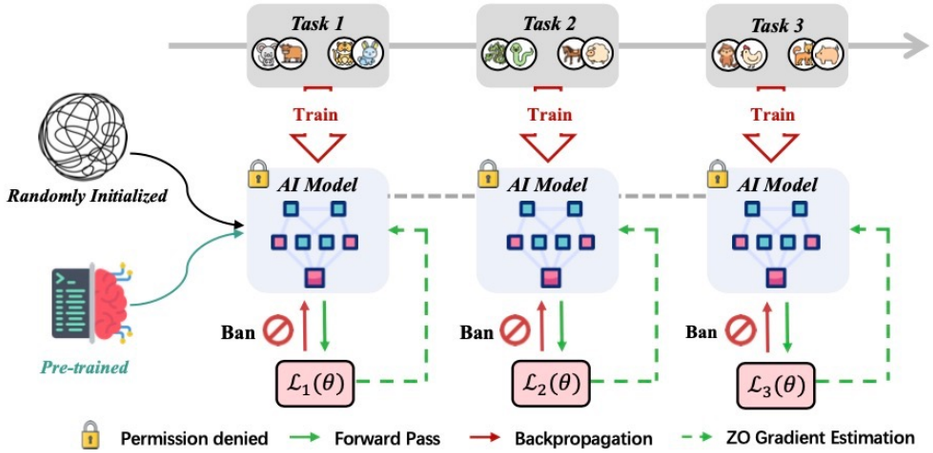
ZeroFlow: Overcoming Catastrophic Forgetting is Easier than You Think

ZeroFlow

Motivation

Catastrophic forgetting remains one of the major challenges on the path to **AGI**:

- **Continual Learning**
- **Fine-tuning Foundation Models**
- **Continual Pre-training**



Gradient bans block the model from learning and memorizing using backpropagation.

Our hypothesis

- Overcome forgetting via **only forward passes**. *Maybe, once is all it takes!*

[ICML 2025] ZeroFlow: Overcoming Catastrophic Forgetting is Easier than You Think".

ZeroFlow

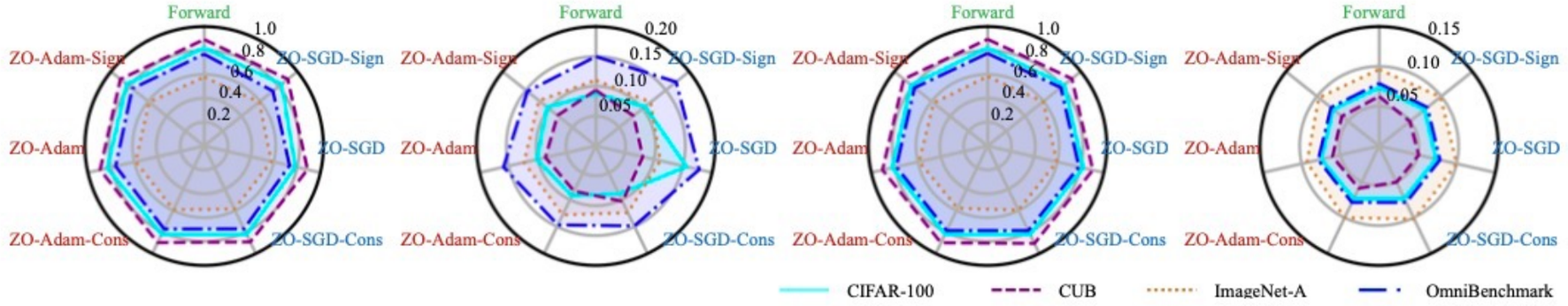
Challenge

Our hypothesis

- Overcome forgetting via **only forward passes.**

Contribution

- We propose the **first benchmark ZeroFlow** for overcoming forgetting under gradient bans.
- We **uncover insights** into how forward passes can mitigate forgetting.
- We **introduce three enhancement** techniques, which further improve the performance.



