NUS-Tsinghua-Southampton
Centre for Extreme Search

Meta-transfer Learning for Few-shot Learning

Yaoyao Liu
Tianjin University and NUS School of Computing

OUTLINE

- Research Background
- Methods
  - Meta-transfer Learning
  - Hard-task Meta Batch
- Experiments and Conclusions

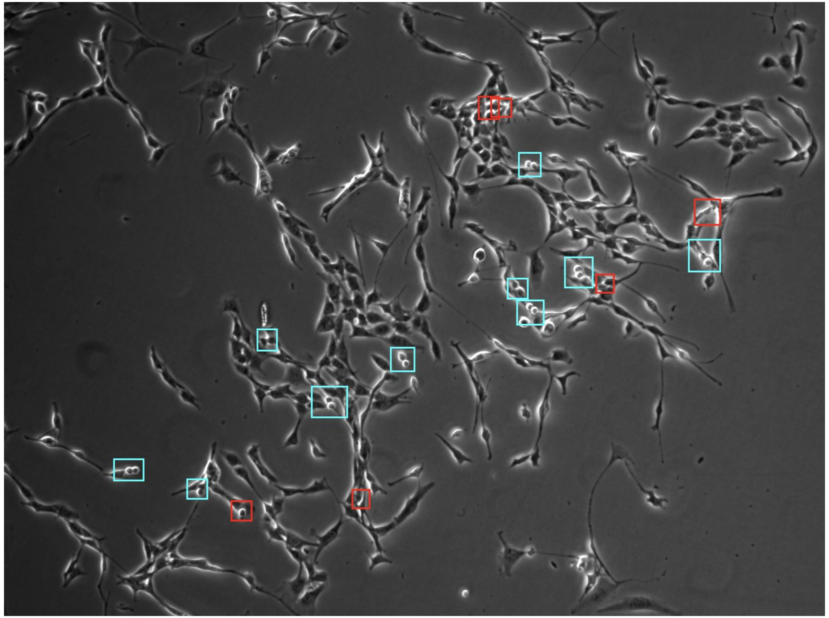
Research Background

- Deep learning achieved a lot of success in many fields:
  Computer Vision, NLP…

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

Research Background

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



medical images

mitosis
有丝分裂

Few-shot learning: learn with limited data

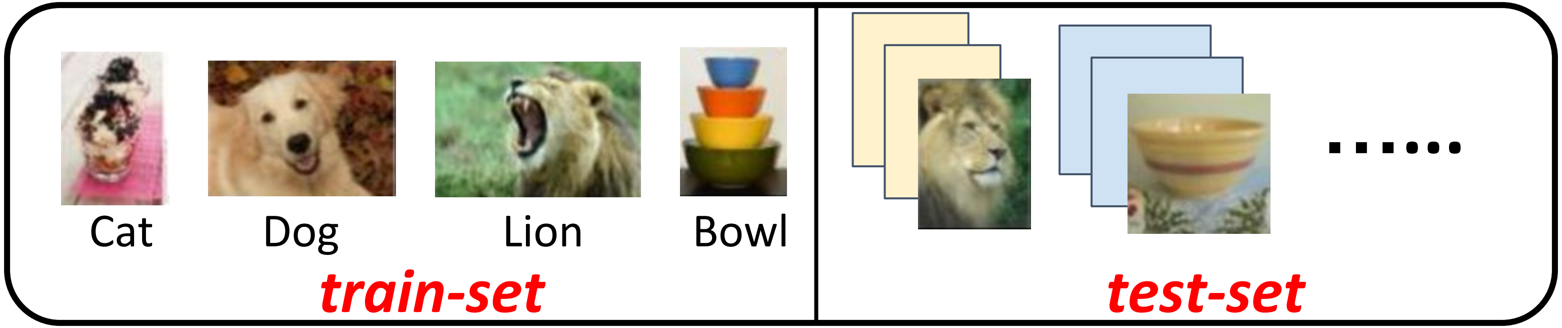
- How to learn a model with limited labeled data?

