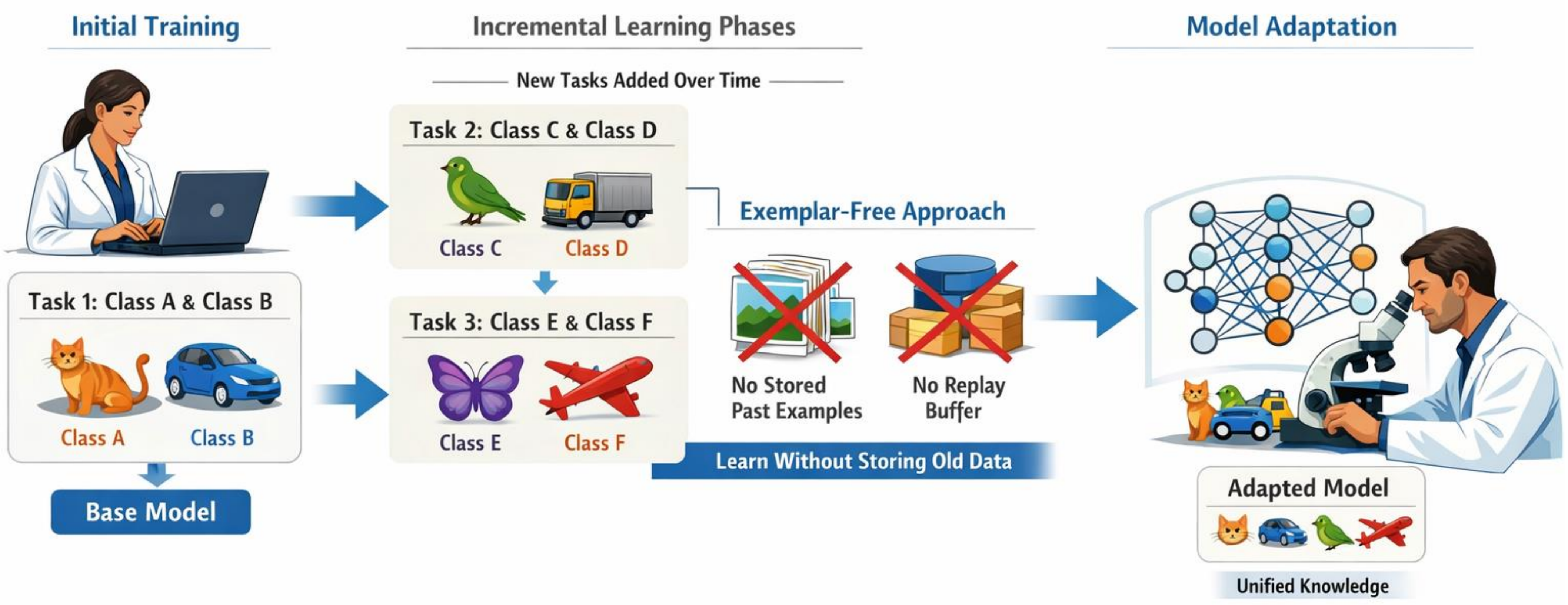
Two-Way Is Better Than One: Bidirectional Alignment with Cycle Consistency for Exemplar-Free Class-Incremental Learning

Hongye Xu
Bartosz Krawczyk

Center for Imaging Science
Rochester Institute of Technology
Rochester NY, USA

Exemplar-Free Class Incremental Learning (EFCIL)

