Meta-Learning with Self-Improving Momentum Target

Jihoon Tack[1], Jongjin Park[1], Hankook Lee[1], Jaeho Lee[2], Jinwoo Shin[1]

Korea Advanced Institute of Science and Technology (KAIST)[1]
Pohang University of Science and Technology (POSTECH)[2]

Meta-Learning

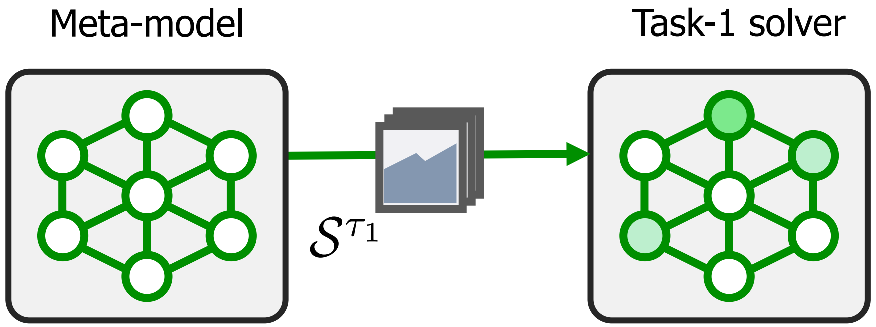
What is meta-learning?

- Extracting and utilizing the **knowledge from the distribution of tasks** to better solve a relevant task

Meta-Learning

What is meta-learning?

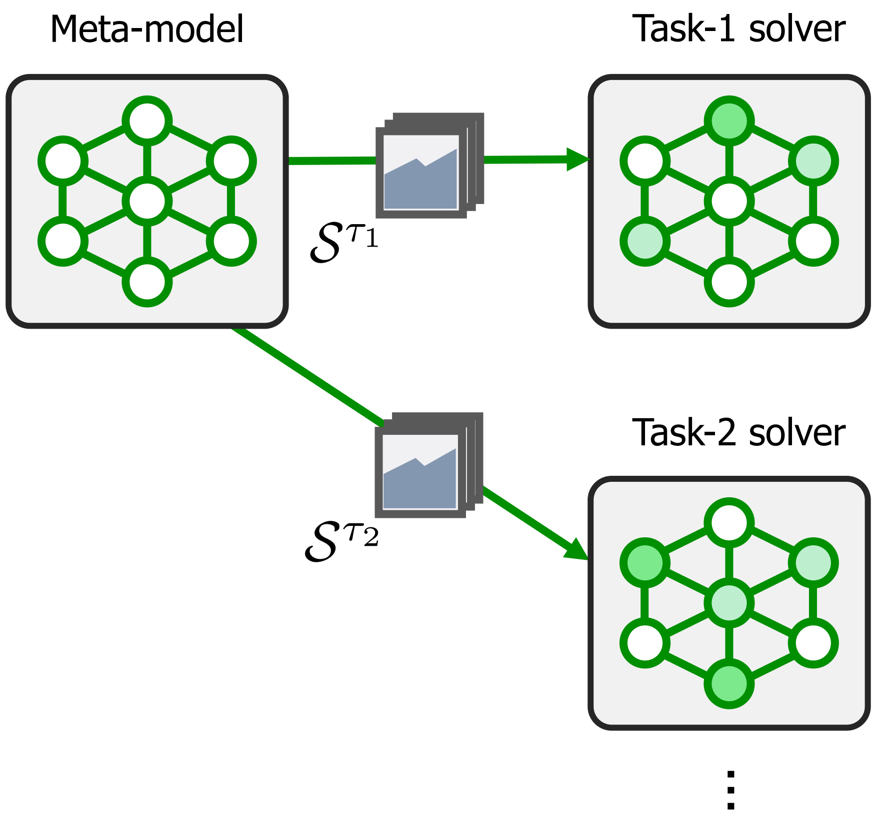
- Extracting and utilizing the **knowledge from the distribution of tasks** to better solve a relevant task



Meta-Learning

What is meta-learning?

