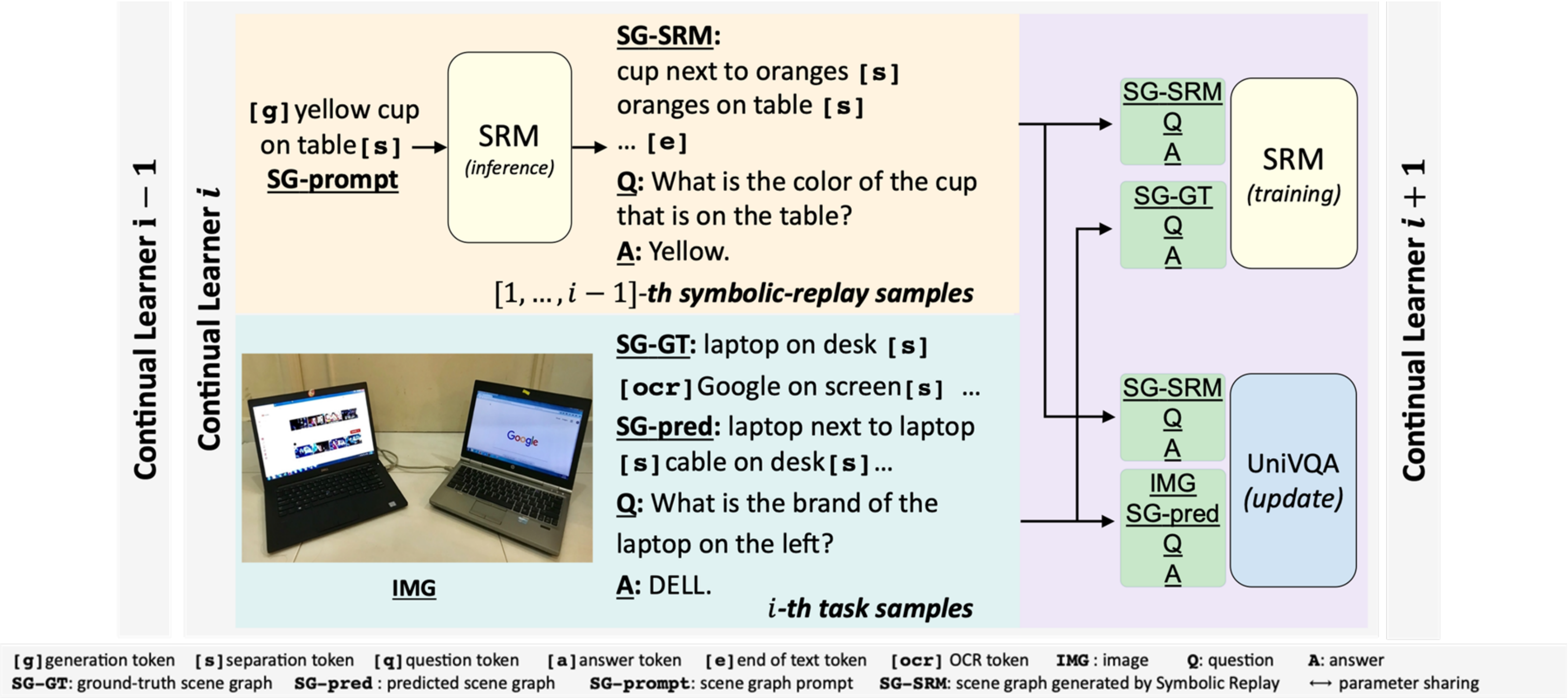


[g] generation token [s] separation token [q] question token [a] answer token [e] end of text token [ocr] OCR token **IMG** : image **Q**: question **A**: answer
SG-GT: ground-truth scene graph **SG-pred** : predicted scene graph **SG-prompt**: scene graph prompt **SG-SRM**: scene graph generated by Symbolic Replay ⟷ parameter sharing

Lei et al. "Symbolic replay: Scene graph as prompt for continual learning on vqa task." AAAI 2023.