Task: Few-shot Learning
Our focus: few-shot image classification

Few-shot Classification

Using only ***a few labeled samples*** to train the classifier

1-shot,
4-class

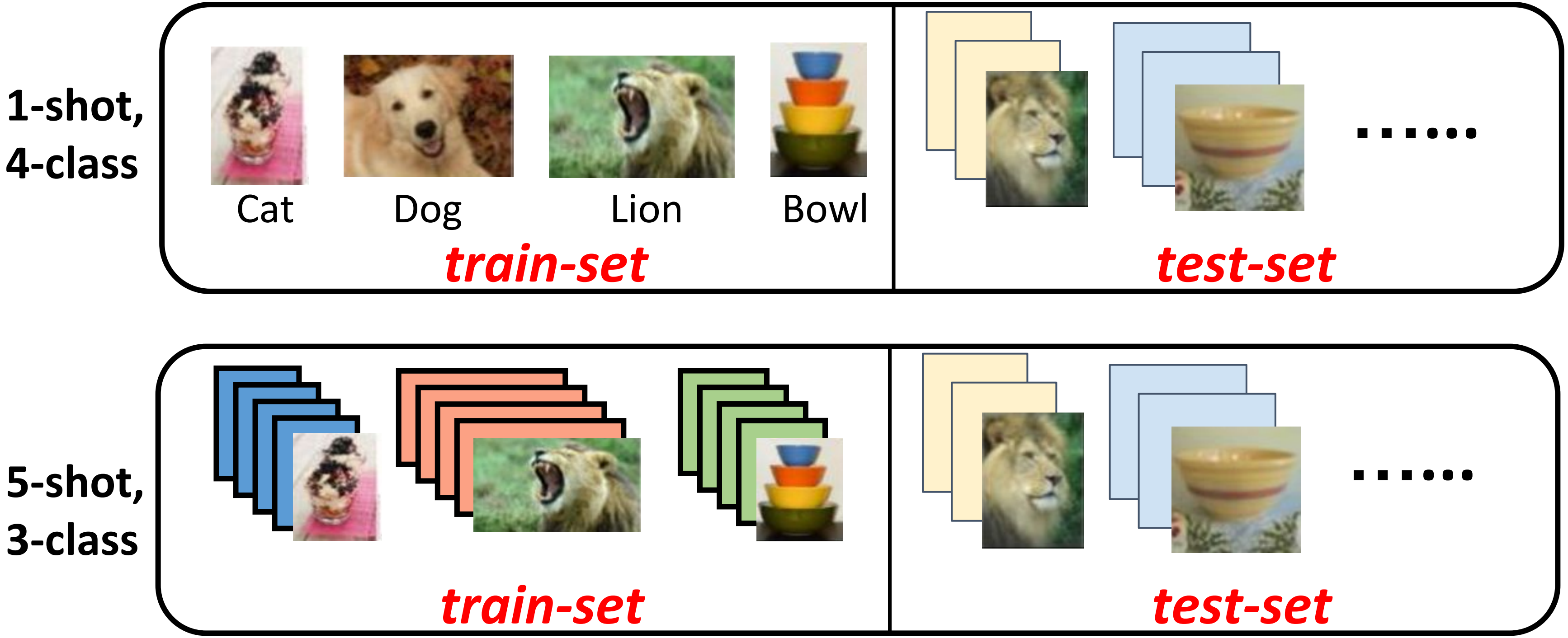


Shot number: how many samples for one class

Class number: how many classes in the small dataset

Few-shot Classification

Using only ***a few labeled samples*** to train the classifier



Literature Review

1. ***Meta learning based:*** ***Design learnable components***
   Meta-LSTM[1], MAML[2], ...
2. ***Metric learning based:*** ***Design distance-based objective functions***
   MatchingNets[3], ProtoNets[4], ...
3. ***Others*** ***(based on augmentation, domain adaptation…):***
   Data Augmentation GAN[5], CCN+[6]...

[1] Ravi et al. "Optimization as a model for few-shot learning." ICLR 2016;
[2] Finn et al. "Model-agnostic meta-learning for fast adaptation of deep networks." ICML 2017;
[3] Vinyals et al. "Matching networks for one shot learning." NIPS 2016;
[4] Snell et al. "Prototypical networks for few-shot learning." NIPS 2017;
[5] Antoniou et al. "Data augmentation generative adversarial networks." In ICLR Workshops 2018;
[6] Hsu et al. "Learning to cluster in order to transfer across domains and tasks." ICLR 2018.

Literature Review

1. ***Meta learning based:***
   Meta-LSTM[1], MAML[2], ...