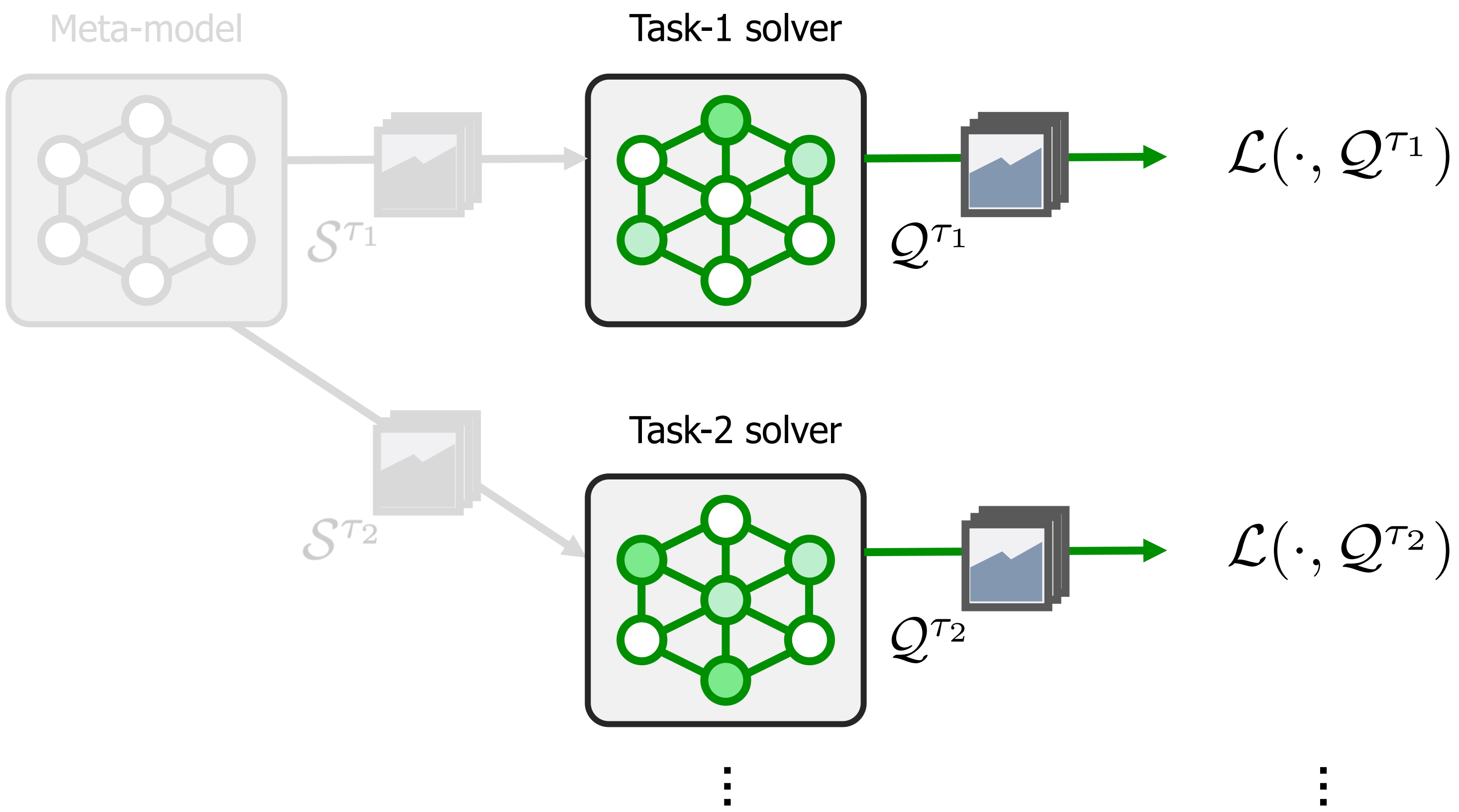
- Extracting and utilizing the **knowledge from the distribution of tasks** to better solve a relevant task



Meta-Learning

What is meta-learning?

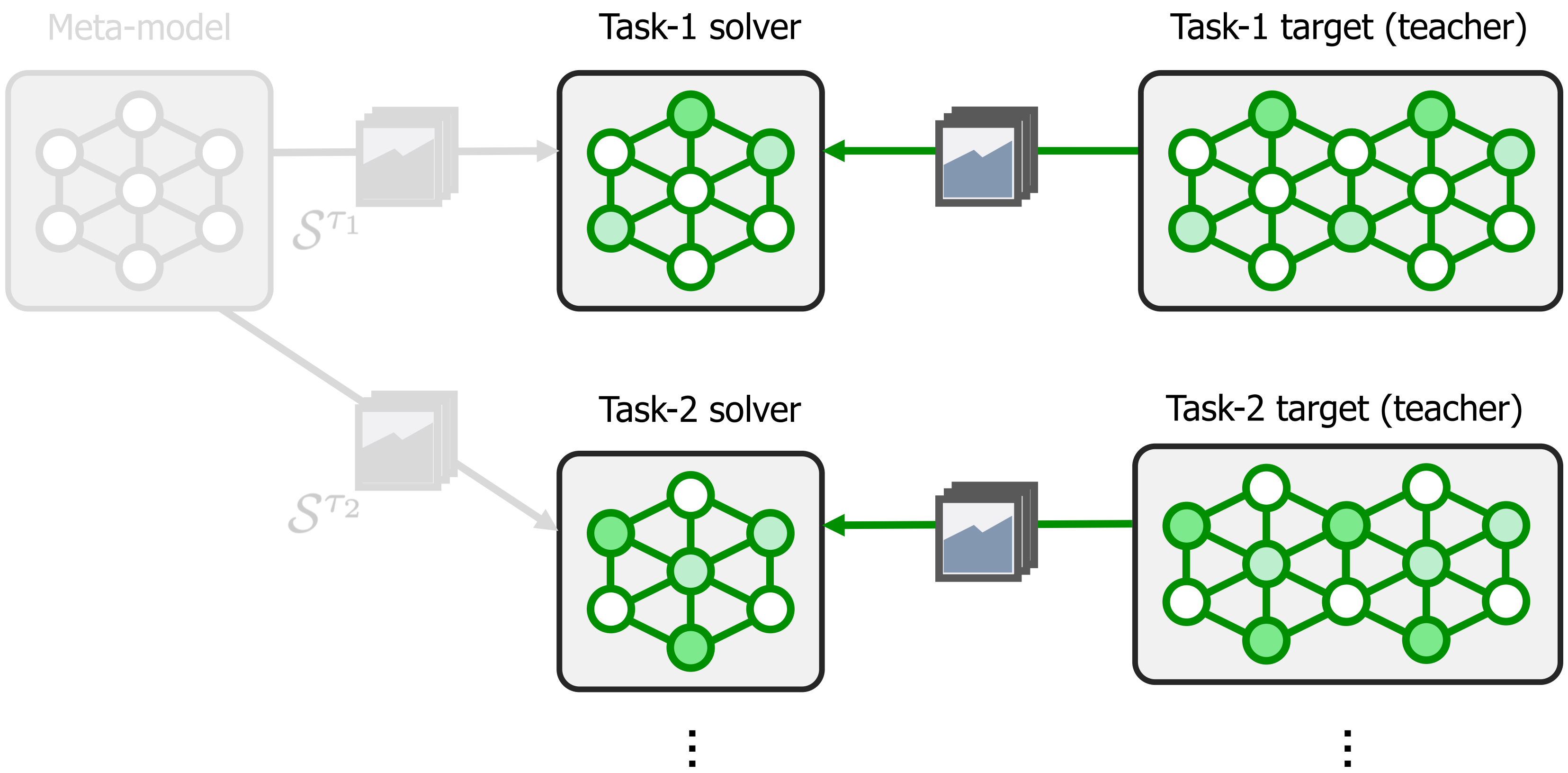
- Extracting and utilizing the **knowledge from the distribution of tasks** to better solve a relevant task