Scene Graph as Prompt for symbolic replay

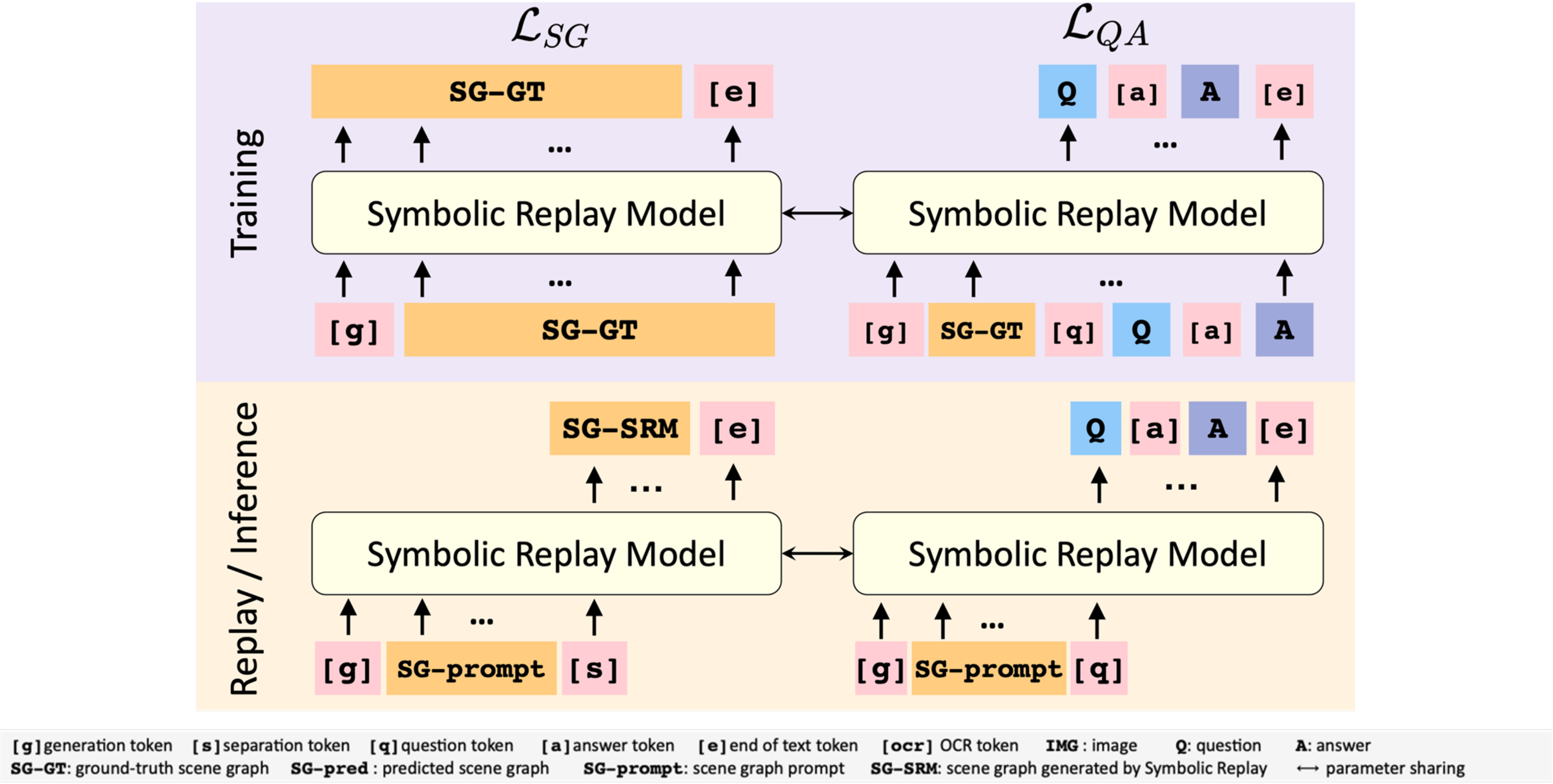
Overall framework



Lei et al. "Symbolic replay: Scene graph as prompt for continual learning on vqa task." AAAI 2023.

