SHANGHAI JIAO TONG UNIVERSITY

IDER: IDempotent Experience Replay for Reliable Continual Learning

Zhanwang Liu *, Yuting Li *, Haoyuan Gao, Yexin Li,
Linghe Kong Lichao Sun, Weiran Huang †

ICLR 2026

SJTU MIFA LAB

ICLR

Motivation

- Continual Learning suffers from **recency bias** and models deployed in real-world scenarios benefit from **uncertainty awareness**.

- Our analysis reveals a strong **positive correlation** between improved calibration and higher accuracy.

- This motivates us to introduce the **idempotent property** into rehearsal-based continual learning for more **reliable prediction**.

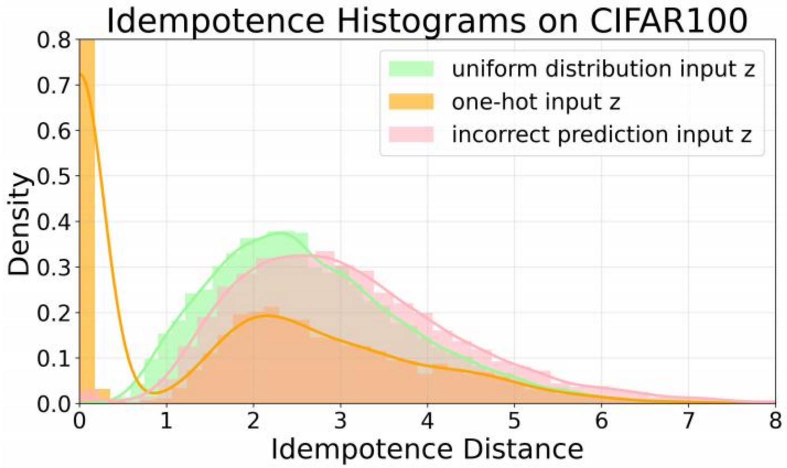
Idempotent property

- **Formulation:**

$$f(f(x)) = f(x)$$

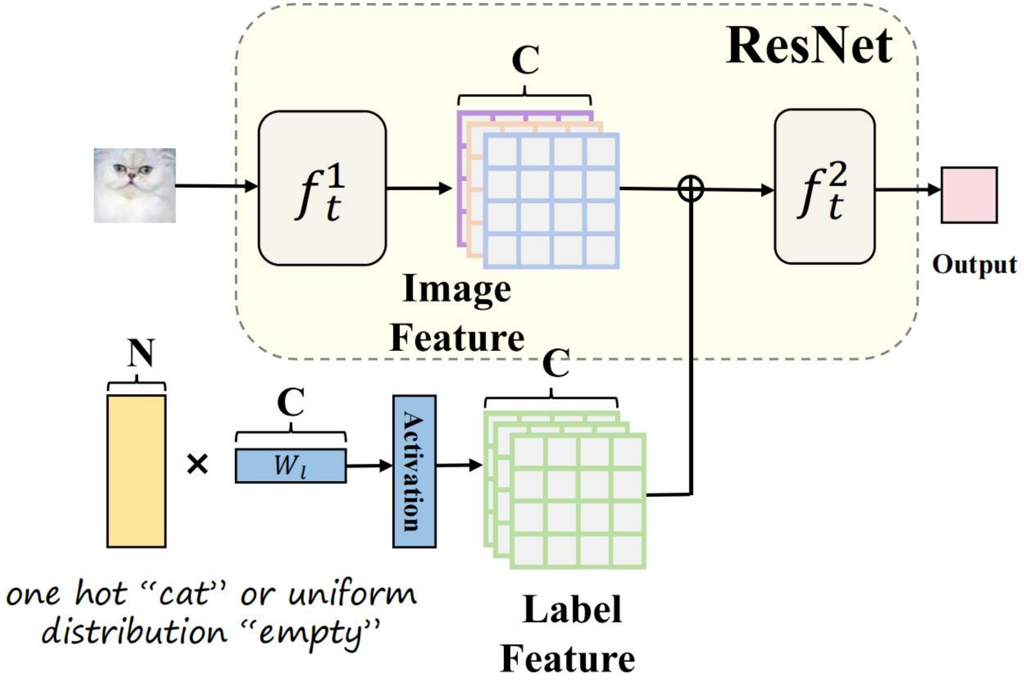
$$f(x,0) = f(x,y) = y$$

- **It can be used to measure prediction reliability.**



Methods

We introduce **idempotence** into continual learning by modifying the backbone to **accept a second input** in addition to the image.