This talk

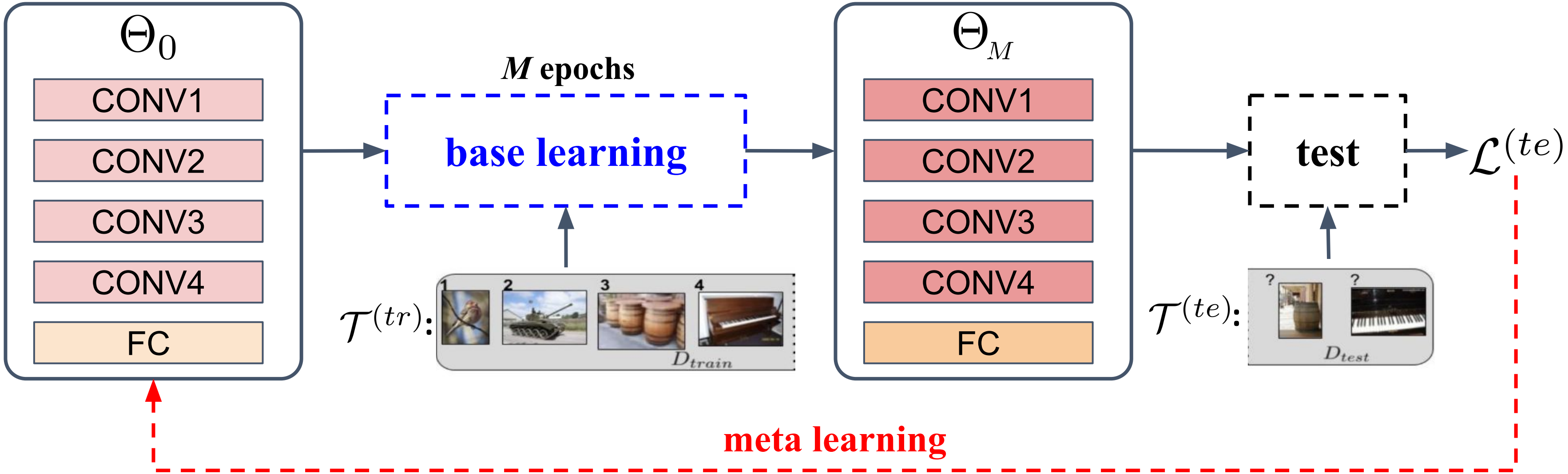
2. ***Metric learning based:***
   MatchingNets[3], ProtoNets[4], ...
3. ***Others** **(based on augmentation, domain adaptation…):***
   Data Augmentation GAN[5], CCN+[6]...

[1] Ravi et al. "Optimization as a model for few-shot learning." ICLR 2016;
[2] Finn et al. "Model-agnostic meta-learning for fast adaptation of deep networks." ICML 2017;
[3] Vinyals et al. "Matching networks for one shot learning." NIPS 2016;
[4] Snell et al. "Prototypical networks for few-shot learning." NIPS 2017;
[5] Antoniou et al. "Data augmentation generative adversarial networks." In ICLR Workshops 2018;
[6] Hsu et al. "Learning to cluster in order to transfer across domains and tasks." ICLR 2018.

OUTLINE

- Research Background
- Methods
  - Meta-transfer Learning
  - Hard-task Meta Batch
- Experiments and Conclusions

