NEURAL INFORMATION PROCESSING SYSTEMS

Agency for Science, Technology and Research SINGAPORE

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

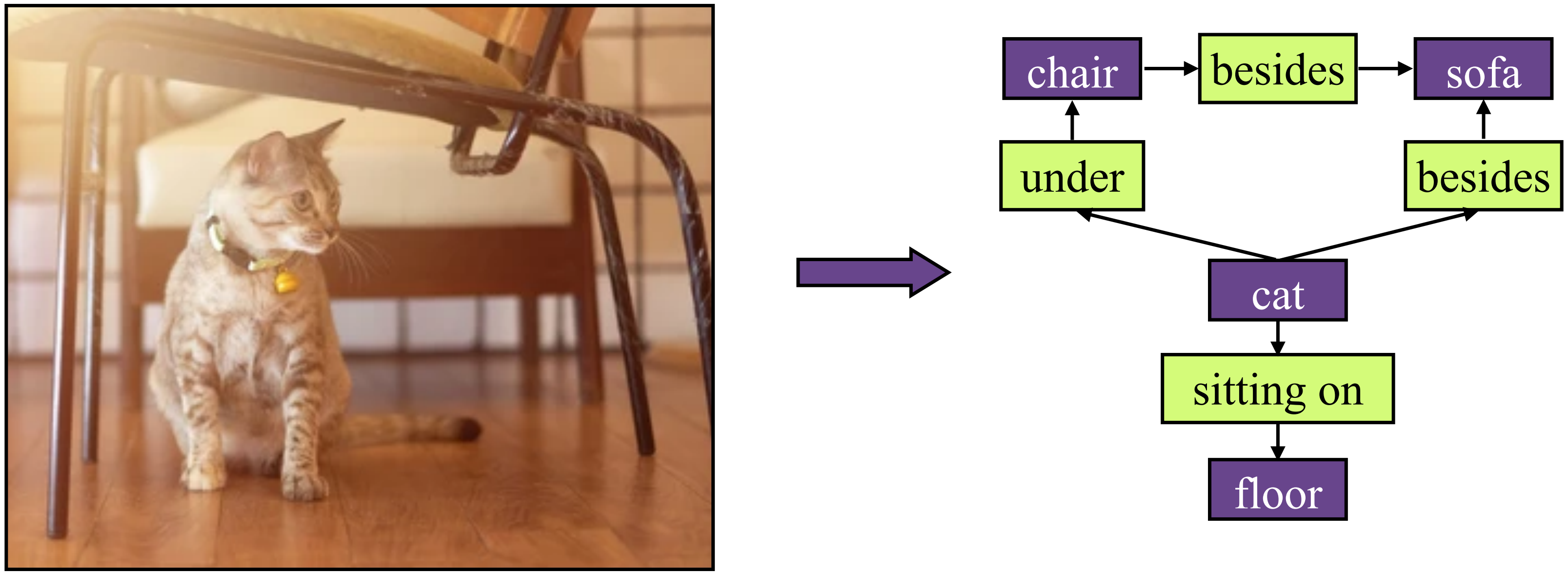
Adaptive Visual Scene Understanding: Incremental Scene Graph Generation

Naitik Khandelwal[1,2], Xiao Liu[1,2], Mengmi Zhang[1,2,*]

[1]College of Computing and Data Science, Nanyang Technological University, Singapore
[2]Deep NeuroCognition Lab, I2R and CFAR, Agency for Science, Technology and Research, Singapore

