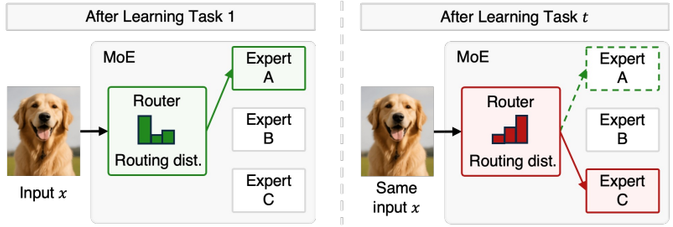
Not All Routing Drift Is Harmful: Trust-Region Projection for Class-Incremental Learning

Tae-Hwan Kim and Sung-Bae Cho

Contact Paper

Motivation

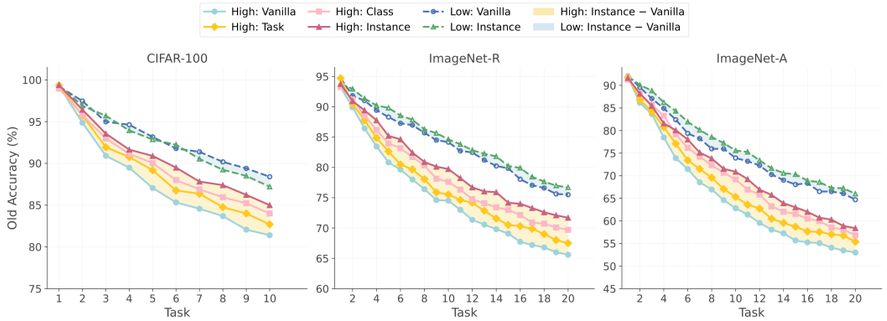


- **Setting**: class-incremental learning with router-selected adapters on a frozen backbone
- **Routing drift**: expert switching for the same input over tasks, a cause of forgetting
- **Limitation of prior work**: freezing or ignoring all drift alike
- **Open question**: whether all routing drift is harmful

Contribution

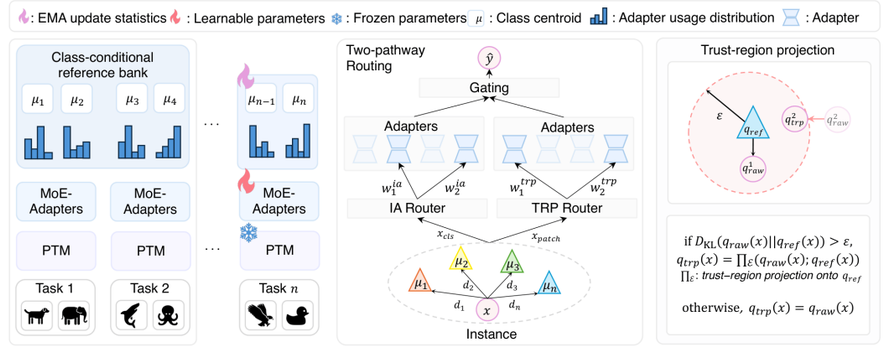
- **Instance-level analysis**: forgetting concentrated in high-drift inputs
- **TRPR**: selective **correction of only excessive drift**, with a provable KL bound
- **Consistent gains**: up to **+3.06%p accuracy** and **−2.48%p forgetting** over 9 PEFT baselines on 4 CIL benchmarks

When Does Routing Drift Hurt?



- **Setup**: Task-1 inputs split by drift size, routing restored per instance or per class
- **Result**: **loss concentrated in high-drift inputs**, nearly recovered by class-level restoration
- **Insight**: correction of only large drift, toward class-level references

Overview of Trust-Region Projected Routing (TRPR)

