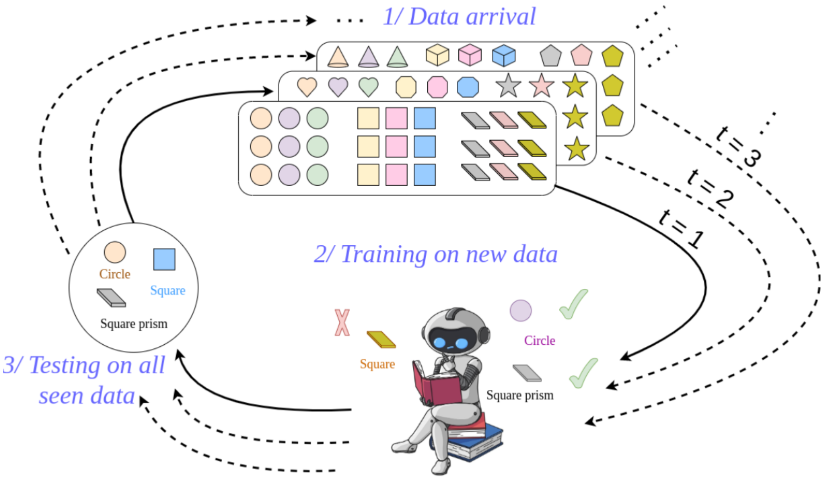
Large-Scale Deep Class-Incremental Learning

Eden Belouadah — Ph.D Defence

Jury:

Nicu Sebe	Professor	University of Trento
David Filliat	Professor	ENSTA Paris
Tinne Tuytelaars	Professor	Katholieke Universiteit
Céline Hudelot	HDR/MDC	CentraleSupélec
Symeon Papadopoulos	Senior Researcher	CERTH-ITI
Adrian POPESCU	Research Engineer	CEA-LIST
Ioannis KANELLOS	Professor	IMT Atlantique

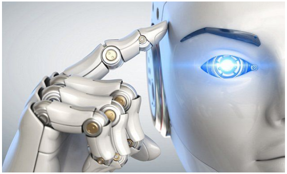
IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom



Continual Learning

Application examples of continual learning

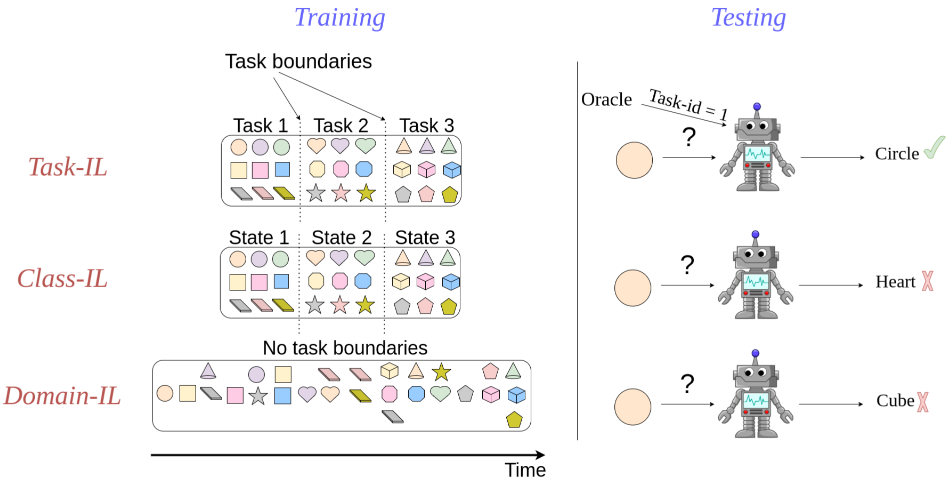
Face Recognition



Robotics

Health

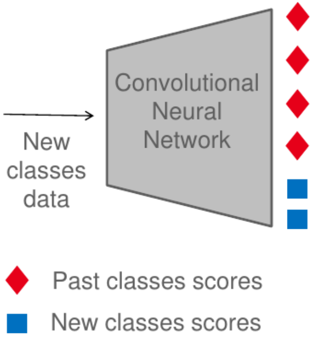
[Images taken from the web]

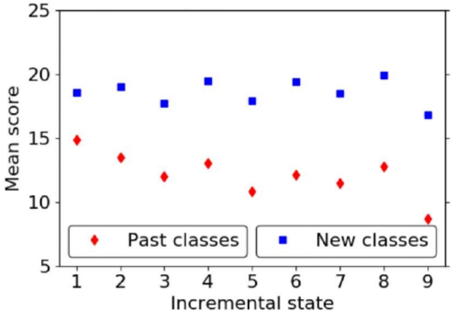


Three Scenarios of Continual Learning

Catastrophic forgetting

- Tendency of neural networks to underfit past data when new one is ingested



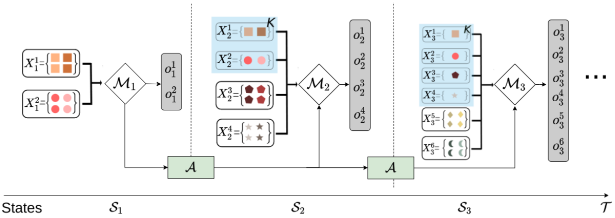


CIL process notations			**Model-related notations in state $\mathcal{S}_t$**		
$\mathcal{A}$	-	CIL algorithm	$\mathcal{M}_t$	-	Model in state $\mathcal{S}_t$
$\mathcal{T}$	-	Total number of states	$\mathcal{F}_t$	-	Feature extractor
$\mathcal{S}_t$	-	State	$\mathcal{C}_t$	-	Classification component
P_t	-	Number of new classes	$\boldsymbol{\mathcal{W}_t}$	-	Last layer weights matrix
N_t	-	Total number of classes until state $\mathcal{S}_t$	$\boldsymbol{W_t^j}$	-	weitgh vector of class j
$\mathcal{K}$	-	Exemplars memory	$\boldsymbol{o_t}$	-	Raw output scores vector
$\mathcal{I}$	-	Secondary memory	o_t^j	-	score of class j

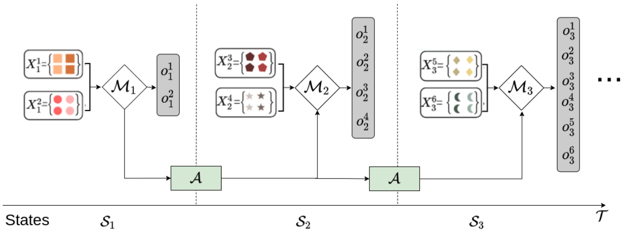
Introduction

Two scenarios of Class-Incremental Learning

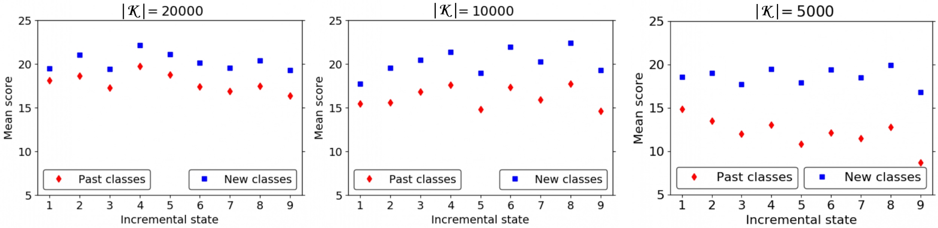
- Incremental Learning with memory



- Incremental Learning without memory



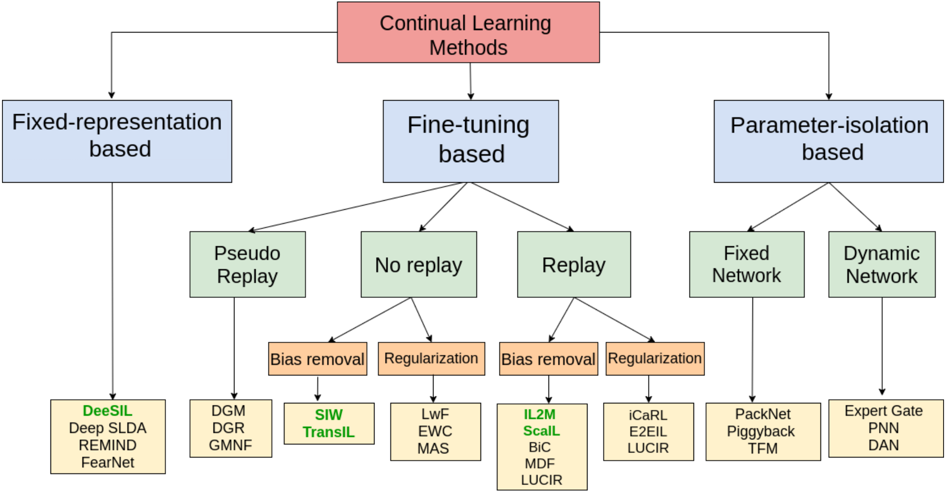
Reduce prediction bias towards new classes



Mean prediction scores of past and new classes with vanilla fine tuning

Table of Contents

State of the art



[Schema inspired by Lange et al., 2019]

[References are in appendix slides]

- **Complexity*** - capacity to integrate new information with a minimal change in terms of the model structure
- **Memory*** - ability to work with or without a fixed-size memory of past classes
- **Accuracy*** - performance for past and new classes should approach that of a non-incremental learning process that has access to all data at all times.
- **Timeliness** - delay needed between the occurrence of new data and its integration in the incremental models
- **Plasticity-Stability** - capacity not only to deal with new classes that are significantly different from the ones learned in the past but also to keep as much knowledge as possible from the past.
- **Scalability** - the aptitude for learning a large number of classes

** [Inspired by Rebuffi et al., 2017]*

State of the art
Pros and Cons

	Complexity	Memory	Accuracy	Timeliness	Plasticity-Stability	Scalability
Fine tuning based	Increases slowly	The bigger the memory, the better the model	Best SoTA results with memory	Retraining is needed at each state	Depends on the availability of the memory	Depends on the size of the memory
Fixed representation based	Increases slowly	Low dependency	Good if the initial model is trained on large dataset	Fast	Bad if incremental classes are different from the initial ones	Heavily depends on the fixed representation
Parameter isolation based	Depends if fixed or dynamic network	Non compulsory	Depends on how much the model architecture can increase	Retraining is needed at each state	Good	Scale well if resources are available

LwF: Learning without Forgetting (Li and Hoiem, 2016)

- Fine Tuning with distillation loss

$$\mathcal{L}_t^d(x) = \sum_{(x,y)\ \in\ \mathcal{D}_t} \sum_{j=1}^{N_{t-1}} -\hat{\sigma}_{t-1}^j(x)\ log[\hat{\sigma}_t^j(x)] \tag{1}$$

where $\hat{\sigma}$ is the softened softmax

- (+) No memory of the past is needed
- (−) The gap with a *Joint* training is large

iCaRL: Incremental Classifier and Representation Learning (Rebuffi et al., 2017)

- *LwF* with memory
- Herding to select exemplars

$$e \leftarrow \underset{x \in X}{\arg\min} \|\mu - \frac{1}{k}[f(x) + \sum_{j=1}^{k-1} f(e_j)]\| \quad (2)$$

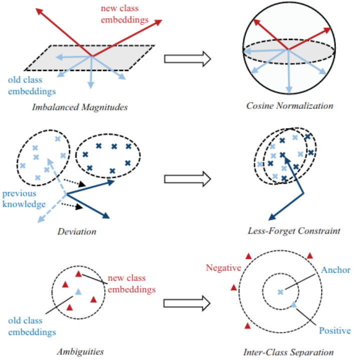
- Nearest Exemplars Mean (NEM)

$$y^* = \underset{y \in [1, N_t]}{\arg\min} \|f(x) - \mu_y\| \quad (3)$$

  - (+) Combination of powerful components
  - (−) Unfair comparison with baselines

LUCIR: Learning a Unified Classifier Incrementally via Rebalancing (Hou et al., 2019)

- Cosine normalization
- Less-forget constraint
- Inter-class separation



- (+) Powerful objective
- (−) Important execution time

BiC: Bias Correction (Wu et al., 2019)

- Distillation loss
- Bias-removal layer

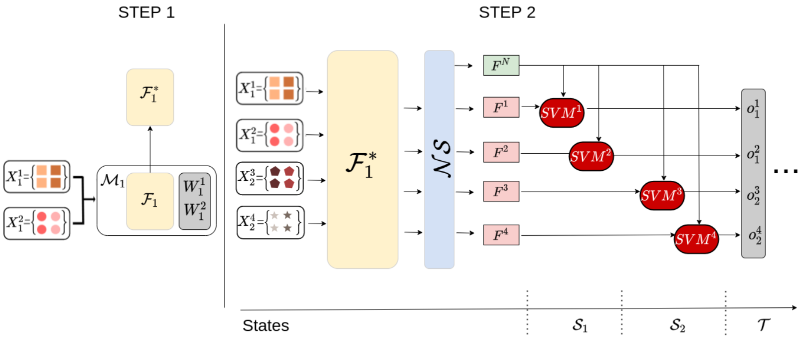
$$BiC(\boldsymbol{o}_t^k) = \begin{cases} \boldsymbol{o}_t^k & \text{if} \quad k \in [1,\ t-1] \\ \alpha_t \boldsymbol{o}_t^k + \beta_t \cdot 1 & \text{if} \quad k = t \end{cases} \tag{4}$$

  - (+) Simple, fast, and accurate
  - (−) Uses a validation set (memory required)

2. Class-incremental learning with memory

DeeSIL: Deep-Shallow Incremental Learning
(Belouadah and Popescu, 2018, W-ECCV)

- Fixed Representation based
- Inspired by transfer learning
- Works with and without memory

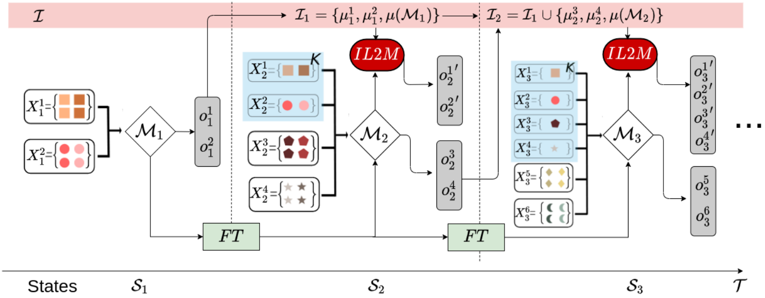


Overview of *DeeSIL*

IL2M: Incremental Learning with Dual memory

(Belouadah and Popescu, 2019, ICCV)

- Fine Tuning based
- Leverages past class statistics



Overview of *IL2M*

- Past class score rectification

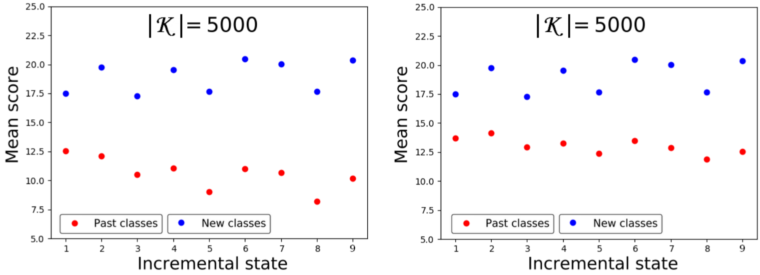
For all past classes ($j = 1, \ldots, N_{t-1}$):

$$o_t^{j\prime} = IL2M(o_t^j) = \begin{cases} o_t^j \times \frac{\mu_i^j}{\mu_t^j} \times \frac{\mu(\mathcal{M}_t)}{\mu(\mathcal{M}_i)}, & \text{if } pred = new \\ o_t^j, & \text{otherwise} \end{cases} \tag{5}$$

with:

- i - the initial state in which the j^{th} class was learned
- t - the current incremental state
- o_t^j - the raw prediction the j^{th} class in the current state t
- μ_i^j and μ_t^j - the mean classification scores of the j^{th} class in states i and t
- $\mu(\mathcal{M}_t)$ and $\mu(\mathcal{M}_i)$ - the model confidences in states t and i

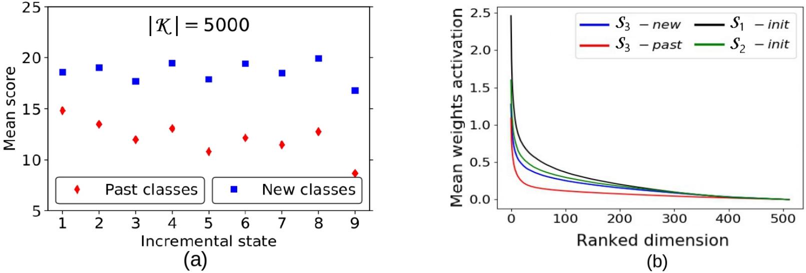
- Effect of *IL2M*



Before   After

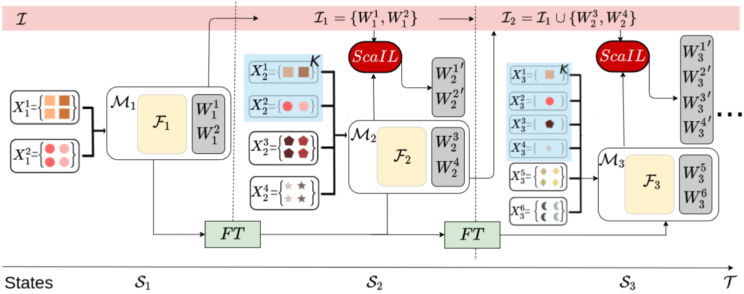
ScaIL: Classifier weights Scaling for Class IL
(Belouadah and Popescu, 2020, WACV)

- Fine Tuning based
- Forgetting happens mainly in the final layer
- Features are usable across incremental states



Mean prediction scores and weights magnitudes

- Past class weights replay



Overview of *ScaIL*

- Sort class weights

$$\widehat{\boldsymbol{W}_{\boldsymbol{t}}^{\boldsymbol{j}}} = sort(|w_j^1|, |w_j^2|, ..., |w_j^d|, ..., |w_j^D|) \ \ ; j \in [N_{t-1}, N_t] \, , d \in [1, D] \quad (6)$$

$\widehat{\boldsymbol{W}_{\boldsymbol{t}}^{\boldsymbol{j}}}$ is the sorted version of the initial weights vector of new class j.

- Compute state mean vector

$$\mu_t^d = \frac{1}{P_t} \times \sum_{j=N_{t-1}}^{N_t} \widehat{w}_j^d \qquad d \in [1, D] \quad (7)$$

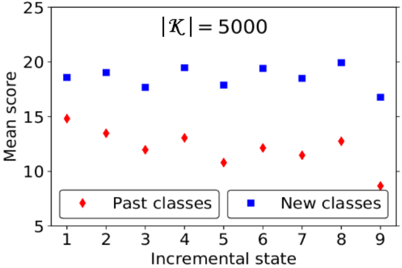
where μ_t (of dimension D) is the mean vector of the ranked new classes' weights in the state $\mathcal{S}_t$, and d is a dimension in the feature vector.

- Normalize past class weights

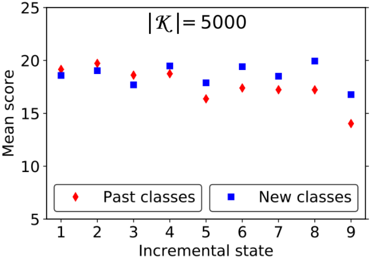
$$w_j^{d'} = \frac{\mu_t^{R(d)}}{\mu_i^{R(d)}} \times w_j^d \quad (8)$$

$w_j^{d'}$ is the scaled version of w_j^d, the d^{th} dimension of the initial classifier W_i^j of the j^{th} past class.

- Effect of *ScaIL*



Before



After

FT+th: Fine tuning with threshold calibration (Belouadah et al., 2020a, W-ECCV)

• Inspired by an imbalanced-learning approach proposed by Buda et al., 2018

• Increases the classification score of the least represented classes (past and new)

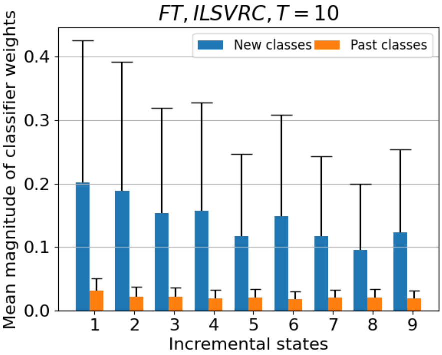
$$p_t^{j\prime} = p_t^j \times \frac{|\mathcal{X}_t \cup \mathcal{K}|}{|X_t^j|} \tag{9}$$

- $|X_t^j|$ is the number of training examples for the class j in state $\mathcal{S}_t$
- $|\mathcal{X}_t^l \cup \mathcal{K}|$ is the total number of training examples in state $\mathcal{S}_t$

2. Class-incremental learning without memory

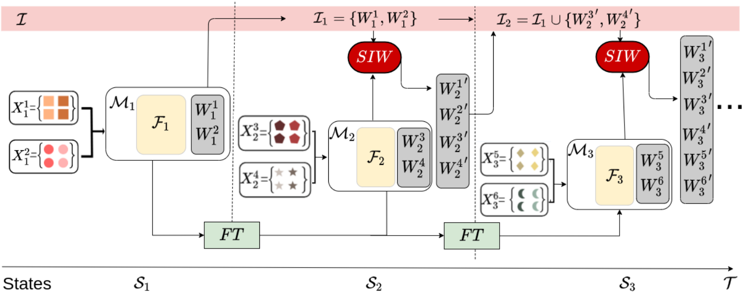
SIW: Standardization of Initial Weights for Class Incremental Learning (Belouadah et al., 2020b, BMVC)

- Fine Tuning based
- Bias in the mean weights magnitudes



Mean weights magnitudes without memory

- Past class weights replay

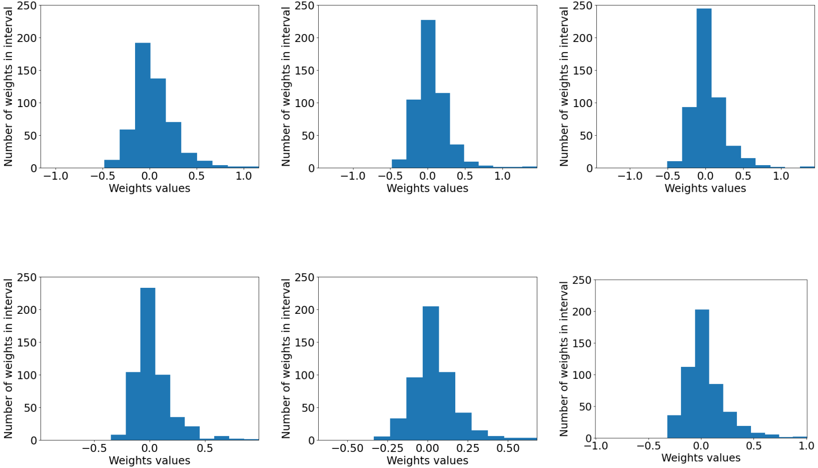


Overview of *SIW*

Class-Incremental Learning without memory

SIW: Standardization of Initial Weights for Class IL (BMVC 2020)

- Classifier weights distribution



- Standardization of Initial Weights

$$w'_d = \frac{w_d - \mu(\boldsymbol{W})}{\sigma(\boldsymbol{W})} \tag{10}$$

with:

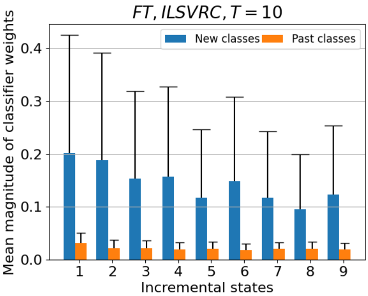
w_d is the d^{th} dimension of an initial classifier $\boldsymbol{W}$, $\mu(\boldsymbol{W})$ and $\sigma(\boldsymbol{W})$ are the mean and standard deviation of $\boldsymbol{W}$.

- State-level calibration

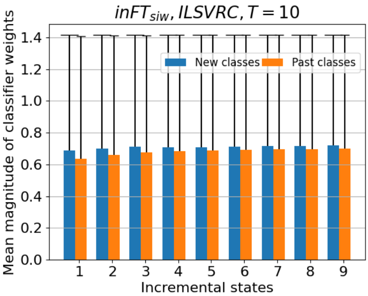
$$o_t^j(x) = (\boldsymbol{f_t}(\boldsymbol{x}) \cdot \boldsymbol{W_t^j}' + b_j^i) \times \frac{\mu(\mathcal{M}_t)}{\mu(\mathcal{M}_i)} \tag{11}$$

$\mu(\mathcal{M}_t)$ and $\mu(\mathcal{M}_i)$ are means of top-1 predictions of models learned in the t^{th} and i^{th} states

- Effect of *SIW* on weights magnitudes



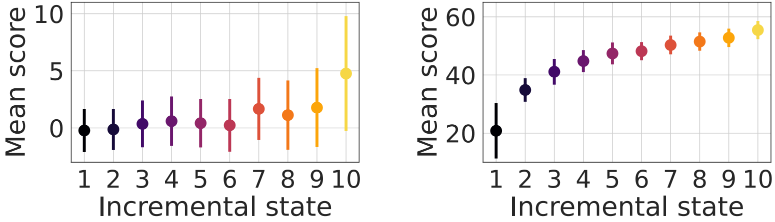
Before



After

TransIL: Dataset Knowledge Transfer for Class IL
(Slim et al., 2022, WACV)

- Fine Tuning based
- Bias in the mean classification scores after *LwF* and *LUCIR*



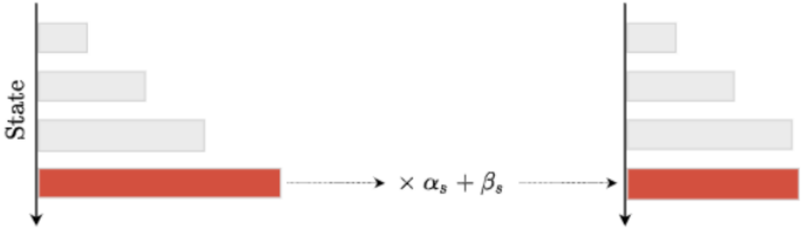
LwF (Li and Hoiem, 2016) LUCIR (Hou et al., 2019)

Mean prediction scores and standard deviation

Enable the use of Bias Correction layers in a memoryless scenario

- *BiC* : Bias Correction (Wu et al., 2019)

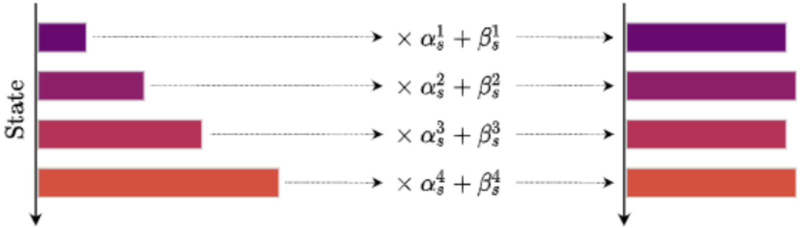
$$BiC(\boldsymbol{o}_t^k) = \begin{cases} \boldsymbol{o}_t^k & \text{if} \quad k \in [1, \ t-1] \\ \alpha_t \boldsymbol{o}_t^k + \beta_t \cdot 1 & \text{if} \quad k = t \end{cases} \tag{12}$$



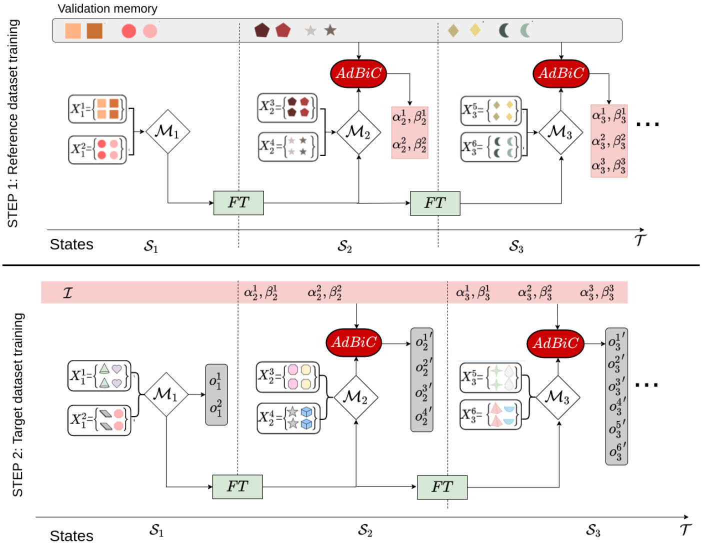
- *adBiC* : Adaptive Bias Correction (proposed)

$$adBiC(\boldsymbol{o_t^k}) = \alpha_t^k \boldsymbol{o_t^k} + \beta_t^k \cdot 1 \; ; \quad k \in [1, t] \qquad (13)$$

where α_t^k, β_t^k are the parameters applied in state $\mathcal{S}_t$ to classes first learned in state $\mathcal{S}_k$.