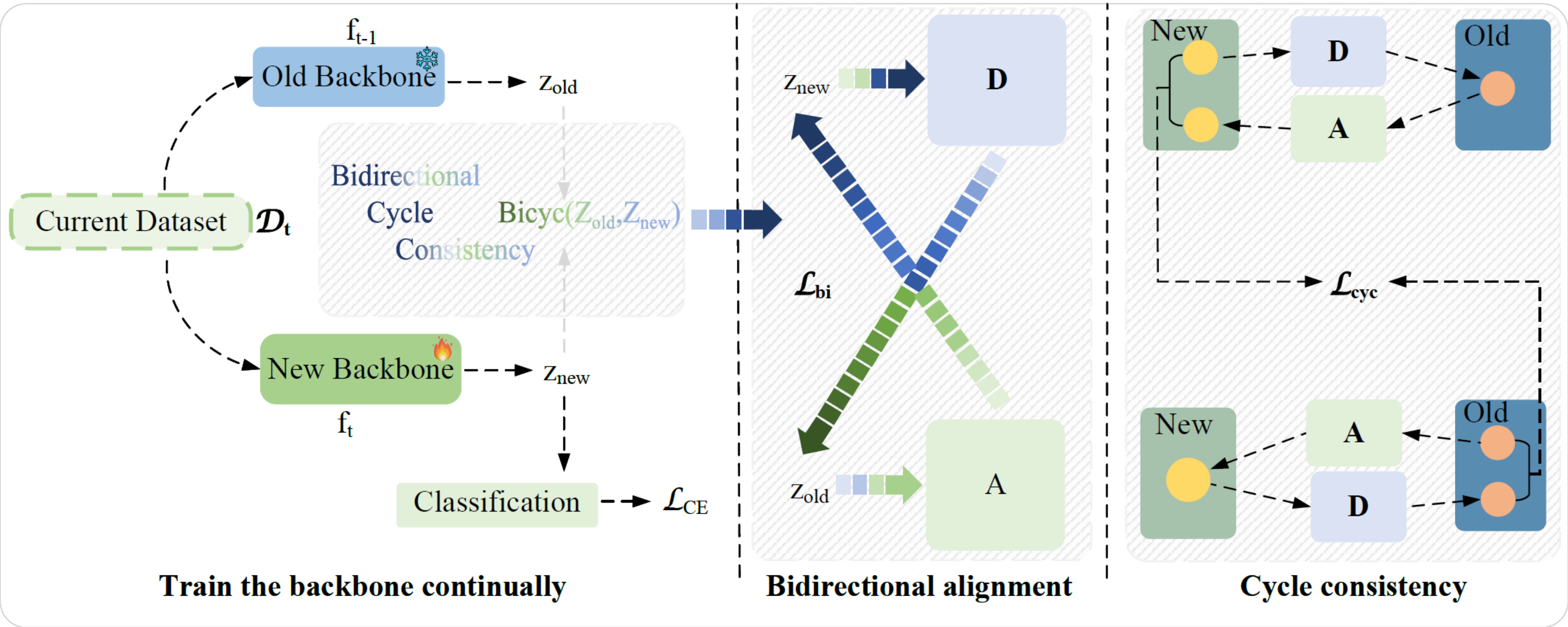


Limitations of current EFCIL approaches

- Current EFCIL is typically two-stage:
  - Stage I: train the new backbone on current-task data with cross-entropy plus distillation/regularization from the frozen old backbone
  - Stage II: after training, freeze both backbones and learn a separate adapter to map old features/prototypes into the new feature space
- Main limitation of the current two-stage recipe:
  - **prototype transport is learned only post hoc**, after representation learning is already finished
  - this creates **mismatch between representation update and prototype transport**
  - **residual alignment errors and cycle inconsistencies can accumulate across tasks**

- What we propose instead:
  - make the forward and backward mappings explicit during training
  - jointly learn:
    - forward adapter A: old -> new
    - backward distiller D: new -> old
  - optimize them together with the backbone during the main task training
- **Why this is better:**
  - **transport and representation co-evolve** instead of being decoupled
  - **old-class geometry is preserved more faithfully**
  - **prototype transport becomes more accurate** by design
  - **reduces forgetting** without sacrificing plasticity

Proposed solution



Overview:

(1) Train: the current backbone learns on new task, while frozen old backbone provides old embeddings with task loss

(2) Bidirectional alignment: jointly learn a distiller and an adapter

(3) Cycle consistency: obtain near-bijective, geometry-preserving transport. Old Gaussian prototypes are mapped forward, and all classes are evaluated in the new space.