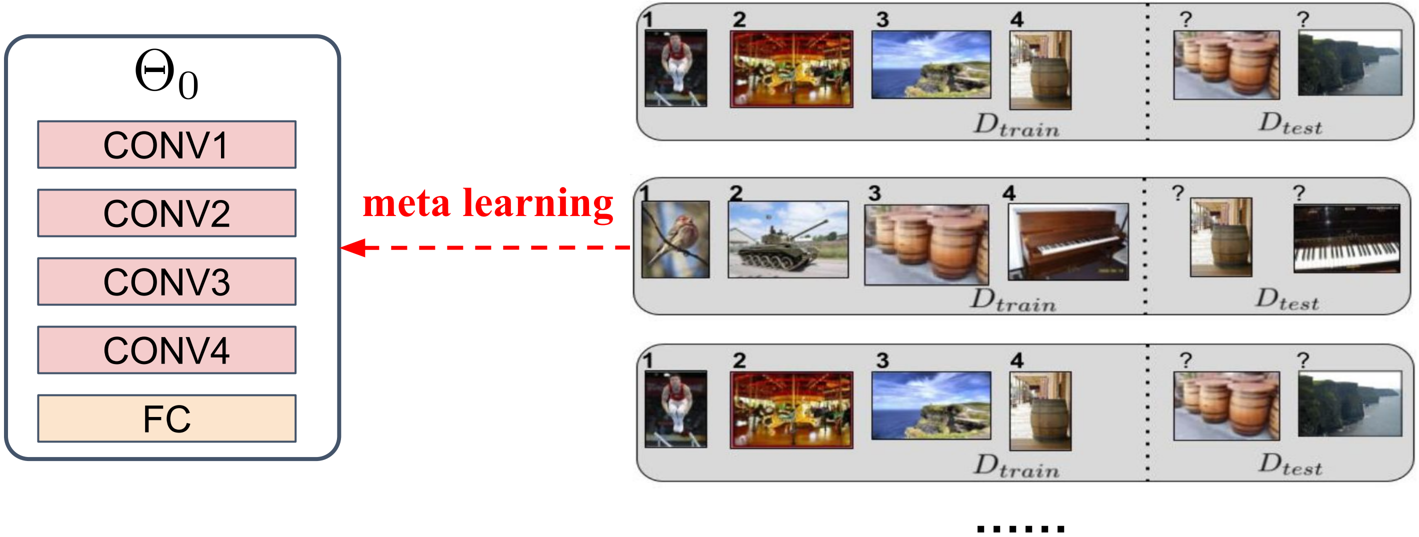
Classic Algorithm: *MAML*



meta-train phase

Finn et al. "Model-agnostic meta-learning for fast adaptation of deep networks." ICML 2017.

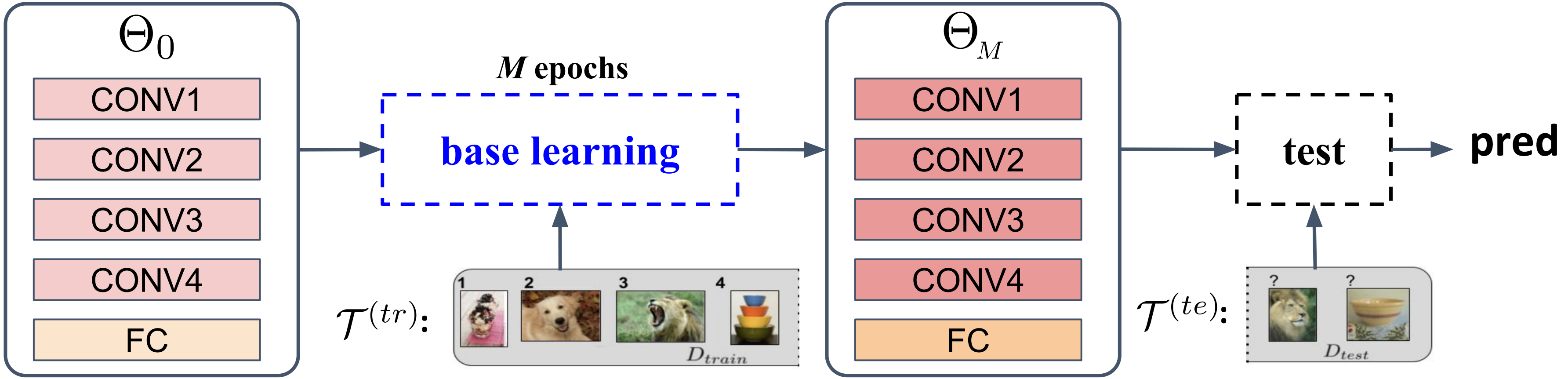
Classic Algorithm: *MAML*



Learn initialization weights for different tasks using meta-learning.

Finn et al. "Model-agnostic meta-learning for fast adaptation of deep networks." ICML 2017.