- Dataset knowledge transfer



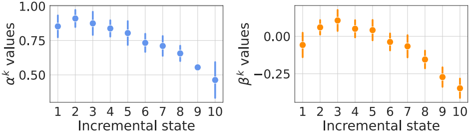
Overview of *TransIL*

Class-Incremental Learning without memory
TransIL: Dataset Knowledge Transfer for Class IL (WACV 2022)

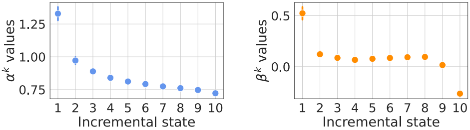
- Optimization of calibration parameters on reference datasets

$$\mathcal{L}(\boldsymbol{q_t}, y) = -\sum_{k=1}^{t}\sum_{j=1}^{|P_k|} \delta_{y=\widehat{y}} \log\left(q_{t,j}^k\right)$$

- y is the ground-truth label
- $\widehat{y}$ is the predicted label
- δ is the Kronecker delta
- $\boldsymbol{q_t}$ is the corrected softmax output of the sample



LwF (Li and Hoiem, 2016)



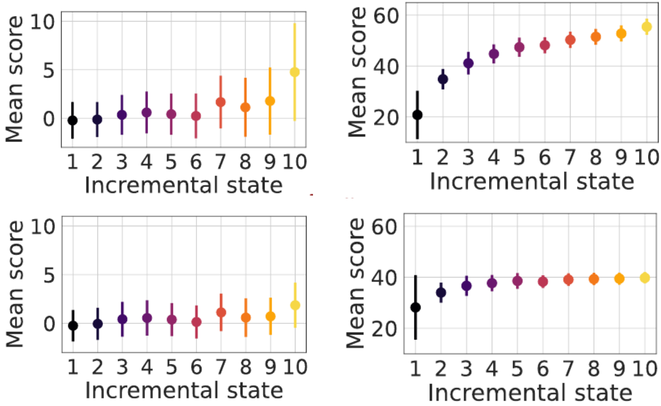
LUCIR (Hou et al., 2019)

Averaged calibration parameters values

- Effect on classification scores

Before

After

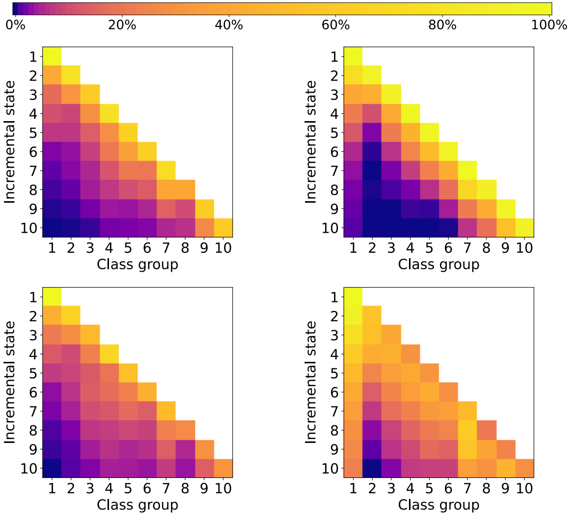


LwF (Li and Hoiem, 2016) *LUCIR* (Hou et al., 2019)

- Effect on the percentage of correct and wrong predictions

Before

After



LwF (Li and Hoiem, 2016)   *LUCIR* (Hou et al., 2019)

Experiments and Results

- Evaluation of *DeeSIL*, *IL2M*, *ScaIL*, FT^{th} and *SIW*

Dataset	#Train	#Test	#Classes	μ(train)	σ(train)
ILSVRC (Russakovsky et al., 2015)	1,231,167	50,000	1,000	1231.2	70.2
VGGFACE2 (Cao et al., 2018)	491,746	50,000	1,000	491.7	49.4
LANDMARKS (Noh et al., 2017)	374,367	20,000	1,000	374.4	103.8
CIFAR-100 (Krizhevsky, 2009)	50,000	10,000	100	500.00	0.00

Summary of the datasets used for evaluation

- Architecture: a ResNet-18 network
- Memory size : $|\mathcal{K}| = \{2\%, 1\%, 0.5\%\}$ of the training set, and no memory.
- Number of states: $\mathcal{T} = \{10, 20, 50\}$

- Evaluation of *TransIL*

10 Reference datasets

10 random 100 leaf classes from ImageNet (Deng et al., 2009)	
Train	500 images per class
Val	200 images per class

4 Test datasets

CIFAR-100 (Krizhevsky, 2009), IMN-100 (Deng et al., 2009), BIRDS-100 (Deng et al., 2009), FOOD-100 (Bossard et al., 2014)	
Train	500 images per class
Test	100 images per class

- Architecture: a ResNet-18 network
- No memory of the past
- Number of states: $\mathcal{T} = \{5, 10, 20\}$

Experiments and Results
Experimental protocol

- Baselines - with memory

Method	works without memory?
FT	✓
FR	✓
iCaRL (Rebuffi et al., 2017)	*LwF* (Li and Hoiem, 2016)
LUCIR (Hou et al., 2019)	✓
BiC (Wu et al., 2019)	×
REMIND (Hayes et al., 2019)	✓

- Baselines - without memory
  - *FT*, *FR*, *LwF* (Li and Hoiem, 2016), *LUCIR* (Hou et al., 2019), *REMIND* (Hayes et al., 2019)
  - *FT+* (Masana et al., 2021)
  - *Deep-SLDA* (Hayes and Kanan, 2019)

- Plugins applied on top of *FT*
  - **init** - use of initial classifiers of past classes (used in *ScaIL* and *SIW*)
  - **L2** - L2 normalization of the weights matrix
  - **mc** - mean state calibration (used in *IL2M* and *SIW*)
  - **th** - threshold calibration (Buda et al., 2018)
  - **BAL** - balanced fine tuning (Castro et al., 2018)
  - **NEM** - nearest exemplars mean (Rebuffi et al., 2017)
- Upper bound of Class IL
  - **Joint** - full training with all data

- Evaluation metrics
  - Average incremental accuracy (Castro et al., 2018)
  - G_{IL} aggregation measure

$$G_{IL} = \frac{1}{C} \times \sum_{c=1}^{C} \frac{A(c) - A(Joint)}{A_{max} - A(Joint)} \tag{14}$$

C - number of tested configurations; $A(c)$ - accuracy of each configuration ; $A(Joint)$ - accuracy of *Joint* ; $A_{max} = 100$

Top-5 results with herding ($\mathcal{T}$ fixed, $|\mathcal{K}|$ variable)

States	$\mathcal{T} = 10$											
Dataset	ILSVRC			VGGFACE2			LANDMARKS			CIFAR-100		
$\|\mathcal{K}\|$	2%	1%	0.5%	2%	1%	0.5%	2%	1%	0.5%	2%	1%	0.5%
iCaRL	79.3	76.5	71.0	96.0	95.3	93.9	95.1	94.0	91.8	66.5	56.1	47.9
FT	79.4	74.4	65.9	96.4	94.5	91.3	96.6	94.7	91.4	82.4	77.9	70.7
FT+init	79.2	76.5	73.0	95.9	95.2	94.6	97.0	95.5	92.7	83.4	80.5	75.2
FT+NEM	81.4	79.0	75.0	96.4	95.4	94.0	96.1	94.6	92.6	85.1	81.7	76.0
FT+BAL	84.0	80.9	76.5	97.0	95.7	92.4	96.9	95.3	92.2	80.0	74.0	69.0
FT+th	84.3	82.1	78.3	97.2	96.3	94.8	97.2	95.8	94.0	86.4	83.9	79.1
LUCIR	79.9	76.4	72.6	97.2	**96.9**	**96.5**	97.2	96.6	96.1	79.8	75.4	69.9
BiC	**85.5**	**82.8**	**79.7**	**97.3**	96.6	95.7	**97.9**	**97.3**	**96.6**	**88.8**	**87.6**	**83.5**
ScaIL	82.0	79.8	76.6	96.5	95.8	95.2	97.3	96.0	94.0	85.6	83.2	79.1
IL2M	80.9	78.1	73.9	96.7	95.4	93.4	96.5	94.7	92.5	81.8	77.0	71.2
FR	76.7	76.6	76.4	91.7	91.5	89.7	93.8	93.5	93.5	79.5	79.4	78.7
DeeSIL	75.5	75.1	74.3	92.7	92.5	92.2	94.0	93.7	93.2	66.9	65.8	64.2
REMIND	80.9	80.7	78.2	94.7	93.2	93.0	96.3	95.8	94.7	60.7	60.7	60.7
Joint	92.3			99.2			99.1			91.2		

Top-5 results with herding ($\mathcal{T}$ fixed, $|\mathcal{K}|$ variable)

States	$\mathcal{T} = 10$											
Dataset	ILSVRC			VGGFACE2			LANDMARKS			CIFAR-100		
$\lvert\mathcal{K}\rvert$	2%	1%	0.5%	2%	1%	0.5%	2%	1%	0.5%	2%	1%	0.5%
iCaRL	79.3	76.5	71.0	96.0	95.3	93.9	95.1	94.0	91.8	66.5	56.1	47.9
FT	79.4	74.4	65.9	96.4	94.5	91.3	96.6	94.7	91.4	82.4	77.9	70.7
FT+init	79.2	76.5	73.0	95.9	95.2	94.6	97.0	95.5	92.7	83.4	80.5	75.2
FT+NEM	81.4	79.0	75.0	96.4	95.4	94.0	96.1	94.6	92.6	85.1	81.7	76.0
FT+BAL	84.0	80.9	76.5	97.0	95.7	92.4	96.9	95.3	92.2	80.0	74.0	69.0
FT+th	84.3	82.1	78.3	97.2	96.3	94.8	97.2	95.8	94.0	86.4	83.9	79.1
LUCIR	79.9	76.4	72.6	97.2	**96.9**	**96.5**	97.2	96.6	96.1	79.8	75.4	69.9
BiC	**85.5**	**82.8**	**79.7**	**97.3**	96.6	95.7	**97.9**	**97.3**	**96.6**	**88.8**	**87.6**	**83.5**
ScaIL	82.0	79.8	76.6	96.5	95.8	95.2	97.3	96.0	94.0	85.6	83.2	79.1
IL2M	80.9	78.1	73.9	96.7	95.4	93.4	96.5	94.7	92.5	81.8	77.0	71.2
FR	76.7	76.6	76.4	91.7	91.5	89.7	93.8	93.5	93.5	79.5	79.4	78.7
DeeSIL	75.5	75.1	74.3	92.7	92.5	92.2	94.0	93.7	93.2	66.9	65.8	64.2
REMIND	80.9	80.7	78.2	94.7	93.2	93.0	96.3	95.8	94.7	60.7	60.7	60.7
Joint	92.3			99.2			99.1			91.2		

Top-5 results with herding ($\mathcal{T}$ fixed, $|\mathcal{K}|$ variable)

States	$\mathcal{T} = 10$											
Dataset	ILSVRC			VGGFACE2			LANDMARKS			CIFAR-100		
$\lvert\mathcal{K}\rvert$	2%	1%	0.5%	2%	1%	0.5%	2%	1%	0.5%	2%	1%	0.5%
iCaRL	79.3	76.5	71.0	96.0	95.3	93.9	95.1	94.0	91.8	66.5	56.1	47.9
FT	79.4	74.4	65.9	96.4	94.5	91.3	96.6	94.7	91.4	82.4	77.9	70.7
FT+init	79.2	76.5	73.0	95.9	95.2	94.6	97.0	95.5	92.7	83.4	80.5	75.2
FT+NEM	81.4	79.0	75.0	96.4	95.4	94.0	96.1	94.6	92.6	85.1	81.7	76.0
FT+BAL	84.0	80.9	76.5	97.0	95.7	92.4	96.9	95.3	92.2	80.0	74.0	69.0
FT+th	84.3	82.1	78.3	97.2	96.3	94.8	97.2	95.8	94.0	86.4	83.9	79.1
LUCIR	79.9	76.4	72.6	97.2	**96.9**	**96.5**	97.2	96.6	96.1	79.8	75.4	69.9
BiC	**85.5**	**82.8**	**79.7**	**97.3**	96.6	95.7	**97.9**	**97.3**	**96.6**	**88.8**	**87.6**	**83.5**
ScaIL	82.0	79.8	76.6	96.5	95.8	95.2	97.3	96.0	94.0	85.6	83.2	79.1
IL2M	80.9	78.1	73.9	96.7	95.4	93.4	96.5	94.7	92.5	81.8	77.0	71.2
FR	76.7	76.6	76.4	91.7	91.5	89.7	93.8	93.5	93.5	79.5	79.4	78.7
DeeSIL	75.5	75.1	74.3	92.7	92.5	92.2	94.0	93.7	93.2	66.9	65.8	64.2
REMIND	80.9	80.7	78.2	94.7	93.2	93.0	96.3	95.8	94.7	60.7	60.7	60.7
Joint	92.3			99.2			99.1			91.2		

Top-5 results with herding ($\mathcal{T}$ fixed, $|\mathcal{K}|$ variable)

States	$\mathcal{T} = 10$											
Dataset	ILSVRC			VGGFACE2			LANDMARKS			CIFAR-100		
$\lvert\mathcal{K}\rvert$	2%	1%	0.5%	2%	1%	0.5%	2%	1%	0.5%	2%	1%	0.5%
iCaRL	79.3	76.5	71.0	96.0	95.3	93.9	95.1	94.0	91.8	66.5	56.1	47.9
FT	79.4	74.4	65.9	96.4	94.5	91.3	96.6	94.7	91.4	82.4	77.9	70.7
FT+init	79.2	76.5	73.0	95.9	95.2	94.6	97.0	95.5	92.7	83.4	80.5	75.2
FT+NEM	81.4	79.0	75.0	96.4	95.4	94.0	96.1	94.6	92.6	85.1	81.7	76.0
FT+BAL	84.0	80.9	76.5	97.0	95.7	92.4	96.9	95.3	92.2	80.0	74.0	69.0
FT+th	84.3	82.1	78.3	97.2	96.3	94.8	97.2	95.8	94.0	86.4	83.9	79.1
LUCIR	79.9	76.4	72.6	97.2	**96.9**	**96.5**	97.2	96.6	96.1	79.8	75.4	69.9
BiC	**85.5**	**82.8**	**79.7**	**97.3**	96.6	95.7	**97.9**	**97.3**	**96.6**	**88.8**	**87.6**	**83.5**
ScaIL	82.0	79.8	76.6	96.5	95.8	95.2	97.3	96.0	94.0	85.6	83.2	79.1
IL2M	80.9	78.1	73.9	96.7	95.4	93.4	96.5	94.7	92.5	81.8	77.0	71.2
FR	76.7	76.6	76.4	91.7	91.5	89.7	93.8	93.5	93.5	79.5	79.4	78.7
DeeSIL	75.5	75.1	74.3	92.7	92.5	92.2	94.0	93.7	93.2	66.9	65.8	64.2
REMIND	80.9	80.7	78.2	94.7	93.2	93.0	96.3	95.8	94.7	60.7	60.7	60.7
Joint	92.3			99.2			99.1			91.2		

Top-5 results with herding ($\mathcal{T}$ fixed, $|\mathcal{K}|$ variable)

States	$\mathcal{T}=10$											
Dataset	ILSVRC			VGGFACE2			LANDMARKS			CIFAR-100		
$\|\mathcal{K}\|$	2%	1%	0.5%	2%	1%	0.5%	2%	1%	0.5%	2%	1%	0.5%
iCaRL	79.3	76.5	71.0	96.0	95.3	93.9	95.1	94.0	91.8	66.5	56.1	47.9
FT	79.4	74.4	65.9	96.4	94.5	91.3	96.6	94.7	91.4	82.4	77.9	70.7
FT+init	79.2	76.5	73.0	95.9	95.2	94.6	97.0	95.5	92.7	83.4	80.5	75.2
FT+NEM	81.4	79.0	75.0	96.4	95.4	94.0	96.1	94.6	92.6	85.1	81.7	76.0
FT+BAL	84.0	80.9	76.5	97.0	95.7	92.4	96.9	95.3	92.2	80.0	74.0	69.0
FT+th	84.3	82.1	78.3	97.2	96.3	94.8	97.2	95.8	94.0	86.4	83.9	79.1
LUCIR	79.9	76.4	72.6	97.2	**96.9**	**96.5**	97.2	96.6	96.1	79.8	75.4	69.9
BiC	**85.5**	**82.8**	**79.7**	**97.3**	96.6	95.7	**97.9**	**97.3**	**96.6**	**88.8**	**87.6**	**83.5**
ScaIL	82.0	79.8	76.6	96.5	95.8	95.2	97.3	96.0	94.0	85.6	83.2	79.1
IL2M	80.9	78.1	73.9	96.7	95.4	93.4	96.5	94.7	92.5	81.8	77.0	71.2
FR	76.7	76.6	76.4	91.7	91.5	89.7	93.8	93.5	93.5	79.5	79.4	78.7
DeeSIL	75.5	75.1	74.3	92.7	92.5	92.2	94.0	93.7	93.2	66.9	65.8	64.2
REMIND	80.9	80.7	78.2	94.7	93.2	93.0	96.3	95.8	94.7	60.7	60.7	60.7
Joint	92.3			99.2			99.1			91.2		

