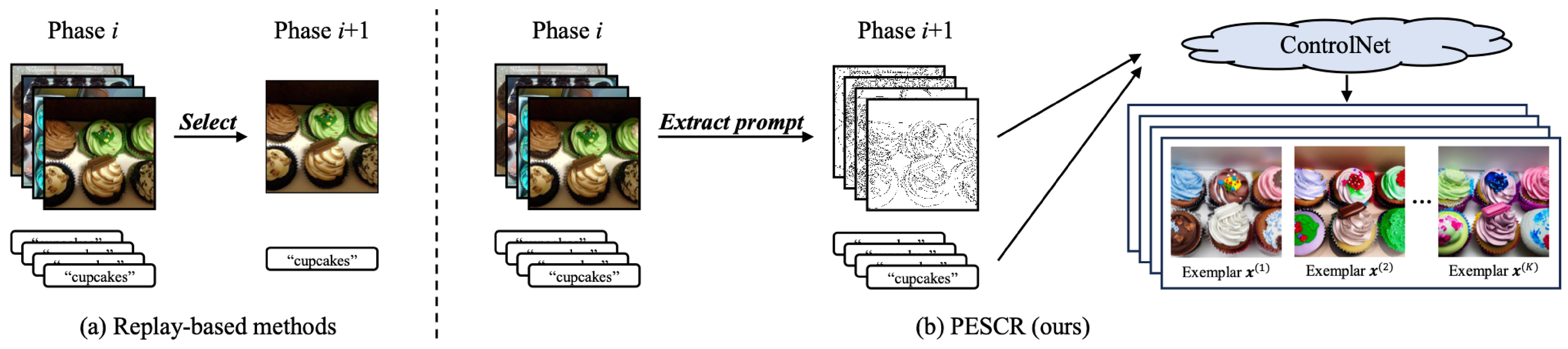
Prompt-Based Exemplar Super-Compression and Regeneration for Class-Incremental Learning

Ruxiao Duan[1], Jieneng Chen[2], Adam Kortylewski[3,4], Alan Yuille[2], Yaoyao Liu[5]

[1]Yale University, [2]Johns Hopkins University, [3]University of Freiburg,
[4]Max Planck Institute for Informatics, [5]University of Illinois Urbana-Champaign