Incremental and last-task average accuracy

Method	CIFAR-100				TinyImageNet			
	T=10		T=20		T=10		T=20	
	A_{last}	A_{inc}	A_{last}	A_{inc}	A_{last}	A_{inc}	A_{last}	A_{inc}
EWC	30.9±1.9	50.4±1.7	17.0±1.6	34.2±2.1	18.5±1.8	34.3±2.3	11.3±1.9	26.8±2.5
LwF$_{\text{ECCV16}}$	31.9±1.1	51.8±1.5	17.6±1.2	39.2±1.7	27.1±1.5	39.6±2.0	15.2±1.6	31.5±2.1
SDC$_{\text{CVPR20}}$	40.6±0.9	56.2±1.3	32.3±1.0	46.6±1.4	29.5±1.1	43.8±1.5	26.3±1.2	40.6±1.7
PASS$_{\text{CVPR21}}$	30.8±1.2	48.3±1.1	17.6±0.8	31.1±1.3	24.5±0.6	39.5±1.0	18.5±1.4	30.4±1.9
FeTrIL$_{\text{WACV23}}$	34.9±0.5	51.2±1.1	23.3±1.4	37.9±1.2	31.0±0.9	45.3±1.8	25.9±1.2	39.9±1.2
FeCAM$_{\text{NeurIPS23}}$	32.4±0.5	48.7±0.9	21.1±1.0	34.5±1.3	30.9±0.9	44.9±1.4	24.9±0.8	37.9±1.4
EFC$_{\text{ICLR24}}$	43.5±0.8	58.1±1.2	32.4±0.9	47.0±1.3	34.5±1.1	47.9±1.5	28.4±1.2	42.1±1.6
ADC$_{\text{CVPR24}}$	46.5±1.2	61.4±1.6	35.1±1.4	51.7±1.8	32.3±1.5	43.0±1.9	18.1±1.6	36.0±2.1
LDC$_{\text{ECCV24}}$	45.4±1.6	59.5±1.9	35.5±1.9	51.9±2.3	34.2±1.1	46.8±1.6	24.9±2.2	38.2±2.7
AdaGauss$_{\text{NeurIPS24}}$	46.8±1.2	60.9±1.0	37.9±1.0	54.4±0.8	32.9±0.9	45.8±1.3	27.5±1.2	39.5±1.1
DPCR$_{\text{ICML2025}}$	50.2±0.7	62.8±1.1	39.8±1.2	54.8±0.9	33.9±1.8	46.9±0.9	25.6±0.7	39.3±0.6
BiCyc (Ours)	**50.6±0.9**	**63.2±1.3**	**41.5±1.1**	**56.5±1.3**	**35.4±0.8**	**49.1±1.4**	**30.2±1.1**	**44.2±1.3**