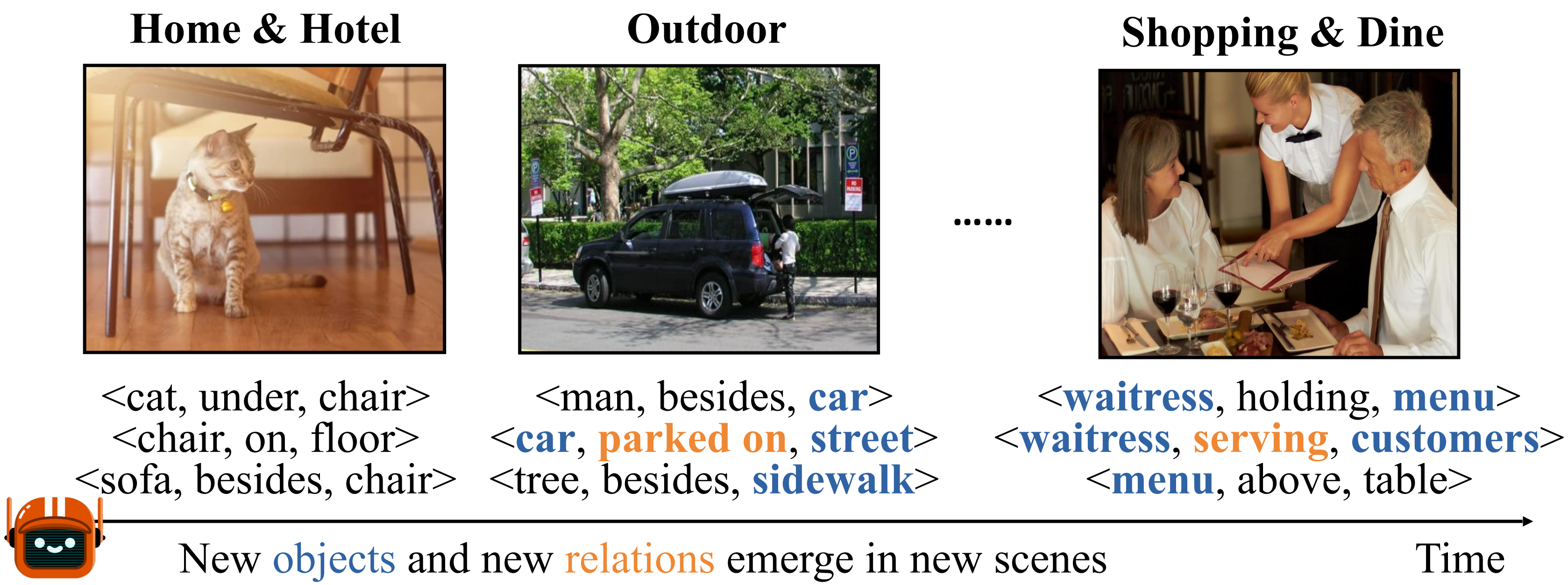
arxiv github

Introduction