Scene Graph as Prompt for symbolic replay

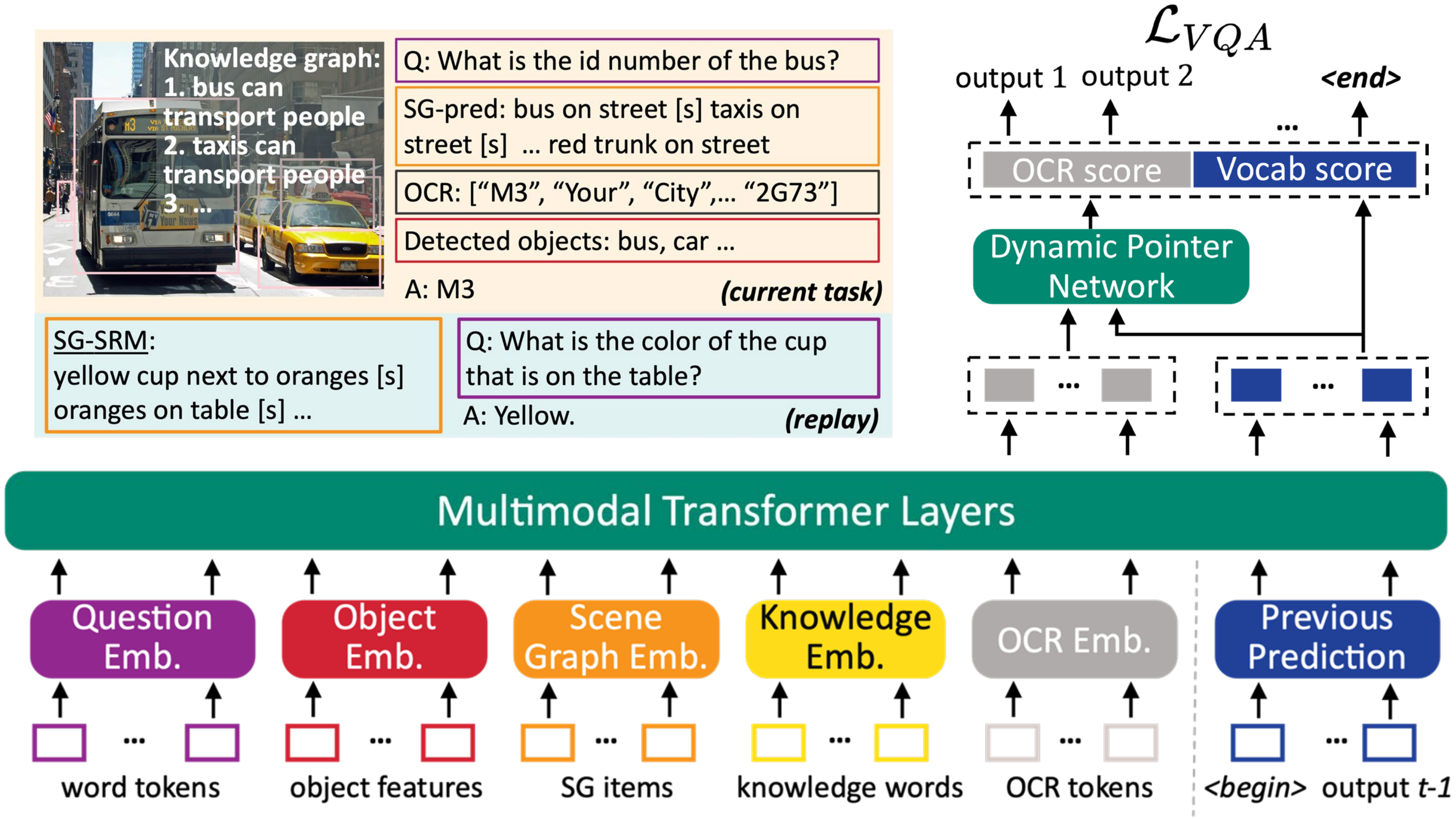
Symbolic Replay Model



Lei et al. "Symbolic replay: Scene graph as prompt for continual learning on vqa task." AAAI 2023.

