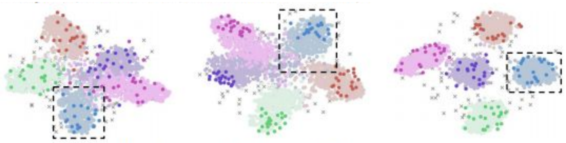
Annotation-Efficient Learning:
Class-Incremental Learning and Few-Shot Learning

Yaoyao Liu

11 Sept. 2020

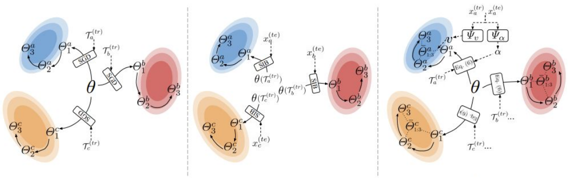
Outline of today's talk



random (baseline) *herding* (related) *mnemonics* (ours)

1. Class-Incremental Learning

Mnemonics Training: Multi-Class Incremental Learning without Forgetting
CVPR 2020



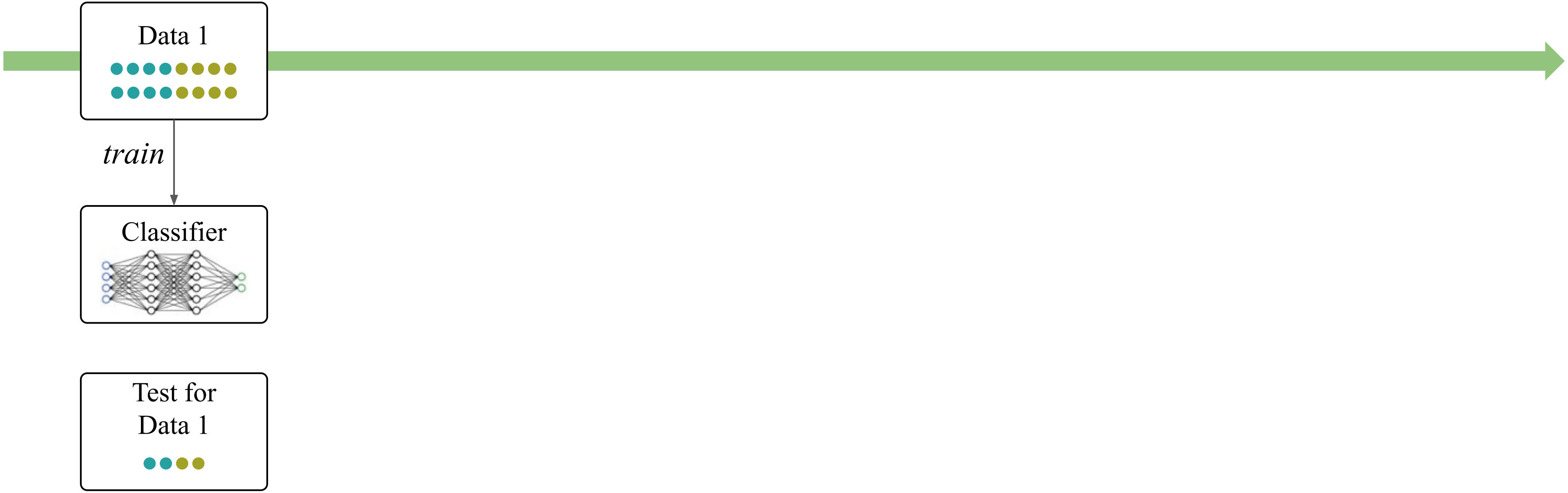
(a) MAML [13] (b) SIB [25] (c) E^3BM (ours)