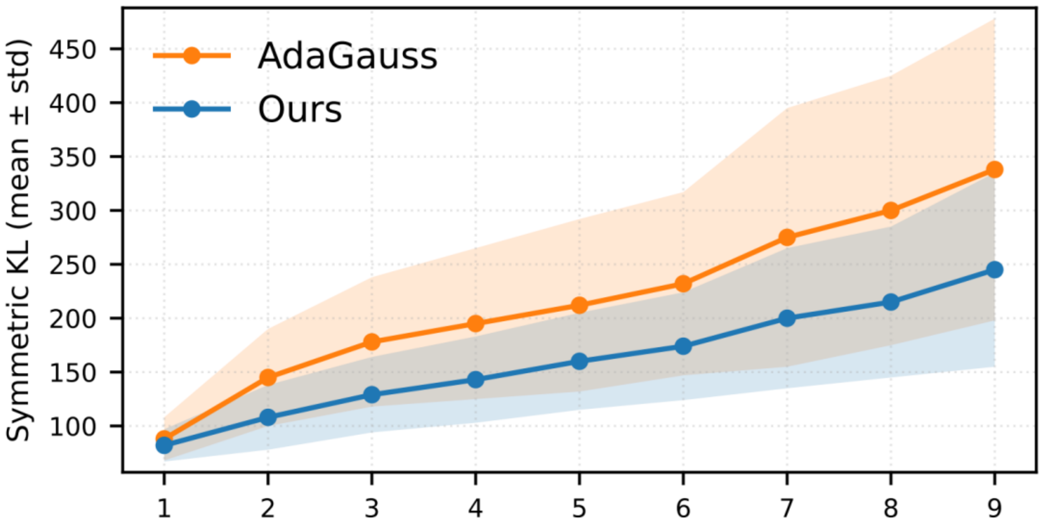
Incremental and last-task average accuracy

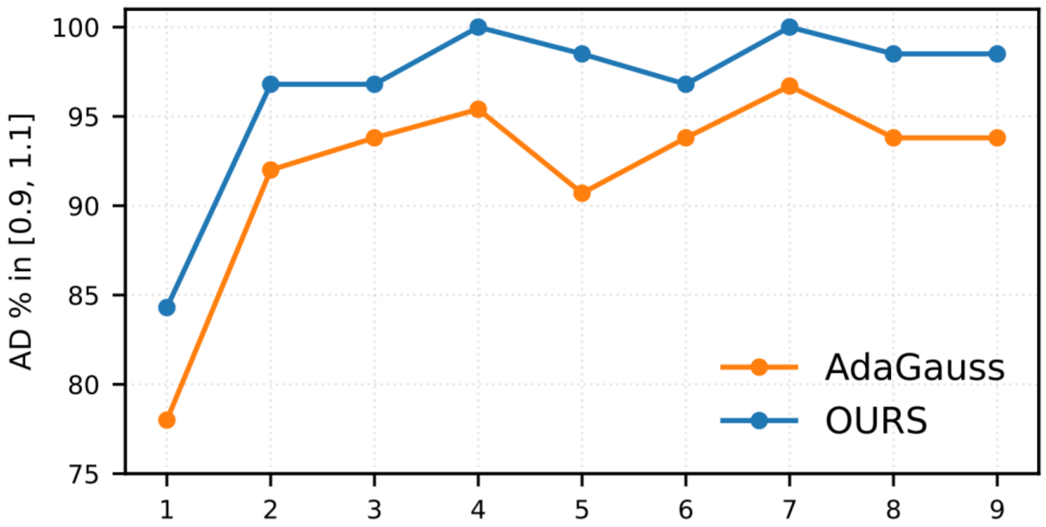
Method	ImageNet-100				CUB-200			
	T=10		T=20		T=10		T=20	
	A_{last}	A_{inc}	A_{last}	A_{inc}	A_{last}	A_{inc}	A_{last}	A_{inc}
EWC	25.1±2.8	40.6±3.3	13.7±2.1	29.2±2.5	15.8±0.7	32.6±0.5	12.3±0.8	27.2±0.6
LwF$_{\text{ECCV16}}$	33.4±2.2	51.5±1.6	18.6±1.6	41.3±1.9	30.4±1.1	46.1±1.0	19.4±1.6	34.7±1.8
SDC$_{\text{CVPR20}}$	35.4±1.9	50.1±1.6	19.4±1.0	36.5±1.4	50.3±1.3	60.5±1.2	27.9±1.4	40.1±1.6
PASS$_{\text{CVPR21}}$	26.4±1.3	45.7±0.2	14.4±1.2	31.7±0.4	27.0±0.9	42.3±0.9	18.1±1.2	36.9±1.1
FeTrIL$_{\text{WACV23}}$	36.2±1.2	52.6±0.6	26.6±1.5	42.4±2.1	36.9±0.7	48.2±0.6	34.6±1.0	45.3±0.9
FeCAM$_{\text{NeurIPS23}}$	38.7±1.0	54.8±0.5	29.0±1.3	44.6±2.0	40.2±0.8	54.9±1.0	36.2±1.1	48.9±1.3
EFC$_{\text{ICLR24}}$	50.9±1.1	61.3±1.2	38.6±1.2	50.5±1.5	51.0±0.6	63.3±0.7	**46.1±1.0**	**59.3±1.3**
ADC$_{\text{CVPR24}}$	38.3±1.2	55.5±1.5	25.1±1.3	43.4±1.7	49.5±0.9	58.8±1.1	35.4±1.4	48.3±1.4
LDC$_{\text{ECCV24}}$	51.4†±1.2†	**69.4†±0.6†**	28.5±1.7	46.5±2.7	47.5±0.7	55.7±1.3	27.2±1.1	39.8±2.1
AdaGauss$_{\text{NeurIPS24}}$	51.1±1.2	65.0±1.4	42.6±1.6	57.4±1.9	52.9±0.8	63.4±1.3	45.0±1.3	57.0±1.0
DPCR$_{\text{ICML2025}}$	49.9±0.8	64.8±1.1	37.3±1.6	54.7±0.7	–	–	–	–
BiCyc (Ours)	**52.7±0.9**	66.8±1.4	**43.8±1.4**	**58.2±1.8**	**53.7±0.7**	**64.0±0.8**	43.7±1.4	55.9±1.2