Meta-Learning with Target Models

Recently an alternative paradigm has gained much attention [1,2]

- Use a **task-specific target model**, i.e., an expert or teacher, to evaluate the performance



[1] Towards enabling meta-learning from target models, Lu et al., NeurIPS 2021
[2] Bootstrapped Meta-Learning, Flennerhag et al., ICLR 2022

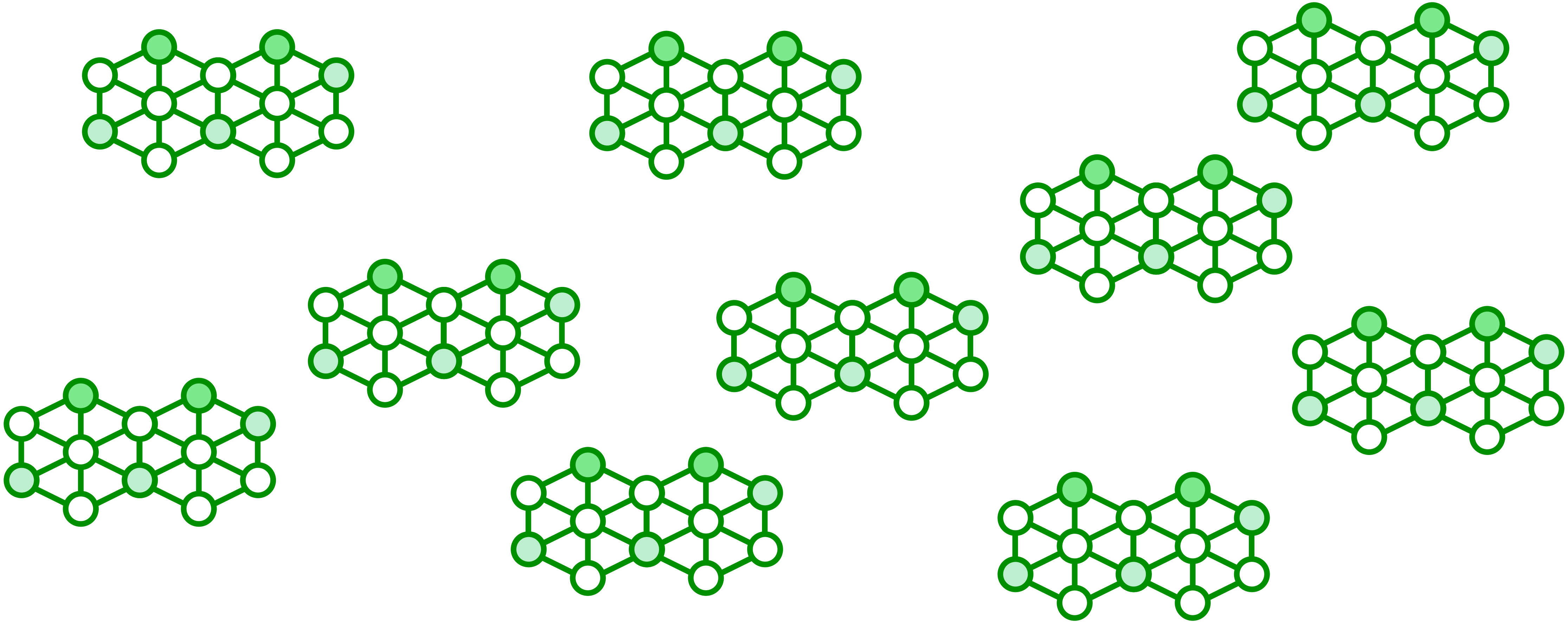
An Obstacle

Q. How can we train target models for **every task**…?

An Obstacle

Q. How can we train target models for **every task**…?

- This suffers from both **computation** and **storage** burden !



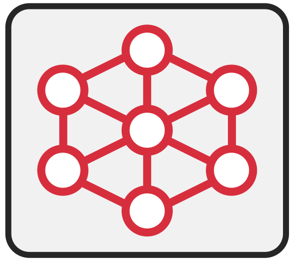
Idea

Q. How can we train target models for **every task**…?

- This suffers from both computation and storage burden !