2. Few-Shot Learning

An Ensemble of Epoch-wise Empirical Bayes for Few-shot Learning
ECCV 2020

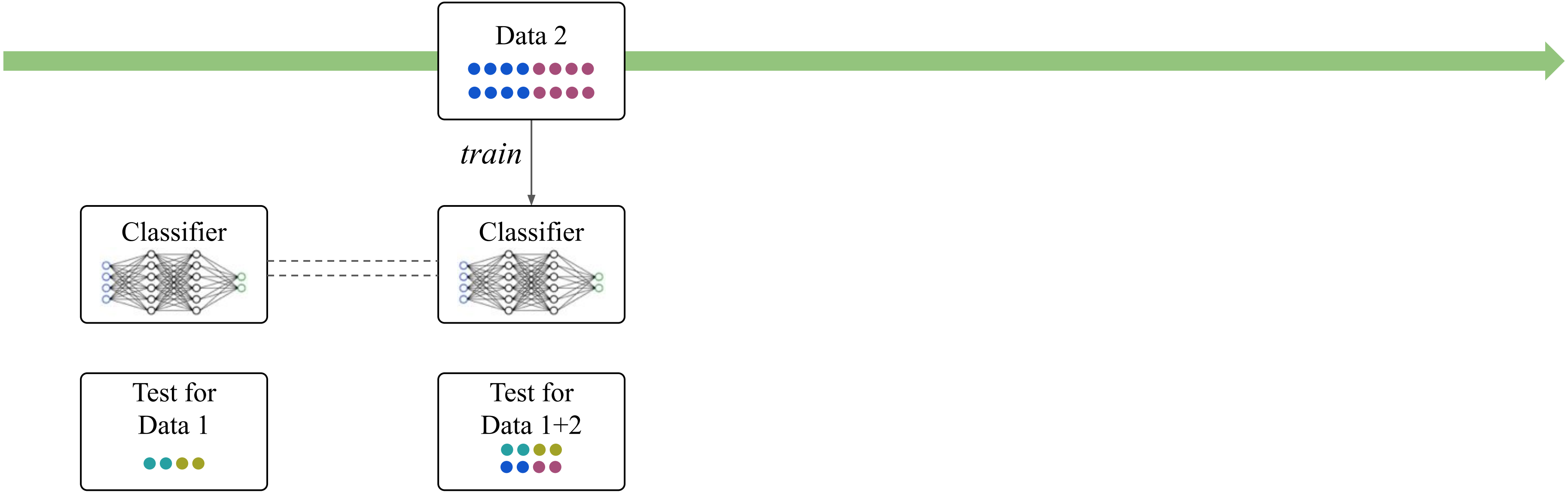
Background: class-incremental learning

Phase 1



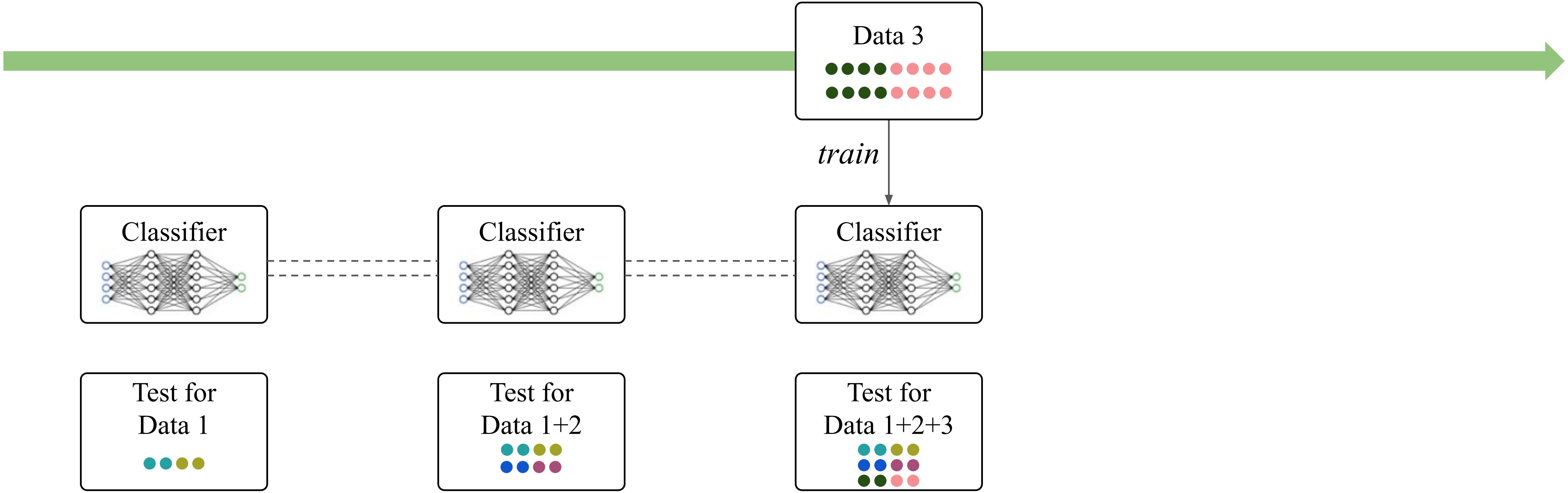
Background: class-incremental learning

Phase 2



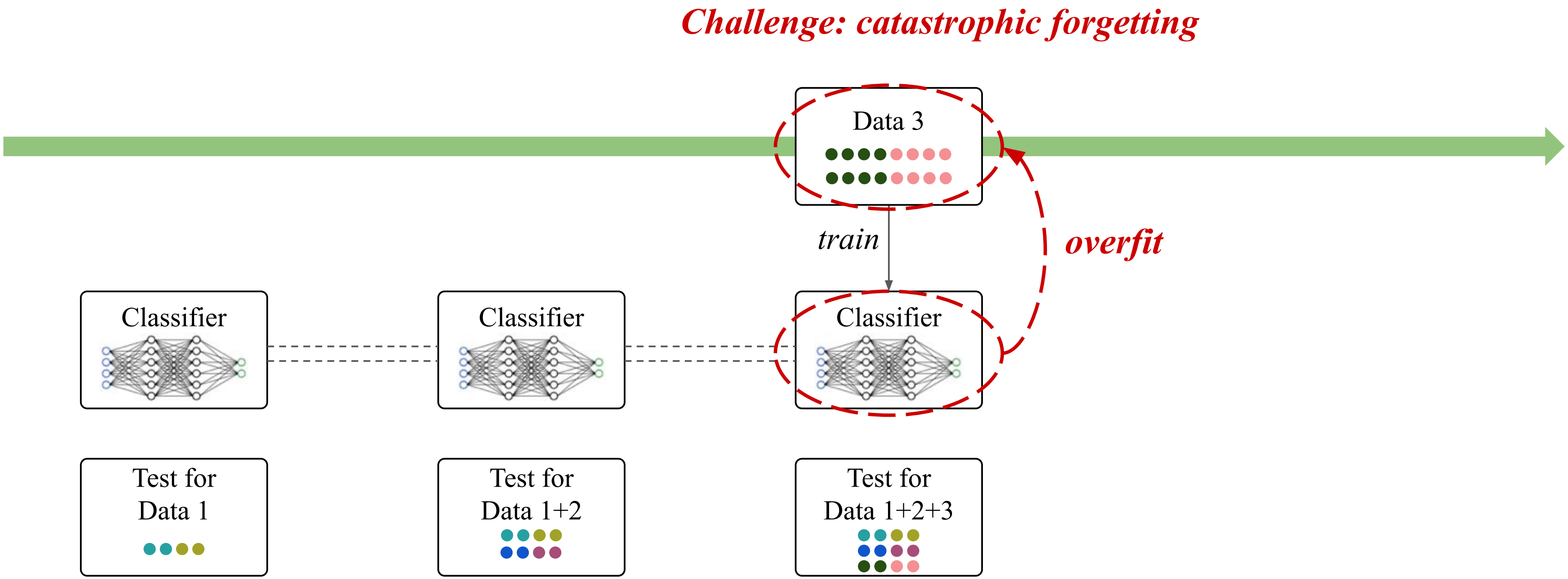
Background: class-incremental learning

Phase 3



Background: Class-Incremental Learning

Phase 3



Literature review

Technique 1: Replay samples for the old classes:

iCaRL[1], IL2M[2], ... ⇨ ***Mnemonics exemplars***

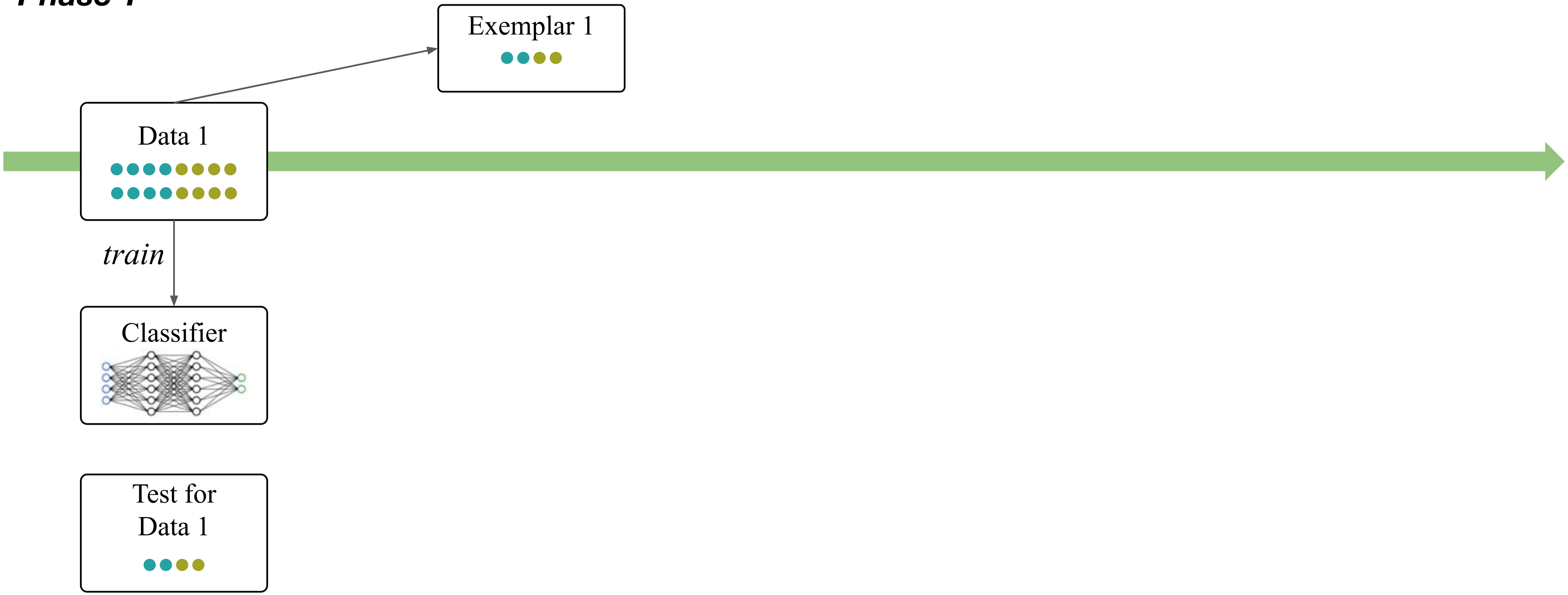
Technique 2: Preserve the knowledge for the old model:

LwF[3], LUCIR[4], PODNet[5]... ⇨ ***Weight transfer operations***

References
[1] Rebuffi, Sylvestre-Alvise, et al. "icarl: Incremental classifier and representation learning." CVPR 2017;
[2] Belouadah, Eden, and Adrian Popescu. "Il2m: Class incremental learning with dual memory." CVPR 2019;
[3] Li, Zhizhong, and Derek Hoiem. "Learning without forgetting." TPAMI 2017;
[4] Hou, Saihui, et al. "Learning a unified classifier incrementally via rebalancing." CVPR 2019;
[5] Douillard, Arthur, et al. "PODNet: Pooled Outputs Distillation for Small-Tasks Incremental Learning." ECCV 2020.

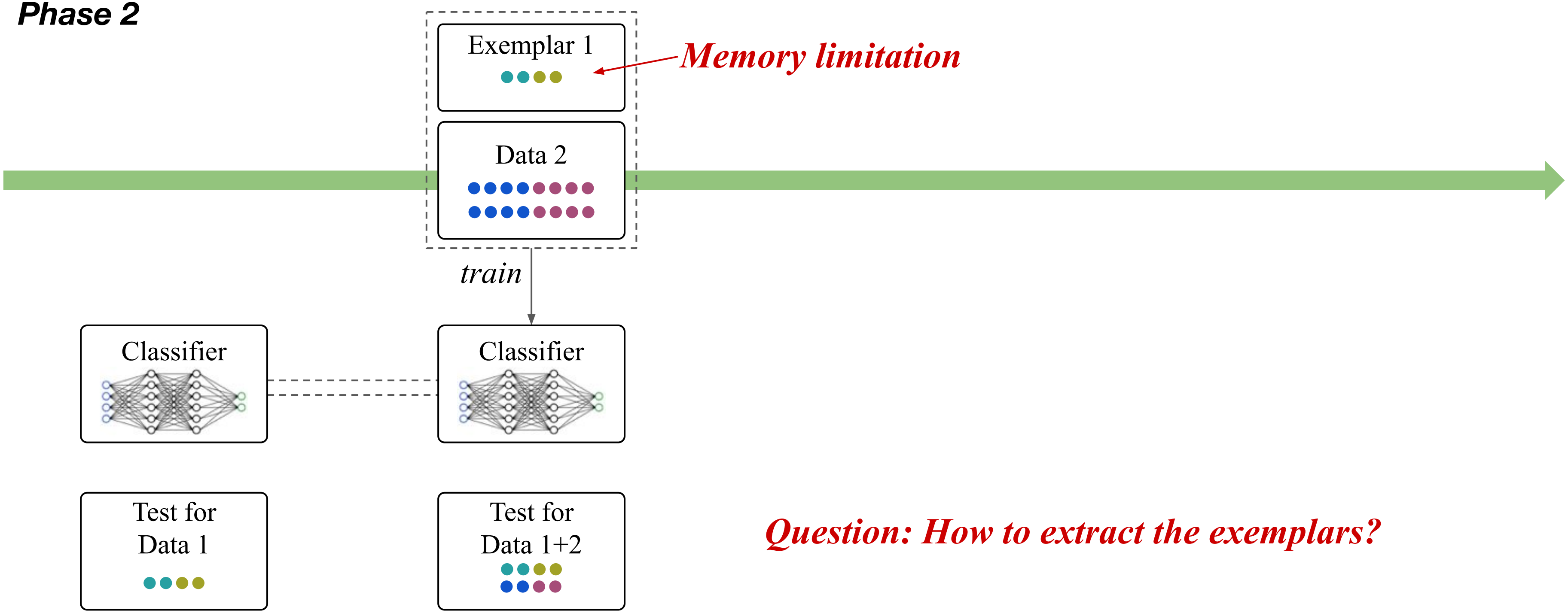
Replay samples for the old classes

Phase 1



Replay samples for the old classes

Phase 2



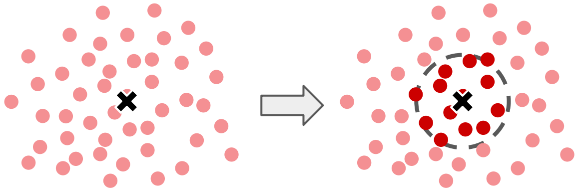
Question: How to extract the exemplars?