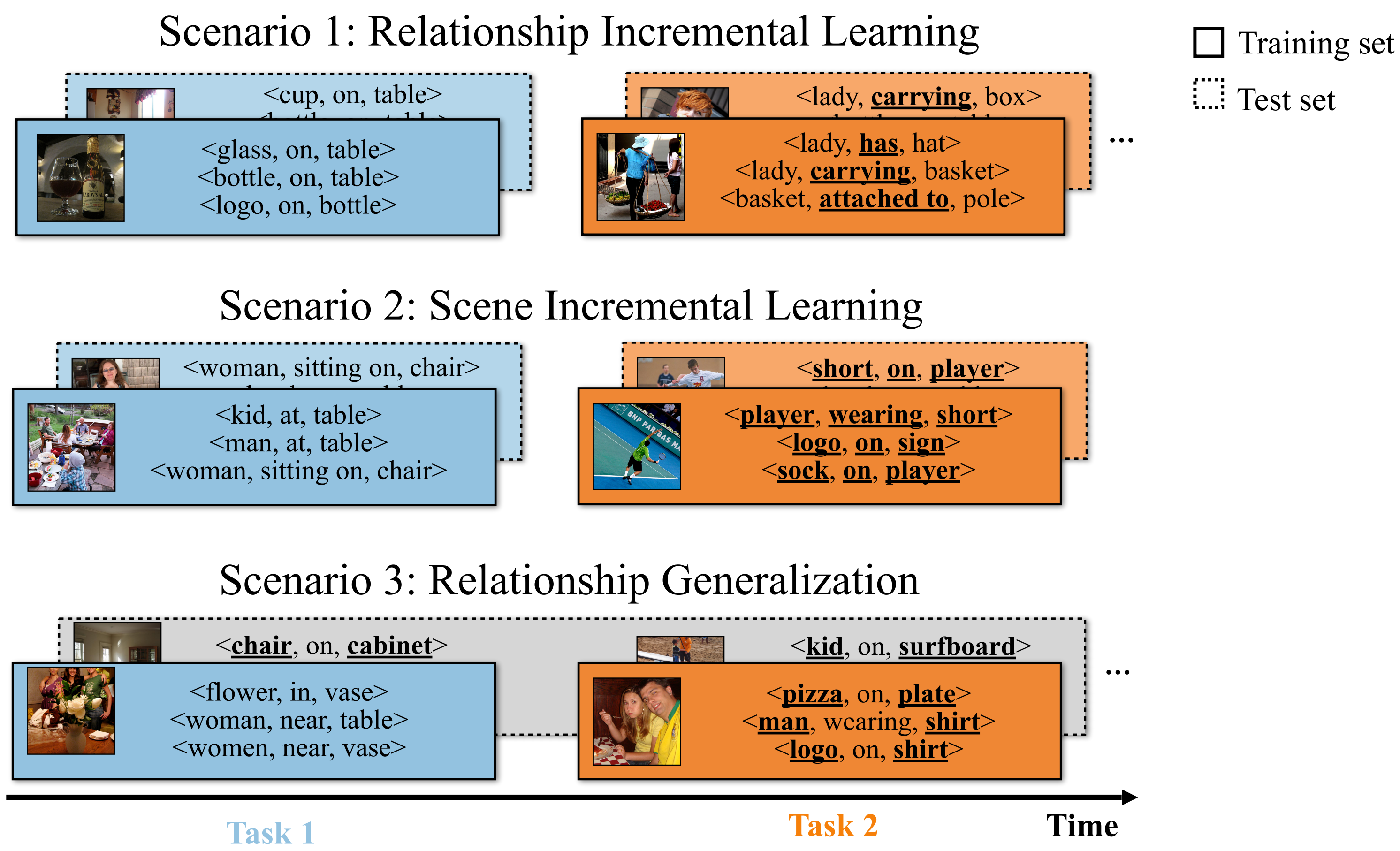
Scene **G**raph **G**eneration (**SGG**)

