NEURAL INFORMATION PROCESSING SYSTEMS

F-OAL: Forward-only Online Analytic Learning with Fast Training and Low Memory Footprint in Class Incremental Learning

Huiping Zhuang [1*] ,Yuchen Liu [2*] ,Run He [1], Kai Tong [1], Ziqian Zeng [1], Cen Chen [1†] ,Yi Wang [3], Lap-Pui Chau [3]

[1] South China University of Technology
[2] The University of Hong Kong
[3] The Hong Kong Polytechnic University

Introduction & motivation

- Online Class Incremental Learning (OCIL) requires to train the model on an online data stream task by task. New classes occur in new task and all training has to be done in one epoch (i.e., one pass data stream).
- The goal of OCIL is to counter Catastrophic Forgetting (CF), the losing of previous knowledge when new data are introduced.

Task 1

Task 2

……

Task k

One pass data stream

Introduction & motivation

To solve CF, several methods are proposed, which can be considered as two type:

- **Exemplar-free:** adds constraint to loss function to avoid forgetting.
- **Replay-based:** stores some historical data and reuse the in new task.

Exemplar-free methods use less training resources while show less competitive results.

Replay-based methods counter CF effectively but requires additional GPU memory and long training time.

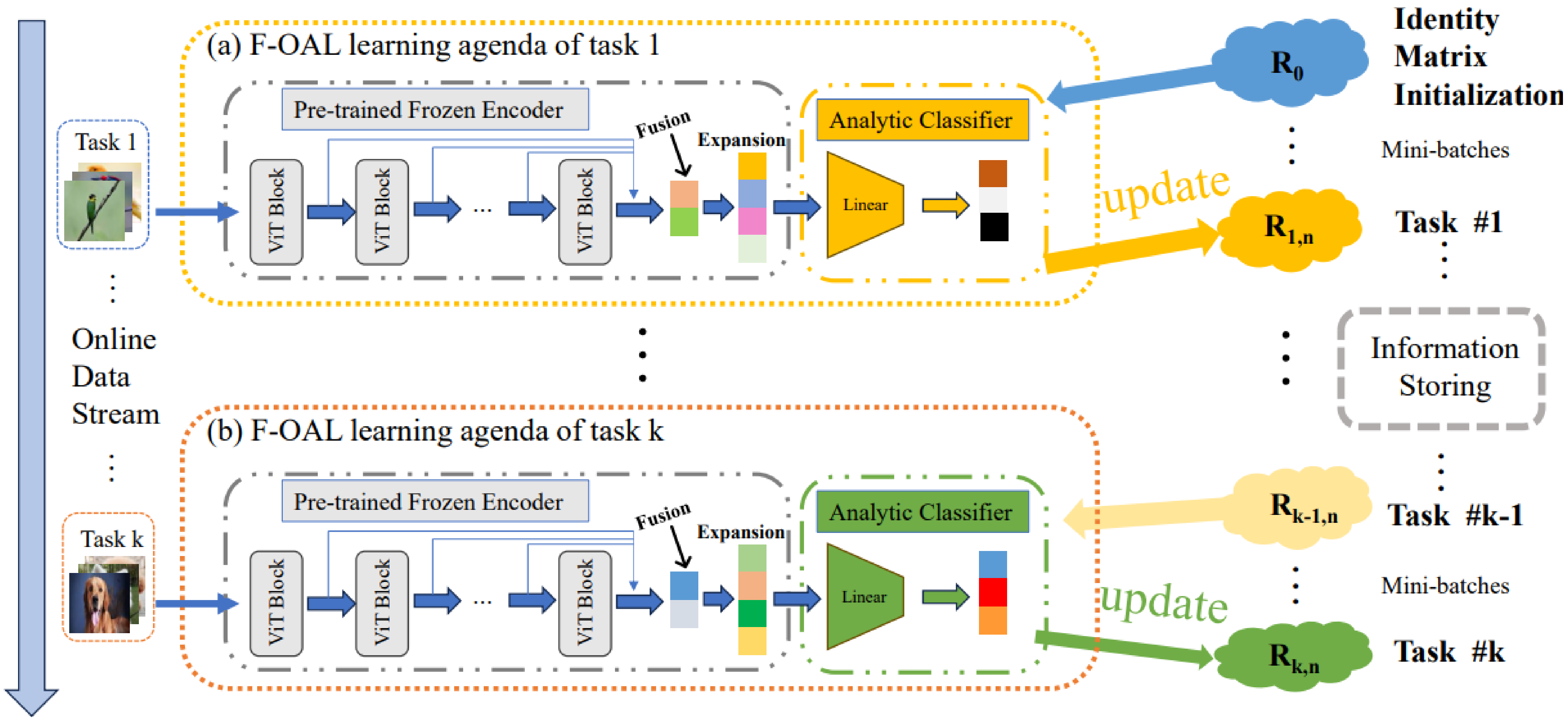
Introduction & motivation

Analytic Continual Learning (ACL):

- A new branch that redefines the Continual Learning, including Class Incremental Learning into recursive least square problem.
- ACL is able to train the model in one epoch without compromising its accuracy.
- ACL has been used in several CL scenario such as classic CIL, few-shot CIL.