Question: how to extract the exemplars?

Existing methods:

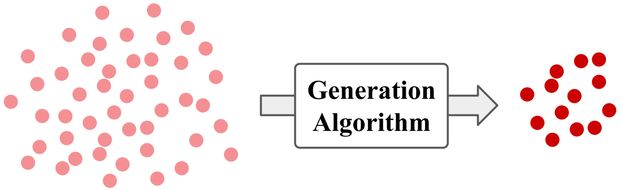
E.g., herding[1, 4, 6] :

select the samples near the average embedding



Our method: Mnemonics exemplars

Question: Can we generate the optimal exemplars?



Limitations for existing methods:

- Heuristic selection, not performance-based

- Select from finite sets (real images)

Benefits for our method:

+ Optimal selection by end-to-end training

+ Select from continuous (infinite) synthetic data

References
[1] Rebuffi, Sylvestre-Alvise, et al. "icarl: Incremental classifier and representation learning." CVPR 2017;
[4] Hou, Saihui, et al. "Learning a unified classifier incrementally via rebalancing." CVPR 2019;
[6] Wu, Yue, et al. “Large scale incremental learning.” CVPR 2019.

Question: how to formulate the optimization of the exemplars?

In the i-th incremental phase,

Data for the current phase

We have: Exemplars for previous phases $\boxed{\mathcal{E}_{0:i-1}} \cup \boxed{D_i} \xrightarrow{train} \Theta_i^*$ *(model)*

$\approx$

We aim to get: Exemplars for the current phase $\boxed{\mathcal{E}_{0:i}} \xrightarrow{train} \Theta_i^{\mathcal{E}}$

Bilevel optimization formulation:

$$\min_{\mathcal{E}_{0:i}} \mathcal{L}\left(\Theta_i^{\mathcal{E}}; \mathcal{E}_{0:i-1} \cup D_i\right)$$

$$\text{s.t.}\ \Theta_i^{\mathcal{E}} = \min_{\mathcal{E}_{0:i}} \mathcal{L}\left(\Theta_i; \mathcal{E}_{0:i}\right)$$

Literature review

Technique 1: Replay samples for the old classes:

iCaRL[1], IL2M[2], ... ⇨ ***Mnemonics exemplars***

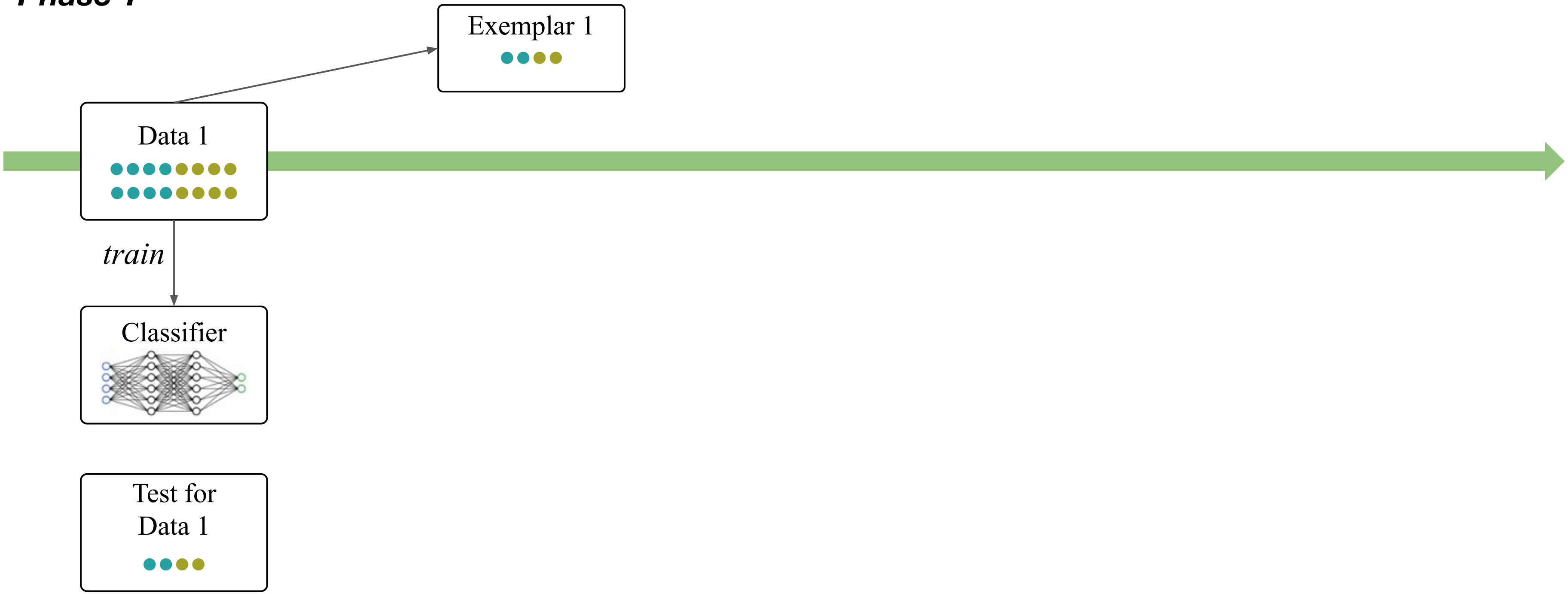
Technique 2: Preserve the knowledge for the old model:

LwF[3], LUCIR[4], PODNet[5]... ⇨ ***Weight transfer operations***

References
[1] Rebuffi, Sylvestre-Alvise, et al. "icarl: Incremental classifier and representation learning." CVPR 2017;
[2] Belouadah, Eden, and Adrian Popescu. "Il2m: Class incremental learning with dual memory." CVPR 2019;
[3] Li, Zhizhong, and Derek Hoiem. "Learning without forgetting." TPAMI 2017;
[4] Hou, Saihui, et al. "Learning a unified classifier incrementally via rebalancing." CVPR 2019;
[5] Douillard, Arthur, et al. "PODNet: Pooled Outputs Distillation for Small-Tasks Incremental Learning." ECCV 2020.

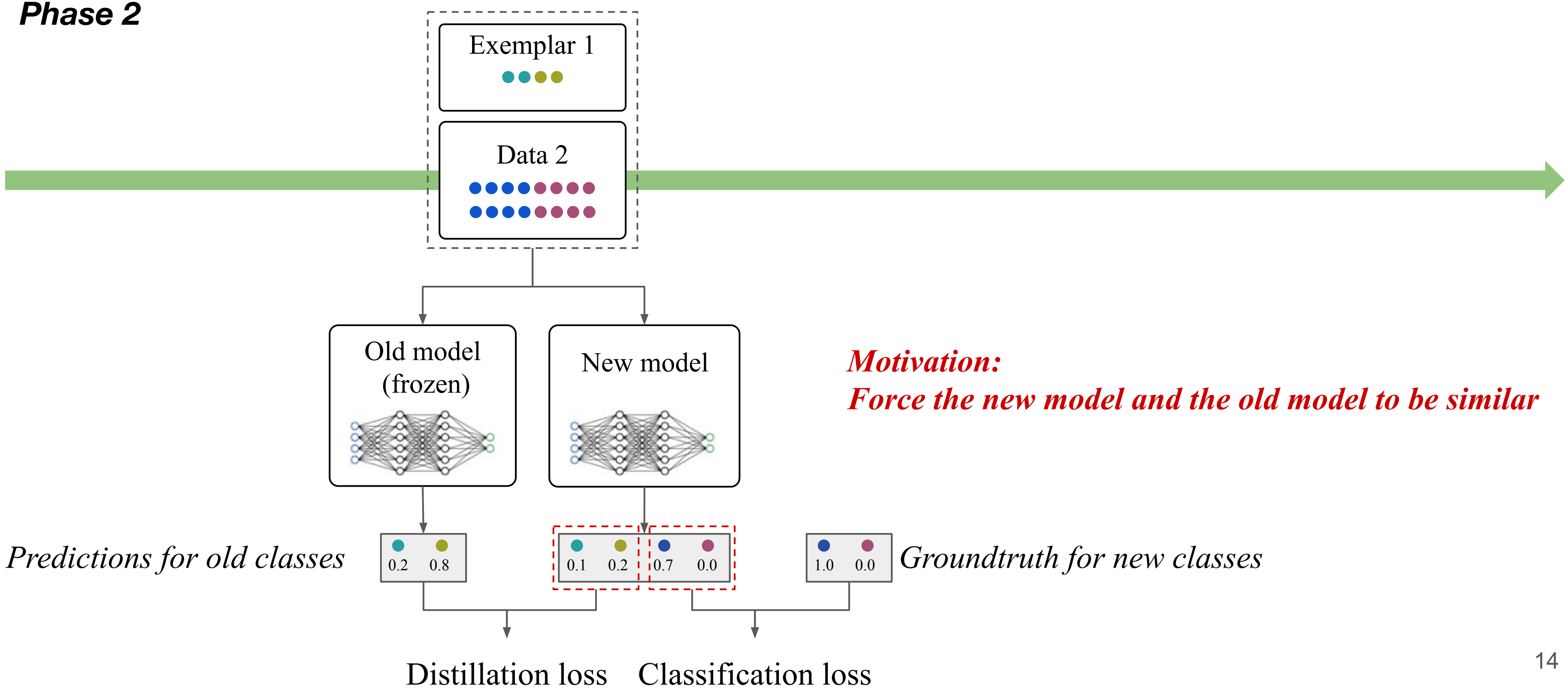
Preserve the knowledge for the old model