Introduction

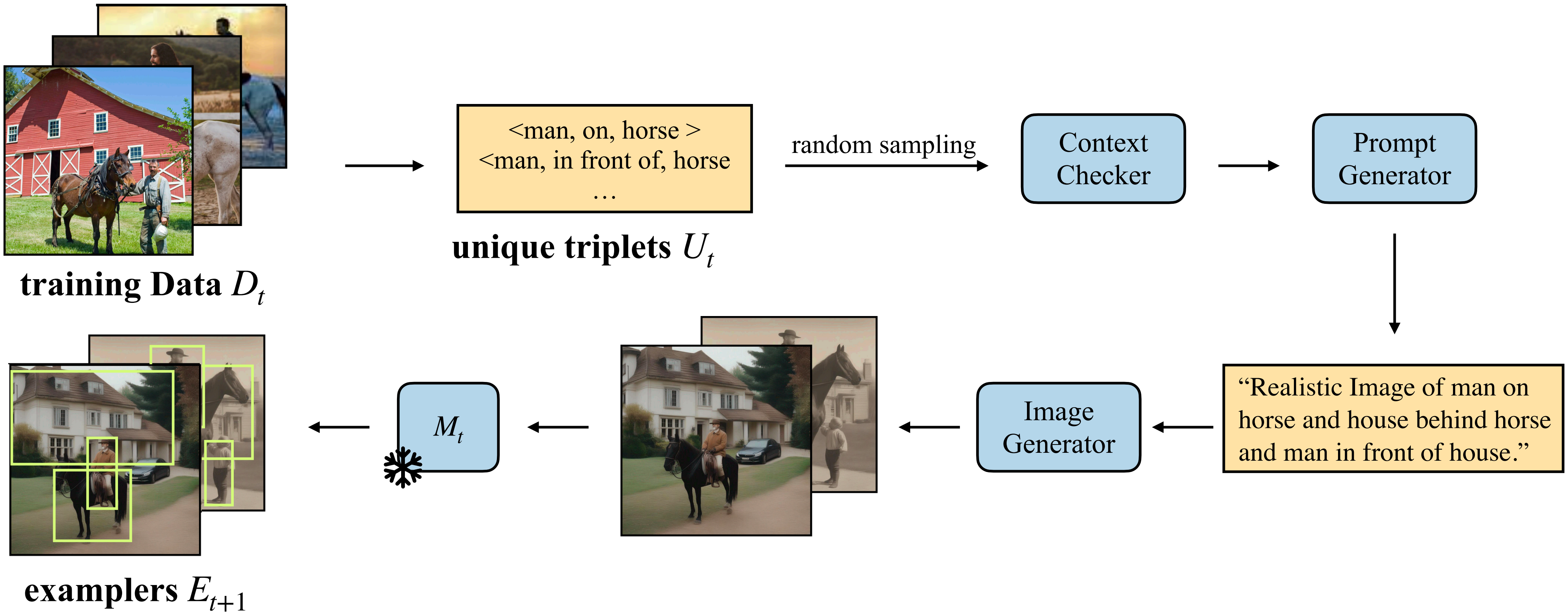


Continual **S**cene **G**raph **G**eneration (**CSEGG**)

CSEGG Benchmark



Replays via Analysis by Synthesis (RAS)



Main Results

SGTR[32]

Methods	Learning Scenario 1 (S1)						Learning Scenario 2 (S2)					
	Avg.R ↑	F↑	mR↑	mF↑	FWT↑	BWT↑	Avg.R↑	F↑	mR↑	mF↑	FWT↑	BWT↑
Joint	20.15	0	4.6	0	-	-	12.64	0	9.84	0	-	-
Replay@100%	16.17	-12.24	3.32	-1.34	-1.77	-11.72	4.56	-4.13	4.56	-5.61	-1.045	-30.25
Naive	1.33	-28.7	0.86	-1.74	-2.03	-60.67	0.51	-23.22	0.05	-11.31	-3.77	-62.34
EWC[24]	1.89	-28.4	0.96	-1.72	-1.17	-52.45	0	-23.22	0	-11.31	-2.65	-50.12
RAS_GT	5.78	-26.51	1.43	-1.54	-1.2	-44.27	0.98	-23.11	0.76	-10.86	-1.6	-43.25
PackNet[44]	7.19	-25.67	1.35	-1.64	-1.03	-42.35	1.67	-22.77	0.9	-10.33	-1.4	-42.45
Replay@10%	8.55	-22.21	4.33	-1.44	**4.29**	-38.35	1.81	-20.72	1.15	-9.64	-0.9	-40.67
Replay@20%	9.25	-20.35	4.78	-1.42	3.21	-31.98	2.57	-17.17	1.56	-8.07	-0.67	-38.27
Ours*	**10.78**	**-18.92**	**5.6**	**-1.39**	2.3	**-25.56**	**3.45**	**-10.23**	**2.75**	**-6.57**	**-0.54**	**-35.67**

TCNN[72]

Methods	Learning Scenario 1 (S1)						Learning Scenario 2 (S2)					
	Avg.R↑	F↑	mR↑	mF↑	FWT↑	BWT↑	Avg.R↑	F↑	mR↑	mF↑	FWT↑	BWT↑
Joint	19.53	0	3.9	0	-	-	4.3	0	3.7	0	-	-
Replay@100%	13.45	-8.83	3.6	-0.35	-1.5	-10.45	12.45	-4.13	3.2	-0.56	-2.1	-20.34
Naive	0.98	-21.2	0.74	-1.35	-3.45	-43.87	0	-18.22	0.45	-2.67	-4.12	-53.12
EWC[24]	2.36	-21.05	0.67	-1.34	-2.34	-39.89	0	-18.22	0.03	0	-3.77	-51.67
PackNet[44]	3.2	-19.7	1.1	-1.13	-1.3	-32.45	1.1	-17.82	0.84	-1.97	-2.84	-40.34
Replay@10%	5.67	-18.9	3.21	-1.05	**1.45**	-28.34	1.81	-16.72	1.03	-1.74	-1.4	-43.56
Replay@20%	6.23	-17.45	3.5	-1.01	1.01	-24.32	2.37	-15.17	1.45	-1.53	-1.1	-38.56
Ours*	**7.8**	**-15.67**	**3.9**	**-0.95**	0.5	**-19.83**	**4.67**	**-11.31**	**2.2**	**-0.89**	**-0.97**	**-29.65**

