Decoupling Learning and Remembering: a Bilevel Memory Framework with Knowledge Projection for Task-Incremental Learning

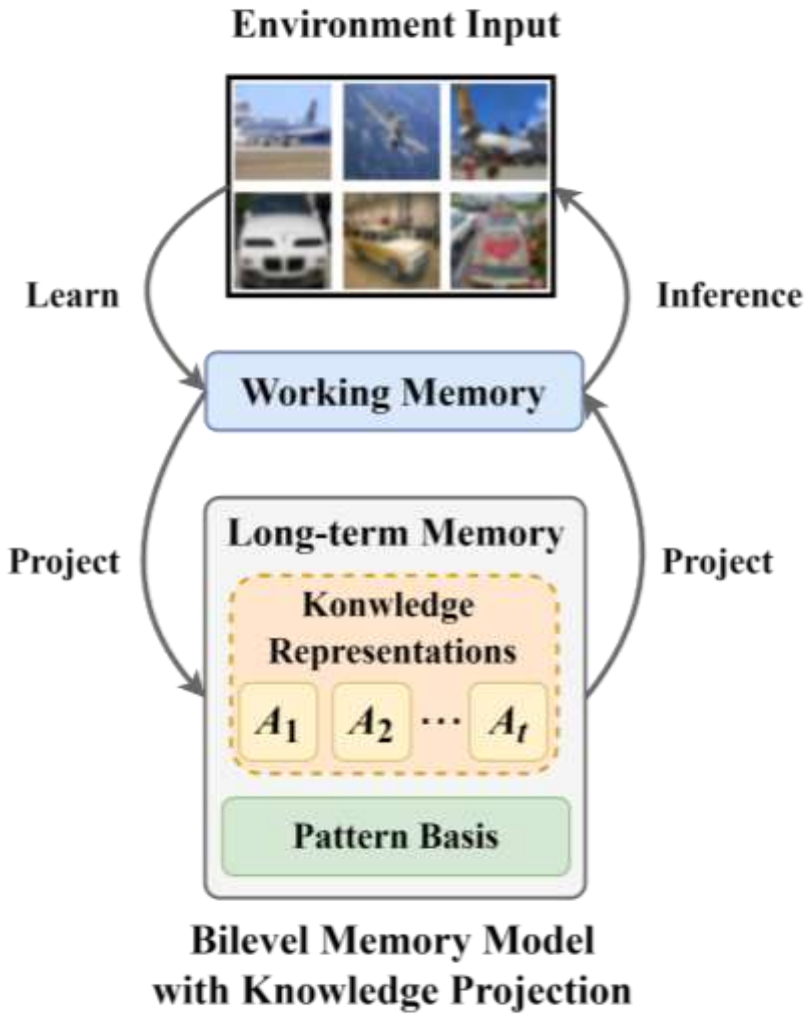
Wenju Sun, Qingyong Li, Jing Zhang, Wen Wang, Yangli-ao Geng

Presented by: Wenju Sun
Beijing Jiaotong University

The IEEE/CVF Conference on Computer Vision and Pattern Recognition 2023

Tag: THU-AM-352   Paper ID: 10244

Overview



Bilevel **M**emory model with **K**nowledge **P**rojection (BMKP)

- **Working memory** responsible for new knowledge learning, to ensure **high plasticity**.
- **Long-term memory** in charge of storing learned knowledge, to guarantee **high stability**.
- **Knowledge projection,** project model knowledge of working memory into the compact representation, and then stored into long-term memory, to achieve **high memory efficiency.**

Content

Introduction

- Incremental learning
- Conventional memory-based IL method vs. Multi-level human memory system

Method

- Bilevel memory framework
- Knowledge spaces
- Incremental learning process of BMKP