A. How about **generating target models** from a **better meta-learner**…?

Better meta-model

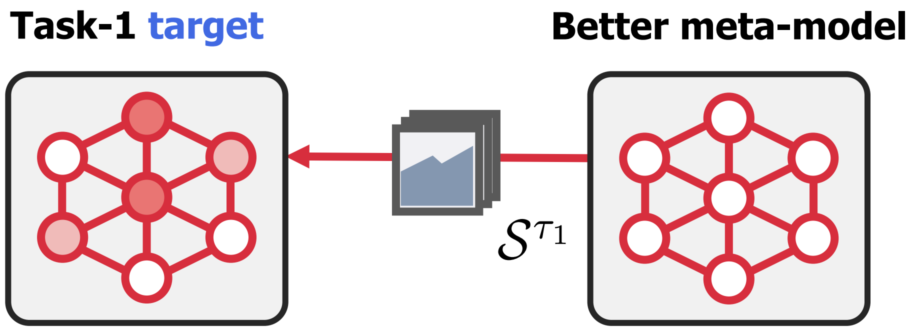


Idea

Q. How can we train target models for **every task**…?

- This suffers from both computation and storage burden !

A. How about **generating target models** from a **better meta-learner**…?

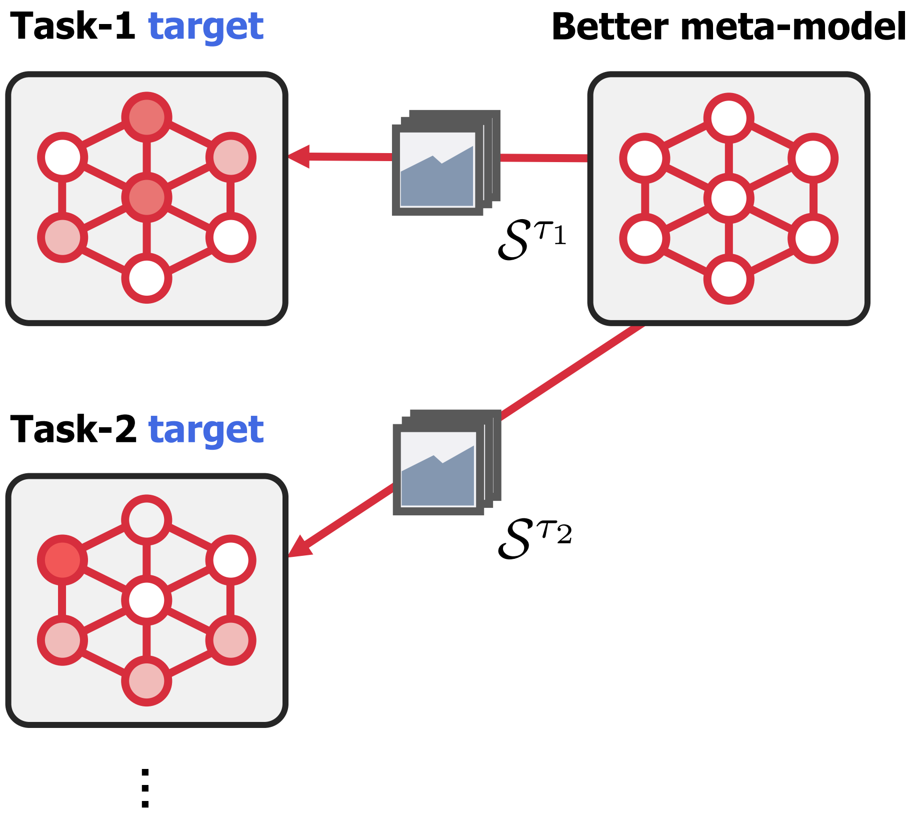


Idea

Q. How can we train target models for **every task**…?

- This suffers from both computation and storage burden !