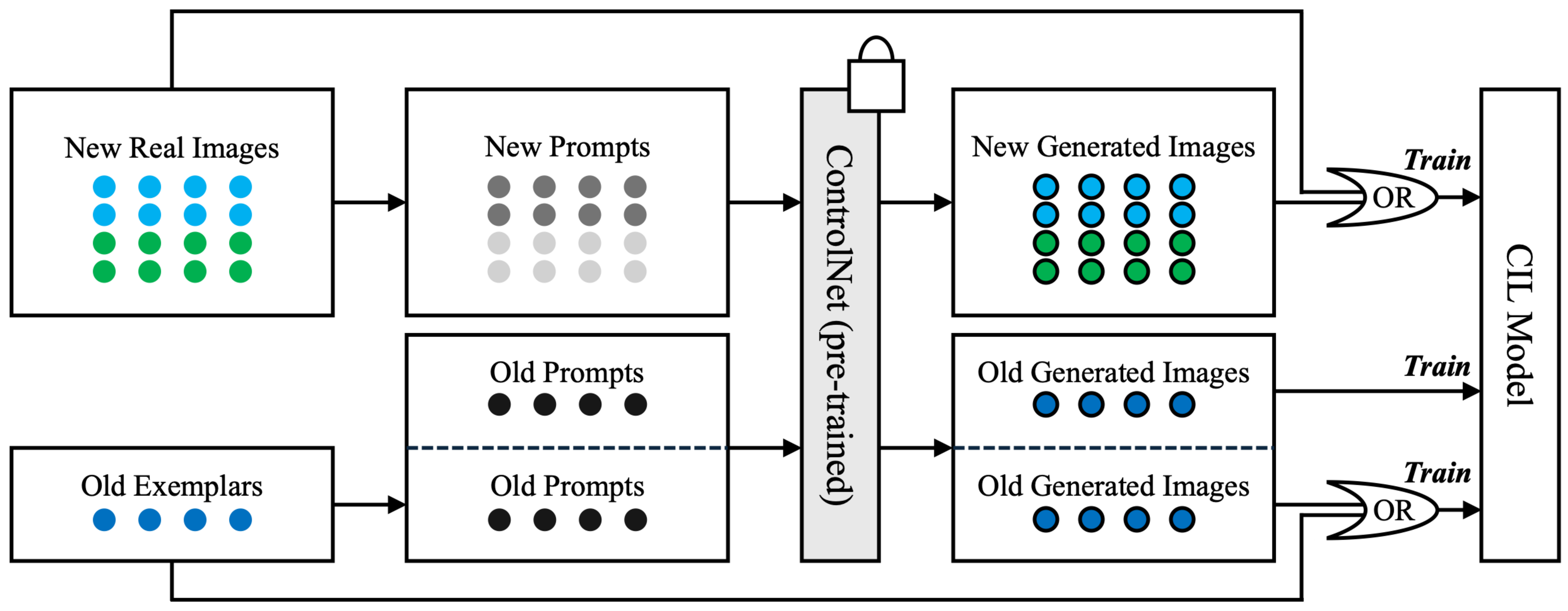
BMVC 2025

Introduction



- Replay-based class-incremental learning (CIL) stores previous samples as exemplars in the memory buffer for rehearsal in subsequent phases
- **Quantity** and **diversity** of exemplars are limited by memory constraint
- Storing samples as RGB images is inefficient, but compressing them into **visual and textual prompts** significantly reduces memory costs

Method: PESCR



- **Prompt-based Exemplar Super-Compression and Regeneration**
  - **Compression**: Exemplars are stored not as RGB images, but as visual prompts (edge maps) and textual prompts (class labels)
  - **Regeneration**: In subsequent phases, diverse RGB images can be regenerated from prompts via a pre-trained ControlNet
- **Boosted Exemplar Quantity and Diversity**
  - **Quantity**: The memory to store an 8-bit 3-channel RGB image can be used to store 24 edge maps instead
  - **Diversity**: Unlimited RGB images can be regenerated from an edge map by altering the random seed of ControlNet
- **Domain Gap Mitigation**
  - **Partial compression**: RGB images and edge maps are both saved in the memory buffer for rehearsal in subsequent stages
  - **Diffusion-based data augmentation**: Model is trained by a mixture of both real and synthetic images to learn synthetic image features

Experiment Results

- **Dataset Information**

Dataset	Caltech-256	Food-101	Places-100	ImageNet-100
# classes	256	101	100	100
Total/average # training images	21,436/84	75,750/750	300,000/3,000	128,856/1,289
Total/average # test images	5,472/21	25,250/250	10,000/100	5,000/50
Median image size ($h \times w$)	289 × 300	512 × 512	512 × 683	375 × 500

Four classification datasets used in the CIL experiments

- **PESCR Achieves State-of-the-art CIL Accuracy**

CIL Method	Learning from Scratch (LFS)				Learning from Half (LFH)							
	Caltech-256		ImageNet-100		Caltech-256		Food-101		Places-100		ImageNet-100	
	N=5	10	5	10	5	10	5	10	5	10	5	10
iCaRL [38]	57.7	48.9	66.2	58.7	53.4	49.1	58.6	50.0	42.2	37.6	59.2	50.3
WA [50]	66.2	54.6	76.2	69.7	60.1	46.8	74.2	64.6	62.0	57.3	73.6	66.2
MEMO [53]	65.7	61.4	78.5	74.0	62.6	60.1	71.1	50.1	53.4	48.8	72.5	70.6
PODNet [13]	67.0	60.9	76.4	68.7	68.4	66.6	79.5	77.0	68.3	66.3	79.4	77.2
FOSTER [44]	41.3	36.4	81.1	78.7	62.4	60.9	81.2	78.9	69.4	68.5	81.6	78.6
CIM [32]	65.5	66.3	82.0	78.0	64.1	65.5	79.5	77.1	71.1	70.5	80.5	79.5
DER [46]	68.1	64.8	81.8	78.5	68.4	66.8	82.0	80.4	70.4	69.5	81.8	80.2
PESCR (ours)	**72.8**	**69.8**	**83.4**	**81.3**	**72.1**	**71.3**	**83.9**	**83.2**	**72.3**	**71.8**	**84.2**	**83.4**