Methods

IDER contains two aspects:

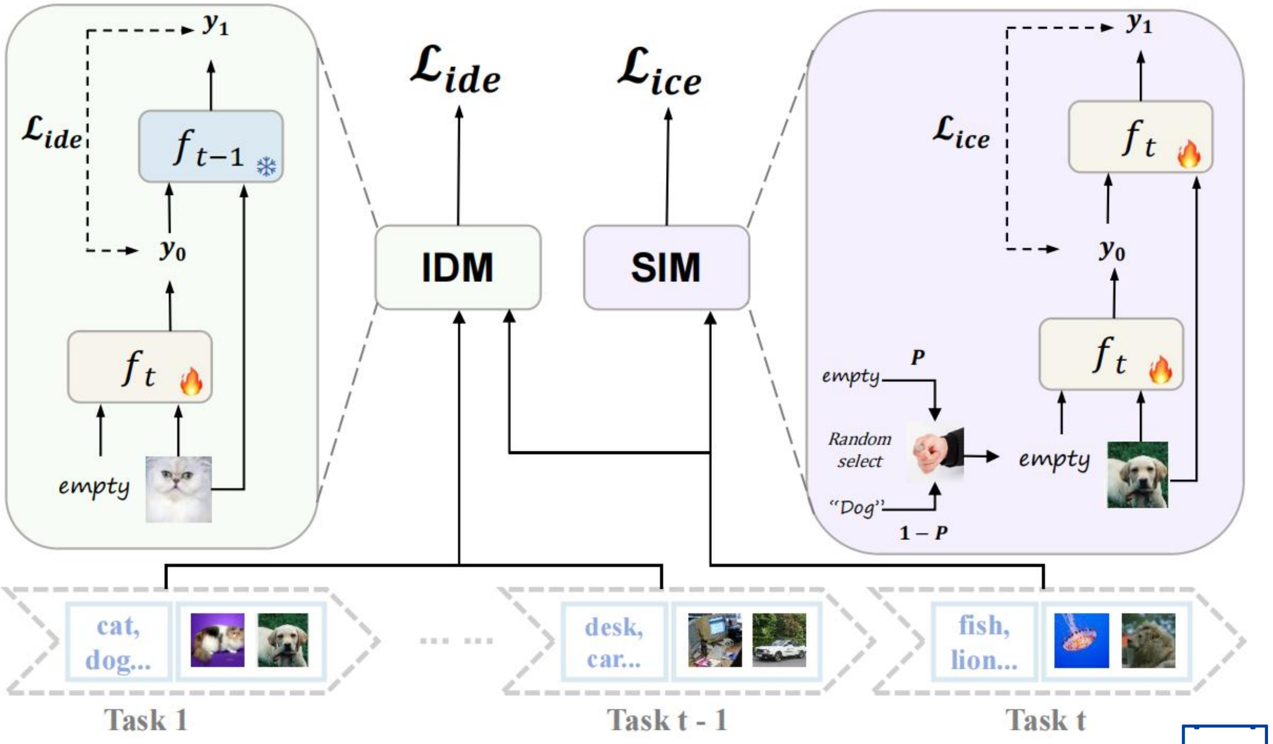
- Idempotent learning on current-task data

$$\mathcal{L}_{ice} = \sum_{(x,y)\in\mathcal{T}_t} [\mathcal{L}_{ce}(f_t(x,y^*),y) + \mathcal{L}_{ce}(f_t(x,f_t(x,y^*)),y)]$$

- Idempotent distillation from the previous model

$$\mathcal{L}_{ide} = \sum_{(x,y)\in\mathcal{T}_t,M} \|f_t(x,0) - f_{t-1}(x,f_t(x,0))\|_2^2$$

$$\mathcal{L}_{rep\text{-}ice} = \sum_{(x,y)\in M} [\mathcal{L}_{ce}(f_t(x,y^*),y) + \mathcal{L}_{ce}(f_t(x,f_t(x,y^*)),y)]$$



Main Experiments

We evaluate our method on class-incremental learning setting.