Classic Algorithm: *MAML*

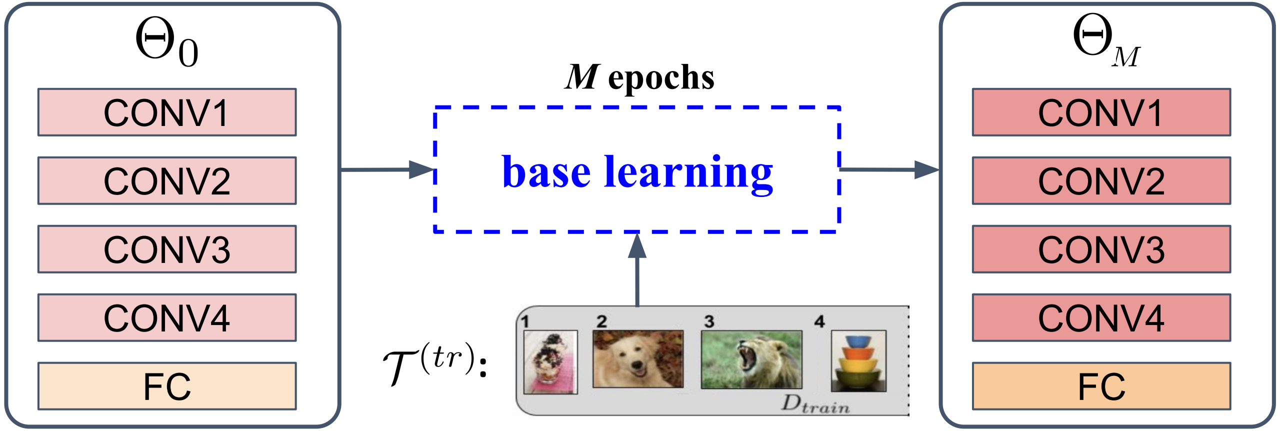


meta-test phase

Finn et al. "Model-agnostic meta-learning for fast adaptation of deep networks." ICML 2017.

Problems of *MAML*

- Failure on deeper networks



Problems of *MAML*

- Failure on deeper networks

- Slow convergence speed

  For the networks with only 4 conv layers, MAML trains ***60k*** iterations.

  It takes more than ***30 hours*** on a NVIDIA V100 GPU.