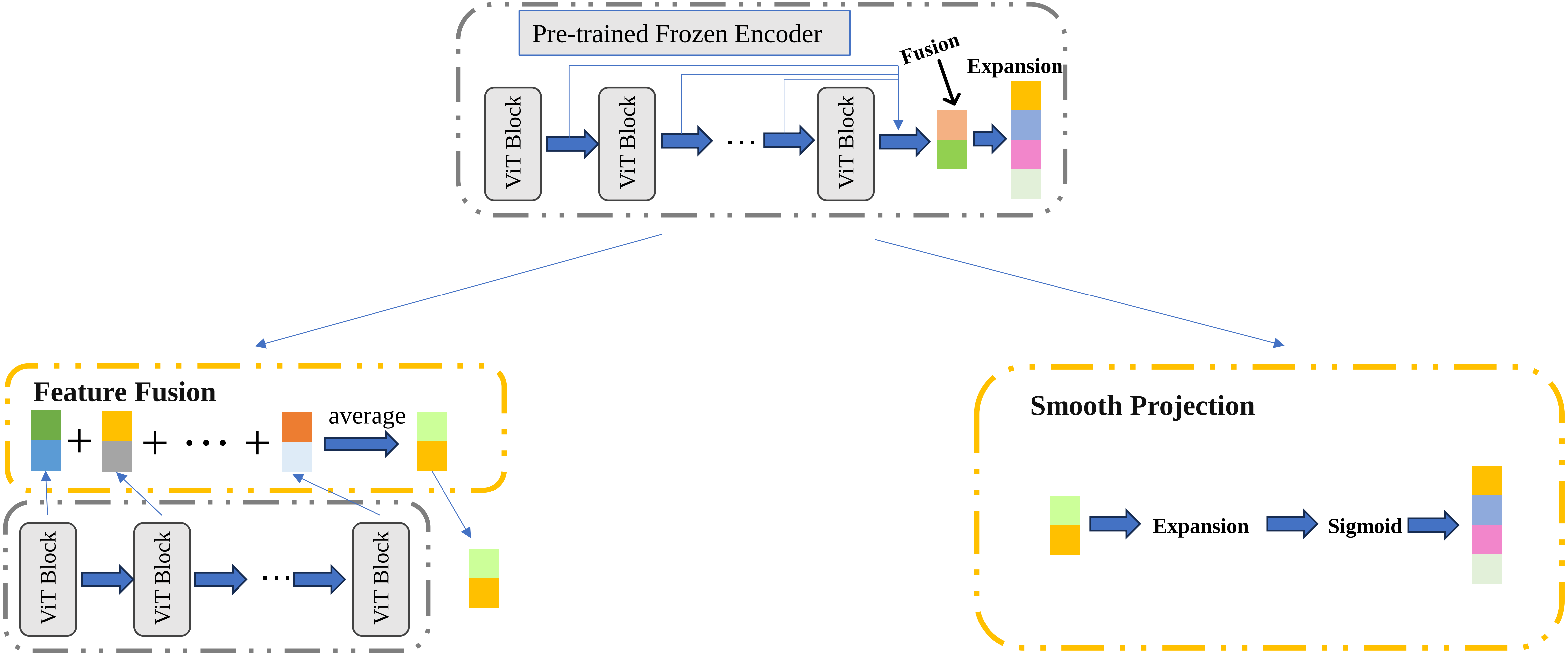
F-OAL

The proposed framework has to parts: **frozen pre-trained encoder** and **Analytic Classifier**.