Last-task average forgetting

Method	CIFAR-100		TinyImageNet		ImageNet-100		CUB-200	
	T=10	T=20	T=10	T=20	T=10	T=20	T=10	T=20
	F_{last}	F_{last}	F_{last}	F_{last}	F_{last}	F_{last}	F_{last}	F_{last}
$\text{LwF}_{\text{ECCV16}}$	23.2±1.7	31.2±1.8	21.9±1.9	33.5±2.4	42.1±2.3	48.1±2.2	16.5±1.1	21.7±1.4
$\text{SDC}_{\text{CVPR20}}$	34.8±1.7	35.9±1.9	25.1±1.4	29.4±2.1	44.6±2.0	54.4±2.3	10.9±1.3	17.3±1.1
$\text{EFC}_{\text{ICLR24}}$	23.1±1.1	24.7±1.8	23.5±2.4	30.1±3.0	21.5±1.9	23.8±2.5	**10.7±0.7**	**14.8±1.7**
$\text{ADC}_{\text{CVPR24}}$	21.9±1.1	31.0±1.6	30.2±2.0	36.8±1.9	32.4±1.6	33.4±1.8	12.8±1.1	21.3±1.5
$\text{LDC}_{\text{ECCV24}}$	21.7±1.9	25.6±2.3	24.7±2.5	30.7±2.1	25.7±1.7	32.9±2.3	13.6±1.2	23.9±1.8
$\text{AdaGauss}_{\text{NeurIPS24}}$	16.7±1.4	21.0±1.5	18.7±1.2	23.1±1.0	20.6±0.9	22.9±1.1	11.6±0.7	16.9±1.3
BiCyc (Ours)	**13.5±1.3**	**16.6±0.9**	**12.0±0.9**	**18.9±1.1**	**18.2±1.6**	**20.8±1.4**	11.3±0.9	17.5±1.3