Experiment

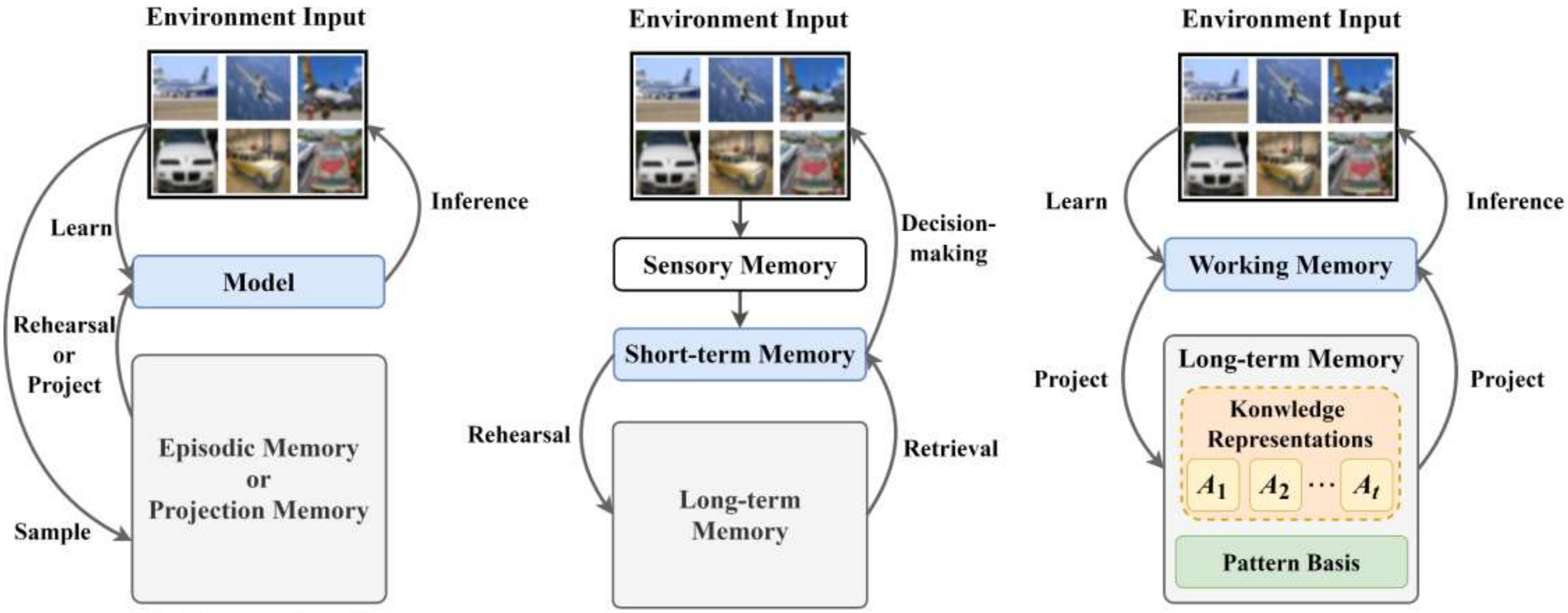
Conclusion

Introduction

Incremental learning

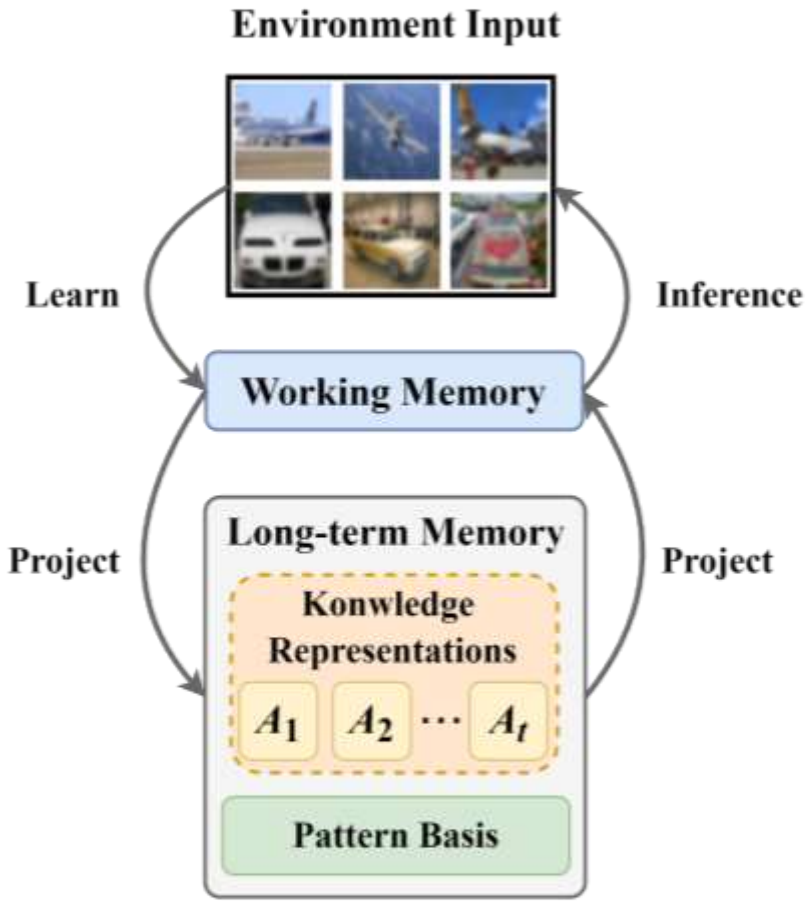
- The ability to continuously learn new knowledge while keeping the memory of the old knowledge.

Introduction



(a) Memory-based Incremental Learning Model (b) Atkinson–Shiffrin Human Memory Model (c) Bilevel Memory Model with Knowledge Projection

Bilevel Memory Framework



Bilevel Memory Model
with Knowledge Projection

Bilevel **M**emory model with **K**nowledge **P**rojection (BMKP)

- **Working memory** responsible for new knowledge learning, to ensure **high plasticity**.
- **Long-term memory** in charge of storing learned knowledge, to guarantee **high stability**.
- **Knowledge projection,** project model knowledge of working memory into the compact representation, and then stored into long-term memory, to achieve **high memory efficiency.**

Knowledge Spaces

Knowledge: the ability to transfer a given input to the expected output.

Parameter Knowledge Space (PKS)

- The Space where the knowledge is represented as the trained parameters.