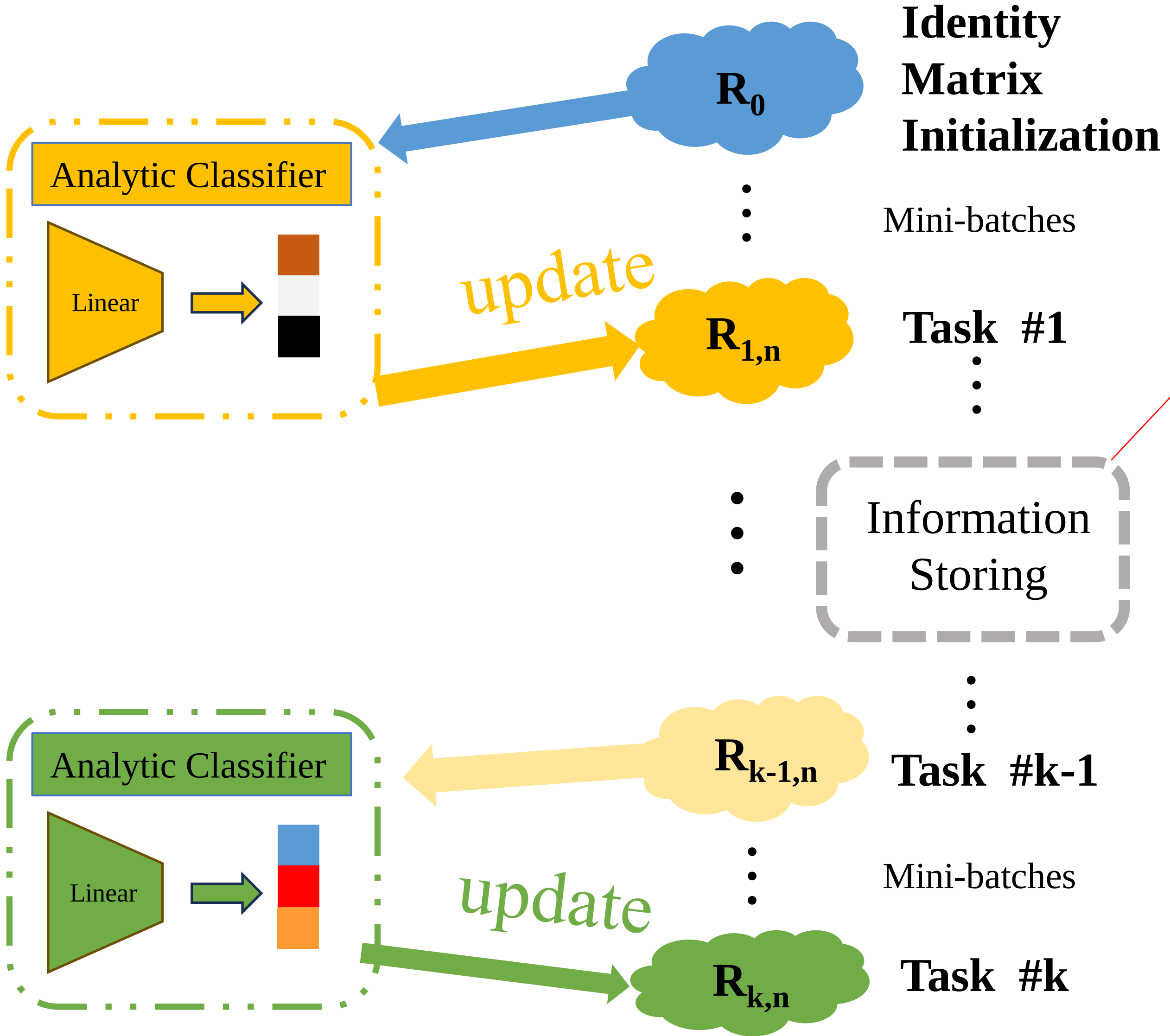
F-OAL

In encoder part, we leverage the pre-trained Vision Transformer and add two module: **Feature fusion** and **Smooth Projection** to produce representative activations.



F-OAL

In Analytic Classifier part, we update the weight matrix of the linear classifier by recursive least square with the activations extracted by encoder.



$$\boldsymbol{R}_{1,i-1} = \left[(\boldsymbol{X}^{(a)}_{1,1:i-1})^{\mathrm{T}}(\boldsymbol{X}^{(a)}_{1,1:i-1}) + \gamma \boldsymbol{I}\right]^{-1}$$

At task k batch 1:

$$\hat{\boldsymbol{W}}^{(k-1,n)'} = \begin{bmatrix} \hat{\boldsymbol{W}}^{(k-1,n)} & \boldsymbol{0} \end{bmatrix}$$

$$\hat{\boldsymbol{W}}^{(k,1)} = \hat{\boldsymbol{W}}^{(k-1,n)'} + \boldsymbol{R}_{k,1}\boldsymbol{X}^{(\mathrm{a})\mathrm{T}}_{k,1}\left(\boldsymbol{Y}^{\mathrm{train}}_{k,1} - \boldsymbol{X}^{(a)}_{k,1}\hat{\boldsymbol{W}}^{(k-1,n)'}\right),$$