Our Methods

- Failure on deeper networks ⇨ *Meta-transfer Learning*

- Slow convergence speed ⇨ *Hard Task Meta Batch*

Overview of the Methods

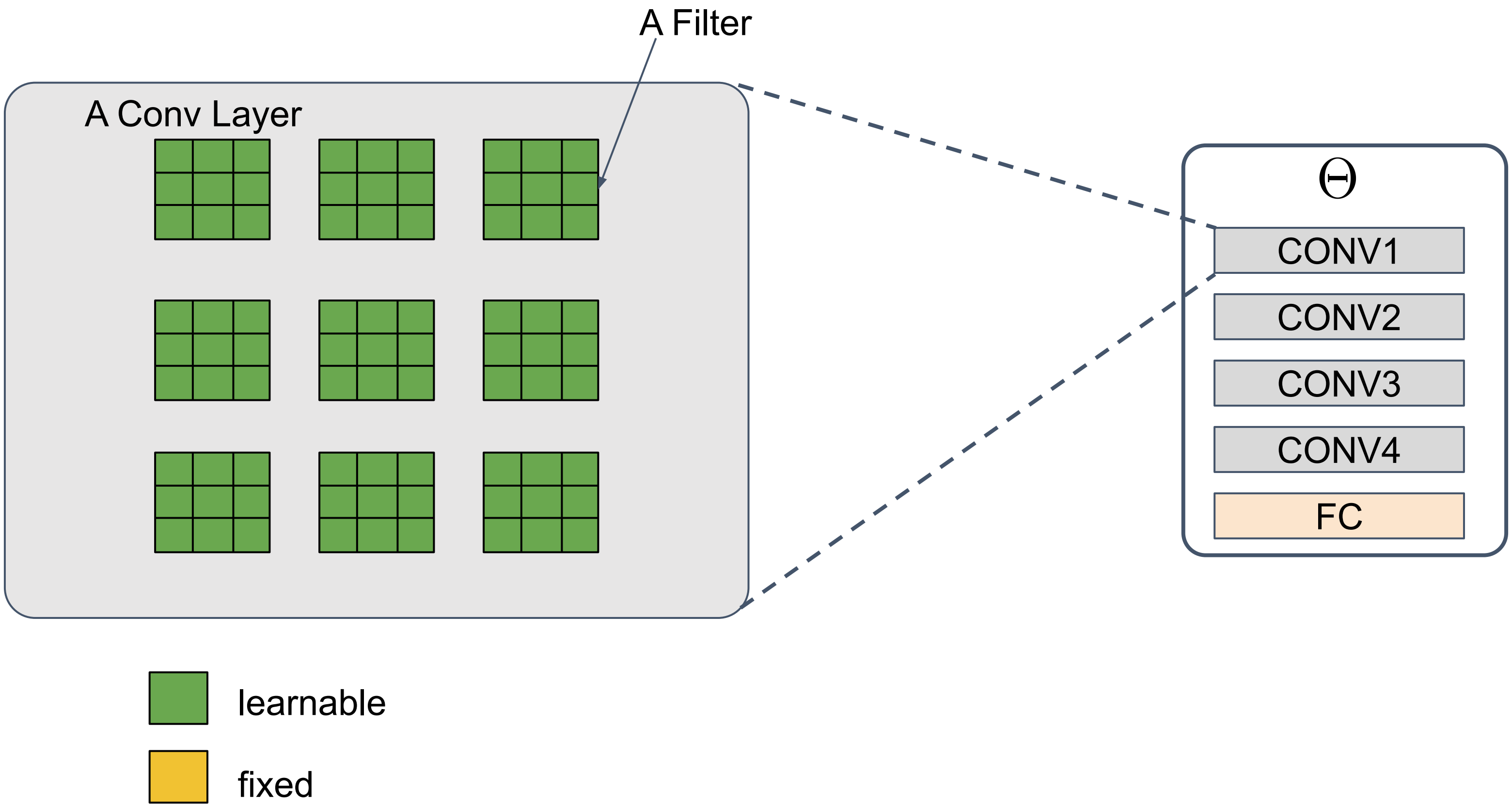
- ***Meta-transfer Learning***

  Explore the structure of the classifier Θ , control the degree of freedom

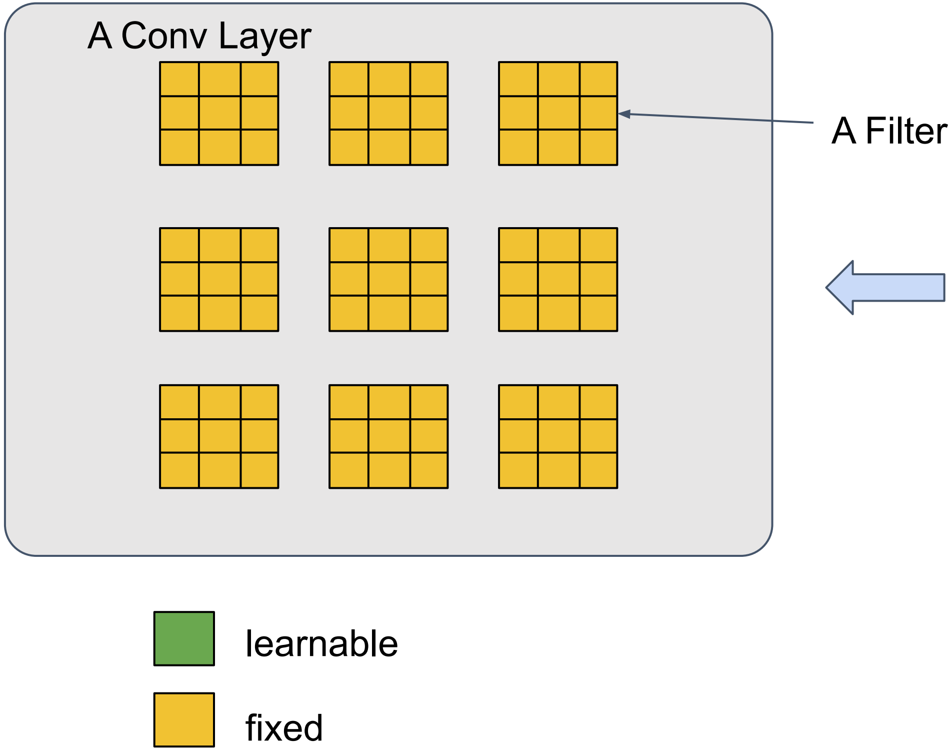
- *Hard Task Meta Batch*

[1] Shrivastava et al. "Training region-based object detectors with online hard example mining." CVPR 2016.