Improved transport of old prototypes

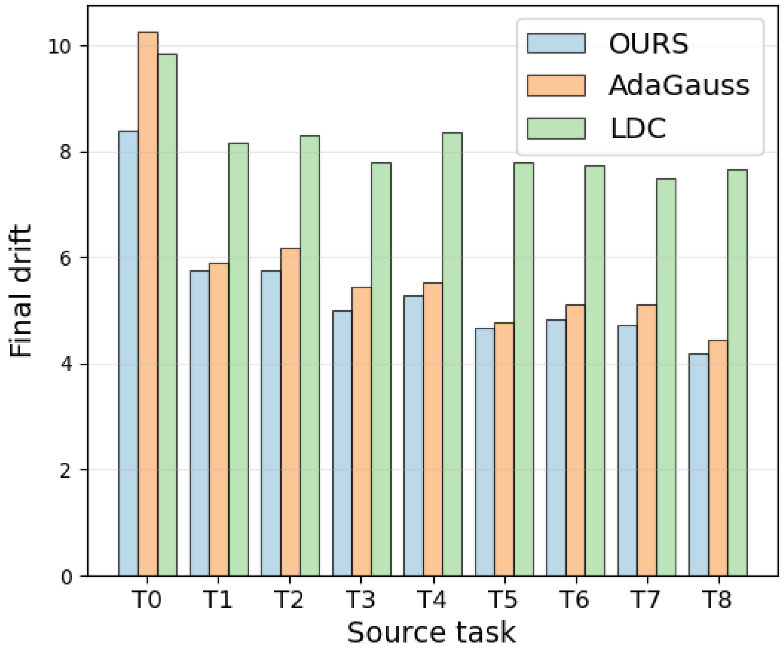


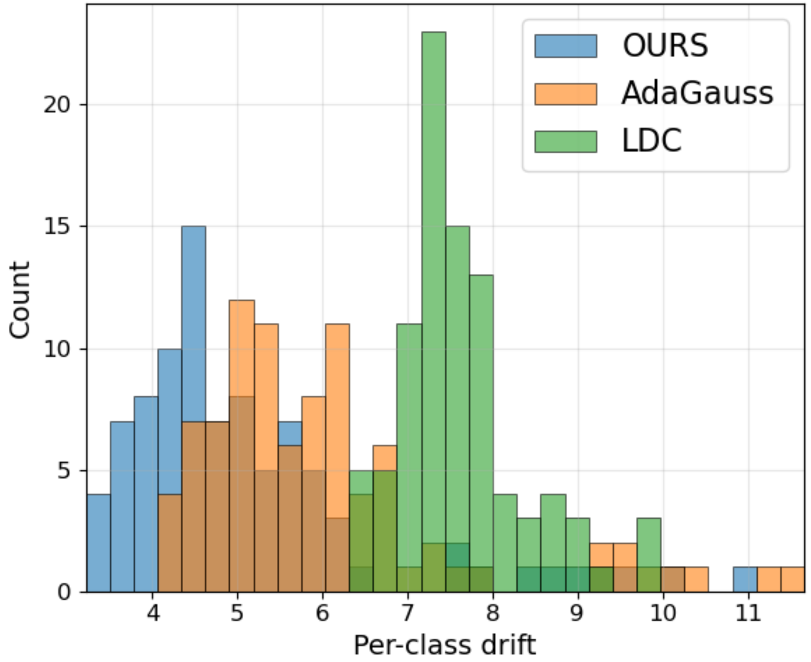


(left) On the fixed task-0 data, we compare Gaussian fits from models after t= 1,…,9 to the task-0 reference using symmetric KL. Our method **maintains a smaller divergence**—i.e., a **closer match to the original distribution**.

(right) Near-isometry on task-0 under continual updates. AD-% in[0.9, 1.1] for models after t= 1,…,9; our method **consistently preserves geometry** better than SOTA.

Reduced prototype drift





Drift between maintained prototypes and oracle prototypes (empirical class means)after completing task 9. For each of the 90 old classes(Tasks 0–8), we measure the ℓ2 distance in feature space between the maintained prototype and its oracle prototype.

(left) **Per-source-task average drift** for the three methods.

(right) **Histogram of per-class drift** over all old classes.

Find out more!

Thank you!