Phase 1



Preserve the knowledge for the old model

Phase 2



Motivation:
Force the new model and the old model to be similar

Preserve the knowledge for the old model

iCaRL[1] (CVPR 2017) ⟹ Distillation on predictions

LUCIR[4] (CVPR 2019) ⟹ Distillation on the final feature maps

PODNet[5] (ECCV 2020) ⟹ Distillation on the feature maps from all layers

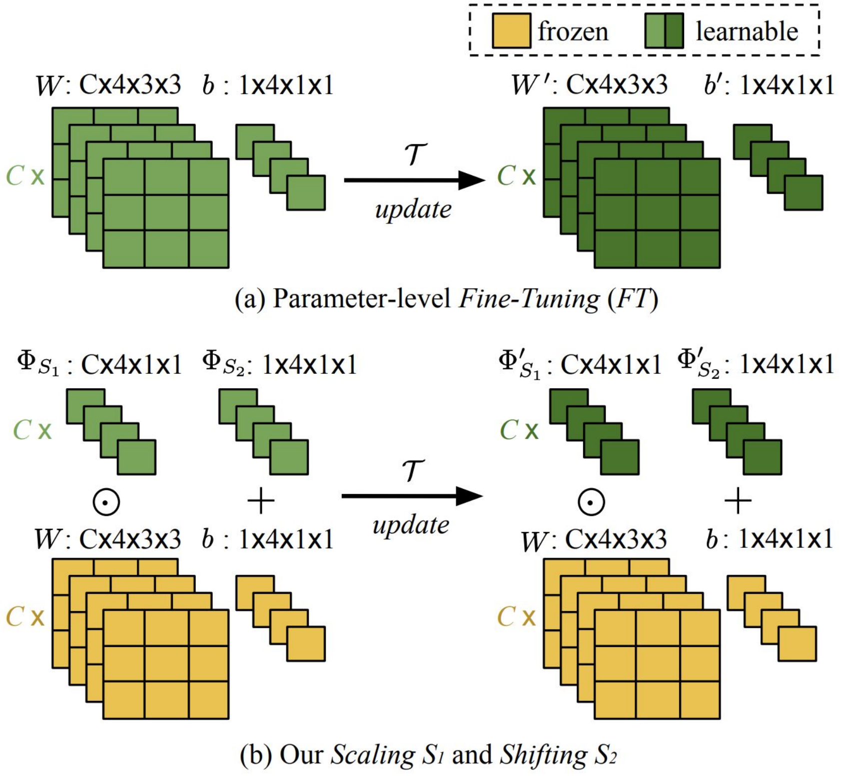
Distillation: preserve high-level knowledge for the old model

It is better to transfer low-level knowledge among tasks…[7]

Q: Can we preserve the low-level knowledge for the old model?

References
[1] Rebuffi, Sylvestre-Alvise, et al. "icarl: Incremental classifier and representation learning." CVPR 2017;
[4] Hou, Saihui, et al. "Learning a unified classifier incrementally via rebalancing." CVPR 2019;
[5] Douillard, Arthur, et al. "PODNet: Pooled Outputs Distillation for Small-Tasks Incremental Learning." ECCV 2020;
[7] Taylor, Matthew E., and Peter Stone. "An introduction to intertask transfer for reinforcement learning." Ai Magazine 32.1 (2011): 15-15.

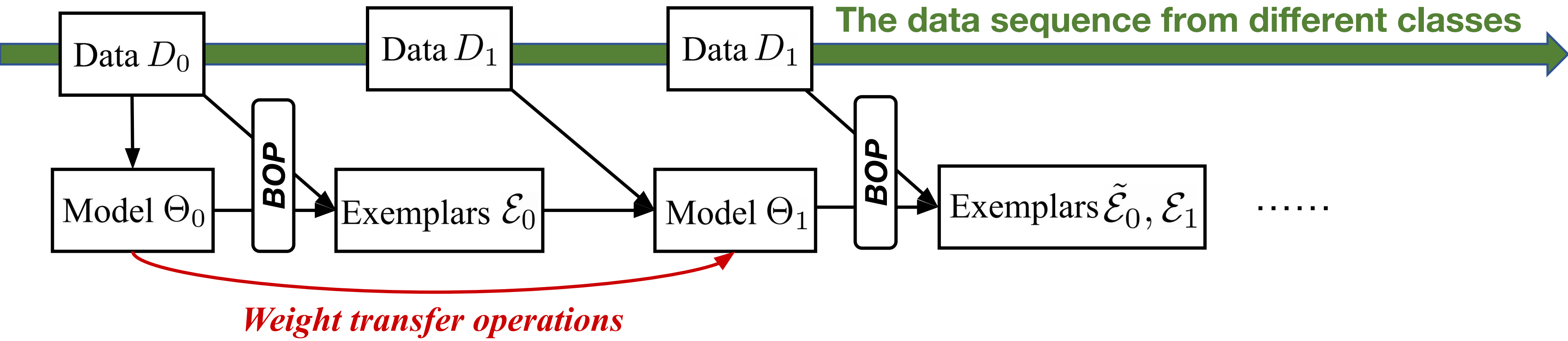
How to transfer low-level knowledge for class-incremental learning?



Weight transfer operations:
Channel-wise masks

Global computing glow

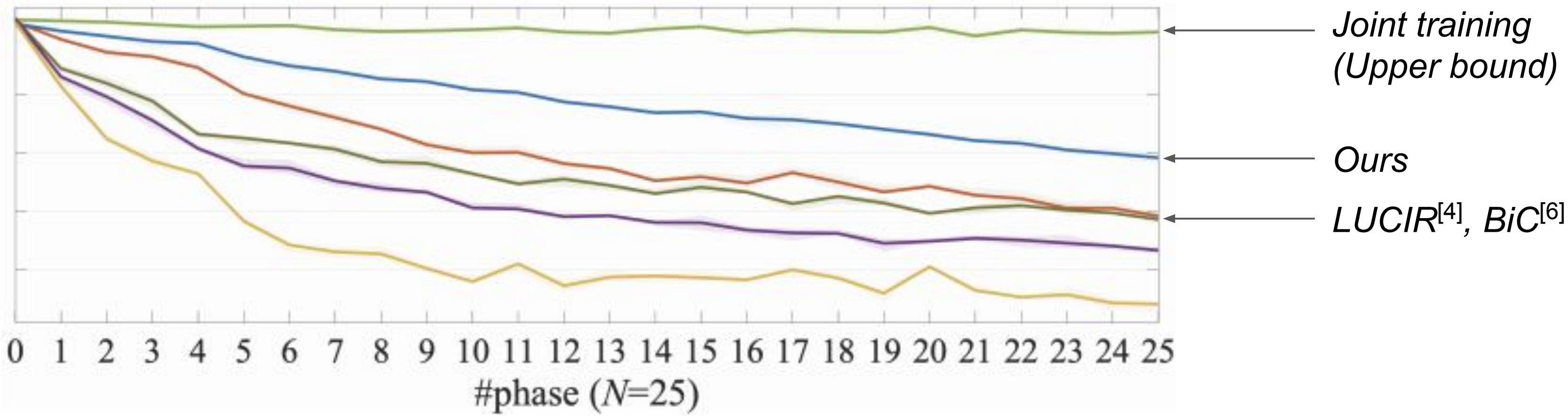
Our method: *Technique 1 + Technique 2*



BOP = Bilevel Optimization Program

Our method boosts the performance

Dataset: ImageNet-Subset



References

[4] Hou, Saihui, et al. "Learning a unified classifier incrementally via rebalancing." CVPR 2019;

[6] Wu, Yue, et al. "Large scale incremental learning." CVPR 2019.