Top-5 results with herding ($\mathcal{T}$ fixed, $|\mathcal{K}|$ variable)

States	$\mathcal{T} = 10$											
Dataset	ILSVRC			VGGFACE2			LANDMARKS			CIFAR-100		
$\lvert\mathcal{K}\rvert$	2%	1%	0.5%	2%	1%	0.5%	2%	1%	0.5%	2%	1%	0.5%
iCaRL	79.3	76.5	71.0	96.0	95.3	93.9	95.1	94.0	91.8	66.5	56.1	47.9
FT	79.4	74.4	65.9	96.4	94.5	91.3	96.6	94.7	91.4	82.4	77.9	70.7
FT+init	79.2	76.5	73.0	95.9	95.2	94.6	97.0	95.5	92.7	83.4	80.5	75.2
FT+NEM	81.4	79.0	75.0	96.4	95.4	94.0	96.1	94.6	92.6	85.1	81.7	76.0
FT+BAL	84.0	80.9	76.5	97.0	95.7	92.4	96.9	95.3	92.2	80.0	74.0	69.0
FT+th	84.3	82.1	78.3	97.2	96.3	94.8	97.2	95.8	94.0	86.4	83.9	79.1
LUCIR	79.9	76.4	72.6	97.2	**96.9**	**96.5**	97.2	96.6	96.1	79.8	75.4	69.9
BiC	**85.5**	**82.8**	**79.7**	**97.3**	96.6	95.7	**97.9**	**97.3**	**96.6**	**88.8**	**87.6**	**83.5**
ScaIL	82.0	79.8	76.6	96.5	95.8	95.2	97.3	96.0	94.0	85.6	83.2	79.1
IL2M	80.9	78.1	73.9	96.7	95.4	93.4	96.5	94.7	92.5	81.8	77.0	71.2
FR	76.7	76.6	76.4	91.7	91.5	89.7	93.8	93.5	93.5	79.5	79.4	78.7
DeeSIL	75.5	75.1	74.3	92.7	92.5	92.2	94.0	93.7	93.2	66.9	65.8	64.2
REMIND	80.9	80.7	78.2	94.7	93.2	93.0	96.3	95.8	94.7	60.7	60.7	60.7
Joint	92.3			99.2			99.1			91.2		

Top-5 results with herding ($|\mathcal{K}|$ fixed, $\mathcal{T}$ variable)

States	$\|\mathcal{K}\| = 0.5\%$								
Dataset	ILSVRC		VGGFACE2		LANDMARKS		CIFAR-100		G_{IL}
$\mathcal{T}$	20	50	20	50	20	50	20	50	
iCaRL	55.9	45.0	88.5	78.2	86.8	82.4	35.5	35.4	-7.36
FT	69.4	64.3	91.6	89.2	90.9	89.0	64.3	54.8	-5.19
FT+init	73.6	67.3	94.6	91.4	91.2	88.5	63.6	44.1	-4.43
FT+NEM	76.5	69.0	94.0	91.1	91.9	89.9	68.8	55.9	-4.28
FT+BAL	75.9	67.1	92.3	89.5	91.2	88.9	62.9	54.2	-4.70
FT+th	**78.6**	**71.2**	94.3	91.6	92.9	**90.7**	**71.4**	**57.9**	**-3.62**
LUCIR	63.9	55.3	93.5	88.3	93.7	90.5	53.5	47.9	-4.13
BiC	74.6	63.9	92.3	85.3	**94.7**	90.5	50.5	19.6	-4.03
ScaIL	76.6	70.9	**95.0**	**92.4**	92.6	90.4	69.8	51.0	-3.70
IL2M	70.9	60.6	92.5	88.4	90.8	88.1	61.5	51.0	-4.95
FR	69.2	58.2	85.8	75.2	89.3	82.8	62.3	33.5	-7.62
DeeSIL	73.0	58.1	87.2	80.0	90.5	85.1	63.9	44.0	-6.92
REMIND	73.9	65.0	87.4	80.1	92.8	88.6	52.8	46.4	-6.02
Joint	92.3		99.2		99.1		91.2		-

Top-5 results with herding ($|\mathcal{K}|$ fixed, $\mathcal{T}$ variable)

States	$\|\mathcal{K}\| = 0.5\%$								
Dataset	ILSVRC		VGGFACE2		LANDMARKS		CIFAR-100		G_{IL}
$\mathcal{T}$	20	50	20	50	20	50	20	50	
iCaRL	55.9	45.0	88.5	78.2	86.8	82.4	35.5	35.4	-7.36
FT	69.4	64.3	91.6	89.2	90.9	89.0	64.3	54.8	-5.19
FT+init	73.6	67.3	94.6	91.4	91.2	88.5	63.6	44.1	-4.43
FT+NEM	76.5	69.0	94.0	91.1	91.9	89.9	68.8	55.9	-4.28
FT+BAL	75.9	67.1	92.3	89.5	91.2	88.9	62.9	54.2	-4.70
FT+th	**78.6**	**71.2**	94.3	91.6	92.9	**90.7**	**71.4**	**57.9**	**-3.62**
LUCIR	63.9	55.3	93.5	88.3	93.7	90.5	53.5	47.9	-4.13
BiC	74.6	63.9	92.3	85.3	**94.7**	90.5	50.5	19.6	-4.03
ScaIL	76.6	70.9	**95.0**	**92.4**	92.6	90.4	69.8	51.0	-3.70
IL2M	70.9	60.6	92.5	88.4	90.8	88.1	61.5	51.0	-4.95
FR	69.2	58.2	85.8	75.2	89.3	82.8	62.3	33.5	-7.62
DeeSIL	73.0	58.1	87.2	80.0	90.5	85.1	63.9	44.0	-6.92
REMIND	73.9	65.0	87.4	80.1	92.8	88.6	52.8	46.4	-6.02
Joint	92.3		99.2		99.1		91.2		-

Top-5 results with herding ($|\mathcal{K}|$ fixed, $\mathcal{T}$ variable)

States	$\|\mathcal{K}\| = 0.5\%$								
Dataset	ILSVRC		VGGFACE2		LANDMARKS		CIFAR-100		G_{IL}
$\mathcal{T}$	20	50	20	50	20	50	20	50	
iCaRL	55.9	45.0	88.5	78.2	86.8	82.4	35.5	35.4	-7.36
FT	69.4	64.3	91.6	89.2	90.9	89.0	64.3	54.8	-5.19
FT+init	73.6	67.3	94.6	91.4	91.2	88.5	63.6	44.1	-4.43
FT+NEM	76.5	69.0	94.0	91.1	91.9	89.9	68.8	55.9	-4.28
FT+BAL	75.9	67.1	92.3	89.5	91.2	88.9	62.9	54.2	-4.70
FT+th	**78.6**	**71.2**	94.3	91.6	92.9	**90.7**	**71.4**	**57.9**	**-3.62**
LUCIR	63.9	55.3	93.5	88.3	93.7	90.5	53.5	47.9	-4.13
BiC	74.6	63.9	92.3	85.3	**94.7**	90.5	50.5	19.6	-4.03
ScaIL	76.6	70.9	**95.0**	**92.4**	92.6	90.4	69.8	51.0	-3.70
IL2M	70.9	60.6	92.5	88.4	90.8	88.1	61.5	51.0	-4.95
FR	69.2	58.2	85.8	75.2	89.3	82.8	62.3	33.5	-7.62
DeeSIL	73.0	58.1	87.2	80.0	90.5	85.1	63.9	44.0	-6.92
REMIND	73.9	65.0	87.4	80.1	92.8	88.6	52.8	46.4	-6.02
Joint	92.3		99.2		99.1		91.2		-

Top-5 results with herding ($|\mathcal{K}|$ fixed, $\mathcal{T}$ variable)

States	$\|\mathcal{K}\| = 0.5\%$								
Dataset	ILSVRC		VGGFACE2		LANDMARKS		CIFAR-100		G_{IL}
$\mathcal{T}$	20	50	20	50	20	50	20	50	
iCaRL	55.9	45.0	88.5	78.2	86.8	82.4	35.5	35.4	-7.36
FT	69.4	64.3	91.6	89.2	90.9	89.0	64.3	54.8	-5.19
FT+init	73.6	67.3	94.6	91.4	91.2	88.5	63.6	44.1	-4.43
FT+NEM	76.5	69.0	94.0	91.1	91.9	89.9	68.8	55.9	-4.28
FT+BAL	75.9	67.1	92.3	89.5	91.2	88.9	62.9	54.2	-4.70
FT+th	**78.6**	**71.2**	94.3	91.6	92.9	**90.7**	**71.4**	**57.9**	**-3.62**
LUCIR	63.9	55.3	93.5	88.3	93.7	90.5	53.5	47.9	-4.13
BiC	74.6	63.9	92.3	85.3	**94.7**	90.5	50.5	19.6	-4.03
ScaIL	76.6	70.9	**95.0**	**92.4**	92.6	90.4	69.8	51.0	-3.70
IL2M	70.9	60.6	92.5	88.4	90.8	88.1	61.5	51.0	-4.95
FR	69.2	58.2	85.8	75.2	89.3	82.8	62.3	33.5	-7.62
DeeSIL	73.0	58.1	87.2	80.0	90.5	85.1	63.9	44.0	-6.92
REMIND	73.9	65.0	87.4	80.1	92.8	88.6	52.8	46.4	-6.02
Joint	92.3		99.2		99.1		91.2		-

Top-5 results with herding ($|\mathcal{K}|$ fixed, $\mathcal{T}$ variable)

States	$\lvert\mathcal{K}\rvert = 0.5\%$								G_{IL}
Dataset	ILSVRC		VGGFACE2		LANDMARKS		CIFAR-100		
$\mathcal{T}$	20	50	20	50	20	50	20	50	
iCaRL	55.9	45.0	88.5	78.2	86.8	82.4	35.5	35.4	-7.36
FT	69.4	64.3	91.6	89.2	90.9	89.0	64.3	54.8	-5.19
FT+init	73.6	67.3	94.6	91.4	91.2	88.5	63.6	44.1	-4.43
FT+NEM	76.5	69.0	94.0	91.1	91.9	89.9	68.8	55.9	-4.28
FT+BAL	75.9	67.1	92.3	89.5	91.2	88.9	62.9	54.2	-4.70
FT+th	**78.6**	**71.2**	94.3	91.6	92.9	**90.7**	**71.4**	**57.9**	**-3.62**
LUCIR	63.9	55.3	93.5	88.3	93.7	90.5	53.5	47.9	-4.13
BiC	74.6	63.9	92.3	85.3	**94.7**	90.5	50.5	19.6	-4.03
ScaIL	76.6	70.9	**95.0**	**92.4**	92.6	90.4	69.8	51.0	-3.70
IL2M	70.9	60.6	92.5	88.4	90.8	88.1	61.5	51.0	-4.95
FR	69.2	58.2	85.8	75.2	89.3	82.8	62.3	33.5	-7.62
DeeSIL	73.0	58.1	87.2	80.0	90.5	85.1	63.9	44.0	-6.92
REMIND	73.9	65.0	87.4	80.1	92.8	88.6	52.8	46.4	-6.02
Joint	92.3		99.2		99.1		91.2		-

Top-5 results without memory

Dataset	ILSVRC			VGGFACE2			LANDMARKS			CIFAR-100			G_{IL}
States $\mathcal{T}$	10	20	50	10	20	50	10	20	50	10	20	50	
FT	20.6	13.4	7.1	21.3	13.6	7.1	21.3	13.6	7.1	21.3	13.7	17.4	-54.91
LwF	45.3	37.6	27.1	53.3	42.6	30.8	58.8	49.2	35.2	79.5	65.3	39.0	-34.72
FT+init	61.0	44.9	23.8	90.9	64.4	33.1	68.8	49.4	22.2	55.1	40.8	19.9	-28.99
FT+init+L2	51.6	43.3	34.5	76.8	66.8	55.1	61.4	52.5	39.2	47.5	39.3	22.5	-26.80
FT+init+L2+mc	53.6	42.7	35.6	86.9	71.4	53.6	66.2	52.6	37.9	52.6	43.1	18.2	**-25.02**
SIW(FT)	64.4	54.3	41.4	88.6	84.1	62.6	79.5	64.5	43.2	59.7	44.3	18.4	-19.38
SIW(LwF)	54.0	45.8	35.1	70.4	59.3	45.2	61.0	53.8	42.2	**80.0**	**68.8**	44.6	-28.06
LUCIR	57.6	39.4	21.9	91.4	68.2	32.2	87.8	63.7	32.3	57.5	35.3	21.0	-24.75
FR	74.0	66.9	49.2	88.7	83.0	54.4	**93.6**	88.1	71.2	73.1	54.8	27.4	-16.30
DeeSIL	**73.9**	**67.5**	53.9	**92.3**	**87.5**	75.1	**93.6**	**91.1**	**82.1**	65.2	63.4	32.3	**-9.22**
REMIND	62.2	56.3	44.4	86.8	81.4	69.2	84.5	79.6	69.0	52.7	40.5	25.7	-22.00
Deep-SLDA	70.3	64.5	**56.0**	90.2	85.4	**78.2**	89.3	86.4	81.3	68.9	64.4	**54.5**	-15.40
Joint	92.3			99.2			99.1			91.2			-

Top-5 results without memory

Dataset	ILSVRC			VGGFACE2			LANDMARKS			CIFAR-100			G_{IL}
States $\mathcal{T}$	10	20	50	10	20	50	10	20	50	10	20	50	
FT	20.6	13.4	7.1	21.3	13.6	7.1	21.3	13.6	7.1	21.3	13.7	17.4	-54.91
LwF	45.3	37.6	27.1	53.3	42.6	30.8	58.8	49.2	35.2	79.5	65.3	39.0	-34.72
FT+init	61.0	44.9	23.8	90.9	64.4	33.1	68.8	49.4	22.2	55.1	40.8	19.9	-28.99
FT+init+L2	51.6	43.3	34.5	76.8	66.8	55.1	61.4	52.5	39.2	47.5	39.3	22.5	-26.80
FT+init+L2+mc	53.6	42.7	35.6	86.9	71.4	53.6	66.2	52.6	37.9	52.6	43.1	18.2	**-25.02**
SIW(FT)	64.4	54.3	41.4	88.6	84.1	62.6	79.5	64.5	43.2	59.7	44.3	18.4	-19.38
SIW(LwF)	54.0	45.8	35.1	70.4	59.3	45.2	61.0	53.8	42.2	**80.0**	**68.8**	44.6	-28.06
LUCIR	57.6	39.4	21.9	91.4	68.2	32.2	87.8	63.7	32.3	57.5	35.3	21.0	-24.75
FR	74.0	66.9	49.2	88.7	83.0	54.4	**93.6**	88.1	71.2	73.1	54.8	27.4	-16.30
DeeSIL	**73.9**	**67.5**	53.9	**92.3**	**87.5**	75.1	**93.6**	**91.1**	**82.1**	65.2	63.4	32.3	**-9.22**
REMIND	62.2	56.3	44.4	86.8	81.4	69.2	84.5	79.6	69.0	52.7	40.5	25.7	-22.00
Deep-SLDA	70.3	64.5	**56.0**	90.2	85.4	**78.2**	89.3	86.4	81.3	68.9	64.4	**54.5**	-15.40
Joint	92.3			99.2			99.1			91.2			-