RAS outperforms all CSEGG baselines in S1 and S2

Main Results

Methods	SGTR[32] Learning Scenario 1 (S1)						Learning Scenario 2 (S2)					
	Avg.R ↑	F↑	mR↑	mF↑	FWT↑	BWT↑	Avg.R↑	F↑	mR↑	mF↑	FWT↑	BWT↑
Joint	20.15	0	4.6	0	-	-	12.64	0	9.84	0	-	-
Replay@100%	16.17	-12.24	3.32	-1.34	-1.77	-11.72	4.56	-4.13	4.56	-5.61	-1.045	-30.25
Naive	1.33	-28.7	0.86	-1.74	-2.03	-60.67	0.51	-23.22	0.05	-11.31	-3.77	-62.34
EWC[24]	1.89	-28.4	0.96	-1.72	-1.17	-52.45	0	-23.22	0	-11.31	-2.65	-50.12
RAS_GT	5.78	-26.51	1.43	-1.54	-1.2	-44.27	0.98	-23.11	0.76	-10.86	-1.6	-43.25
PackNet[44]	7.19	-25.67	1.35	-1.64	-1.03	-42.35	1.67	-22.77	0.9	-10.33	-1.4	-42.45
Replay@10%	8.55	-22.21	4.33	-1.44	**4.29**	-38.35	1.81	-20.72	1.15	-9.64	-0.9	-40.67
Replay@20%	9.25	-20.35	4.78	-1.42	3.21	-31.98	2.57	-17.17	1.56	-8.07	-0.67	-38.27
Ours*	**10.78**	**-18.92**	**5.6**	**-1.39**	2.3	**-25.56**	**3.45**	**-10.23**	**2.75**	**-6.57**	**-0.54**	**-35.67**

Methods	TCNN[72] Learning Scenario 1 (S1)						Learning Scenario 2 (S2)					
	Avg.R↑	F↑	mR↑	mF↑	FWT↑	BWT↑	Avg.R↑	F↑	mR↑	mF↑	FWT↑	BWT↑
Joint	19.53	0	3.9	0	-	-	4.3	0	3.7	0	-	-
Replay@100%	13.45	-8.83	3.6	-0.35	-1.5	-10.45	12.45	-4.13	3.2	-0.56	-2.1	-20.34
Naive	0.98	-21.2	0.74	-1.35	-3.45	-43.87	0	-18.22	0.45	-2.67	-4.12	-53.12
EWC[24]	2.36	-21.05	0.67	-1.34	-2.34	-39.89	0	-18.22	0.03	0	-3.77	-51.67
PackNet[44]	3.2	-19.7	1.1	-1.13	-1.3	-32.45	1.1	-17.82	0.84	-1.97	-2.84	-40.34
Replay@10%	5.67	-18.9	3.21	-1.05	**1.45**	-28.34	1.81	-16.72	1.03	-1.74	-1.4	-43.56
Replay@20%	6.23	-17.45	3.5	-1.01	1.01	-24.32	2.37	-15.17	1.45	-1.53	-1.1	-38.56
Ours*	**7.8**	**-15.67**	**3.9**	**-0.95**	0.5	**-19.83**	**4.67**	**-11.31**	**2.2**	**-0.89**	**-0.97**	**-29.65**

Decomposing scene graphs into smaller, diverse components with clear prompts is more effective than directly storing and using ground truth scene graphs for image generation

Main Results

SGTR[32]												
Methods	Learning Scenario 1 (S1)						Learning Scenario 2 (S2)					
	Avg.R ↑	F↑	mR↑	mF↑	FWT↑	BWT↑	Avg.R↑	F↑	mR↑	mF↑	FWT↑	BWT↑
Joint	20.15	0	4.6	0	-	-	12.64	0	9.84	0	-	-
Replay@100%	16.17	-12.24	3.32	-1.34	-1.77	-11.72	4.56	-4.13	4.56	-5.61	-1.045	-30.25
Naive	1.33	-28.7	0.86	-1.74	-2.03	-60.67	0.51	-23.22	0.05	-11.31	-3.77	-62.34
EWC[24]	1.89	-28.4	0.96	-1.72	-1.17	-52.45	0	-23.22	0	-11.31	-2.65	-50.12
RAS_GT	5.78	-26.51	1.43	-1.54	-1.2	-44.27	0.98	-23.11	0.76	-10.86	-1.6	-43.25
PackNet[44]	7.19	-25.67	1.35	-1.64	-1.03	-42.35	1.67	-22.77	0.9	-10.33	-1.4	-42.45
Replay@10%	8.55	-22.21	4.33	-1.44	**4.29**	-38.35	1.81	-20.72	1.15	-9.64	-0.9	-40.67
Replay@20%	9.25	-20.35	4.78	-1.42	3.21	-31.98	2.57	-17.17	1.56	-8.07	-0.67	-38.27
Ours*	**10.78**	**-18.92**	**5.6**	**-1.39**	2.3	**-25.56**	**3.45**	**-10.23**	**2.75**	**-6.57**	**-0.54**	**-35.67**