(a) EASE on average accuracy (b) EASE on forgetting (c) APER on average accuracy (d) APER on forgetting

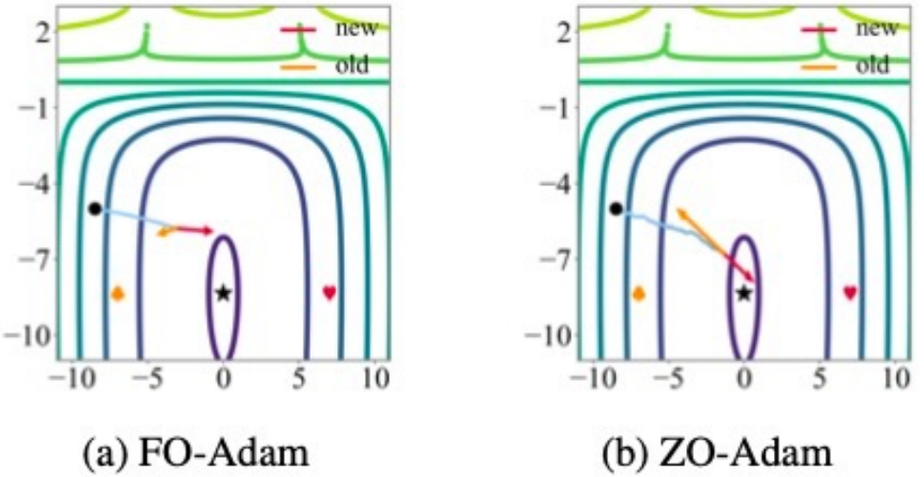
[ICML 2025] ZeroFlow: Overcoming Catastrophic Forgetting is Easier than You Think".

ZeroFlow
Framework

$$\mathcal{L}_{total} = \frac{1}{N_{context}}\mathcal{L}_{cur} + (1 - \frac{1}{N_{context}})\mathcal{L}_{replay}$$

$$\mathcal{L}_{total} = \mathcal{L}_{cur} + \alpha\mathcal{L}_{reg}$$

- **Trajectory of FO and ZO Optimization during Overcoming Forgetting.**
  - Trends of ZO optimization hold the potential to manage forgetting and learning.



(a) FO-Adam (b) ZO-Adam

[ICML 2025] ZeroFlow: Overcoming Catastrophic Forgetting is Easier than You Think".

ZeroFlow

Make Continual Learning Easier

Method	Optimizer	Strategy	CIFAR-100			CUB			ImageNet-A			OmniBenchmark		
			Avg	Last	Fgt	Avg	Last	Fgt	Avg	Last	Fgt	Avg	Last	Fgt
EASE	SGD	FO	91.23	85.96	7.32	89.31	83.76	9.61	61.24	51.02	10.84	74.73	67.40	15.11
		ZO	78.62	68.40	15.64	88.94	82.91	8.08	57.87	48.32	11.08	73.50	66.60	17.78
		Sign	**83.21**	**75.88**	10.58	**89.81**	**84.61**	8.10	**59.15**	**49.31**	11.77	73.81	66.75	17.21
		Conserve	82.22	**75.88**	8.93	89.21	83.42	10.31	58.61	48.58	12.41	**77.07**	**70.73**	14.87
	Adam	FO	90.56	84.82	7.69	84.44	77.10	10.51	59.60	47.20	19.08	74.27	66.28	15.63
		ZO	**83.36**	**76.09**	10.16	89.49	84.14	8.67	58.90	48.72	12.35	76.15	69.69	15.87
		Sign	83.14	76.01	10.44	**89.82**	**84.65**	8.21	58.97	**48.85**	12.20	77.12	**71.08**	14.68
		Conserve	82.15	75.65	9.24	**89.82**	84.61	8.40	**59.23**	**48.85**	12.81	**77.19**	70.99	14.68
	-	Forward	82.26	76.05	8.74	89.26	83.67	9.35	57.76	48.19	11.03	77.00	70.74	14.99
APER	SGD	FO	82.31	76.21	7.33	90.56	85.16	5.19	59.50	49.37	9.91	78.61	72.21	7.87
		ZO	**82.33**	76.21	7.36	90.53	85.20	5.12	59.58	49.51	10.02	78.60	72.21	7.85
		Sign	82.32	**76.23**	7.32	90.42	**85.28**	4.96	59.65	**49.77**	9.89	78.60	**72.26**	7.78
		Conserve	82.31	76.21	7.33	**90.62**	**85.28**	5.05	**59.68**	49.70	10.18	**78.61**	72.21	7.87
	Adam	FO	82.31	76.21	7.33	90.56	85.16	5.19	59.60	49.77	10.06	76.60	72.21	7.85
		ZO	82.12	75.45	7.47	**90.33**	84.31	6.01	**58.89**	**49.24**	9.32	78.44	72.10	7.87
		Sign	82.01	75.60	7.38	89.86	84.18	5.99	57.82	48.12	9.72	78.26	72.05	7.75
		Conserve	82.21	75.98	7.34	89.96	**84.48**	5.90	57.86	47.53	10.00	**78.61**	**72.21**	7.87
	-	Forward	**82.32**	**76.22**	7.32	89.47	83.38	6.24	58.25	47.99	9.62	77.61	71.45	7.87