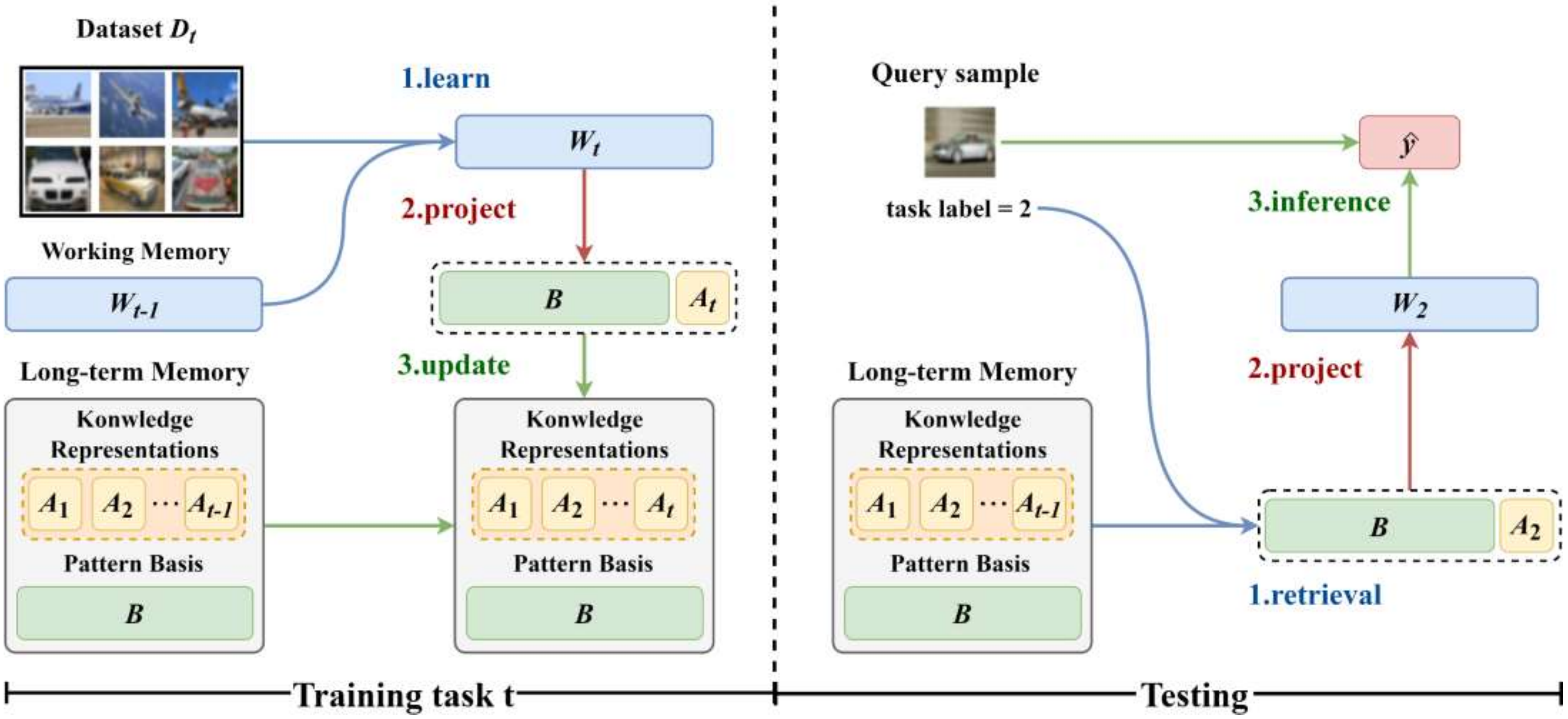
$$Z^l = W^l X^l$$

Core Knowledge Space (CKS)

- in which the knowledge can be organized in a quite compact form without loss of performance.

$$\begin{aligned} Z^l = B^l {B^l}^\top Z^l &= B^l {B^l}^\top W^l X^l \\ &= \hat{W}^l X^l \\ &= B^l A^l X^l, \end{aligned}$$

BMKP Training & Inference



Working Memory Learning

- Following the principle of minimum energy consumption, BMKP encourages the working memory to learn new knowledge with respect to the pattern basis *B*:

$$L_{reg}(W) = \sum_{l=1}^{L} \frac{\mathrm{Trace}\left((W^l - \tilde{W}^l)^\top (W^l - \tilde{W}^l) \right)}{\left\| (W^l - \tilde{W}^l)^\top (W^l - \tilde{W}^l) \right\|_1}$$

  Where $\tilde{W}^l = B^l {B^l}^\top W^l$ denotes the orthogonal projection of W^l into CKS.

- The overall loss for task *t* learning is:

$$W_t \longleftarrow \arg\min_{W} L_{task}(W, D_t) + \lambda L_{reg}(W)$$

Knowledge Projection

- We first extend CKS with new pattern basis:

$$U_t^l \Sigma_t^l {V_t^l}^\top \longleftarrow \mathrm{SVD}\left(Z_t^l - B^l {B^l}^\top Z_t^l\right)$$

$$B^l \longleftarrow \begin{bmatrix} B^l & {U_t^l}^\top \end{bmatrix}$$

- Then project knowledge into CKS:

$$A_t^l \longleftarrow {B^l}^\top W_t^l$$

Long-term Memory Updating

- The knowledge projection may not be perfect since some minor basis are dropped through threshold selection, We introduce a recall mechanism:

$$A_t \longleftarrow \arg\min_{A_t} L_{task}(BA_t, D_t)$$

北京交通大學
BEIJING JIAOTONG UNIVERSITY

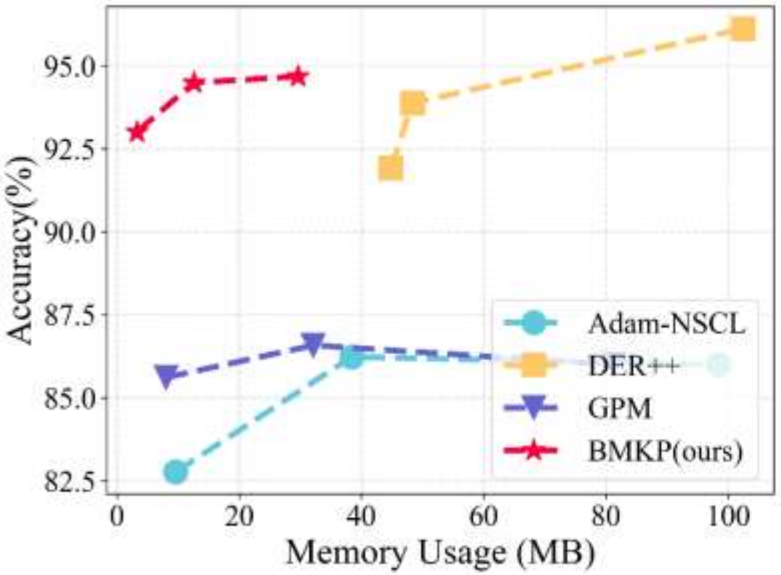
Experiment

Methods	Venue	CIFAR-10	CIFAR-100	Tiny-ImageNet	Average
Joint*	-	98.07	91.18	82.01	90.42
LwF [16]	TPAMI2017	91.91±0.7	63.78±4.3	58.61±1.8	71.43
SI [38]	ICML2017	76.15±2.6	62.21±2.6	60.91±1.3	66.42
DGR [29]	NIPS2017	91.06±7.4	44.53±2.5	-	-
GEM [17]	NIPS2017	85.14±2.1	62.80±2.7	44.66±1.7	64.20
oEWC [27]	ICML2018	64.17±4.8	38.40±1.9	31.91±0.9	44.83
LwM [9]	CVPR2019	78.01±0.8	68.88±0.9	45.57±0.2	64.15
DI [35]	CVPR2020	<u>94.46±0.6</u>	68.43±2.1	66.12±0.9	76.34
DER [3]	NIPS2020	93.13±0.3	73.26±1.3	51.22±1.5	72.54
DER++ [3]	NIPS2020	93.71±0.4	74.86±1.1	53.00±0.4	73.86
DER++† [3]	NIPS2020	93.88±0.5	-	51.91±0.7	-
HAL [4]	AAAI2021	82.34±1.5	43.91±3.6	-	-
PASS [39]	CVPR2021	86.07±0.2	77.30±0.4	62.87±0.4	75.41
GPM [26]	ICLR2021	86.58±0.9	70.93±0.9	59.84±0.2	72.45
GPM† [26]	ICLR2021	-	72.48	-	-
Adam-NSCL [33]	CVPR2021	87.23±0.4	65.69±0.2	59.98±0.7	70.97
CLS-ER [2]	ICLR2022	93.53±0.3	72.11±0.5	57.36±0.7	74.33
WSN [12]	ICML2022	92.99±0.4	**81.10±0.7**	<u>67.50±0.7</u>	<u>80.53</u>
CF-IL† [23]	ICLR2022	93.12	-	67.42	-
FAS [22]	ICLR2022	90.89±1.3	70.89±0.6	60.10±0.2	73.96
BMKP (ours)	-	**94.49±0.2**	<u>79.62±0.8</u>	**70.36±0.2**	**81.49**