A. How about **generating target models** from a **better meta-learner**…?

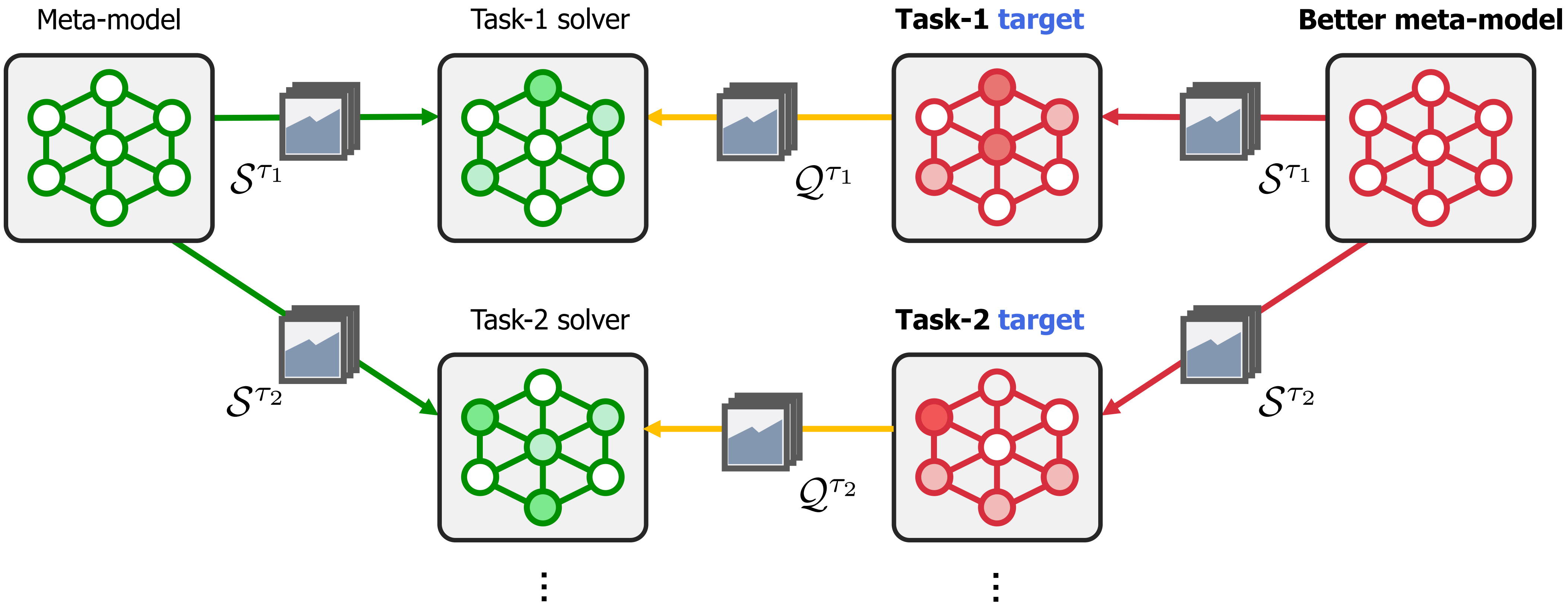


Idea

Q. How can we train target models for **every task**…?

- This suffers from both computation and storage burden !

A. How about **generating target models** from a **better meta-learner**…?



Utilize Momentum Network !

We draw inspiration from recent observations in **semi/self-supervised learning literature** [3,4]

- The **momentum network** can be an effective teacher of the original model

[3] Temporal ensembling for semi-supervised learning, Laine et al., ICLR 2017
[4] Emerging properties in self-supervised vision transformers, Caron et al., ICCV 2021

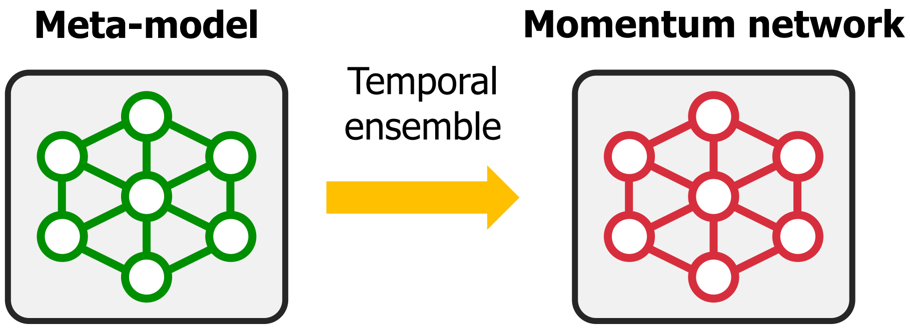
Utilize Momentum Network !

We draw inspiration from recent observations in **semi/self-supervised learning literature** [3,4]

- The **momentum network** can be an effective teacher of the original model

A similar phenomenon happens in meta-learning !

- Momentum network of the meta-model shows an **effective adaptation performance**



$$\theta_{\texttt{moment}} \leftarrow \eta \cdot \theta_{\texttt{moment}} + (1-\eta) \cdot \theta$$

[3] Temporal ensembling for semi-supervised learning, Laine et al., ICLR 2017
[4] Emerging properties in self-supervised vision transformers, Caron et al., ICCV 2021