Scene Graph as Prompt for symbolic replay

Unified VQA Transformer (UniVQA)



Lei et al. "Symbolic replay: Scene graph as prompt for continual learning on vqa task." AAAI 2023.

Scene Graph as Prompt for symbolic replay

Unified VQA Transformer (UniVQA)

Method	CLOVE-scene							CLOVE-function						
	abcdef	*bdfcae*	*beacfd*	*beadcf*	*bedfca*	*ecdfab*	Avg.	*oarlks*	*roslak*	*rklsao*	*rsolak*	*lkosra*	*kaorls*	Avg.
Finetune	27.53	27.98	28.39	27.71	24.49	25.42	26.92	27.60	29.33	21.12	30.65	25.43	22.82	26.16
EWC	27.59	27.64	28.47	29.18	24.03	25.48	27.07	29.26	30.87	21.87	28.69	23.58	23.27	26.26
MAS	27.41	27.15	28.19	27.34	25.40	26.78	27.05	28.73	31.59	28.62	28.57	24.26	26.73	28.08
VQG	29.15	29.74	30.02	30.27	27.28	28.66	29.19	32.78	33.16	29.55	33.82	30.17	28.67	31.36
LAMOL-m	29.40	28.52	29.45	29.86	26.52	27.82	28.60	28.42	29.04	24.16	32.17	26.94	26.92	27.94
SGP (Ours)	**32.21**	**33.72**	**34.37**	**33.18**	**31.84**	**32.98**	**33.05**	**45.97**	**41.80**	**39.05**	**42.95**	**38.65**	**43.62**	**42.01**
Real-rnd	36.60	37.69	35.50	36.51	35.86	36.84	36.50	44.83	42.62	39.28	43.37	40.85	40.08	41.84
Real-kmeans	36.91	38.15	37.01	38.30	37.93	34.86	37.19	40.28	41.19	38.49	42.21	38.39	36.29	39.48
Offline	48.45							57.53						

- SGP outperforms other real-data-free CL methods
- SGP is on par with real-data replay under CLOVE-function
- CLOVE is challenging

Lei et al. "Symbolic replay: Scene graph as prompt for continual learning on vqa task." AAAI 2023.

Scene Graph as Prompt for symbolic replay

Ablation study

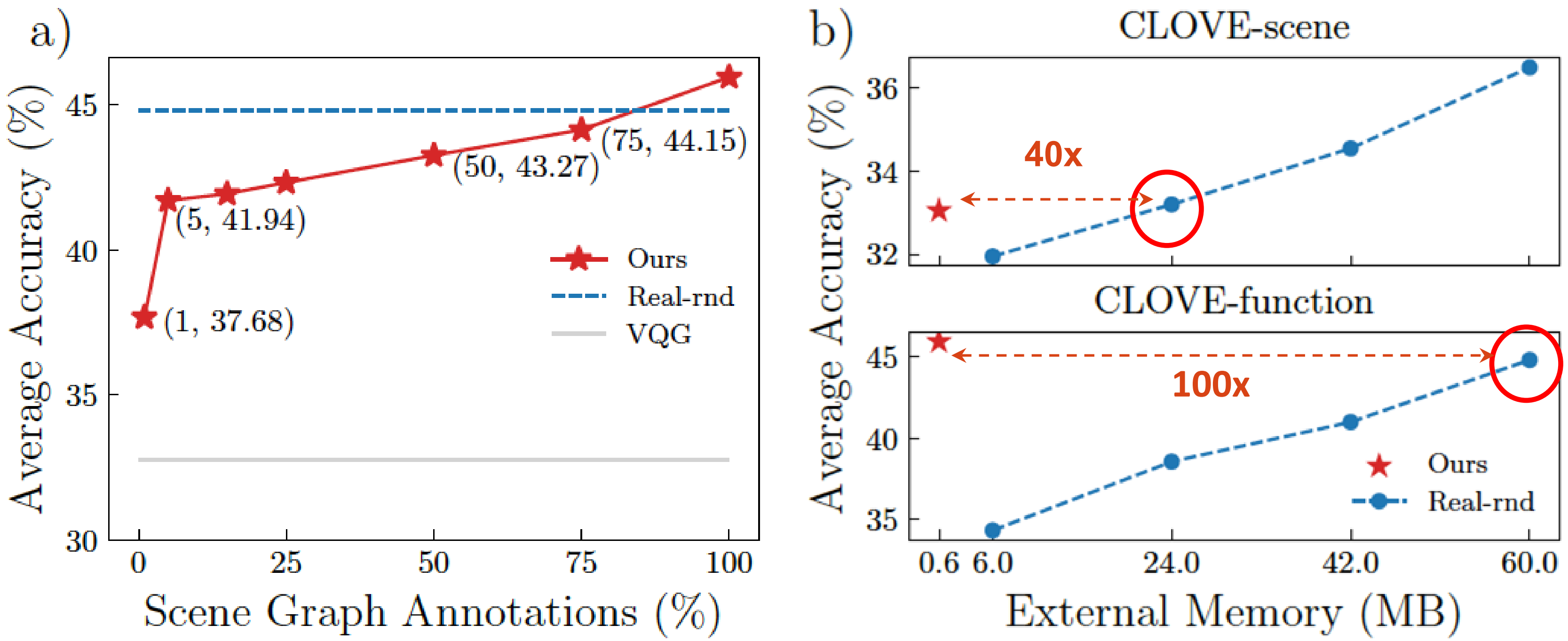
No.	Prompt type	Replay elements	CLOVE-Scene	CLOVE-Function
#1	Random	Q + A	29.52	40.24
#2	Random	SG + Q + A	32.08	44.21
#3	GT	SG + Q + A	35.09	47.01