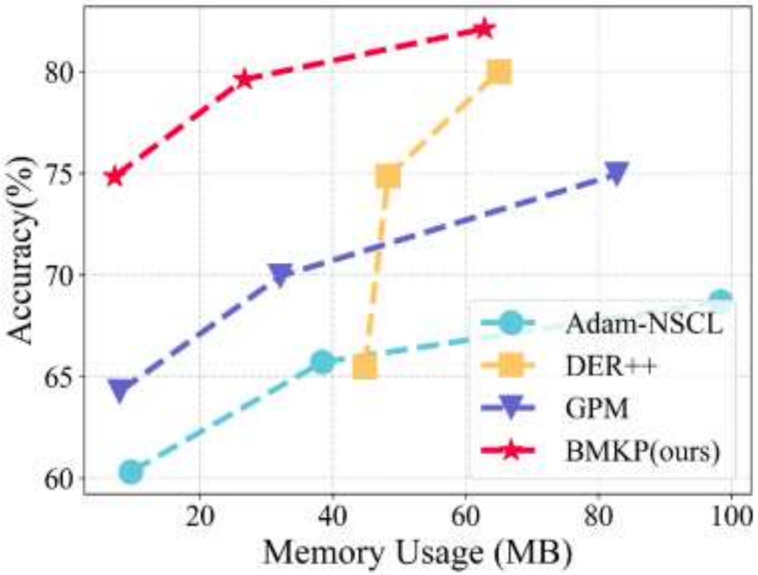
Ablation Study

Methods	Split CIFAR-10	Split CIFAR-100	Split Tiny-ImageNet
BMKP w/o basis updating	79.44±2.7	43.00±1.9	28.27±1.1
BMKP w/o retraining	94.07±0.3	78.73±0.6	68.12±0.8
BMKP	94.49±0.2	79.62±0.8	70.36±0.2

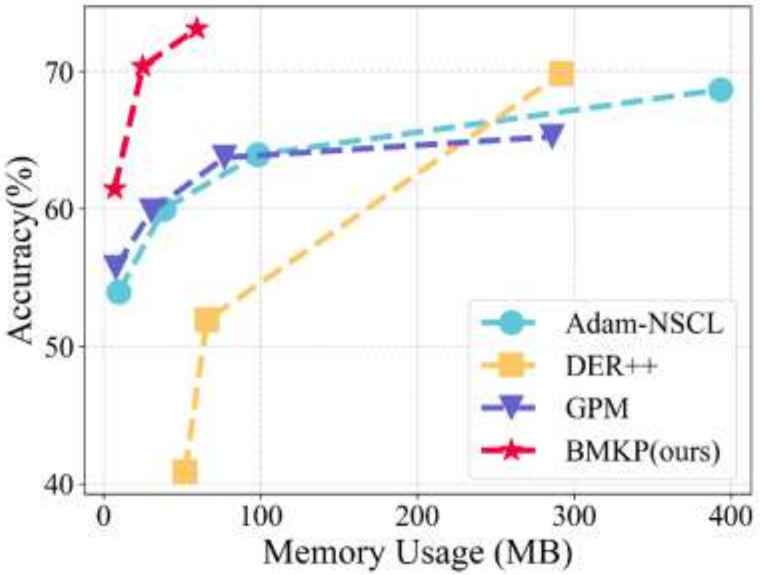
Memory Efficiency Analysis



CIFAR-10



CIFAR-100



Tiny-ImageNet

Conclusion

1. Inspired by the multi-level human memory system, we propose a bilevel-memory framework for incremental learning, which benefits from both high plasticity and stability.
2. We propose a knowledge projection process to project knowledge from PKS into compact representation in CKS, which not only improves memory utilization efficiency but also enables forward knowledge transfer for incremental learning.
3. We design a regularizer to encourage the working memory to reuse previously learned knowledge, which enhances both the memory efficiency and the performance of BMKP.
4. The experimental results show that BMKP achieves state-of-the-art performance in most cases with lower memory usage.

THANKS FOR WATCHING

Decoupling Learning and Remembering: a Bilevel Memory Framework with Knowledge Projection for Task-Incremental Learning

Presented by: Wenju Sun
Beijing Jiaotong University

The IEEE/CVF Conference on Computer Vision and Pattern Recognition 2023

北京交通大學
BEIJING JIAOTONG UNIVERSITY

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

THANKS FOR WATCHING

Decoupling Learning and Remembering: a Bilevel Memory Framework with Knowledge Projection for Task-Incremental Learning

The code is available at
https://github.com/SunWenJu123/BMKP
Contact:
SunWenJu@bjtu.edu.cn