TCNN[72]												
Methods	Learning Scenario 1 (S1)						Learning Scenario 2 (S2)					
	Avg.R↑	F↑	mR↑	mF↑	FWT↑	BWT↑	Avg.R↑	F↑	mR↑	mF↑	FWT↑	BWT↑
Joint	19.53	0	3.9	0	-	-	4.3	0	3.7	0	-	-
Replay@100%	13.45	-8.83	3.6	-0.35	-1.5	-10.45	12.45	-4.13	3.2	-0.56	-2.1	-20.34
Naive	0.98	-21.2	0.74	-1.35	-3.45	-43.87	0	-18.22	0.45	-2.67	-4.12	-53.12
EWC[24]	2.36	-21.05	0.67	-1.34	-2.34	-39.89	0	-18.22	0.03	0	-3.77	-51.67
PackNet[44]	3.2	-19.7	1.1	-1.13	-1.3	-32.45	1.1	-17.82	0.84	-1.97	-2.84	-40.34
Replay@10%	5.67	-18.9	3.21	-1.05	**1.45**	-28.34	1.81	-16.72	1.03	-1.74	-1.4	-43.56
Replay@20%	6.23	-17.45	3.5	-1.01	1.01	-24.32	2.37	-15.17	1.45	-1.53	-1.1	-38.56
Ours*	**7.8**	**-15.67**	**3.9**	**-0.95**	0.5	**-19.83**	**4.67**	**-11.31**	**2.2**	**-0.89**	**-0.97**	**-29.65**

RAS outperforms Replay@20% (~2 Gb) while requiring less storage (~1.2 Gb)