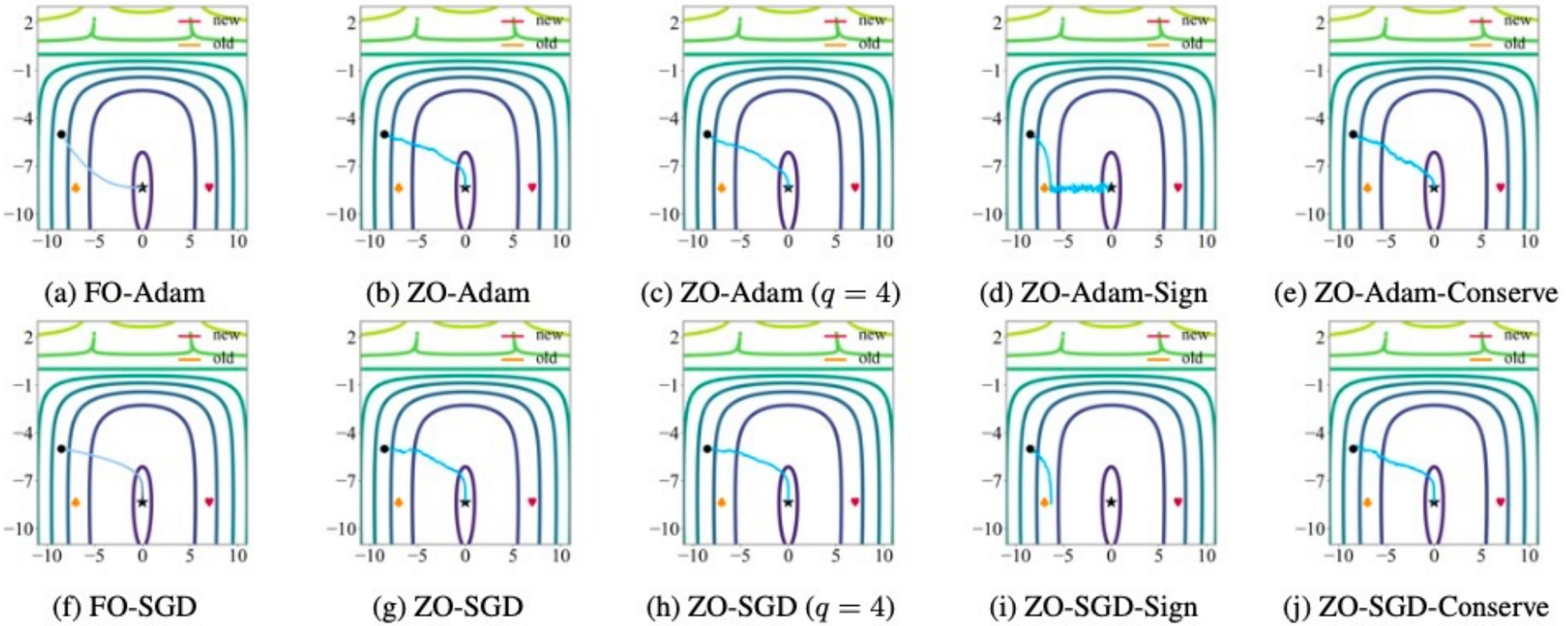
- Include **7** forward pass optimization methods.
- Span **several** forgetting scenarios and datasets.