Our method boosts the performance

Metric	Method	CIFAR-100			ImageNet-Subset			ImageNet		
		N=5	10	25	5	10	25	5	10	25
Average acc. (%) ↑ $\bar{\mathcal{A}} = \frac{1}{N+1}\sum_{i=0}^{N}\mathcal{A}_i$	LwF$^{\diamond}$ (2016) [2]	49.59	46.98	45.51	53.62	47.64	44.32	44.35	38.90	36.87
	LwF *w/* ours	54.21	52.72	51.59	60.94	59.25	59.71	52.70	50.37	50.79
	iCaRL (2017) [1]	57.12	52.66	48.22	65.44	59.88	52.97	51.50	46.89	43.14
	iCaRL *w/* ours	60.00	57.37	54.13	72.34	70.50	67.12	60.61	58.62	53.46
	BiC (2019) [6]	59.36	54.20	50.00	70.07	64.96	57.73	62.65	58.72	53.47
	BiC *w/* ours	60.67	58.11	55.51	71.92	70.73	69.22	**64.63**	62.71	60.20
	LUCIR (2019) [4]	63.17	60.14	57.54	70.84	68.32	61.44	64.45	61.57	56.56
	LUCIR *w/* ours	**63.34**	**62.28**	**60.96**	**72.58**	**71.37**	**69.74**	64.54	**63.01**	**61.00**

- *Generic*
- *Boost the performance for **FOUR** different baselines*

References

[1] Rebuffi, Sylvestre-Alvise, et al. "icarl: Incremental classifier and representation learning." CVPR 2017;

[3] Li, Zhizhong, and Derek Hoiem. "Learning without forgetting." TPAMI 2017;

[4] Hou, Saihui, et al. "Learning a unified classifier incrementally via rebalancing." CVPR 2019;

[6] Wu, Yue, et al. “Large scale incremental learning.” CVPR 2019.

Ablation study

Method	*CIFAR-100*			*ImagNet-Subset*		
	N=5	10	25	5	10	25
Baseline (LUCIR [4])	63.17	60.14	57.54	70.84	68.32	61.44
+ *weight transfer operations*	62.98	61.23	60.36	71.66	71.02	69.40
+ *weight transfer operations* and *mnemonics exemplars*	**63.34**	**62.28**	**60.96**	**72.58**	**71.37**	**69.74**

References

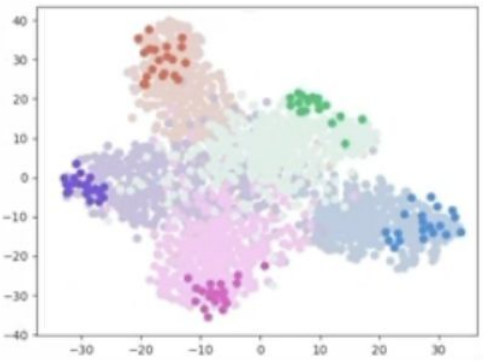
[4] Hou, Saihui, et al. "Learning a unified classifier incrementally via rebalancing." CVPR 2019.

t-SNE results: clearer separation in the data

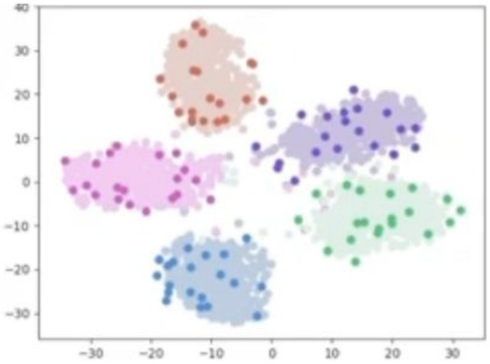
Phase 25

One region for one class
Light color: original data
Deep color: exemplars

Dataset: ImageNet



Herding [1, 4]



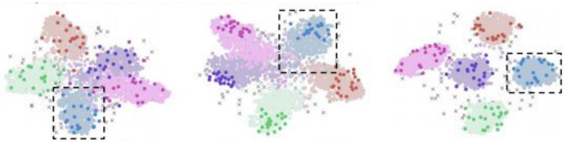
Mnemonics (ours)

Our method:

- *Clearer separation in data*
- *Exemplars locate on the class boundaries*

References
[1] Rebuffi, Sylvestre-Alvise, et al. "icarl: Incremental classifier and representation learning." CVPR 2017;
[4] Hou, Saihui, et al. "Learning a unified classifier incrementally via rebalancing." CVPR 2019.

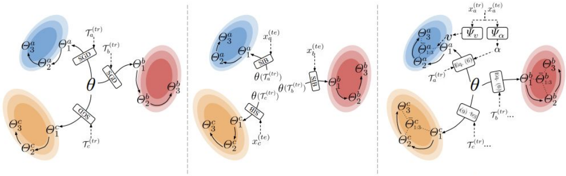
Outline of today's talk



random (baseline) herding (related) mnemonics (ours)

1. **Class-Incremental Learning**

   Mnemonics Training: Multi-Class Incremental Learning without Forgetting
   CVPR 2020



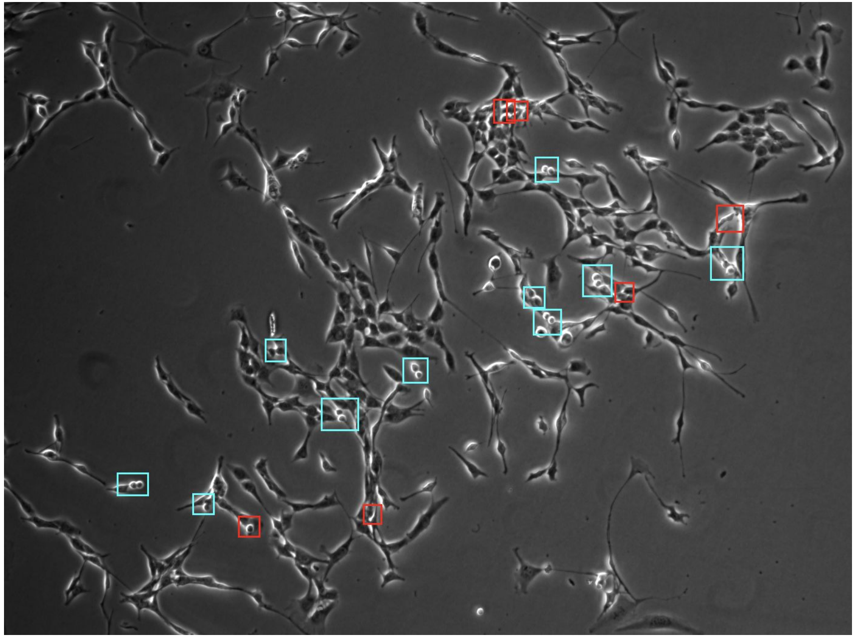
(a) MAML [13] (b) SIB [25] (c) E^{3}BM (ours)

2. **Few-Shot Learning**

   An Ensemble of Epoch-wise Empirical Bayes for Few-shot Learning
   ECCV 2020

Research background

- ***Limitation***: most algorithms are based on ***supervised learning***, so we need lots of ***labeled samples*** to train the model



Medical images:
expensive to label the data

Mitosis detection
有丝分裂检测

(Image from Yao Lu)

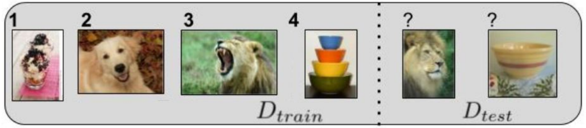
Few-shot learning: learning with limited data

Question: how to learn a model with limited labeled data?