$$\boldsymbol{R}_{k,1} = \boldsymbol{R}_{k-1,n} - \boldsymbol{R}_{k-1,n}\boldsymbol{X}^{(\mathrm{a})\mathrm{T}}_{k,1}\left(\boldsymbol{I} + \boldsymbol{X}^{(\mathrm{a})}_{k,1}\boldsymbol{R}_{k-1,n}\boldsymbol{X}^{(\mathrm{a})\mathrm{T}}_{k,1}\right)^{-1}\boldsymbol{X}^{(\mathrm{a})}_{k,1}\boldsymbol{R}_{k-1,n}.$$

At task k batch i:

$$\hat{\boldsymbol{W}}^{(k,i)} = \hat{\boldsymbol{W}}^{(k,i-1)} + \boldsymbol{R}_{k,i}\boldsymbol{X}^{(\mathrm{a})\mathrm{T}}_{k,i}\left(\boldsymbol{Y}^{\mathrm{train}}_{k,i} - \boldsymbol{X}^{(a)}_{k,i}\hat{\boldsymbol{W}}^{(k,i-1)}\right),$$

$$\boldsymbol{R}_{k,i} = \boldsymbol{R}_{k,i-1} - \boldsymbol{R}_{k,i-1}\boldsymbol{X}^{(\mathrm{a})\mathrm{T}}_{k,i}\left(\boldsymbol{I} + \boldsymbol{X}^{(\mathrm{a})}_{k,i}\boldsymbol{R}_{k,i-1}\boldsymbol{X}^{(\mathrm{a})\mathrm{T}}_{k,i}\right)^{-1}\boldsymbol{X}^{(\mathrm{a})}_{k,i}\boldsymbol{R}_{k,i-1}.$$

Results

Metrics:
Average accuracy, last task accuracy and forgetting rate.