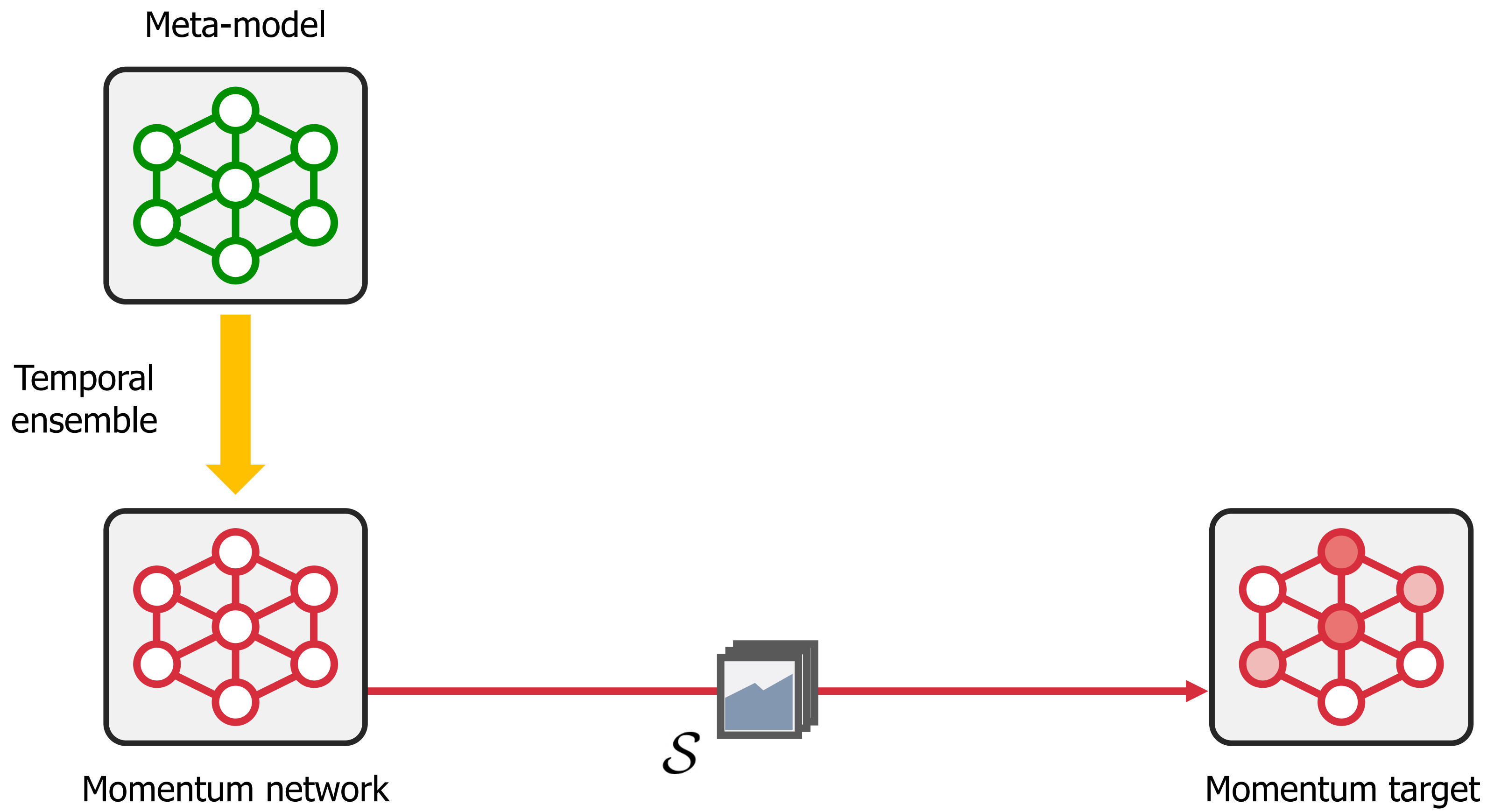
SiMT: Self-improving Momentum Target

We propose **SiMT**, a simple yet effective framework for meta-learning

SiMT: Self-improving Momentum Target

We propose **SiMT**, a simple yet effective framework for meta-learning

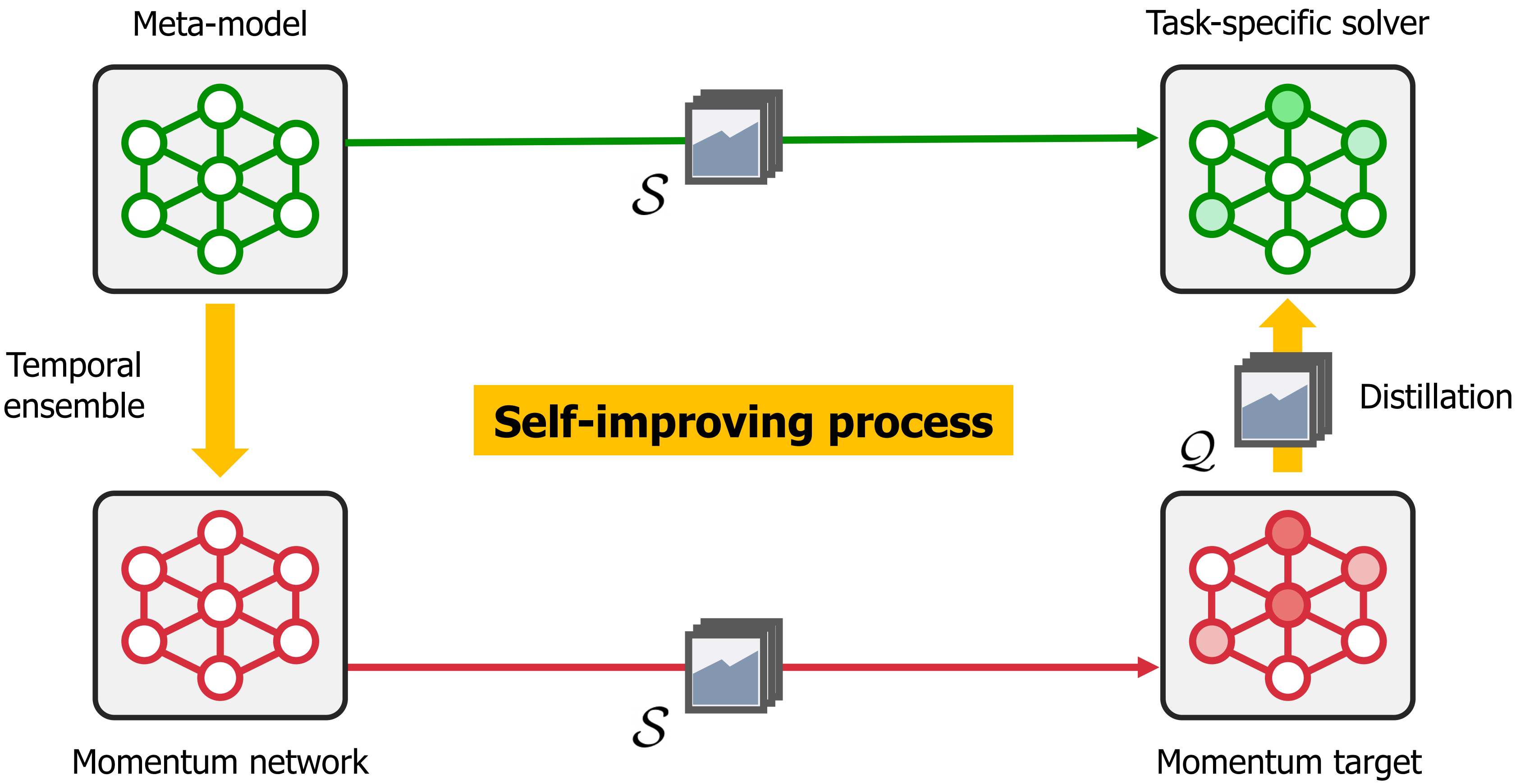
- **Momentum target**: Generating the target model from the momentum network to teach the original model