Method	CIFAR-10		CIFAR-100		Tiny-ImageNet	
	Buffer 200	Buffer 500	Buffer 500	Buffer 2000	Buffer 500	Buffer 4000
Joint (upper bound)	91.93±0.29		71.15±0.51		59.52±0.33	
iCaRL (Rebuffi et al., 2017)	58.37±3.51	62.49±5.42	46.81±0.41	52.51±0.44	22.53±0.62	26.38±0.23
ER (Riemer et al., 2019)	44.46±2.87	58.84±3.85	23.41±1.15	40.47±0.95	10.13±0.39	25.12±0.56
BiC (Wu et al., 2019)	52.61±5.37	71.95±1.82	37.82±1.67	47.17±1.17	15.36±1.31	18.67±0.57
LUCIR (Hou et al., 2019)	49.18±7.61	65.26±2.54	37.91±1.18	50.42±0.76	28.79±0.51	31.64±0.51
DER (Buzzega et al., 2020)	57.92±1.91	68.65±1.82	34.83±2.09	50.12±0.75	15.14±1.29	20.35±0.35
DER++ (Buzzega et al., 2020)	62.19±1.94	70.10±1.65	37.69±0.97	51.82±1.04	19.43± 1.63	36.89± 1.16
ER-ACE (Caccia et al., 2021)	62.19±1.67	71.15±1.08	37.81±0.54	49.77±0.34	20.42±0.39	37.76±0.53
XDER (Boschini et al., 2022)	64.10±1.08	67.42±2.16	48.14±0.34	57.57±0.84	29.12±0.47	46.12±0.46
CLS-ER (Arani et al., 2022)	64.56±2.63	74.27±0.81	43.92±0.62	54.84±1.30	30.91±0.59	45.17±0.89
SCoMMER (Sarfraz et al., 2023)	66.95±1.52	73.64±0.43	39.05±0.79	49.42±0.85	21.47±0.54	37.2±0.70
BFP (Gu et al., 2023)	68.64±2.23	73.51±1.54	46.70±1.45	57.39±0.75	28.71±0.55	43.17±1.89
SARL (Sarfraz et al., 2025)	68.87±1.37	73.98±0.46	46.69±0.79	57.06±0.48	28.44±2.30	38.83±0.81
ER+ID(Ours)	71.02±1.98	74.74±0.42	44.82±0.85	56.59±0.35	29.88±1.15	43.05±1.40
BFP+ID (Ours)	71.99±0.98	76.65±0.63	48.53±0.95	57.74±0.64	30.62±0.47	43.51±0.59
CLS-ER+ID (Ours)	70.32±1.12	75.48±0.91	47.44±2.0	56.36±0.78	31.62±0.57	46.17±0.22

Further Experiments

We evaluate our method on generalized class-incremental learning setting.

Method	**Uniform**				**Longtail**			
	Buffer 200	Δ	Buffer 500	Δ	Buffer 200	Δ	Buffer 500	Δ
Joint (upper bound)	58.36±1.02				56.94±1.56			
DER++ (Buzzega et al., 2020)	19.36±0.65		33.66±0.96		27.05±1.11		25.98±0.81	
SCoMMER (Sarfraz et al., 2023)	28.56±2.26		35.70±0.86		28.47±1.12		32.99±0.49	
ER (Riemer et al., 2019)	16.34±0.74		28.76±0.66		19.55±0.69		20.02±1.05	
Ours (ER+ID)	26.66±0.63	**+10.32**	40.54±0.46	**+11.78**	30.04±0.58	**+10.49**	35.92±0.35	**+15.90**
CLS-ER (Arani et al., 2022)	22.37±0.48		36.80±0.34		28.34±0.99		28.35±0.72	
Ours (CLS-ER+ID)	31.17±1.62	**+8.80**	37.57±1.81	**+0.77**	34.08±0.45	**+5.74**	36.75±0.62	**+8.40**
SARL (Sarfraz et al., 2025)	36.20±0.46		38.73±0.66		34.13±1.07		34.64±0.49	
Ours (SARL+ID)	36.45±0.37	**+0.25**	39.65±0.43	**+0.92**	35.04±0.54	**+0.91**	35.67±0.74	**+1.03**

Further Experiments

Idempotence improves prediction reliability.