Task: few-shot image classification

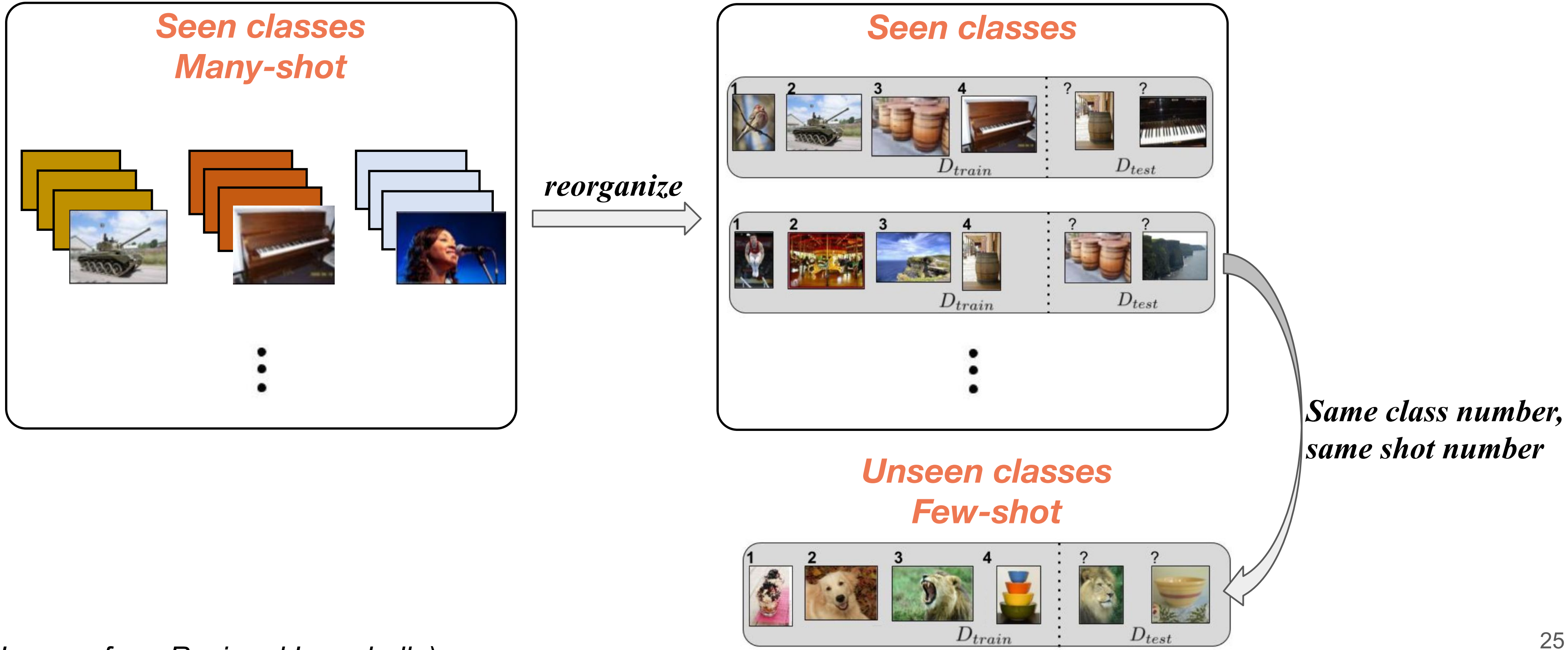
Seen classes
Many-shot

Unseen classes
Few-shot



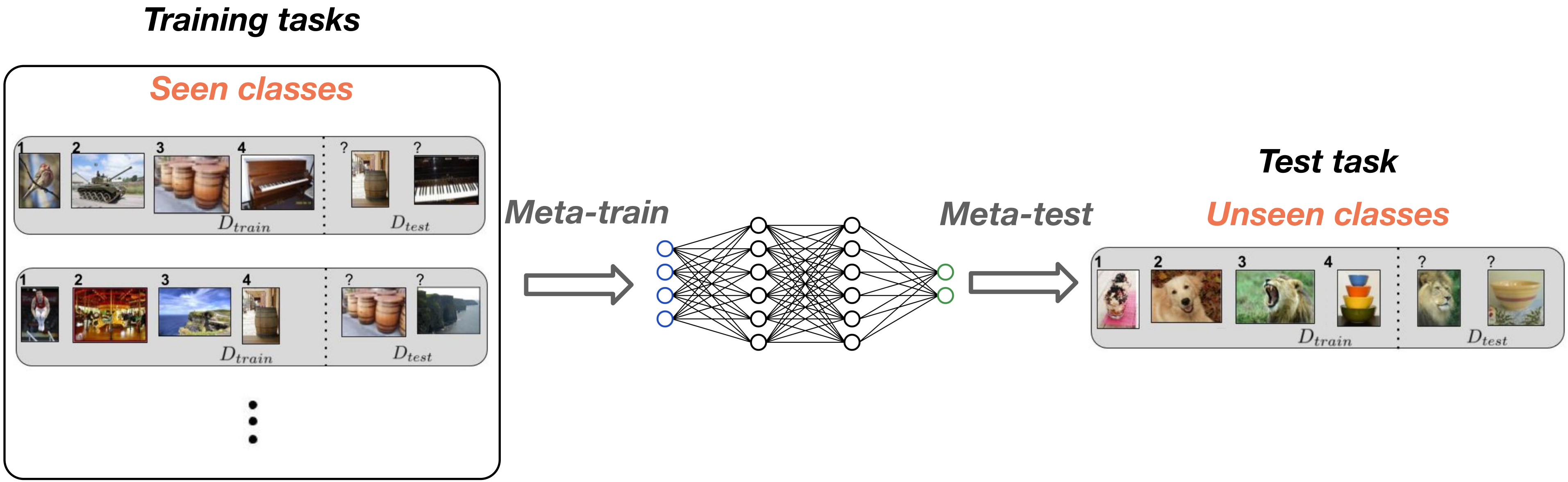
(Images from Ravi and Larochelle)

Review: meta-learning



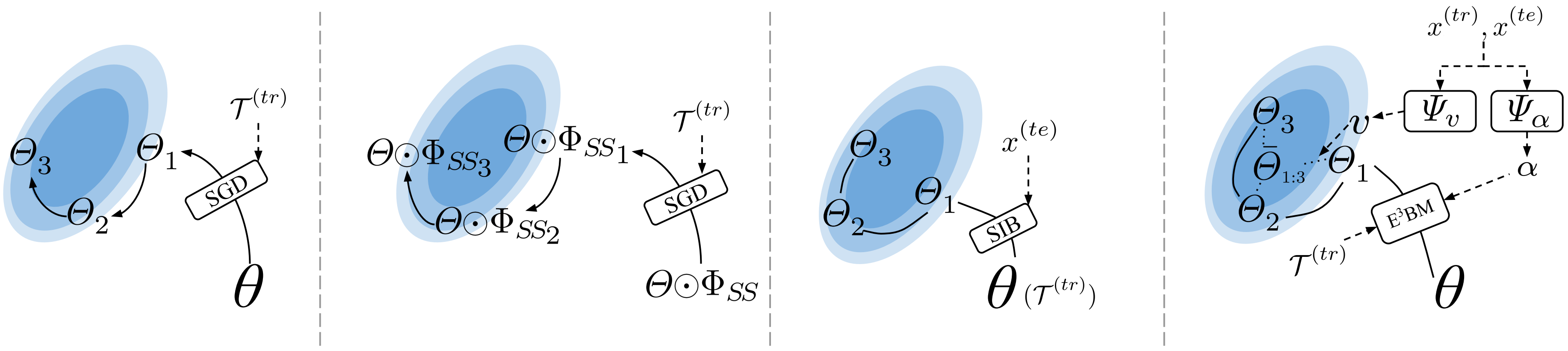
(Images from Ravi and Larochelle)

Review: meta-learning



(Images from Ravi, Larochelle)

Existing methods vs. our E³BM



(a) MAML [8] (b) MTL [9] (c) SIB [10] (d) E³BM (ours)

Existing methods:

- *A single base-learner*
- *Arbitrary base-learning hyperparameters*

\- *Unstable*

Our E³BM:

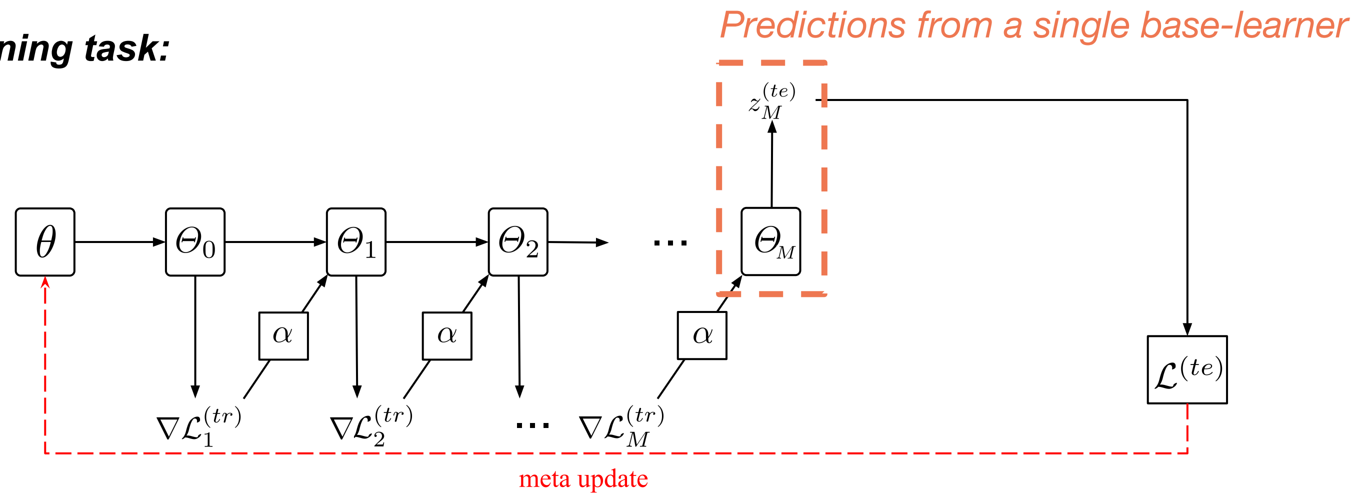
- *An ensemble of multiple base-learners*
- *Task-specific base-learning hyperparameters*

\+ *Stable and robust*

References
[8] Finn, Chelsea, Pieter Abbeel, and Sergey Levine. "Model-agnostic meta-learning for fast adaptation of deep networks." ICML 2017;
[9] Sun, Qianru, et al. "Meta-transfer learning for few-shot learning." CVPR 2019;
[10] Hu, Shell Xu, et al. "Empirical Bayes Transductive Meta-Learning with Synthetic Gradients." ICLR 2020.