SiMT: Self-improving Momentum Target

We propose **SiMT**, a simple yet effective framework for meta-learning

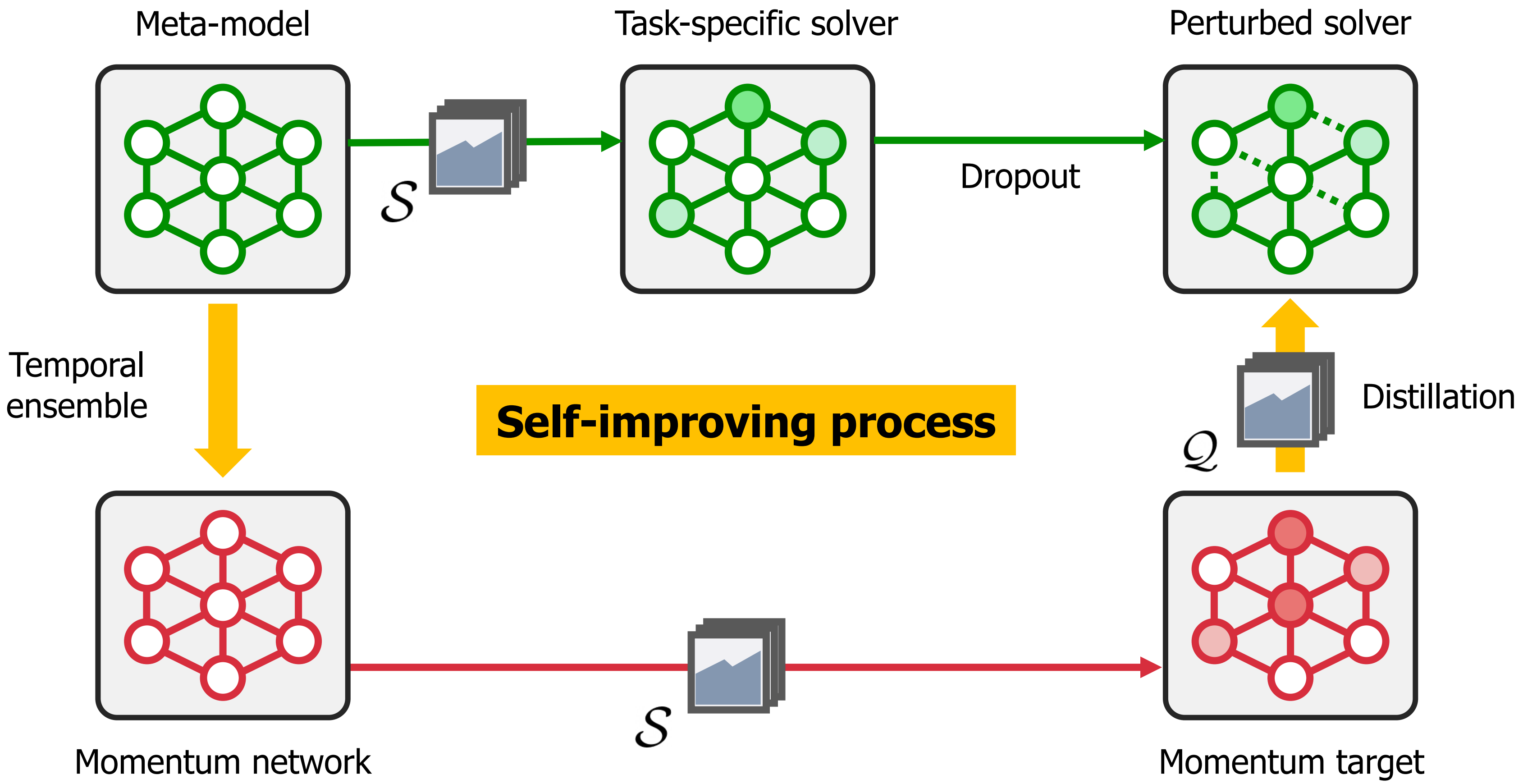
- **Momentum target**: Generating the target model from the momentum network to teach the original model
- **Self-improving process**: Improving the meta-model recursively improves the momentum network



SiMT: Self-improving Momentum Target

We propose **SiMT**, a simple yet effective framework for meta-learning

- **Momentum target**: Generating the target model from the momentum network to teach the original model
- **Self-improving process**: Improving the meta-model recursively improves the momentum network



Experiments: Few-shot Learning

We verify the effectiveness of SiMT in various

- **few-shot learning** and **meta-reinforcement learning scenarios**

SiMT **consistently and significantly improves** the backbone meta-learning algorithms