Top-5 results without memory

Dataset	ILSVRC			VGGFACE2			LANDMARKS			CIFAR-100			G_{IL}
States $\mathcal{T}$	10	20	50	10	20	50	10	20	50	10	20	50	
FT	20.6	13.4	7.1	21.3	13.6	7.1	21.3	13.6	7.1	21.3	13.7	17.4	-54.91
LwF	45.3	37.6	27.1	53.3	42.6	30.8	58.8	49.2	35.2	79.5	65.3	39.0	-34.72
FT+init	61.0	44.9	23.8	90.9	64.4	33.1	68.8	49.4	22.2	55.1	40.8	19.9	-28.99
FT+init+L2	51.6	43.3	34.5	76.8	66.8	55.1	61.4	52.5	39.2	47.5	39.3	22.5	-26.80
FT+init+L2+mc	53.6	42.7	35.6	86.9	71.4	53.6	66.2	52.6	37.9	52.6	43.1	18.2	**-25.02**
SIW(FT)	64.4	54.3	41.4	88.6	84.1	62.6	79.5	64.5	43.2	59.7	44.3	18.4	-19.38
SIW(LwF)	54.0	45.8	35.1	70.4	59.3	45.2	61.0	53.8	42.2	**80.0**	**68.8**	44.6	-28.06
LUCIR	57.6	39.4	21.9	91.4	68.2	32.2	87.8	63.7	32.3	57.5	35.3	21.0	-24.75
FR	74.0	66.9	49.2	88.7	83.0	54.4	**93.6**	88.1	71.2	73.1	54.8	27.4	-16.30
DeeSIL	**73.9**	**67.5**	53.9	**92.3**	**87.5**	75.1	**93.6**	**91.1**	**82.1**	65.2	63.4	32.3	**-9.22**
REMIND	62.2	56.3	44.4	86.8	81.4	69.2	84.5	79.6	69.0	52.7	40.5	25.7	-22.00
Deep-SLDA	70.3	64.5	**56.0**	90.2	85.4	**78.2**	89.3	86.4	81.3	68.9	64.4	**54.5**	-15.40
Joint	92.3			99.2			99.1			91.2			-

Top-5 results without memory

Dataset	ILSVRC			VGGFACE2			LANDMARKS			CIFAR-100			G_{IL}
States $\mathcal{T}$	10	20	50	10	20	50	10	20	50	10	20	50	
FT	20.6	13.4	7.1	21.3	13.6	7.1	21.3	13.6	7.1	21.3	13.7	17.4	-54.91
LwF	45.3	37.6	27.1	53.3	42.6	30.8	58.8	49.2	35.2	79.5	65.3	39.0	-34.72
FT+init	61.0	44.9	23.8	90.9	64.4	33.1	68.8	49.4	22.2	55.1	40.8	19.9	-28.99
FT+init+L2	51.6	43.3	34.5	76.8	66.8	55.1	61.4	52.5	39.2	47.5	39.3	22.5	-26.80
FT+init+L2+mc	53.6	42.7	35.6	86.9	71.4	53.6	66.2	52.6	37.9	52.6	43.1	18.2	**-25.02**
SIW(FT)	64.4	54.3	41.4	88.6	84.1	62.6	79.5	64.5	43.2	59.7	44.3	18.4	-19.38
SIW(LwF)	54.0	45.8	35.1	70.4	59.3	45.2	61.0	53.8	42.2	**80.0**	**68.8**	44.6	-28.06
LUCIR	57.6	39.4	21.9	91.4	68.2	32.2	87.8	63.7	32.3	57.5	35.3	21.0	-24.75
FR	74.0	66.9	49.2	88.7	83.0	54.4	**93.6**	88.1	71.2	73.1	54.8	27.4	-16.30
DeeSIL	**73.9**	**67.5**	53.9	**92.3**	**87.5**	75.1	**93.6**	**91.1**	**82.1**	65.2	63.4	32.3	**-9.22**
REMIND	62.2	56.3	44.4	86.8	81.4	69.2	84.5	79.6	69.0	52.7	40.5	25.7	-22.00
Deep-SLDA	70.3	64.5	**56.0**	90.2	85.4	**78.2**	89.3	86.4	81.3	68.9	64.4	**54.5**	-15.40
Joint	92.3			99.2			99.1			91.2			-

Top-5 results without memory

Dataset	ILSVRC			VGGFACE2			LANDMARKS			CIFAR-100			G_{IL}
States $\mathcal{T}$	10	20	50	10	20	50	10	20	50	10	20	50	
FT	20.6	13.4	7.1	21.3	13.6	7.1	21.3	13.6	7.1	21.3	13.7	17.4	-54.91
LwF	45.3	37.6	27.1	53.3	42.6	30.8	58.8	49.2	35.2	79.5	65.3	39.0	-34.72
FT+init	61.0	44.9	23.8	90.9	64.4	33.1	68.8	49.4	22.2	55.1	40.8	19.9	-28.99
FT+init+L2	51.6	43.3	34.5	76.8	66.8	55.1	61.4	52.5	39.2	47.5	39.3	22.5	-26.80
FT+init+L2+mc	53.6	42.7	35.6	86.9	71.4	53.6	66.2	52.6	37.9	52.6	43.1	18.2	**-25.02**
SIW(FT)	64.4	54.3	41.4	88.6	84.1	62.6	79.5	64.5	43.2	59.7	44.3	18.4	-19.38
SIW(LwF)	54.0	45.8	35.1	70.4	59.3	45.2	61.0	53.8	42.2	**80.0**	**68.8**	44.6	-28.06
LUCIR	57.6	39.4	21.9	91.4	68.2	32.2	87.8	63.7	32.3	57.5	35.3	21.0	-24.75
FR	74.0	66.9	49.2	88.7	83.0	54.4	**93.6**	88.1	71.2	73.1	54.8	27.4	-16.30
DeeSIL	**73.9**	**67.5**	53.9	**92.3**	**87.5**	75.1	**93.6**	**91.1**	**82.1**	65.2	63.4	32.3	**-9.22**
REMIND	62.2	56.3	44.4	86.8	81.4	69.2	84.5	79.6	69.0	52.7	40.5	25.7	-22.00
Deep-SLDA	70.3	64.5	**56.0**	90.2	85.4	**78.2**	89.3	86.4	81.3	68.9	64.4	**54.5**	-15.40
Joint	92.3			99.2			99.1			91.2			-

Top-5 results without memory

Dataset	ILSVRC			VGGFACE2			LANDMARKS			CIFAR-100			G_{IL}
States $\mathcal{T}$	10	20	50	10	20	50	10	20	50	10	20	50	
FT	20.6	13.4	7.1	21.3	13.6	7.1	21.3	13.6	7.1	21.3	13.7	17.4	-54.91
LwF	45.3	37.6	27.1	53.3	42.6	30.8	58.8	49.2	35.2	79.5	65.3	39.0	-34.72
FT+init	61.0	44.9	23.8	90.9	64.4	33.1	68.8	49.4	22.2	55.1	40.8	19.9	-28.99
FT+init+L2	51.6	43.3	34.5	76.8	66.8	55.1	61.4	52.5	39.2	47.5	39.3	22.5	-26.80
FT+init+L2+mc	53.6	42.7	35.6	86.9	71.4	53.6	66.2	52.6	37.9	52.6	43.1	18.2	**-25.02**
SIW(FT)	64.4	54.3	41.4	88.6	84.1	62.6	79.5	64.5	43.2	59.7	44.3	18.4	-19.38
SIW(LwF)	54.0	45.8	35.1	70.4	59.3	45.2	61.0	53.8	42.2	**80.0**	**68.8**	44.6	-28.06
LUCIR	57.6	39.4	21.9	91.4	68.2	32.2	87.8	63.7	32.3	57.5	35.3	21.0	-24.75
FR	74.0	66.9	49.2	88.7	83.0	54.4	**93.6**	88.1	71.2	73.1	54.8	27.4	-16.30
DeeSIL	**73.9**	**67.5**	53.9	**92.3**	**87.5**	75.1	**93.6**	**91.1**	**82.1**	65.2	63.4	32.3	**-9.22**
REMIND	62.2	56.3	44.4	86.8	81.4	69.2	84.5	79.6	69.0	52.7	40.5	25.7	-22.00
Deep-SLDA	70.3	64.5	**56.0**	90.2	85.4	**78.2**	89.3	86.4	81.3	68.9	64.4	**54.5**	-15.40
Joint	92.3			99.2			99.1			91.2			-

Top-1 results without memory

Method	CIFAR-100			BIRDS-100			FOOD-100		
States	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$
LwF	53.0	44.0	29.1	53.7	41.8	30.1	42.9	31.8	22.2
w/ BiC	54.0 + 1.0	45.5 + 1.5	30.8 + 1.7	54.6 + 0.9	43.1 + 1.3	31.8 + 1.7	43.4 + 0.5	32.6 + 0.8	23.8 + 1.6
w/ AdBiC	54.3 + 1.3	**46.4** + 2.4	**32.3** + 3.2	55.0 + 1.3	44.0 + 2.2	**32.8** + 2.7	43.5 + 0.6	33.3 + 1.5	**24.7** + 2.5
w/ AdBiC + $\mathbb{O}$	54.9 + 1.9	47.3 + 3.3	32.6 + 3.5	55.8 + 2.1	44.8 + 3.0	33.3 + 3.2	44.0 + 1.1	34.2 + 2.4	25.3 + 3.1
LUCIR	50.1	33.7	19.5	50.8	31.4	17.9	44.2	26.4	15.5
w/ BiC	52.5 + 2.4	37.1 + 3.4	22.4 + 2.9	56.0 + 5.2	37.7 + 6.3	20.6 + 2.7	49.9 + 5.7	31.5 + 5.1	17.2 + 1.7
w/ AdBiC	**54.8** + 4.7	42.2 + 8.5	28.4 + 8.9	**58.5** + 7.7	**45.4** + 14.0	27.3 + 9.4	**52.0** + 7.8	**37.1** + 10.7	17.7 + 2.2
w/ AdBiC + $\mathbb{O}$	55.5 + 5.4	43.6 + 9.9	31.2 + 11.7	59.0 + 8.2	46.0 + 14.6	28.8 + 10.9	52.6 + 8.4	38.2 + 11.8	21.0 + 5.5
SIW	29.9	22.7	14.8	30.6	23.2	14.9	29.4	21.6	14.1
w/ BiC	31.4 + 1.5	22.8 + 0.1	14.7 - 0.1	32.8 + 2.2	22.7 - 0.5	12.8 - 2.1	29.1 - 0.3	20.3 - 1.3	12.1 - 2.0
w/ AdBiC	31.7 + 1.8	24.1 + 1.4	15.8 + 1.0	33.0 + 2.4	25.2 + 2.0	15.3 + 0.4	30.9 + 1.5	21.3 - 0.3	14.5 + 0.4
w/ AdBiC + $\mathbb{O}$	32.8 + 2.9	25.0 + 2.3	16.5 + 1.7	34.4 + 3.8	26.2 + 3.0	16.3 + 1.4	31.5 + 2.1	22.6 + 1.0	15.1 + 1.0
FT+	28.9	22.6	14.5	29.7	23.3	13.5	28.7	21.1	13.3
w/ BiC	30.7 + 1.8	22.5 - 0.1	14.8 + 0.3	32.3 + 2.6	22.5 - 0.8	12.4 - 1.1	28.6 - 0.1	20.6 - 0.5	11.8 - 1.5
w/ AdBiC	31.9 + 3.0	23.6 + 1.0	15.0 + 0.5	34.0 + 4.3	25.0 + 1.7	14.2 + 0.7	30.8 + 2.1	22.2 + 1.1	14.2 + 0.9
w/ AdBiC + $\mathbb{O}$	32.5 + 3.6	24.6 + 2.0	15.9 + 1.4	34.5 + 4.8	25.7 + 2.4	15.4 + 1.9	31.3 + 2.6	22.7 + 1.6	14.5 + 1.2
Joint	72.7			80.9			71.03		

gains, losses

Results and discussion

Class-Incremental Learning without memory

Top-1 results without memory

Method / States	CIFAR-100 $\mathcal{T}=5$	CIFAR-100 $\mathcal{T}=10$	CIFAR-100 $\mathcal{T}=20$	BIRDS-100 $\mathcal{T}=5$	BIRDS-100 $\mathcal{T}=10$	BIRDS-100 $\mathcal{T}=20$	FOOD-100 $\mathcal{T}=5$	FOOD-100 $\mathcal{T}=10$	FOOD-100 $\mathcal{T}=20$
LwF	53.0	44.0	29.1	53.7	41.8	30.1	42.9	31.8	22.2
w/ BiC	54.0 + 1.0	45.5 + 1.5	30.8 + 1.7	54.6 + 0.9	43.1 + 1.3	31.8 + 1.7	43.4 + 0.5	32.6 + 0.8	23.8 + 1.6
w/ AdBiC	54.3 + 1.3	**46.4** + 2.4	**32.3** + 3.2	55.0 + 1.3	44.0 + 2.2	**32.8** + 2.7	43.5 + 0.6	33.3 + 1.5	**24.7** + 2.5
w/ AdBiC + $\mathbb{O}$	54.9 + 1.9	47.3 + 3.3	32.6 + 3.5	55.8 + 2.1	44.8 + 3.0	33.3 + 3.2	44.0 + 1.1	34.2 + 2.4	25.3 + 3.1
LUCIR	50.1	33.7	19.5	50.8	31.4	17.9	44.2	26.4	15.5
w/ BiC	52.5 + 2.4	37.1 + 3.4	22.4 + 2.9	56.0 + 5.2	37.7 + 6.3	20.6 + 2.7	49.9 + 5.7	31.5 + 5.1	17.2 + 1.7
w/ AdBiC	**54.8** + 4.7	42.2 + 8.5	28.4 + 8.9	**58.5** + 7.7	**45.4** + 14.0	27.3 + 9.4	**52.0** + 7.8	**37.1** + 10.7	17.7 + 2.2
w/ AdBiC + $\mathbb{O}$	55.5 + 5.4	43.6 + 9.9	31.2 + 11.7	59.0 + 8.2	46.0 + 14.6	28.8 + 10.9	52.6 + 8.4	38.2 + 11.8	21.0 + 5.5
SIW	29.9	22.7	14.8	30.6	23.2	14.9	29.4	21.6	14.1
w/ BiC	31.4 + 1.5	22.8 + 0.1	14.7 - 0.1	32.8 + 2.2	22.7 - 0.5	12.8 - 2.1	29.1 - 0.3	20.3 - 1.3	12.1 - 2.0
w/ AdBiC	31.7 + 1.8	24.1 + 1.4	15.8 + 1.0	33.0 + 2.4	25.2 + 2.0	15.3 + 0.4	30.9 + 1.5	21.3 - 0.3	14.5 + 0.4
w/ AdBiC + $\mathbb{O}$	32.8 + 2.9	25.0 + 2.3	16.5 + 1.7	34.4 + 3.8	26.2 + 3.0	16.3 + 1.4	31.5 + 2.1	22.6 + 1.0	15.1 + 1.0
FT+	28.9	22.6	14.5	29.7	23.3	13.5	28.7	21.1	13.3
w/ BiC	30.7 + 1.8	22.5 - 0.1	14.8 + 0.3	32.3 + 2.6	22.5 - 0.8	12.4 - 1.1	28.6 - 0.1	20.6 - 0.5	11.8 - 1.5
w/ AdBiC	31.9 + 3.0	23.6 + 1.0	15.0 + 0.5	34.0 + 4.3	25.0 + 1.7	14.2 + 0.7	30.8 + 2.1	22.2 + 1.1	14.2 + 0.9
w/ AdBiC + $\mathbb{O}$	32.5 + 3.6	24.6 + 2.0	15.9 + 1.4	34.5 + 4.8	25.7 + 2.4	15.4 + 1.9	31.3 + 2.6	22.7 + 1.6	14.5 + 1.2
Joint	72.7			80.9			71.03		