- Replay scene graph can prevent forgetting of past knowledge (#1 and #2)
- Using better prompts is promising (#2 and #3)

Lei et al. "Symbolic replay: Scene graph as prompt for continual learning on vqa task." AAAI 2023.

Scene Graph as Prompt for symbolic replay

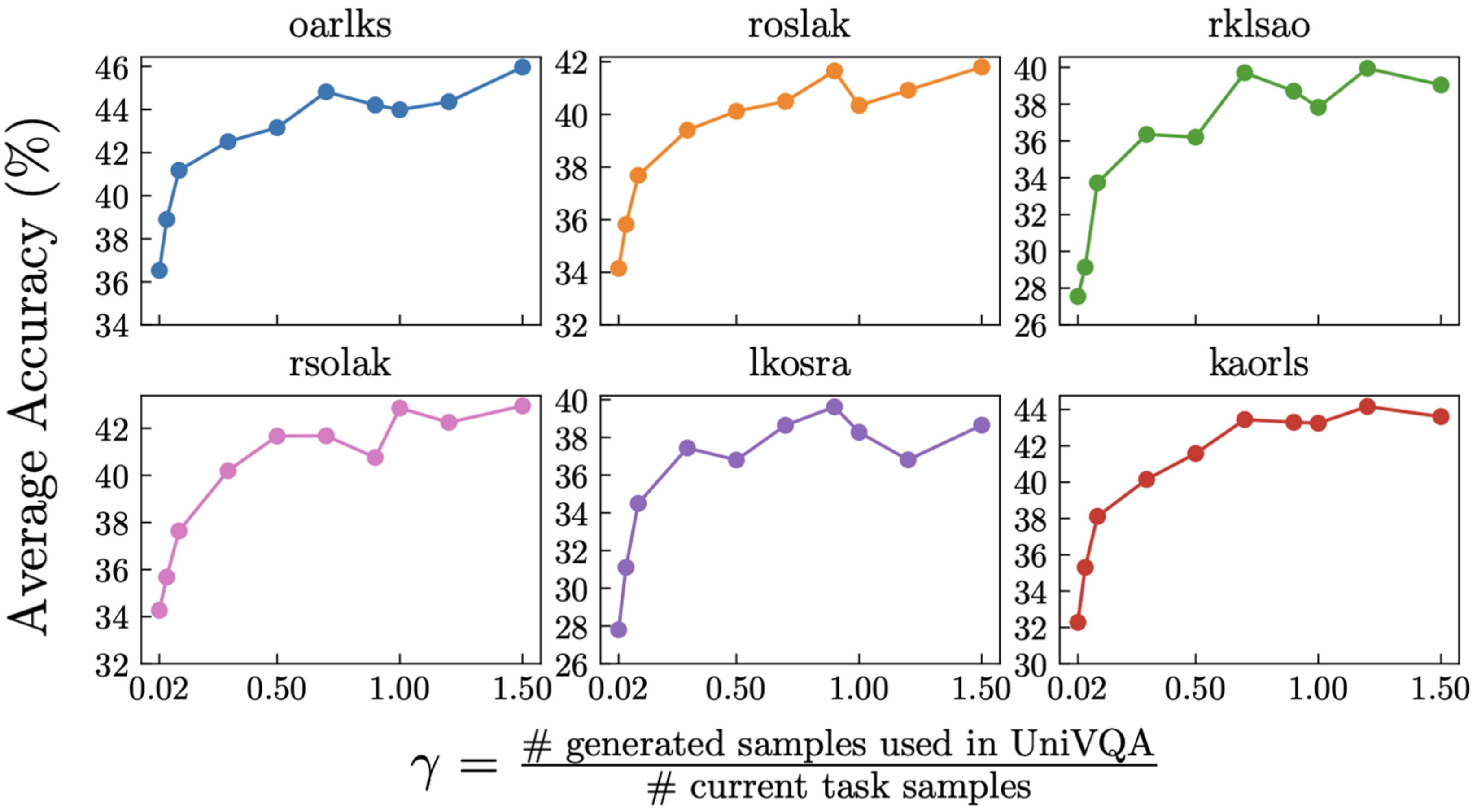
SGP is label-efficient and memory-efficient