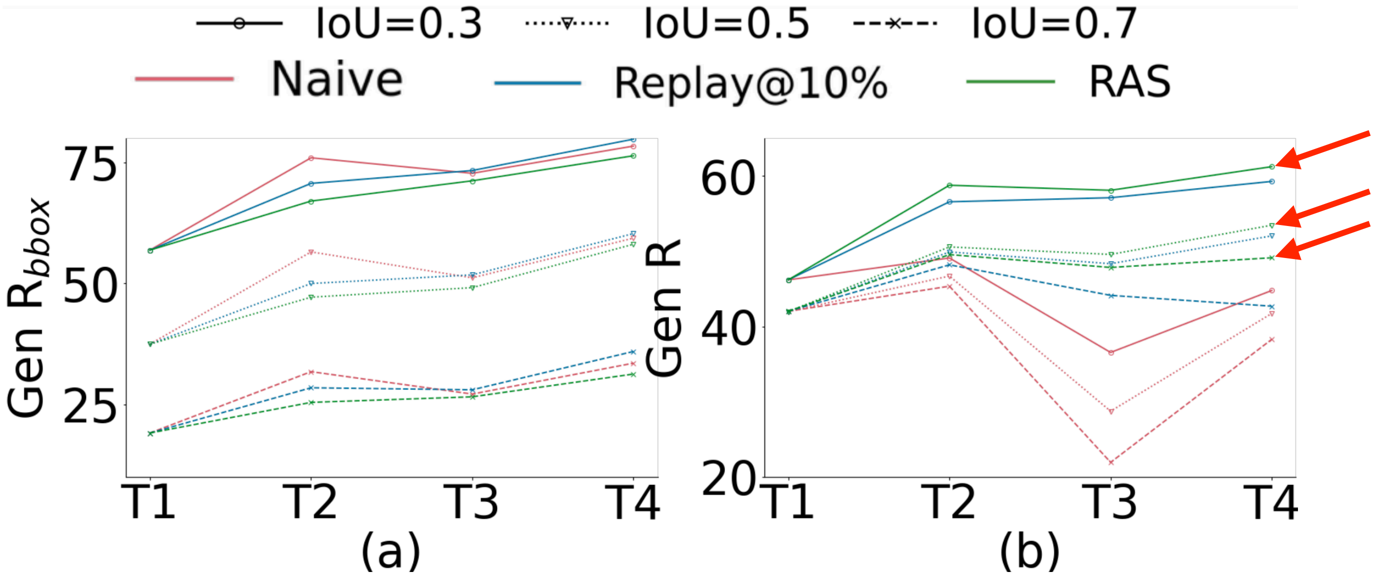
Main Results

SGTR[32]												
Methods	Learning Scenario 1 (S1)						Learning Scenario 2 (S2)					
	Avg.R ↑	F↑	mR↑	mF↑	FWT↑	BWT↑	Avg.R↑	F↑	mR↑	mF↑	FWT↑	BWT↑
Joint	20.15	0	4.6	0	-	-	12.64	0	9.84	0	-	-
Replay@100%	16.17	-12.24	3.32	-1.34	-1.77	-11.72	4.56	-4.13	4.56	-5.61	-1.045	-30.25
Naive	1.33	-28.7	0.86	-1.74	-2.03	-60.67	0.51	-23.22	0.05	-11.31	-3.77	-62.34
EWC[24]	1.89	-28.4	0.96	-1.72	-1.17	-52.45	0	-23.22	0	-11.31	-2.65	-50.12
RAS_GT	5.78	-26.51	1.43	-1.54	-1.2	-44.27	0.98	-23.11	0.76	-10.86	-1.6	-43.25
PackNet[44]	7.19	-25.67	1.35	-1.64	-1.03	-42.35	1.67	-22.77	0.9	-10.33	-1.4	-42.45
Replay@10%	8.55	-22.21	4.33	-1.44	**4.29**	-38.35	1.81	-20.72	1.15	-9.64	-0.9	-40.67
Replay@20%	9.25	-20.35	4.78	-1.42	3.21	-31.98	2.57	-17.17	1.56	-8.07	-0.67	-38.27
Ours*	**10.78**	**-18.92**	**5.6**	**-1.39**	2.3	**-25.56**	**3.45**	**-10.23**	**2.75**	**-6.57**	**-0.54**	**-35.67**

TCNN[72]												
Methods	Learning Scenario 1 (S1)						Learning Scenario 2 (S2)					
	Avg.R↑	F↑	mR↑	mF↑	FWT↑	BWT↑	Avg.R↑	F↑	mR↑	mF↑	FWT↑	BWT↑
Joint	19.53	0	3.9	0	-	-	4.3	0	3.7	0	-	-
Replay@100%	13.45	-8.83	3.6	-0.35	-1.5	-10.45	12.45	-4.13	3.2	-0.56	-2.1	-20.34
Naive	0.98	-21.2	0.74	-1.35	-3.45	-43.87	0	-18.22	0.45	-2.67	-4.12	-53.12
EWC[24]	2.36	-21.05	0.67	-1.34	-2.34	-39.89	0	-18.22	0.03	0	-3.77	-51.67
PackNet[44]	3.2	-19.7	1.1	-1.13	-1.3	-32.45	1.1	-17.82	0.84	-1.97	-2.84	-40.34
Replay@10%	5.67	-18.9	3.21	-1.05	**1.45**	-28.34	1.81	-16.72	1.03	-1.74	-1.4	-43.56
Replay@20%	6.23	-17.45	3.5	-1.01	1.01	-24.32	2.37	-15.17	1.45	-1.53	-1.1	-38.56
Ours*	**7.8**	**-15.67**	**3.9**	**-0.95**	0.5	**-19.83**	**4.67**	**-11.31**	**2.2**	**-0.89**	**-0.97**	**-29.65**

CSEGG models still find it challenging to learn new scenes incrementally

Main Results



RAS is more proficient in generalizing to classify relationships among unknown objects

Future Work

- ➢ Explore CSEGG on video-based datasets.
- ➢ Develop a synthetic CSEGG dataset to analyze continual learning under controlled conditions.
- ➢ Integrate a generative model with fine-grained control signals

We support Open Science!
Scan QR code for
Papers, code, data

PhD students and postdocs wanted! Join us!

Agency for Science, Technology and Research
SINGAPORE

AI SINGAPORE

NRF SINGAPORE