gains, losses

Top-1 results without memory

Method	CIFAR-100			BIRDS-100			FOOD-100		
States	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$
LwF	53.0	44.0	29.1	53.7	41.8	30.1	42.9	31.8	22.2
w/ BiC	54.0 + 1.0	45.5 + 1.5	30.8 + 1.7	54.6 + 0.9	43.1 + 1.3	31.8 + 1.7	43.4 + 0.5	32.6 + 0.8	23.8 + 1.6
w/ AdBiC	54.3 + 1.3	**46.4** + 2.4	**32.3** + 3.2	55.0 + 1.3	44.0 + 2.2	**32.8** + 2.7	43.5 + 0.6	33.3 + 1.5	**24.7** + 2.5
w/ AdBiC + $\mathbb{O}$	54.9 + 1.9	47.3 + 3.3	32.6 + 3.5	55.8 + 2.1	44.8 + 3.0	33.3 + 3.2	44.0 + 1.1	34.2 + 2.4	25.3 + 3.1
LUCIR	50.1	33.7	19.5	50.8	31.4	17.9	44.2	26.4	15.5
w/ BiC	52.5 + 2.4	37.1 + 3.4	22.4 + 2.9	56.0 + 5.2	37.7 + 6.3	20.6 + 2.7	49.9 + 5.7	31.5 + 5.1	17.2 + 1.7
w/ AdBiC	**54.8** + 4.7	42.2 + 8.5	28.4 + 8.9	**58.5** + 7.7	**45.4** + 14.0	27.3 + 9.4	**52.0** + 7.8	**37.1** + 10.7	17.7 + 2.2
w/ AdBiC + $\mathbb{O}$	55.5 + 5.4	43.6 + 9.9	31.2 + 11.7	59.0 + 8.2	46.0 + 14.6	28.8 + 10.9	52.6 + 8.4	38.2 + 11.8	21.0 + 5.5
SIW	29.9	22.7	14.8	30.6	23.2	14.9	29.4	21.6	14.1
w/ BiC	31.4 + 1.5	22.8 + 0.1	14.7 - 0.1	32.8 + 2.2	22.7 - 0.5	12.8 - 2.1	29.1 - 0.3	20.3 - 1.3	12.1 - 2.0
w/ AdBiC	31.7 + 1.8	24.1 + 1.4	15.8 + 1.0	33.0 + 2.4	25.2 + 2.0	15.3 + 0.4	30.9 + 1.5	21.3 - 0.3	14.5 + 0.4
w/ AdBiC + $\mathbb{O}$	32.8 + 2.9	25.0 + 2.3	16.5 + 1.7	34.4 + 3.8	26.2 + 3.0	16.3 + 1.4	31.5 + 2.1	22.6 + 1.0	15.1 + 1.0
FT+	28.9	22.6	14.5	29.7	23.3	13.5	28.7	21.1	13.3
w/ BiC	30.7 + 1.8	22.5 - 0.1	14.8 + 0.3	32.3 + 2.6	22.5 - 0.8	12.4 - 1.1	28.6 - 0.1	20.6 - 0.5	11.8 - 1.5
w/ AdBiC	31.9 + 3.0	23.6 + 1.0	15.0 + 0.5	34.0 + 4.3	25.0 + 1.7	14.2 + 0.7	30.8 + 2.1	22.2 + 1.1	14.2 + 0.9
w/ AdBiC + $\mathbb{O}$	32.5 + 3.6	24.6 + 2.0	15.9 + 1.4	34.5 + 4.8	25.7 + 2.4	15.4 + 1.9	31.3 + 2.6	22.7 + 1.6	14.5 + 1.2
Joint	72.7			80.9			71.03		

gains, losses

Top-1 results without memory

Method	CIFAR-100			BIRDS-100			FOOD-100		
States	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$
LwF	53.0	44.0	29.1	53.7	41.8	30.1	42.9	31.8	22.2
w/ BiC	54.0 + 1.0	45.5 + 1.5	30.8 + 1.7	54.6 + 0.9	43.1 + 1.3	31.8 + 1.7	43.4 + 0.5	32.6 + 0.8	23.8 + 1.6
w/ AdBiC	54.3 + 1.3	**46.4** + 2.4	**32.3** + 3.2	55.0 + 1.3	44.0 + 2.2	**32.8** + 2.7	43.5 + 0.6	33.3 + 1.5	**24.7** + 2.5
w/ AdBiC + $\mathbb{O}$	54.9 + 1.9	47.3 + 3.3	32.6 + 3.5	55.8 + 2.1	44.8 + 3.0	33.3 + 3.2	44.0 + 1.1	34.2 + 2.4	25.3 + 3.1
LUCIR	50.1	33.7	19.5	50.8	31.4	17.9	44.2	26.4	15.5
w/ BiC	52.5 + 2.4	37.1 + 3.4	22.4 + 2.9	56.0 + 5.2	37.7 + 6.3	20.6 + 2.7	49.9 + 5.7	31.5 + 5.1	17.2 + 1.7
w/ AdBiC	**54.8** + 4.7	42.2 + 8.5	28.4 + 8.9	**58.5** + 7.7	**45.4** + 14.0	27.3 + 9.4	**52.0** + 7.8	**37.1** + 10.7	17.7 + 2.2
w/ AdBiC + $\mathbb{O}$	55.5 + 5.4	43.6 + 9.9	31.2 + 11.7	59.0 + 8.2	46.0 + 14.6	28.8 + 10.9	52.6 + 8.4	38.2 + 11.8	21.0 + 5.5
SIW	29.9	22.7	14.8	30.6	23.2	14.9	29.4	21.6	14.1
w/ BiC	31.4 + 1.5	22.8 + 0.1	14.7 - 0.1	32.8 + 2.2	22.7 - 0.5	12.8 - 2.1	29.1 - 0.3	20.3 - 1.3	12.1 - 2.0
w/ AdBiC	31.7 + 1.8	24.1 + 1.4	15.8 + 1.0	33.0 + 2.4	25.2 + 2.0	15.3 + 0.4	30.9 + 1.5	21.3 - 0.3	14.5 + 0.4
w/ AdBiC + $\mathbb{O}$	32.8 + 2.9	25.0 + 2.3	16.5 + 1.7	34.4 + 3.8	26.2 + 3.0	16.3 + 1.4	31.5 + 2.1	22.6 + 1.0	15.1 + 1.0
FT+	28.9	22.6	14.5	29.7	23.3	13.5	28.7	21.1	13.3
w/ BiC	30.7 + 1.8	22.5 - 0.1	14.8 + 0.3	32.3 + 2.6	22.5 - 0.8	12.4 - 1.1	28.6 - 0.1	20.6 - 0.5	11.8 - 1.5
w/ AdBiC	31.9 + 3.0	23.6 + 1.0	15.0 + 0.5	34.0 + 4.3	25.0 + 1.7	14.2 + 0.7	30.8 + 2.1	22.2 + 1.1	14.2 + 0.9
w/ AdBiC + $\mathbb{O}$	32.5 + 3.6	24.6 + 2.0	15.9 + 1.4	34.5 + 4.8	25.7 + 2.4	15.4 + 1.9	31.3 + 2.6	22.7 + 1.6	14.5 + 1.2
Joint	72.7			80.9			71.03		

gains, losses

Top-1 results without memory

Method	CIFAR-100			BIRDS-100			FOOD-100		
States	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$
LwF	53.0	44.0	29.1	53.7	41.8	30.1	42.9	31.8	22.2
w/ BiC	54.0 + 1.0	45.5 + 1.5	30.8 + 1.7	54.6 + 0.9	43.1 + 1.3	31.8 + 1.7	43.4 + 0.5	32.6 + 0.8	23.8 + 1.6
w/ AdBiC	54.3 + 1.3	**46.4** + 2.4	**32.3** + 3.2	55.0 + 1.3	44.0 + 2.2	**32.8** + 2.7	43.5 + 0.6	33.3 + 1.5	**24.7** + 2.5
w/ AdBiC + $\mathbb{O}$	54.9 + 1.9	47.3 + 3.3	32.6 + 3.5	55.8 + 2.1	44.8 + 3.0	33.3 + 3.2	44.0 + 1.1	34.2 + 2.4	25.3 + 3.1
LUCIR	50.1	33.7	19.5	50.8	31.4	17.9	44.2	26.4	15.5
w/ BiC	52.5 + 2.4	37.1 + 3.4	22.4 + 2.9	56.0 + 5.2	37.7 + 6.3	20.6 + 2.7	49.9 + 5.7	31.5 + 5.1	17.2 + 1.7
w/ AdBiC	**54.8** + 4.7	42.2 + 8.5	28.4 + 8.9	**58.5** + 7.7	**45.4** + 14.0	27.3 + 9.4	**52.0** + 7.8	**37.1** + 10.7	17.7 + 2.2
w/ AdBiC + $\mathbb{O}$	55.5 + 5.4	43.6 + 9.9	31.2 + 11.7	59.0 + 8.2	46.0 + 14.6	28.8 + 10.9	52.6 + 8.4	38.2 + 11.8	21.0 + 5.5
SIW	29.9	22.7	14.8	30.6	23.2	14.9	29.4	21.6	14.1
w/ BiC	31.4 + 1.5	22.8 + 0.1	14.7 - 0.1	32.8 + 2.2	22.7 - 0.5	12.8 - 2.1	29.1 - 0.3	20.3 - 1.3	12.1 - 2.0
w/ AdBiC	31.7 + 1.8	24.1 + 1.4	15.8 + 1.0	33.0 + 2.4	25.2 + 2.0	15.3 + 0.4	30.9 + 1.5	21.3 - 0.3	14.5 + 0.4
w/ AdBiC + $\mathbb{O}$	32.8 + 2.9	25.0 + 2.3	16.5 + 1.7	34.4 + 3.8	26.2 + 3.0	16.3 + 1.4	31.5 + 2.1	22.6 + 1.0	15.1 + 1.0
FT+	28.9	22.6	14.5	29.7	23.3	13.5	28.7	21.1	13.3
w/ BiC	30.7 + 1.8	22.5 - 0.1	14.8 + 0.3	32.3 + 2.6	22.5 - 0.8	12.4 - 1.1	28.6 - 0.1	20.6 - 0.5	11.8 - 1.5
w/ AdBiC	31.9 + 3.0	23.6 + 1.0	15.0 + 0.5	34.0 + 4.3	25.0 + 1.7	14.2 + 0.7	30.8 + 2.1	22.2 + 1.1	14.2 + 0.9
w/ AdBiC + $\mathbb{O}$	32.5 + 3.6	24.6 + 2.0	15.9 + 1.4	34.5 + 4.8	25.7 + 2.4	15.4 + 1.9	31.3 + 2.6	22.7 + 1.6	14.5 + 1.2
Joint	72.7			80.9			71.03		

gains, losses

Top-1 results without memory

Method	CIFAR-100			BIRDS-100			FOOD-100		
States	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$
LwF	53.0	44.0	29.1	53.7	41.8	30.1	42.9	31.8	22.2
w/ BiC	54.0 + 1.0	45.5 + 1.5	30.8 + 1.7	54.6 + 0.9	43.1 + 1.3	31.8 + 1.7	43.4 + 0.5	32.6 + 0.8	23.8 + 1.6
w/ AdBiC	54.3 + 1.3	**46.4** + 2.4	**32.3** + 3.2	55.0 + 1.3	44.0 + 2.2	**32.8** + 2.7	43.5 + 0.6	33.3 + 1.5	**24.7** + 2.5
w/ AdBiC + $\mathbb{O}$	54.9 + 1.9	47.3 + 3.3	32.6 + 3.5	55.8 + 2.1	44.8 + 3.0	33.3 + 3.2	44.0 + 1.1	34.2 + 2.4	25.3 + 3.1
LUCIR	50.1	33.7	19.5	50.8	31.4	17.9	44.2	26.4	15.5
w/ BiC	52.5 + 2.4	37.1 + 3.4	22.4 + 2.9	56.0 + 5.2	37.7 + 6.3	20.6 + 2.7	49.9 + 5.7	31.5 + 5.1	17.2 + 1.7
w/ AdBiC	**54.8** + 4.7	42.2 + 8.5	28.4 + 8.9	**58.5** + 7.7	**45.4** + 14.0	27.3 + 9.4	**52.0** + 7.8	**37.1** + 10.7	17.7 + 2.2
w/ AdBiC + $\mathbb{O}$	55.5 + 5.4	43.6 + 9.9	31.2 + 11.7	59.0 + 8.2	46.0 + 14.6	28.8 + 10.9	52.6 + 8.4	38.2 + 11.8	21.0 + 5.5
SIW	29.9	22.7	14.8	30.6	23.2	14.9	29.4	21.6	14.1
w/ BiC	31.4 + 1.5	22.8 + 0.1	14.7 - 0.1	32.8 + 2.2	22.7 - 0.5	12.8 - 2.1	29.1 - 0.3	20.3 - 1.3	12.1 - 2.0
w/ AdBiC	31.7 + 1.8	24.1 + 1.4	15.8 + 1.0	33.0 + 2.4	25.2 + 2.0	15.3 + 0.4	30.9 + 1.5	21.3 - 0.3	14.5 + 0.4
w/ AdBiC + $\mathbb{O}$	32.8 + 2.9	25.0 + 2.3	16.5 + 1.7	34.4 + 3.8	26.2 + 3.0	16.3 + 1.4	31.5 + 2.1	22.6 + 1.0	15.1 + 1.0
FT+	28.9	22.6	14.5	29.7	23.3	13.5	28.7	21.1	13.3
w/ BiC	30.7 + 1.8	22.5 - 0.1	14.8 + 0.3	32.3 + 2.6	22.5 - 0.8	12.4 - 1.1	28.6 - 0.1	20.6 - 0.5	11.8 - 1.5
w/ AdBiC	31.9 + 3.0	23.6 + 1.0	15.0 + 0.5	34.0 + 4.3	25.0 + 1.7	14.2 + 0.7	30.8 + 2.1	22.2 + 1.1	14.2 + 0.9
w/ AdBiC + $\mathbb{O}$	32.5 + 3.6	24.6 + 2.0	15.9 + 1.4	34.5 + 4.8	25.7 + 2.4	15.4 + 1.9	31.3 + 2.6	22.7 + 1.6	14.5 + 1.2
Joint	72.7			80.9			71.03		