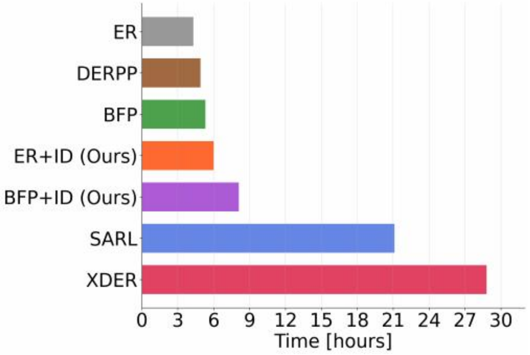
Method	CIFAR-10				CIFAR-100				Tiny-ImageNet			
	Buffer 200	Δ	Buffer 500	Δ	Buffer 500	Δ	Buffer 2000	Δ	Buffer 500	Δ	Buffer 4000	Δ
DER (Buzzega et al., 2020)	29.91		16.20		24.84		10.79		22.80		10.52	
NPCL (Jha et al., 2023)	21.03		-		19.95		-		-		-	
RC (Li et al., 2024a)	16.39		12.84		19.43		19.31		21.32		16.49	
T-CIL (Hwang et al., 2025)	22.50		10.51		15.79		8.67		14.50		10.30	
ER (Riemer et al., 2018)	45.53		32.69		64.59		45.64		67.50		51.37	
Ours (ER+ID)	12.36	-33.17	11.73	-20.96	13.65	-50.94	12.87	-32.77	21.55	-49.45	11.14	-40.23
BFP (Gu et al., 2023)	9.83		9.40		11.93		9.28		9.45		8.25	
Ours (BFP+ID)	9.30	-0.53	8.63	-0.77	8.92	-3.01	8.29	-0.99	7.77	-1.68	6.35	-1.9

Further Analysis

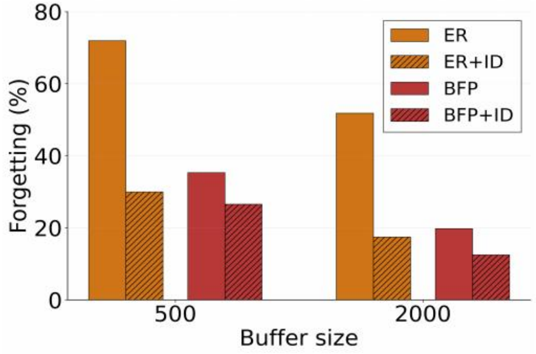
The modified backbone has **negligible** impact on the performance.

Model	Method	Accuracy (%)	Forgetting (%)
Normal ResNet-18	Finetune	8.29	90.52
	ER	24.36	71.30
Modified ResNet-18	Finetune	8.23	90.58
	ER	24.73	70.61

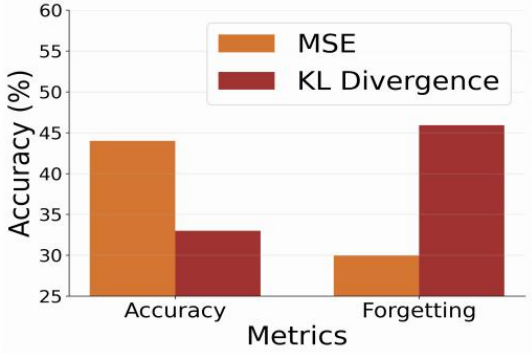
IDER substantially **reduces forgetting** with **minimal additional training cost**, while **MSE** outperforms KL divergence.



(a) Training Times



(b) Forgetting on CIFAR-100



(c) Distance Metrics

Thank you

Paper

Github