Average N-phase classification accuracy (%) of different CIL methods

- **PESCR Enhances All CIL Methods**

	iCaRL	WA	MEMO	PODNet	FOSTER	DER
Baseline	50.3	66.2	70.6	77.2	78.6	80.2
PESCR	66.4	71.7	77.6	80.6	79.4	83.4
Improvements	+16.1	+5.5	+7.0	+3.4	+0.8	+3.2

Average CIL accuracy (%) on ImageNet-100 of CIL methods without and with PESCR

- **Quantity vs. Quality: More Exemplars, Larger Domain Gap**

p	20+0	19+24	18+48	17+72	16+96	15+120	14+144	13+168	12+192
0.0	81.8	82.4	82.4	82.5	82.8	82.6	81.6	81.4	80.8
0.1	82.4	82.9	83.4	83.4	83.5	83.4	83.8	83.4	83.5
0.2	82.8	83.0	83.3	83.6	83.9	**84.2**	83.7	83.4	83.9
0.3	82.5	83.4	83.6	84.1	83.8	84.0	**84.2**	84.0	84.0
0.4	82.9	83.1	83.5	84.0	83.7	83.9	83.8	84.0	83.6
0.5	82.4	82.2	83.1	82.8	83.5	83.5	83.4	83.7	83.4

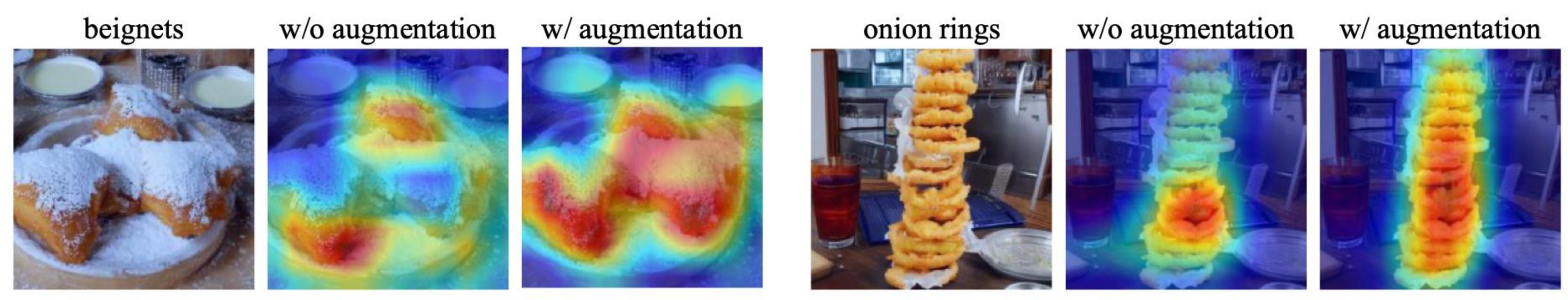
Average CIL accuracy (%) of PESCR on ImageNet-100, with real+synthetic exemplars per class and a diffusion-based data augmentation probability p

- **Performance Saturation with Sufficient Exemplars**

K	0	1	5	10	15	20	25
Accuracy	81.8	83.1	83.7	83.4	**83.8**	83.7	83.4

Average CIL accuracy (%) on ImageNet-100 with K generated images per edge map

- **Effect of Diffusion-based Data Augmentation on Image Recognition**



Generated exemplars with Grad-CAM given by the classification model trained without and with diffusion-based data augmentation

- **Consistent Improvement with Various Budget Settings**

	Food-101			Places-100			ImageNet-100		
	b=5	10	20	5	10	20	5	10	20
DER	77.3	79.1	80.4	67.1	68.6	69.5	77.9	79.3	80.2
DER+PESCR	80.1	81.3	83.2	70.4	70.9	71.8	81.1	81.4	83.4
Improvements	+2.8	+2.2	+2.8	+3.3	+2.3	+2.3	+3.2	+2.1	+3.2

Average CIL accuracy (%) of PESCR, with b memory units per class

Conclusion

- Traditional CIL faces bottlenecks on exemplar **quantity** and **diversity**
- By storing samples as edge maps instead of RGB images, 24 times more exemplars can be saved under the same memory budget
- By regenerating images with different random seeds via ControlNet, diverse exemplars can be produced for rehearsal
- PESCR consistently enhances the performance of all replay-based CIL methods by boosting their exemplar quantity and diversity