Lei et al. "Symbolic replay: Scene graph as prompt for continual learning on vqa task." AAAI 2023.

Scene Graph as Prompt for symbolic replay

generated SG



Lei et al. "Symbolic replay: Scene graph as prompt for continual learning on vqa task." AAAI 2023.

Takeaways

Label efficiency

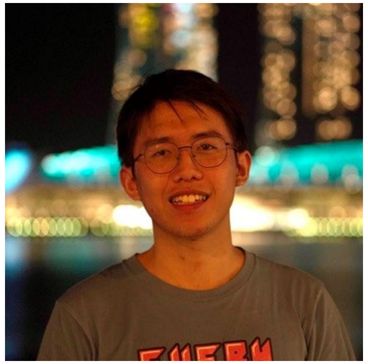
- LEOCOD: a new, challenging and important setting for real-world applications
- Efficient-CLS: a plug-and-play module that learns efficiently and effectively with less supervision and minimal forgetting

Memory efficiency

- CLOVE benchmark for continual learning in VQA
- Scene Graph as Prompt, a real-data-free replayed CL method

Wu et al. "Label-efficient online continual object detection in streaming video." ICCV 2023.
Lei et al. "Symbolic replay: Scene graph as prompt for continual learning on vqa task." AAAI 2023.

Acknowledgement



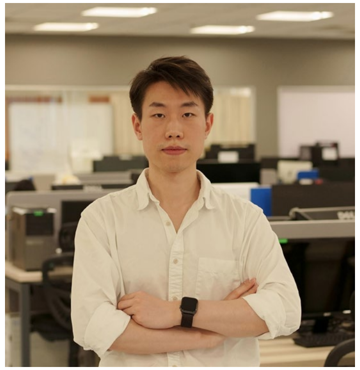
David Junhao Zhang

Stan Weixian Lei

Difei Gao

Wynne Hsu

Mengmi Zhang



Mike Shou

NUS National University of Singapore

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Agency for Science, Technology and Research SINGAPORE

Wu et al. "Label-efficient online continual object detection in streaming video." ICCV 2023.
Lei et al. "Symbolic replay: Scene graph as prompt for continual learning on vqa task." AAAI 2023.