Existing method: MAML[8]

For one training task:



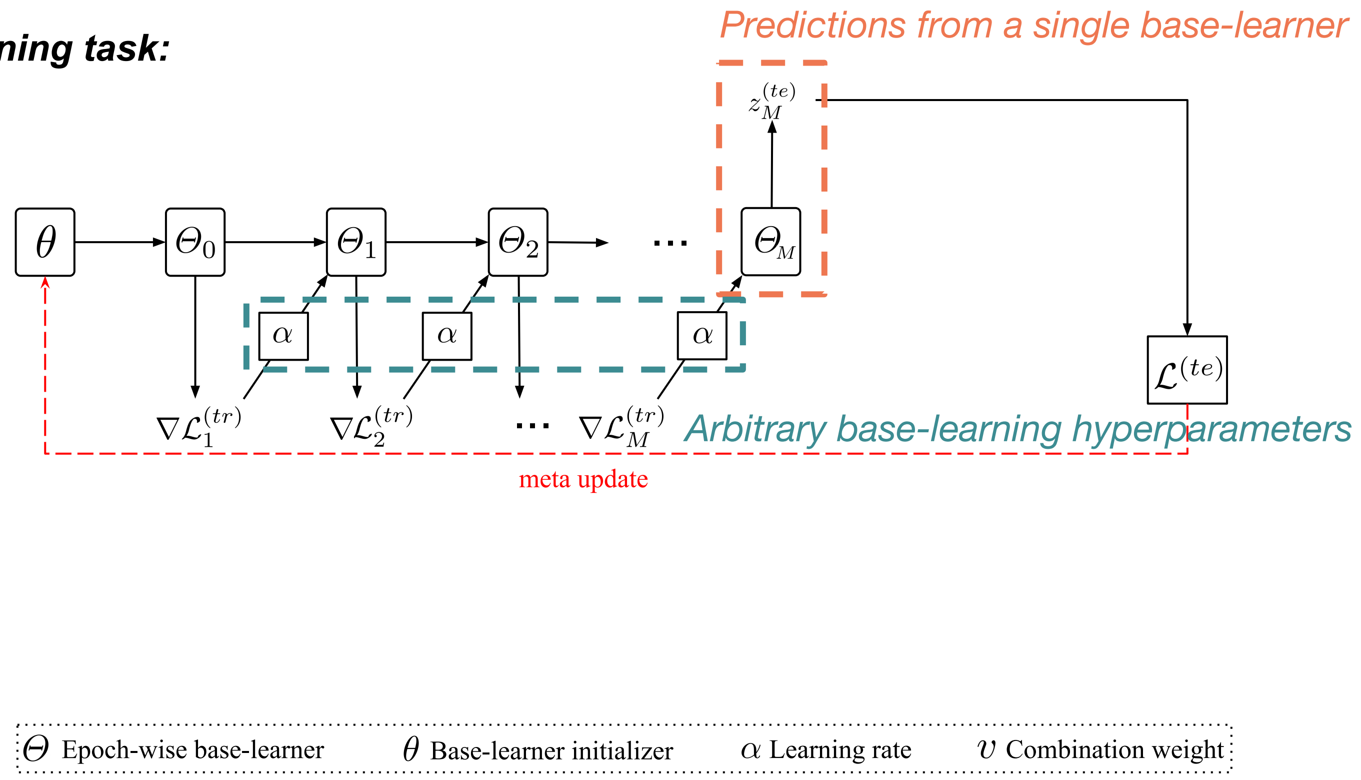
Θ Epoch-wise base-learner $\quad$ θ Base-learner initializer $\quad$ α Learning rate $\quad$ v Combination weight

References

[8] Finn, Chelsea, Pieter Abbeel, and Sergey Levine. "Model-agnostic meta-learning for fast adaptation of deep networks." ICML 2017.

Existing method: MAML[8]

For one training task:

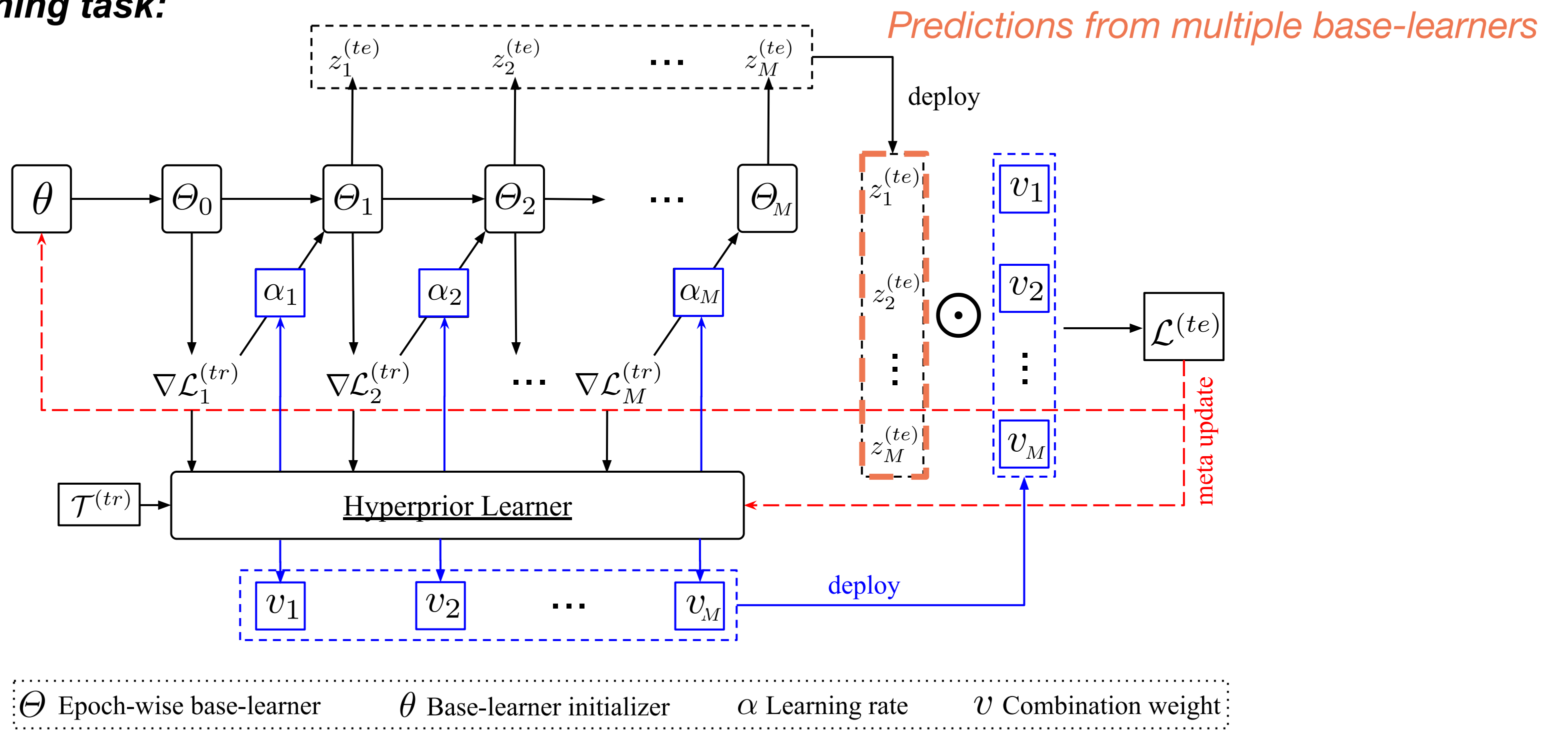


References

[8] Finn, Chelsea, Pieter Abbeel, and Sergey Levine. “Model-agnostic meta-learning for fast adaptation of deep networks.” ICML 2017.

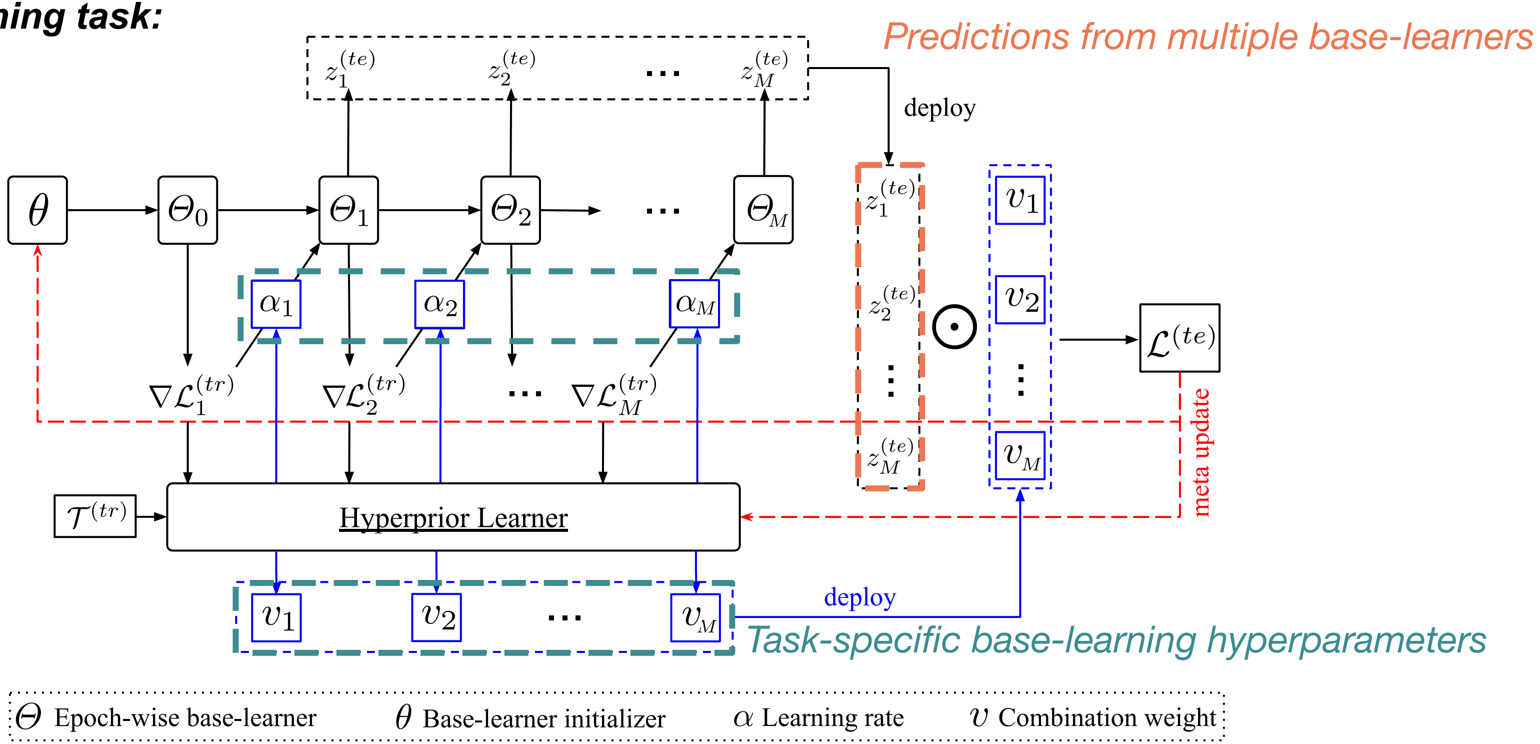
Our method: E³BM framework

For one training task:



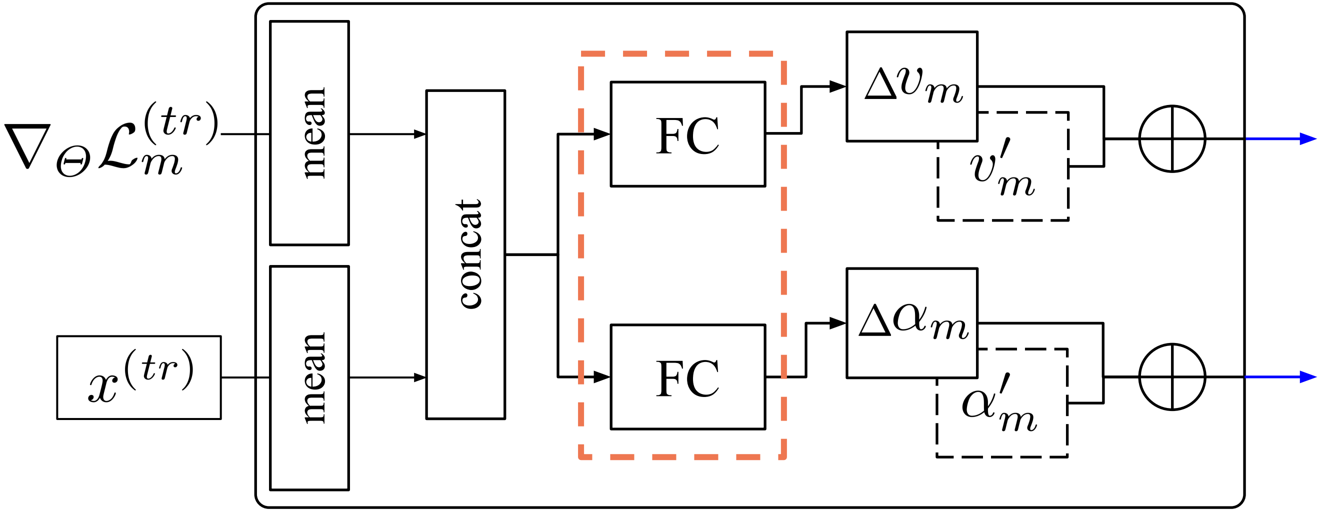
Our method: E³BM framework

For one training task:



The architecture of the hyperprior learner

For the m-th base epoch:



(a) Epoch-independent

Boost the performance on THREE baselines

The 5-class few-shot classification results (%).

Methods	Backbone	*mini*ImageNet 1-shot	*mini*ImageNet 5-shot	*tiered*ImageNet 1-shot	*tiered*ImageNet 5-shot	FC100 1-shot	FC100 5-shot
MAML	4CONV	48.70	63.11	49.0	66.5	38.1	50.4
MTL	ResNet-25	63.4	80.1	69.1	84.2	43.7	**60.1**
MAML+E^{3}BM	4CONV	53.2(↑4.5)	65.1(↑2.0)	52.1(↑3.1)	70.2(↑3.7)	39.9(↑1.8)	52.6(↑2.2)
(+time, +param)	–	(8.9, 2.2)	(9.7, 2.2)	(10.6, 2.2)	(9.3, 2.2)	(7.8, 2.2)	(12.1, 2.2)
MTL+E^{3}BM	ResNet-25	**64.3**(↑0.9)	**81.0**(↑0.9)	**70.0**(↑0.9)	**85.0**(↑0.8)	**45.0**(↑1.3)	**60.5**(↑0.4)
(+time, +param)	–	(5.9, 0.7)	(10.2, 0.7)	(6.7, 0.7)	(9.5, 0.7)	(5.7, 0.7)	(7.9, 0.7)

(a) Inductive Methods

Methods	Backbone	*mini*ImageNet 1-shot	*mini*ImageNet 5-shot	*tiered*ImageNet 1-shot	*tiered*ImageNet 5-shot	FC100 1-shot	FC100 5-shot
SIB	WRN-28-10	**70.0**	79.2	**72.9**	82.8	**45.2**	**55.9**
SIB+E^{3}BM	WRN-28-10	**71.4**(↑1.4)	**81.2**(↑2.0)	**75.6**(↑2.7)	**84.3**(↑1.5)	**46.0**(↑0.8)	**57.1**(↑1.2)
(+time, +param)	–	(2.1, 0.04)	(5.7, 0.04)	(5.2, 0.04)	(4.9, 0.04)	(6.1, 0.04)	(7.3, 0.04)

(b) Transductive Methods

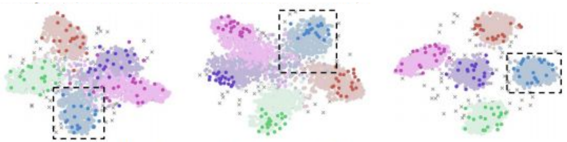
References

[8] Finn, Chelsea, Pieter Abbeel, and Sergey Levine. “Model-agnostic meta-learning for fast adaptation of deep networks.” ICML 2017;

[9] Sun, Qianru, et al. “Meta-transfer learning for few-shot learning.” CVPR 2019;

[10] Hu, Shell Xu, et al. “Empirical Bayes Transductive Meta-Learning with Synthetic Gradients.” ICLR 2020.

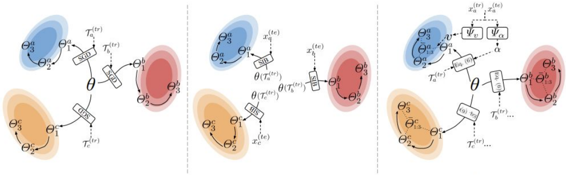
Open-source resources



random (baseline) *herding* (related) *mnemonics* (ours)

1. **Class-Incremental Learning**

   Mnemonics Training: Multi-Class Incremental Learning without Forgetting
   GitHub: https://github.com/yaoyao-liu/mnemonics-training



(a) MAML [13] (b) SIB [25] (c) E^3BM (ours)

2. **Few-Shot Learning**

   An Ensemble of Epoch-wise Empirical Bayes for Few-shot Learning
   GitHub: https://github.com/yaoyao-liu/e3bm

Thanks!
Any questions?

Yaoyao Liu
yaoyao.liu@mpi-inf.mpg.de

References

[1] Rebuffi, Sylvestre-Alvise, et al. "iCaRL: Incremental classifier and representation learning." CVPR 2017;
[2] Belouadah, Eden, and Adrian Popescu. "Il2m: Class incremental learning with dual memory." CVPR 2019;
[3] Li, Zhizhong, and Derek Hoiem. "Learning without forgetting." TPAMI 2017;
[4] Hou, Saihui, et al. "Learning a unified classifier incrementally via rebalancing." CVPR 2019;
[5] Douillard, Arthur, et al. "PODNet: Pooled Outputs Distillation for Small-Tasks Incremental Learning." ECCV 2020;
[6] Wu, Yue, et al. “Large scale incremental learning.” CVPR 2019;
[7] Taylor, Matthew E., and Peter Stone. "An introduction to intertask transfer for reinforcement learning." Ai Magazine 32.1 (2011): 15-15;
[8] Finn, Chelsea, Pieter Abbeel, and Sergey Levine. “Model-agnostic meta-learning for fast adaptation of deep networks.” ICML 2017;
[9] Sun, Qianru, et al. “Meta-transfer learning for few-shot learning.” CVPR 2019;
[10] Hu, Shell Xu, et al. “Empirical Bayes Transductive Meta-Learning with Synthetic Gradients.” ICLR 2020;
[11] Sun, Qianru, et al. "Meta-transfer learning for few-shot learning." CVPR 2019;
[12] Liu, Yaoyao, et al. "Mnemonics Training: Multi-Class Incremental Learning without Forgetting." CVPR 2020;
[13] Sun, Qianru, et al. "Meta-Transfer Learning through Hard Tasks." TPAMI 2020;
[14] Liu, Yaoyao, Bernt Schiele, and Qianru Sun. "An Ensemble of Epoch-wise Empirical Bayes for Few-shot Learning." ECCV 2020.