gains, losses

Top-1 results without memory

Method	CIFAR-100			BIRDS-100			FOOD-100		
States	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$
LwF	53.0	44.0	29.1	53.7	41.8	30.1	42.9	31.8	22.2
w/ BiC	54.0 + 1.0	45.5 + 1.5	30.8 + 1.7	54.6 + 0.9	43.1 + 1.3	31.8 + 1.7	43.4 + 0.5	32.6 + 0.8	23.8 + 1.6
w/ AdBiC	54.3 + 1.3	**46.4** + 2.4	**32.3** + 3.2	55.0 + 1.3	44.0 + 2.2	**32.8** + 2.7	43.5 + 0.6	33.3 + 1.5	**24.7** + 2.5
w/ AdBiC + $\mathbb{O}$	54.9 + 1.9	47.3 + 3.3	32.6 + 3.5	55.8 + 2.1	44.8 + 3.0	33.3 + 3.2	44.0 + 1.1	34.2 + 2.4	25.3 + 3.1
LUCIR	50.1	33.7	19.5	50.8	31.4	17.9	44.2	26.4	15.5
w/ BiC	52.5 + 2.4	37.1 + 3.4	22.4 + 2.9	56.0 + 5.2	37.7 + 6.3	20.6 + 2.7	49.9 + 5.7	31.5 + 5.1	17.2 + 1.7
w/ AdBiC	**54.8** + 4.7	42.2 + 8.5	28.4 + 8.9	**58.5** + 7.7	**45.4** + 14.0	27.3 + 9.4	**52.0** + 7.8	**37.1** + 10.7	17.7 + 2.2
w/ AdBiC + $\mathbb{O}$	55.5 + 5.4	43.6 + 9.9	31.2 + 11.7	59.0 + 8.2	46.0 + 14.6	28.8 + 10.9	52.6 + 8.4	38.2 + 11.8	21.0 + 5.5
SIW	29.9	22.7	14.8	30.6	23.2	14.9	29.4	21.6	14.1
w/ BiC	31.4 + 1.5	22.8 + 0.1	14.7 - 0.1	32.8 + 2.2	22.7 - 0.5	12.8 - 2.1	29.1 - 0.3	20.3 - 1.3	12.1 - 2.0
w/ AdBiC	31.7 + 1.8	24.1 + 1.4	15.8 + 1.0	33.0 + 2.4	25.2 + 2.0	15.3 + 0.4	30.9 + 1.5	21.3 - 0.3	14.5 + 0.4
w/ AdBiC + $\mathbb{O}$	32.8 + 2.9	25.0 + 2.3	16.5 + 1.7	34.4 + 3.8	26.2 + 3.0	16.3 + 1.4	31.5 + 2.1	22.6 + 1.0	15.1 + 1.0
FT+	28.9	22.6	14.5	29.7	23.3	13.5	28.7	21.1	13.3
w/ BiC	30.7 + 1.8	22.5 - 0.1	14.8 + 0.3	32.3 + 2.6	22.5 - 0.8	12.4 - 1.1	28.6 - 0.1	20.6 - 0.5	11.8 - 1.5
w/ AdBiC	31.9 + 3.0	23.6 + 1.0	15.0 + 0.5	34.0 + 4.3	25.0 + 1.7	14.2 + 0.7	30.8 + 2.1	22.2 + 1.1	14.2 + 0.9
w/ AdBiC + $\mathbb{O}$	32.5 + 3.6	24.6 + 2.0	15.9 + 1.4	34.5 + 4.8	25.7 + 2.4	15.4 + 1.9	31.3 + 2.6	22.7 + 1.6	14.5 + 1.2
Joint	72.7			80.9			71.03		

gains, losses

Top-1 results without memory

Method	CIFAR-100			BIRDS-100			FOOD-100		
States	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$
LwF	53.0	44.0	29.1	53.7	41.8	30.1	42.9	31.8	22.2
w/ BiC	54.0 + 1.0	45.5 + 1.5	30.8 + 1.7	54.6 + 0.9	43.1 + 1.3	31.8 + 1.7	43.4 + 0.5	32.6 + 0.8	23.8 + 1.6
w/ AdBiC	54.3 + 1.3	**46.4** + 2.4	**32.3** + 3.2	55.0 + 1.3	44.0 + 2.2	**32.8** + 2.7	43.5 + 0.6	33.3 + 1.5	**24.7** + 2.5
w/ AdBiC + $\mathbb{O}$	54.9 + 1.9	47.3 + 3.3	32.6 + 3.5	55.8 + 2.1	44.8 + 3.0	33.3 + 3.2	44.0 + 1.1	34.2 + 2.4	25.3 + 3.1
LUCIR	50.1	33.7	19.5	50.8	31.4	17.9	44.2	26.4	15.5
w/ BiC	52.5 + 2.4	37.1 + 3.4	22.4 + 2.9	56.0 + 5.2	37.7 + 6.3	20.6 + 2.7	49.9 + 5.7	31.5 + 5.1	17.2 + 1.7
w/ AdBiC	**54.8** + 4.7	42.2 + 8.5	28.4 + 8.9	**58.5** + 7.7	**45.4** + 14.0	27.3 + 9.4	**52.0** + 7.8	**37.1** + 10.7	17.7 + 2.2
w/ AdBiC + $\mathbb{O}$	55.5 + 5.4	43.6 + 9.9	31.2 + 11.7	59.0 + 8.2	46.0 + 14.6	28.8 + 10.9	52.6 + 8.4	38.2 + 11.8	21.0 + 5.5
SIW	29.9	22.7	14.8	30.6	23.2	14.9	29.4	21.6	14.1
w/ BiC	31.4 + 1.5	22.8 + 0.1	14.7 - 0.1	32.8 + 2.2	22.7 - 0.5	12.8 - 2.1	29.1 - 0.3	20.3 - 1.3	12.1 - 2.0
w/ AdBiC	31.7 + 1.8	24.1 + 1.4	15.8 + 1.0	33.0 + 2.4	25.2 + 2.0	15.3 + 0.4	30.9 + 1.5	21.3 - 0.3	14.5 + 0.4
w/ AdBiC + $\mathbb{O}$	32.8 + 2.9	25.0 + 2.3	16.5 + 1.7	34.4 + 3.8	26.2 + 3.0	16.3 + 1.4	31.5 + 2.1	22.6 + 1.0	15.1 + 1.0
FT+	28.9	22.6	14.5	29.7	23.3	13.5	28.7	21.1	13.3
w/ BiC	30.7 + 1.8	22.5 - 0.1	14.8 + 0.3	32.3 + 2.6	22.5 - 0.8	12.4 - 1.1	28.6 - 0.1	20.6 - 0.5	11.8 - 1.5
w/ AdBiC	31.9 + 3.0	23.6 + 1.0	15.0 + 0.5	34.0 + 4.3	25.0 + 1.7	14.2 + 0.7	30.8 + 2.1	22.2 + 1.1	14.2 + 0.9
w/ AdBiC + $\mathbb{O}$	32.5 + 3.6	24.6 + 2.0	15.9 + 1.4	34.5 + 4.8	25.7 + 2.4	15.4 + 1.9	31.3 + 2.6	22.7 + 1.6	14.5 + 1.2
Joint	72.7			80.9			71.03		

gains, losses

Top-1 results without memory

Method	CIFAR-100			BIRDS-100			FOOD-100		
States	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$
LwF	53.0	44.0	29.1	53.7	41.8	30.1	42.9	31.8	22.2
w/ BiC	54.0 + 1.0	45.5 + 1.5	30.8 + 1.7	54.6 + 0.9	43.1 + 1.3	31.8 + 1.7	43.4 + 0.5	32.6 + 0.8	23.8 + 1.6
w/ AdBiC	54.3 + 1.3	**46.4** + 2.4	**32.3** + 3.2	55.0 + 1.3	44.0 + 2.2	**32.8** + 2.7	43.5 + 0.6	33.3 + 1.5	**24.7** + 2.5
w/ AdBiC + $\mathbb{O}$	54.9 + 1.9	47.3 + 3.3	32.6 + 3.5	55.8 + 2.1	44.8 + 3.0	33.3 + 3.2	44.0 + 1.1	34.2 + 2.4	25.3 + 3.1
LUCIR	50.1	33.7	19.5	50.8	31.4	17.9	44.2	26.4	15.5
w/ BiC	52.5 + 2.4	37.1 + 3.4	22.4 + 2.9	56.0 + 5.2	37.7 + 6.3	20.6 + 2.7	49.9 + 5.7	31.5 + 5.1	17.2 + 1.7
w/ AdBiC	**54.8** + 4.7	42.2 + 8.5	28.4 + 8.9	**58.5** + 7.7	**45.4** + 14.0	27.3 + 9.4	**52.0** + 7.8	**37.1** + 10.7	17.7 + 2.2
w/ AdBiC + $\mathbb{O}$	55.5 + 5.4	43.6 + 9.9	31.2 + 11.7	59.0 + 8.2	46.0 + 14.6	28.8 + 10.9	52.6 + 8.4	38.2 + 11.8	21.0 + 5.5
SIW	29.9	22.7	14.8	30.6	23.2	14.9	29.4	21.6	14.1
w/ BiC	31.4 + 1.5	22.8 + 0.1	14.7 - 0.1	32.8 + 2.2	22.7 - 0.5	12.8 - 2.1	29.1 - 0.3	20.3 - 1.3	12.1 - 2.0
w/ AdBiC	31.7 + 1.8	24.1 + 1.4	15.8 + 1.0	33.0 + 2.4	25.2 + 2.0	15.3 + 0.4	30.9 + 1.5	21.3 - 0.3	14.5 + 0.4
w/ AdBiC + $\mathbb{O}$	32.8 + 2.9	25.0 + 2.3	16.5 + 1.7	34.4 + 3.8	26.2 + 3.0	16.3 + 1.4	31.5 + 2.1	22.6 + 1.0	15.1 + 1.0
FT+	28.9	22.6	14.5	29.7	23.3	13.5	28.7	21.1	13.3
w/ BiC	30.7 + 1.8	22.5 - 0.1	14.8 + 0.3	32.3 + 2.6	22.5 - 0.8	12.4 - 1.1	28.6 - 0.1	20.6 - 0.5	11.8 - 1.5
w/ AdBiC	31.9 + 3.0	23.6 + 1.0	15.0 + 0.5	34.0 + 4.3	25.0 + 1.7	14.2 + 0.7	30.8 + 2.1	22.2 + 1.1	14.2 + 0.9
w/ AdBiC + $\mathbb{O}$	32.5 + 3.6	24.6 + 2.0	15.9 + 1.4	34.5 + 4.8	25.7 + 2.4	15.4 + 1.9	31.3 + 2.6	22.7 + 1.6	14.5 + 1.2
Joint	72.7			80.9			71.03		

gains, losses

Top-1 results without memory

Method	CIFAR-100			BIRDS-100			FOOD-100		
States	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$
LwF	53.0	44.0	29.1	53.7	41.8	30.1	42.9	31.8	22.2
w/ BiC	54.0 + 1.0	45.5 + 1.5	30.8 + 1.7	54.6 + 0.9	43.1 + 1.3	31.8 + 1.7	43.4 + 0.5	32.6 + 0.8	23.8 + 1.6
w/ AdBiC	54.3 + 1.3	**46.4** + 2.4	**32.3** + 3.2	55.0 + 1.3	44.0 + 2.2	**32.8** + 2.7	43.5 + 0.6	33.3 + 1.5	**24.7** + 2.5
w/ AdBiC $+ \mathbb{O}$	54.9 + 1.9	47.3 + 3.3	32.6 + 3.5	55.8 + 2.1	44.8 + 3.0	33.3 + 3.2	44.0 + 1.1	34.2 + 2.4	25.3 + 3.1
LUCIR	50.1	33.7	19.5	50.8	31.4	17.9	44.2	26.4	15.5
w/ BiC	52.5 + 2.4	37.1 + 3.4	22.4 + 2.9	56.0 + 5.2	37.7 + 6.3	20.6 + 2.7	49.9 + 5.7	31.5 + 5.1	17.2 + 1.7
w/ AdBiC	**54.8** + 4.7	42.2 + 8.5	28.4 + 8.9	**58.5** + 7.7	**45.4** + 14.0	27.3 + 9.4	**52.0** + 7.8	**37.1** + 10.7	17.7 + 2.2
w/ AdBiC $+ \mathbb{O}$	55.5 + 5.4	43.6 + 9.9	31.2 + 11.7	59.0 + 8.2	46.0 + 14.6	28.8 + 10.9	52.6 + 8.4	38.2 + 11.8	21.0 + 5.5
SIW	29.9	22.7	14.8	30.6	23.2	14.9	29.4	21.6	14.1
w/ BiC	31.4 + 1.5	22.8 + 0.1	14.7 - 0.1	32.8 + 2.2	22.7 - 0.5	12.8 - 2.1	29.1 - 0.3	20.3 - 1.3	12.1 - 2.0
w/ AdBiC	31.7 + 1.8	24.1 + 1.4	15.8 + 1.0	33.0 + 2.4	25.2 + 2.0	15.3 + 0.4	30.9 + 1.5	21.3 - 0.3	14.5 + 0.4
w/ AdBiC $+ \mathbb{O}$	32.8 + 2.9	25.0 + 2.3	16.5 + 1.7	34.4 + 3.8	26.2 + 3.0	16.3 + 1.4	31.5 + 2.1	22.6 + 1.0	15.1 + 1.0
FT+	28.9	22.6	14.5	29.7	23.3	13.5	28.7	21.1	13.3
w/ BiC	30.7 + 1.8	22.5 - 0.1	14.8 + 0.3	32.3 + 2.6	22.5 - 0.8	12.4 - 1.1	28.6 - 0.1	20.6 - 0.5	11.8 - 1.5
w/ AdBiC	31.9 + 3.0	23.6 + 1.0	15.0 + 0.5	34.0 + 4.3	25.0 + 1.7	14.2 + 0.7	30.8 + 2.1	22.2 + 1.1	14.2 + 0.9
w/ AdBiC $+ \mathbb{O}$	32.5 + 3.6	24.6 + 2.0	15.9 + 1.4	34.5 + 4.8	25.7 + 2.4	15.4 + 1.9	31.3 + 2.6	22.7 + 1.6	14.5 + 1.2
Joint	72.7			80.9			71.03		

gains, losses

Top-1 results without memory

Method	CIFAR-100			BIRDS-100			FOOD-100		
States	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=20$
LwF	53.0	44.0	29.1	53.7	41.8	30.1	42.9	31.8	22.2
w/ BiC	54.0 + 1.0	45.5 + 1.5	30.8 + 1.7	54.6 + 0.9	43.1 + 1.3	31.8 + 1.7	43.4 + 0.5	32.6 + 0.8	23.8 + 1.6
w/ AdBiC	54.3 + 1.3	**46.4** + 2.4	**32.3** + 3.2	55.0 + 1.3	44.0 + 2.2	**32.8** + 2.7	43.5 + 0.6	33.3 + 1.5	**24.7** + 2.5
w/ AdBiC + $\mathbb{O}$	54.9 + 1.9	47.3 + 3.3	32.6 + 3.5	55.8 + 2.1	44.8 + 3.0	33.3 + 3.2	44.0 + 1.1	34.2 + 2.4	25.3 + 3.1
LUCIR	50.1	33.7	19.5	50.8	31.4	17.9	44.2	26.4	15.5
w/ BiC	52.5 + 2.4	37.1 + 3.4	22.4 + 2.9	56.0 + 5.2	37.7 + 6.3	20.6 + 2.7	49.9 + 5.7	31.5 + 5.1	17.2 + 1.7
w/ AdBiC	**54.8** + 4.7	42.2 + 8.5	28.4 + 8.9	**58.5** + 7.7	**45.4** + 14.0	27.3 + 9.4	**52.0** + 7.8	**37.1** + 10.7	17.7 + 2.2
w/ AdBiC + $\mathbb{O}$	55.5 + 5.4	43.6 + 9.9	31.2 + 11.7	59.0 + 8.2	46.0 + 14.6	28.8 + 10.9	52.6 + 8.4	38.2 + 11.8	21.0 + 5.5
SIW	29.9	22.7	14.8	30.6	23.2	14.9	29.4	21.6	14.1
w/ BiC	31.4 + 1.5	22.8 + 0.1	14.7 - 0.1	32.8 + 2.2	22.7 - 0.5	12.8 - 2.1	29.1 - 0.3	20.3 - 1.3	12.1 - 2.0
w/ AdBiC	31.7 + 1.8	24.1 + 1.4	15.8 + 1.0	33.0 + 2.4	25.2 + 2.0	15.3 + 0.4	30.9 + 1.5	21.3 - 0.3	14.5 + 0.4
w/ AdBiC + $\mathbb{O}$	32.8 + 2.9	25.0 + 2.3	16.5 + 1.7	34.4 + 3.8	26.2 + 3.0	16.3 + 1.4	31.5 + 2.1	22.6 + 1.0	15.1 + 1.0
FT+	28.9	22.6	14.5	29.7	23.3	13.5	28.7	21.1	13.3
w/ BiC	30.7 + 1.8	22.5 - 0.1	14.8 + 0.3	32.3 + 2.6	22.5 - 0.8	12.4 - 1.1	28.6 - 0.1	20.6 - 0.5	11.8 - 1.5
w/ AdBiC	31.9 + 3.0	23.6 + 1.0	15.0 + 0.5	34.0 + 4.3	25.0 + 1.7	14.2 + 0.7	30.8 + 2.1	22.2 + 1.1	14.2 + 0.9
w/ AdBiC + $\mathbb{O}$	32.5 + 3.6	24.6 + 2.0	15.9 + 1.4	34.5 + 4.8	25.7 + 2.4	15.4 + 1.9	31.3 + 2.6	22.7 + 1.6	14.5 + 1.2
Joint	72.7			80.9			71.03		

gains, losses

Conclusions and future work

Conclusions

- In fine tuning, the classification layer is the most affected by catastrophic forgetting