Optimizer	Memory ⇓	CIFAR-100	CUB	ImageNet-A
FO-SGD	12.08 GB	59.3s	16.1s	12.2s
ZO-SGD ($q = 1$)	2.41 GB	32.4s	8.3s	6.8s
ZO-SGD ($q = 4$)	2.41 GB	111.7s	28.7s	18.0s
ZO-SGD-Sign	2.41 GB	32.4s	8.3s	6.8s
ZO-SGD-Conserve	2.41 GB	70.1s	15.7s	12.4s
Forward-Grad	3.94 GB	45.9s	11.1s	9.0s

- Less memory usage (↓6x)
- Less runtime (↓2x)

[ICML 2025] ZeroFlow: Overcoming Catastrophic Forgetting is Easier than You Think".

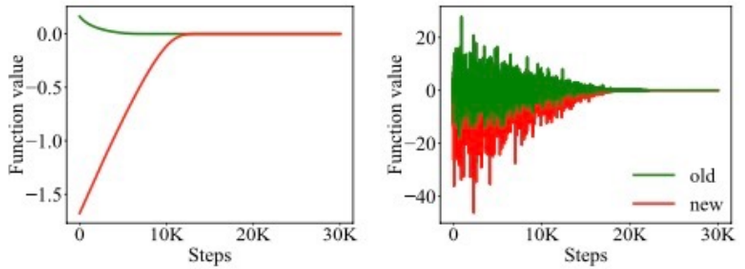
ZeroFlow
Visualization



- Trajectory of 7 forward pass optimization methods during overcoming forgetting.

[ICML 2025] ZeroFlow: Overcoming Catastrophic Forgetting is Easier than You Think".

ZeroFlow
New Enhancements



(a) FO-SGD (b) ZO-SGD

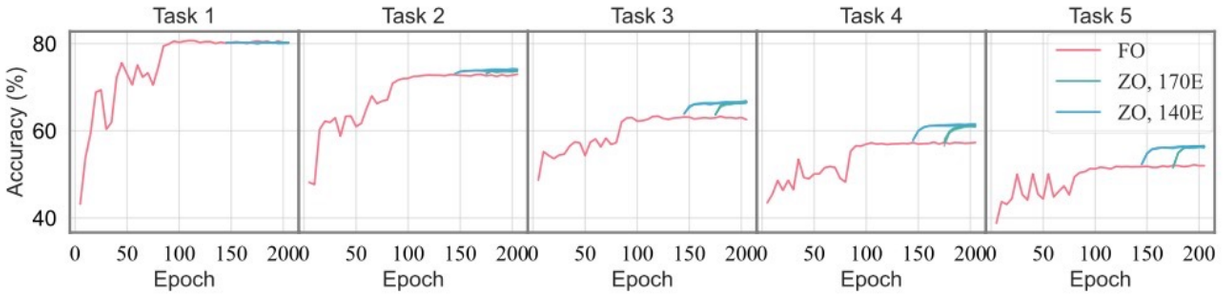
High Variation of Forward Pass

Metrics	Proportion				
	0%	20%	40%	60%	80%
Avg	57.87	**58.90**	58.76	58.34	57.83
Last	48.32	**49.04**	48.84	48.42	48.10
Fgt	11.08	11.79	11.78	11.60	11.57

Enhancement 1

50%	60%	70%	80%	90%
59.45	59.26	59.39	59.38	**59.47**
49.24	49.11	49.05	49.11	**49.24**
12.37	12.36	12.54	12.46	12.33

Enhancement 2



Effectiveness of Enhancement 3

Optimizer	Hybrid	Historical	Sparsity	Avg	Last
FO-SGD	-	-	-	61.24	51.02
ZO-SGD	-	-	-	57.87	48.32
	✓			61.40(+3.53)	51.34(+3.02)
		✓		58.90(+1.03)	49.04(+0.72)
			✓	59.47(+1.60)	49.24(+0.92)
	✓	✓	✓	62.07(+4.20)	51.94(+3.62)

Combining Enhancement 1/2/3

[ICML 2025] ZeroFlow: Overcoming Catastrophic Forgetting is Easier than You Think".

Thank you for listening.