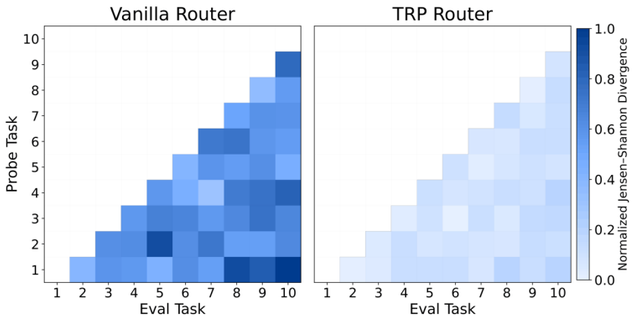


- **Class reference bank**: per-class routing anchor, EMA-updated, retrieved by nearest class
- **Trust-region projection**: **small drift kept**, large drift projected
- **Instance-adaptive path**: per-sample flexibility via a learned gate

Experiments

Method	CIFAR-100			5-Task IN-R			10-Task IN-R			20-Task IN-R			ImageNet-A			VTAB		
	$\bar{\mathcal{A}}$	$\mathcal{A}_N$	$\mathcal{F}$	$\bar{\mathcal{A}}$	$\mathcal{A}_N$	$\mathcal{F}$	$\bar{\mathcal{A}}$	$\mathcal{A}_N$	$\mathcal{F}$	$\bar{\mathcal{A}}$	$\mathcal{A}_N$	$\mathcal{F}$	$\bar{\mathcal{A}}$	$\mathcal{A}_N$	$\mathcal{F}$	$\bar{\mathcal{A}}$	$\mathcal{A}_N$	$\mathcal{F}$
FT Adapter	48.17	30.72	31.25	53.42	40.48	16.84	44.89	30.41	19.48	37.64	23.57	22.17	30.28	18.14	17.24	59.63	42.91	21.42
L2P	87.43	82.29	7.84	78.77	74.28	5.92	65.39	61.55	7.44	75.77	70.20	9.87	48.73	38.91	12.42	80.12	77.32	6.80
DualPrompt	88.72	83.32	5.90	74.54	69.88	5.26	70.10	63.84	8.86	72.55	66.02	11.23	54.80	42.31	14.88	82.92	78.70	14.92
CODA-P	91.55	86.30	4.84	79.11	75.00	4.72	80.32	73.72	6.80	75.90	70.32	8.48	56.46	45.65	11.21	78.92	81.70	5.68
InfLoRA	91.87	87.12	5.16	77.23	74.98	4.25	80.76	74.21	6.25	77.52	71.90	7.82	58.14	47.43	10.81	89.32	85.47	5.24
LAE	90.12	85.18	6.24	78.14	74.62	5.02	75.48	68.91	7.17	79.24	72.51	9.33	57.34	46.28	12.16	86.28	84.12	6.16
EASE	92.09	87.67	6.12	82.10	77.72	4.98	81.67	75.86	7.51	81.54	75.78	10.26	62.20	52.35	11.65	90.85	83.28	5.47
SimpleCIL	87.13	81.26	7.87	65.88	61.33	5.35	67.02	61.24	7.28	67.59	61.35	8.64	60.05	49.24	12.31	89.68	83.88	4.90
ADAM	91.23	87.51	4.82	79.48	74.07	4.12	74.97	66.94	7.93	77.23	70.48	8.95	61.21	52.16	10.42	90.70	85.90	4.80
SEMA	91.21	86.80	5.21	83.66	78.17	4.58	82.98	76.48	6.40	80.48	72.68	8.12	61.05	49.77	11.78	91.21	87.40	4.64
TRPR (Ours)	**93.38**	**88.27**	**4.11**	**84.34**	**79.54**	**3.40**	**84.12**	**78.47**	**4.85**	**82.33**	**77.29**	**5.34**	**63.59**	**53.12**	**9.47**	**92.21**	**90.46**	**3.85**

- **Main results**: **highest accuracy and lowest forgetting** in all 6 settings
- **More results**: gains on 50-task streams and over 5 mitigation methods



- **Routing stability**: suppressed routing drift across tasks with TRPR

This work was supported by the IITP grant funded by the Korea government (MSIT) (RS-2022-II220113, Developing a Sustainable Collaborative Multi-modal Lifelong Learning Framework), and Air Force Defense Research Sciences Program (No. FA2386-25-1-4062) funded by Air Force Office of Scientific Research.