Conclusions

- In fine tuning, the classification layer is the most affected by catastrophic forgetting
- Fine-tuning-based methods are the best option when a memory is allowed

Conclusions

- In fine tuning, the classification layer is the most affected by catastrophic forgetting
- Fine-tuning-based methods are the best option when a memory is allowed
- Fixed representations are an appropriate choice without memory

Conclusions

- In fine tuning, the classification layer is the most affected by catastrophic forgetting
- Fine-tuning-based methods are the best option when a memory is allowed
- Fixed representations are an appropriate choice without memory
- Usefulness of distillation is reduced at large scale

Conclusions

- In fine tuning, the classification layer is the most affected by catastrophic forgetting
- Fine-tuning-based methods are the best option when a memory is allowed
- Fixed representations are an appropriate choice without memory
- Usefulness of distillation is reduced at large scale
- We reduce the model's footprint by half compared to distillation-based methods

Future work

- Propose an alternative of the distillation loss

Future work

- Propose an alternative of the distillation loss
- Focus more on class-incremental learning without memory

Future work

- Propose an alternative of the distillation loss
- Focus more on class-incremental learning without memory
- Explore methods that make a better compromise between stability and plasticity of the network

Future work

- Propose an alternative of the distillation loss
- Focus more on class-incremental learning without memory
- Explore methods that make a better compromise between stability and plasticity of the network
- Work on feature transferability between states

Publications

Journal papers

- Belouadah, E., Popescu, A., Kanellos, I. **A Comprehensive Study of Class Incremental Learning Algorithms for Visual Tasks**. Neural Networks, t. 135, pp. 38-54
- Aggarwal, U., Popescu, A., Belouadah, E. and Hudelot, C., 2020. **A Comparative Study of Calibration Methods for Imbalanced Class Incremental Learning**. Multimedia Tools and Applications.

Conference papers (1/2)

- Belouadah, E., Popescu, A. **IL2M: Class Incremental Learning with Dual Memory**. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2019, p583-592.
- Belouadah, E., Popescu, A. **ScaIL: Classifier Weights Scaling for Class Incremental Learning**. The IEEE Winter Conference on Applications of Computer Vision (WACV), 2020, p1266-1275.
- Belouadah, E., Popescu, A., Kanellos, I. **Initial Classifier Weights Replay for Memoryless Class Incremental Learning**. British Machine Vision Conference (BMVC) 2020.

Conference papers (2/2)

- Slim, H.†, Belouadah, E.†, Popescu, A., Onchis, D. **Dataset Knowledge Transfer for Class-Incremental Learning without Memory**. The IEEE Winter Conference on Applications of Computer Vision (WACV), 2022.

†: equal contribution

Workshop papers

- Belouadah, E., Popescu, A. **DeeSIL: Deep-Shallow Incremental Learning**. Proceeding of the European Conference on Computer Vision workshops (W-ECCV 2018)
- Belouadah, E., Popescu, A., Aggarwal, U., Saci L. **Active Class Incremental Learning for Imbalanced Datasets**. IPCV workshop of the European Conference on Computer Vision (ECCV) 2020.

Codes and dataset details:

https://github.com/EdenBelouadah/class-incremental-learning

eden.belouadah@yahoo.fr

Appendix

Fixed-Representation-based methods:

DeeSIL (Belouadah and Popescu, 2018), *Deep-SLDA* (Hayes and Kanan, 2019), *REMIND* (Hayes et al., 2019), *FearNet* Kemker and Kanan, 2018.

Fine-Tuning-based methods:

DGM (Ostapenko et al., 2019), *DGR* (Shin et al., 2017), *GMNF* (Cong et al., 2020), *LwF* (Li and Hoiem, 2016), *EWC* (Kirkpatrick et al., 2016), *MAS* (Aljundi et al., 2018), *BiC* (Wu et al., 2019), *MDF* (Zhao et al., 2020), *LUCIR* (Hou et al., 2019), *iCaRL* (Rebuffi et al., 2017), *E2EIL* (Castro et al., 2018).

Parameter-isolation-based methods:

PackNet (Mallya and Lazebnik, 2018), *PiggyBack* (Mallya et al., 2018), *TFM* (Masana et al., 2020), $Expert - Gate$ (Aljundi et al., 2017), *PNN* (Rusu et al., 2016), *DAN* (Rosenfeld and Tsotsos, 2017).

Bibliography

[1] Rahaf Aljundi et al. "Expert Gate: Lifelong Learning with a Network of Experts". In: *Conference on Computer Vision and Pattern Recognition*. CVPR. 2017.

[2] Rahaf Aljundi et al. "Memory Aware Synapses: Learning What (not) to Forget". In: *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part III*. Ed. by Vittorio Ferrari et al. Vol. 11207. Lecture Notes in Computer Science. Springer, 2018, pp. 144–161.

[3] Eden Belouadah and Adrian Popescu. "DeeSIL: Deep-Shallow Incremental Learning". In: *TaskCV Workshop @ ECCV 2018.* (2018).

[4] Eden Belouadah and Adrian Popescu. "Il2m: Class incremental learning with dual memory". In: *Proceedings of the IEEE International Conference on Computer Vision*. 2019, pp. 583–592.

[5] Eden Belouadah and Adrian Popescu. "ScaIL: Classifier Weights Scaling for Class Incremental Learning". In: *The IEEE Winter Conference on Applications of Computer Vision (WACV)*. Mar. 2020.

Bibliography (cont.)

[6] Eden Belouadah et al. "Active Class Incremental Learning for Imbalanced Datasets". In: *Computer Vision - ECCV 2020 Workshops - Glasgow, UK, August 23-28, 2020, Proceedings, Part VI*. Ed. by Adrien Bartoli and Andrea Fusiello. Vol. 12540. Lecture Notes in Computer Science. Springer, 2020, pp. 146–162.

[7] Eden Belouadah et al. "Initial Classifier Weights Replay for Memoryless Class Incremental Learning". In: *British Machine Vision Conference (BMVC)*. 2020.

[8] Lukas Bossard et al. "Food-101 – Mining Discriminative Components with Random Forests". In: *European Conference on Computer Vision*. 2014.

[9] Mateusz Buda et al. "A systematic study of the class imbalance problem in convolutional neural networks". In: *Neural Networks* 106 (2018), pp. 249–259.

[10] Qiong Cao et al. "VGGFace2: A Dataset for Recognising Faces across Pose and Age". In: *13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018, Xian, China, May 15-19, 2018*. 2018, pp. 67–74.

[11] Francisco M. Castro et al. "End-to-End Incremental Learning". In: *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XII*. 2018, pp. 241–257.

[12] Yulai Cong et al. "GAN Memory with No Forgetting". In: *CoRR* abs/2006.07543 (2020).

[13] Jia Deng et al. "ImageNet: A large-scale hierarchical image database". In: *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA*. 2009, pp. 248–255.

[14] Tyler L. Hayes and Christopher Kanan. "Lifelong Machine Learning with Deep Streaming Linear Discriminant Analysis". In: *CoRR* abs/1909.01520 (2019).

[15] Tyler L. Hayes et al. "REMIND Your Neural Network to Prevent Catastrophic Forgetting". In: *CoRR* abs/1910.02509 (2019).

Bibliography (cont.)

[16] Saihui Hou et al. “Learning a Unified Classifier Incrementally via Rebalancing”. In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*. 2019, pp. 831–839.

[17] Ronald Kemker and Christopher Kanan. “FearNet: Brain-Inspired Model for Incremental Learning”. In: *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. 2018.

[18] James Kirkpatrick et al. “Overcoming catastrophic forgetting in neural networks”. In: *CoRR* abs/1612.00796 (2016).

[19] Alex Krizhevsky. *Learning multiple layers of features from tiny images*. Tech. rep. University of Toronto, 2009.

[20] Matthias De Lange et al. “Continual learning: A comparative study on how to defy forgetting in classification tasks”. In: *CoRR* abs/1909.08383 (2019).

[21] Zhizhong Li and Derek Hoiem. “Learning Without Forgetting”. In: *European Conference on Computer Vision*. ECCV. 2016.

[22] Arun Mallya and Svetlana Lazebnik. "PackNet: Adding Multiple Tasks to a Single Network by Iterative Pruning". In: *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*. 2018, pp. 7765–7773.

[23] Arun Mallya et al. "Piggyback: Adapting a Single Network to Multiple Tasks by Learning to Mask Weights". In: *ECCV (4)*. Vol. 11208. Lecture Notes in Computer Science. Springer, 2018, pp. 72–88.

[24] Marc Masana et al. *Class-incremental learning: survey and performance evaluation on image classification*. 2021. arXiv: 2010.15277 [cs.LG].

[25] Marc Masana et al. "Ternary Feature Masks: continual learning without any forgetting". In: *CoRR* abs/2001.08714 (2020). URL: https://arxiv.org/abs/2001.08714.

[26] Hyeonwoo Noh et al. "Large-Scale Image Retrieval with Attentive Deep Local Features". In: *ICCV*. IEEE Computer Society, 2017, pp. 3476–3485.

[27] Oleksiy Ostapenko et al. "Learning to Remember: A Synaptic Plasticity Driven Framework for Continual Learning". In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*. Computer Vision Foundation / IEEE, 2019, pp. 11321–11329.

[28] Sylvestre-Alvise Rebuffi et al. "iCaRL: Incremental Classifier and Representation Learning". In: *Conference on Computer Vision and Pattern Recognition*. CVPR. 2017.

[29] Amir Rosenfeld and John K. Tsotsos. "Incremental Learning Through Deep Adaptation". In: *CoRR* abs/1705.04228 (2017).

[30] Olga Russakovsky et al. "ImageNet Large Scale Visual Recognition Challenge". In: *International Journal of Computer Vision* 115.3 (2015), pp. 211–252.

[31] Andrei A. Rusu et al. "Progressive Neural Networks". In: *CoRR* abs/1606.04671 (2016).

[32] Hanul Shin et al. "Continual Learning with Deep Generative Replay". In: *NIPS*. 2017, pp. 2994–3003.

Bibliography (cont.)

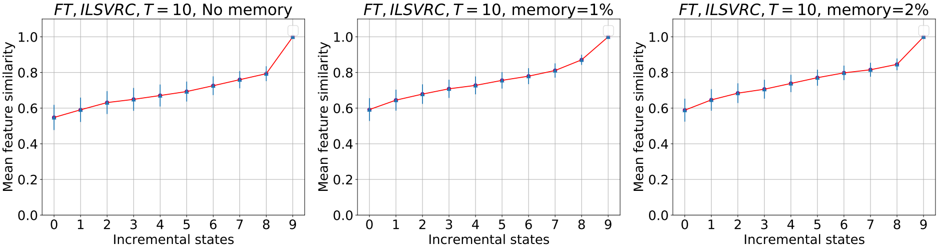
[33] Habib Slim et al. "Dataset Knowledge Transfer for Class-Incremental Learning Without Memory". In: *The IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2022.

[34] Yue Wu et al. "Large Scale Incremental Learning". In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*. 2019, pp. 374–382.

[35] Bowen Zhao et al. "Maintaining Discrimination and Fairness in Class Incremental Learning". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*. IEEE, 2020, pp. 13205–13214.

Backup Slides

Mean Features Similarity



(a) (b) (c)

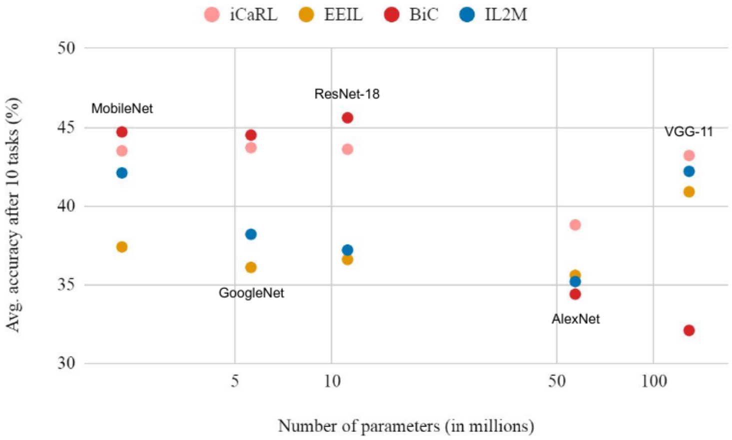
Mean feature similarities between incremental states for test images of the first state.

Additional Storage of our methods

Method	Additional Storage (AS) in float	AS for $N_{\mathcal{T}} = 1000$				
		$\mathcal{T} = 5$	$\mathcal{T} = 10$	$\mathcal{T} = 20$	$\mathcal{T} = 50$	$\mathcal{T} = 100$
DeeSIL	0	0	0	0	0	0
IL2M	$\mathcal{T} + N_{\mathcal{T}}$	4.02 KB	4.04 KB	4.08 KB	4.2 KB	4.4 KB
ScaIL	$N_{\mathcal{T}} \times D$	2.05 MB	2.05 MB	2.05 MB	2.05 MB	2.05 MB
SIW	$\mathcal{T} + N_{\mathcal{T}} \times D$	2.05 MB	2.05 MB	2.05 MB	2.05 MB	2.05 MB
TransIL (*AdBiC*)	$R \times (\mathcal{T} + 2) \times (\mathcal{T} - 1)$	1.12 KB	4.32 KB	16.72 KB	101.92 KB	403.92 KB
TransIL (*BiC*)	$2 \times R \times (\mathcal{T} - 1)$	320 B	720 B	1.52 KB	3.92 KB	7.92 KB

Additional Storage (AS) of our proposed IL approaches

Results with other deep architectures



Results with other architectures (Masana et al., 2021)