Results:
Competitive on coarse-grained datasets.

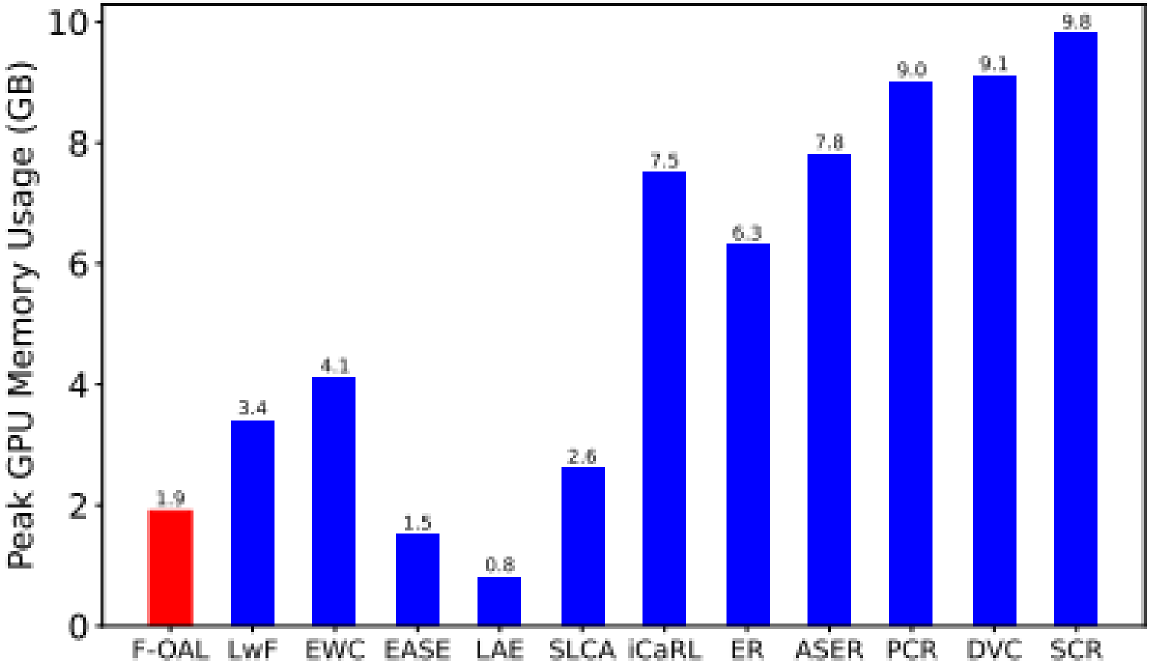
Dominant on fine-grained datasets.

