CODE-CL: Conceptor-Based Gradient Projection for Deep Continual Learning

Marco Apolinario, Sakshi Choudhary, Kaushik Roy

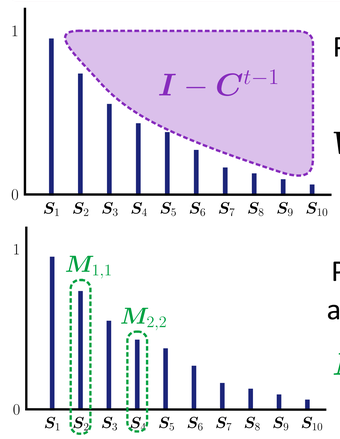
PURDUE UNIVERSITY

ICCV OCT 19-23, 2025 HONOLULU HAWAII

Motivation

- Deep networks forget past tasks when trained sequentially. Orthogonal-projection methods prevent interference but often block **forward transfer** on correlated tasks. CODE-CL balances *stability* and *plasticity*.

Method Overview



Prevent forgetting by projecting gradients onto $I - C^{t-1}$

$$W := W - \eta \nabla_W \mathcal{L}(I - C^{t-1})$$

Promote FWT by learning a linear combination of U^*

$$M := M - \eta \nabla_M \mathcal{L}$$

Pseudo-orthogonal updates avoid interference ($I - C_{t-1}$) while a *learned combination* of shared directions U^* (via M) promotes transfer.

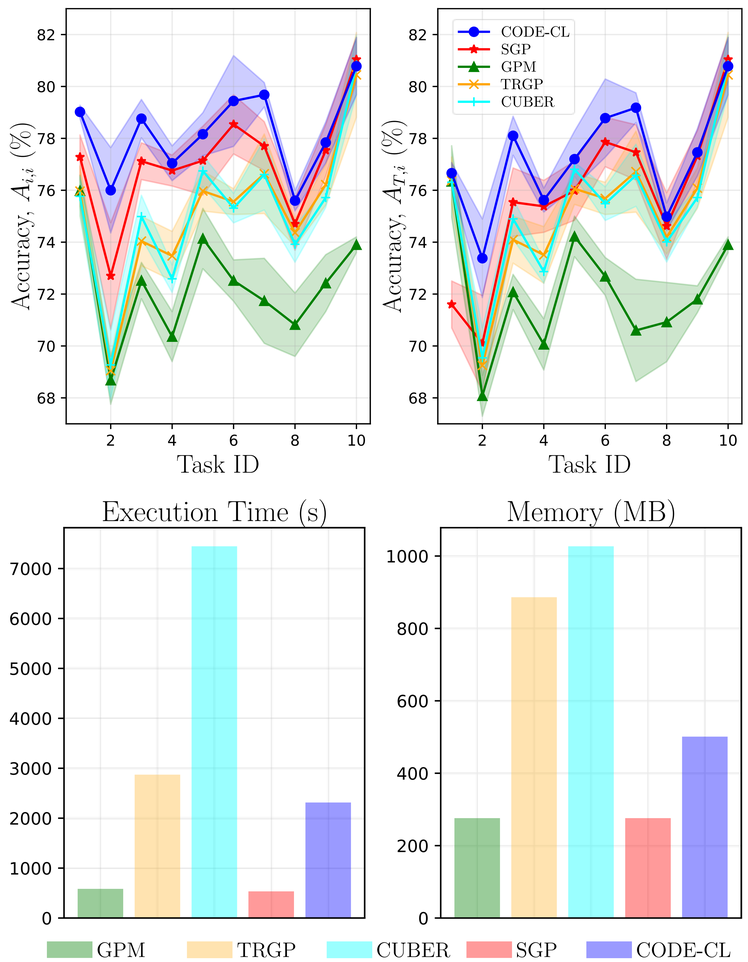
Key Ideas

- Encode each layer's input feature space with a **conceptor** $C = USU^\top$.
- Constrain gradients to the **pseudo-orthogonal** subspace: $\nabla W \leftarrow \nabla W(I - C^{t-1})$
- When tasks overlap, **free** top-K common directions U^* and learn a **linear combination** M for improved forward transfer (FWT).

Experiments

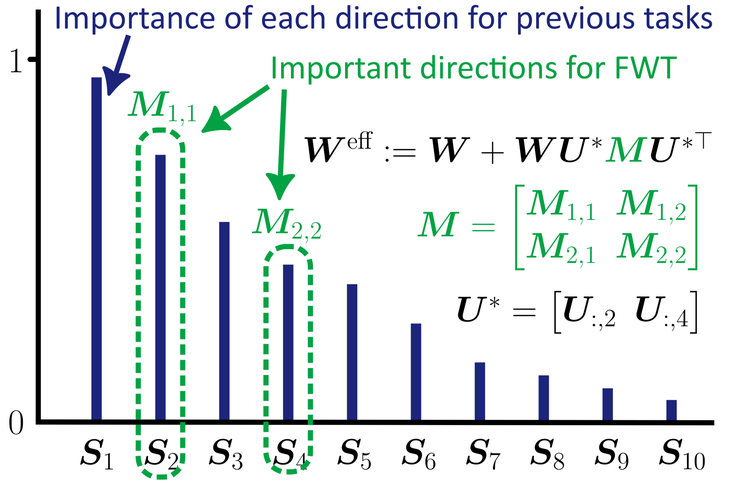
CODE-CL achieves the best trade-off between accuracy and forward transfer, especially on correlated tasks (Split CIFAR-100, miniImageNet).

Method	Split CIFAR100		Split MiniImageNet	
	ACC (%)↑	FWT (%)↑	ACC (%)↑	FWT (%)↑
GPM	72.06 ± 0.29	0	66.26 ± 1.18	0
TRGP	75.24 ± 0.29	2.86 ± 0.26	65.08 ± 0.94	-1.56 ± 0.67
CUBER	75.30 ± 0.43	2.86 ± 0.49	64.25 ± 0.75	-2.22 ± 0.70
SGP	75.69 ± 0.38	4.74 ± 0.37	68.50 ± 2.09	3.37 ± 0.88
CODE-CL	$\mathbf{77.21 \pm 0.32}$	$\mathbf{5.92 \pm 0.34}$	$\mathbf{71.16 \pm 0.32}$	$\mathbf{4.17 \pm 0.41}$



Advantages Over Existing Techniques

- **Selectivity**: releases only shared, beneficial directions (fine-grained vs. whole-task subspaces).
- **Safety**: projections avoid editing past-critical directions.
- **Efficiency**: store conceptors + K free dims per task ($K \ll N$).



Takeaways

- Better accuracy (ACC) with minimal forgetting, improved forward transfer (FWT) on correlated tasks.
- Fine-grained reuse of subspace directions beats coarse task-level gating.
- Light memory overhead vs. methods storing full per-task gradients.

Acknowledgment

This work was supported in part by the Center for Co-design of Cognitive Systems (CoCoSys), one of the seven centers in JUMP 2.0, a Semiconductor Research Corporation (SRC) program, and in part by the Department of Energy (DoE).

CoCoSys CENTER FOR THE CO-DESIGN OF COGNITIVE SYSTEMS
Semiconductor Research Corporation