Model	Method	mini-ImageNet 1-shot	mini-ImageNet 5-shot	tiered-ImageNet 1-shot	tiered-ImageNet 5-shot
Conv4 [55]	MAML [10]	47.33±0.45	63.27±0.14	50.19±0.21	66.05±0.19
	MAML [10] + SiMT	**51.49±0.18**	**68.74±0.12**	**52.51±0.21**	**69.58±0.11**
	ANIL [36]	47.71±0.47	63.13±0.43	49.57±0.04	66.34±0.28
	ANIL [36] + SiMT	**50.81±0.56**	**67.99±0.19**	**51.66±0.26**	**68.88±0.08**
	MetaSGD [31]	50.66±0.18	65.55±0.54	52.48±1.22	71.06±0.20
	MetaSGD [31] + SiMT	**51.70±0.80**	**69.13±1.40**	**52.98±0.07**	**71.46±0.12**
	ProtoNet [45]	47.97±0.29	65.16±0.67	51.90±0.55	71.51±0.25
	ProtoNet [45] + SiMT	**51.25±0.55**	**68.71±0.35**	**53.25±0.27**	**72.69±0.27**
ResNet-12 [34]	MAML [10]	52.66±0.60	68.69±0.33	57.32±0.59	73.78±0.27
	MAML [10] + SiMT	**56.28±0.63**	**72.01±0.26**	**59.72±0.22**	**74.40±0.90**
	ANIL [36]	51.80±0.59	68.38±0.20	57.52±0.68	73.50±0.35
	ANIL [36] + SiMT	**54.44±0.27**	**69.98±0.66**	**58.18±0.31**	**75.59±0.50**
	MetaSGD [31]	54.95±0.11	70.65±0.43	58.97±0.89	76.37±0.11
	MetaSGD [31] + SiMT	**55.72±0.96**	**74.01±0.79**	**61.03±0.05**	**78.04±0.48**
	ProtoNet [45]	52.84±0.21	68.35±0.29	61.16±0.17	79.94±0.20
	ProtoNet [45] + SiMT	**55.84±0.57**	**72.45±0.32**	**62.01±0.42**	**81.82±0.12**

In-domain adaptation accuracy (%)

Problem	Method	mini-ImageNet → CUB	mini-ImageNet → Cars	tiered-ImageNet → CUB	tiered-ImageNet → Cars
1-shot	MAML [10]	39.50±0.91	32.87±0.20	42.32±0.69	36.62±0.12
	MAML [10] + SiMT	**42.32±0.62**	**33.73±0.63**	**44.33±0.43**	**37.21±0.35**
	ANIL [36]	37.30±0.89	31.28±1.03	42.29±0.33	36.27±0.58
	ANIL [36] + SiMT	**38.86±0.98**	**32.34±0.95**	**44.53±1.21**	**36.92±0.56**
	MetaSGD [31]	41.98±0.18	**34.52±0.56**	46.48±2.10	38.09±1.21
	MetaSGD [31] + SiMT	**43.50±0.89**	33.92±0.30	**46.62±0.41**	**38.69±0.26**
	ProtoNet [45]	41.22±0.81	32.79±0.61	47.75±0.56	37.59±0.80
	ProtoNet [45] + SiMT	**44.13±0.30**	**34.53±0.40**	**48.89±0.65**	**38.07±0.42**
5-shot	MAML [10]	56.17±0.92	44.56±0.79	65.00±0.89	51.08±0.28
	MAML [10] + SiMT	**59.22±0.39**	**46.59±0.21**	**67.58±0.61**	**51.88±0.52**
	ANIL [36]	53.42±0.97	41.65±0.67	62.48±0.85	50.50±1.18
	ANIL [36] + SiMT	**56.03±1.40**	**45.88±0.82**	**66.30±0.99**	**54.60±0.91**
	MetaSGD [31]	58.90±1.30	47.44±1.55	70.38±0.27	56.28±0.07
	MetaSGD [31] + SiMT	**65.07±1.89**	**49.86±0.84**	**73.93±0.42**	**57.97±1.34**
	ProtoNet [45]	57.87±0.77	48.06±1.10	74.35±0.93	57.23±0.25
	ProtoNet [45] + SiMT	**63.85±0.76**	**51.67±0.29**	**75.97±0.09**	**59.01±0.50**

Cross-domain adaptation accuracy (%)

Experiments: Comparison with Other Target Models

We also compare SiMT with **other target models**

- (i) Bootstrapped target model [2] and (ii) Pre-trained target model [1]

Method	mini-ImageNet 1-shot	mini-ImageNet 5-shot	tiered-ImageNet 1-shot	tiered-ImageNet 5-shot
MAML [10]	47.33±0.45	63.27±0.14	50.19±0.21	66.05±0.19
MAML [10] + Bootstrap [16]	48.68±0.33	68.45±0.40	49.34±0.26	68.84±0.37
MAML [10] + SiMT	**51.49±0.18**	**68.74±0.12**	**52.51±0.21**	**69.58±0.11**
ANIL [36]	47.71±0.47	63.13±0.43	49.57±0.04	66.34±0.28
ANIL [36] + Bootstrap [16]	47.74±0.44	65.16±0.04	48.85±0.34	66.09±0.07
ANIL [36] + SiMT	**50.81±0.56**	**67.99±0.19**	**51.66±0.26**	**68.88±0.08**

(i) Comparison with the Bootstrapped target model [2]

Method	1-shot train cost (GPU hours)	mini-ImageNet 1-shot	mini-ImageNet 5-shot	tiered-ImageNet 1-shot	tiered-ImageNet 5-shot
MAML [10]*	1.31	58.84±0.25	74.62±0.38	63.02±0.30	67.26±0.32
MAML [10] + Lu et al. [32] - 5%*	5.04	59.14±0.33	75.77±0.29	64.52±0.30	68.39±0.34
MAML [10] + Lu et al. [32] - 10%*	8.32	60.06±0.35	76.34±0.42	**65.23**±0.45	70.02±0.33
MAML [10] + SiMT	1.64	**62.05**±0.39	**78.77**±0.45	63.91±0.32	**77.43**±0.47

(ii) Comparison with the Pre-trained target model [1]

[1] Towards enabling meta-learning from target models, Lu et al., NeurIPS 2021
[2] Bootstrapped Meta-Learning, Flennerhag et al., ICLR 2022

Thank you for your attention ☺

For any more questions, please send us an email!

Email: jihoontack@kaist.ac.kr

Paper: https://arxiv.org/abs/2210.05185

Code: https://github.com/jihoontack/SiMT