Metric	Method	Replay?	CIFAR-100	CORe50	FGCVAircraft	DTD	Tiny-ImageNet	Country211
$A_{avg}(\%)\uparrow$	iCaRL(CVPR 2017)[31]	✓	91.6	95.6	36.4	74.1	91.3	12.2
	ER(ICRA 2019)[12]	✓	90.1	94.8	35.7	65.4	87.3	14.0
	ASER(AAAI 2021)[33]	✓	87.2	87.1	25.2	57.4	85.8	13.2
	SCR(CVPR 2021)[25]	✓	91.9	95.3	55.6	75.0	82.6	14.7
	DVC(CVPR 2022)[9]	✓	92.4	97.1	33.7	67.3	91.5	16.1
	PCR(CVPR 2023)[21]	✓	89.1	95.7	10.1	35.0	91.0	9.7
	LwF(TPAMI 2018)[20]	✗	69.3	47.0	14.2	40.2	82.5	1.4
	EWC(PNAS 2017)[17]	✗	49.9	47.9	12.0	27.6	60.5	6.1
	EASE(CVPR 2024)[42]	✗	91.1	85.0	38.2	76.0	**92.0**	15.9
	LAE(ICCV 2023)[8]	✗	79.1	73.3	13.5	63.5	86.7	14.5
	SLCA(ICCV 2023)[40]	✗	90.4	93.7	34.3	70.9	88.6	17.8
	F-OAL	✗	**91.1**	**96.3**	**62.2**	**82.8**	91.2	**24.4**
$A_{last}(\%)\uparrow$	iCaRL(CVPR 2017)	✓	87.5	93.2	29.8	66.3	87.8	6.5
	ER(ICRA 2019)	✓	84.6	92.1	28.6	54.3	81.6	6.8
	ASER(AAAI 2021)	✓	82.0	82.1	14.8	49.4	80.0	6.7
	SCR(CVPR 2021)	✓	87.7	93.6	50.3	68.7	75.8	8.0
	DVC(CVPR 2022)	✓	87.8	96.0	27.0	57.2	87.2	9.2
	PCR(CVPR 2023)	✓	81.4	93.9	9.0	34.6	86.1	6.1
	LwF(TPAMI 2018)	✗	64.8	26.3	5.8	18.3	72.5	0.5
	EWC(PNAS 2017)	✗	25.2	21.1	3.0	13.3	44.6	1.9
	EASE(CVPR 2024)	✗	85.4	78.3	29.3	67.6	**89.3**	10.5
	LAE(ICCV 2023)	✗	75.6	67.1	6.3	53.6	82.4	9.3
	SLCA(ICCV 2023)	✗	85.6	88.2	32.1	63.3	85.4	12.9
	F-OAL	✗	**86.5**	**92.5**	**54.0**	**75.9**	87.3	**17.5**
$F(\%)\downarrow$	iCaRL(CVPR 2017)	✓	3.2	2.3	7.1	7.8	2.7	6.7
	ER(ICRA 2019)	✓	20.7	4.3	34.0	29.8	13.3	21.0
	ASER(AAAI 2021)	✓	16.5	9.9	35.7	29.3	16.3	19.4
	SCR(CVPR 2021)	✓	6.2	3.8	14.5	11.6	7.7	6.4
	DVC(CVPR 2022)	✓	8.2	2.3	29.7	21.7	8.9	18.9
	PCR(CVPR 2023)	✓	9.2	4.2	2.7	1.4	8.0	1.7
	LwF(TPAMI 2018)	✗	**1.3**	**0.4**	**3.1**	**4.5**	**1.0**	**0**
	EWC(PNAS 2017)	✗	67.4	81.0	38.8	68.3	20.7	51.5
	EASE(CVPR 2024)	✗	6.1	10.7	19.2	12.5	2.8	16.8
	LAE(ICCV 2023)	✗	11.8	13.8	12.2	25.0	5.4	16.7
	SLCA(ICCV 2023)	✗	7.1	3.4	10.2	12.7	4.2	14.9
	F-OAL	✗	5.5	3.9	10.0	10.1	5.0	6.9

Results

- **Fast training time**: Less trainable parameters, only R and W
- **Low GPU footprint**: no gradients needed.

Methods	CIFAR-100	CORe50	FGVCAircraft	DTD	Tiny-ImageNet	Country211
LwF	412	877	135	73	841	256
EWC	451	922	115	62	1,190	249
iCaRL	832	1,716	53	24	1671	513
ER	652	1,433	40	21	1,315	404
ASER	5,608	7,700	91	43	18,597	20,611
SCR	2,843	5,939	88	42	62,996	810
DVC	4,191	9,351	287	130	10,940	2,622
PCR	1,624	3,742	113	53	3,274	1,028
EASE	383	760	147	139	638	304
LAE	**252**	**458**	156	140	**500**	355
SLCA	726	1,416	289	278	1,185	551
F-OAL	261	570	**16**	**8**	507	**157**



Thank you!