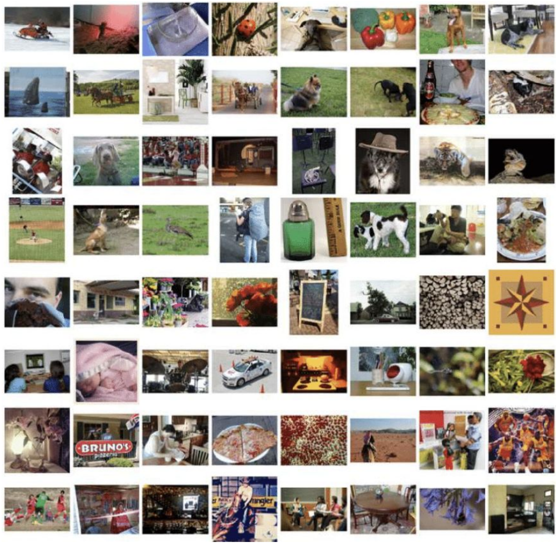
Convolution Networks in *MAML*



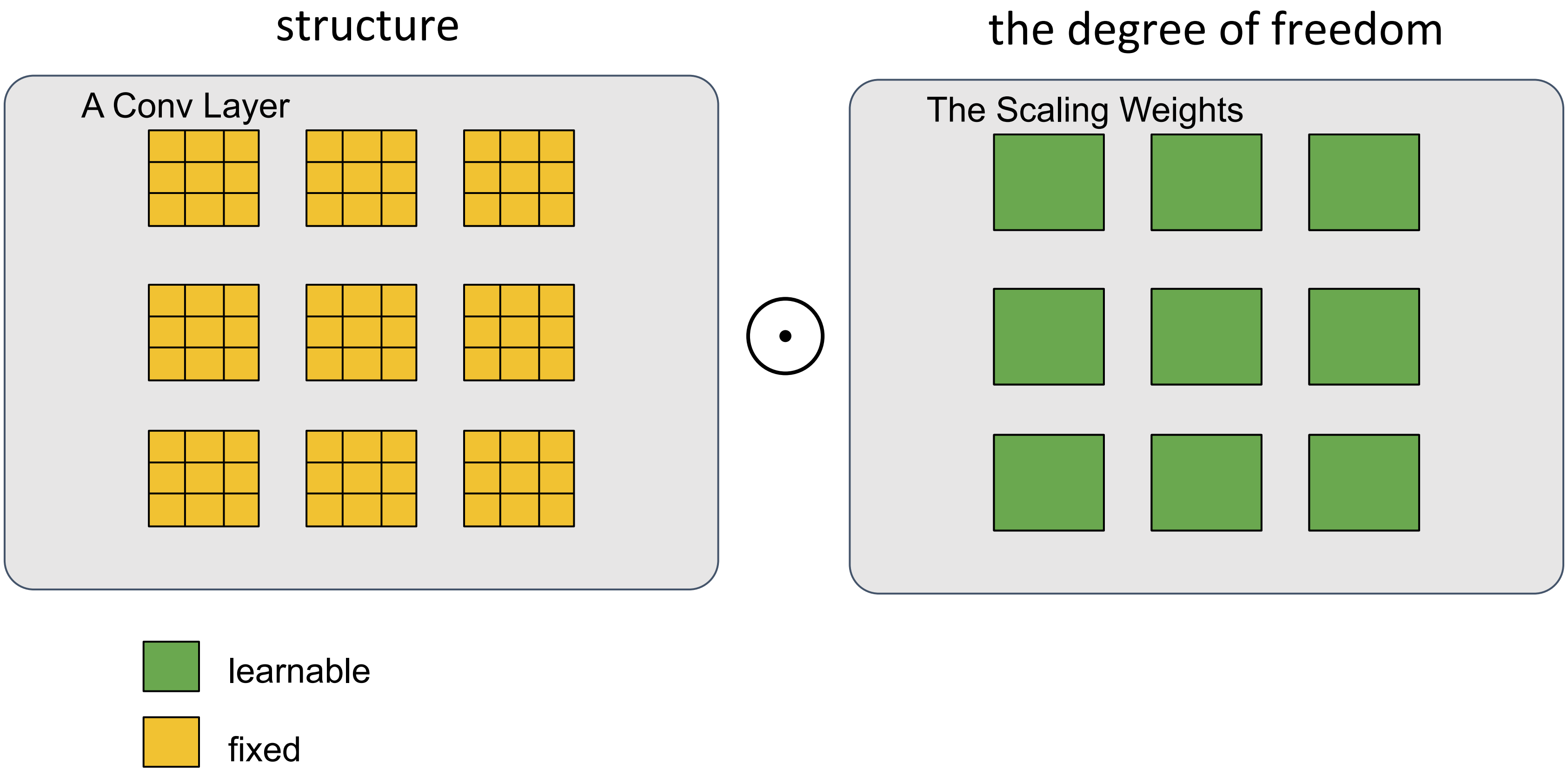
Learn the Structure by Many-shot Classification



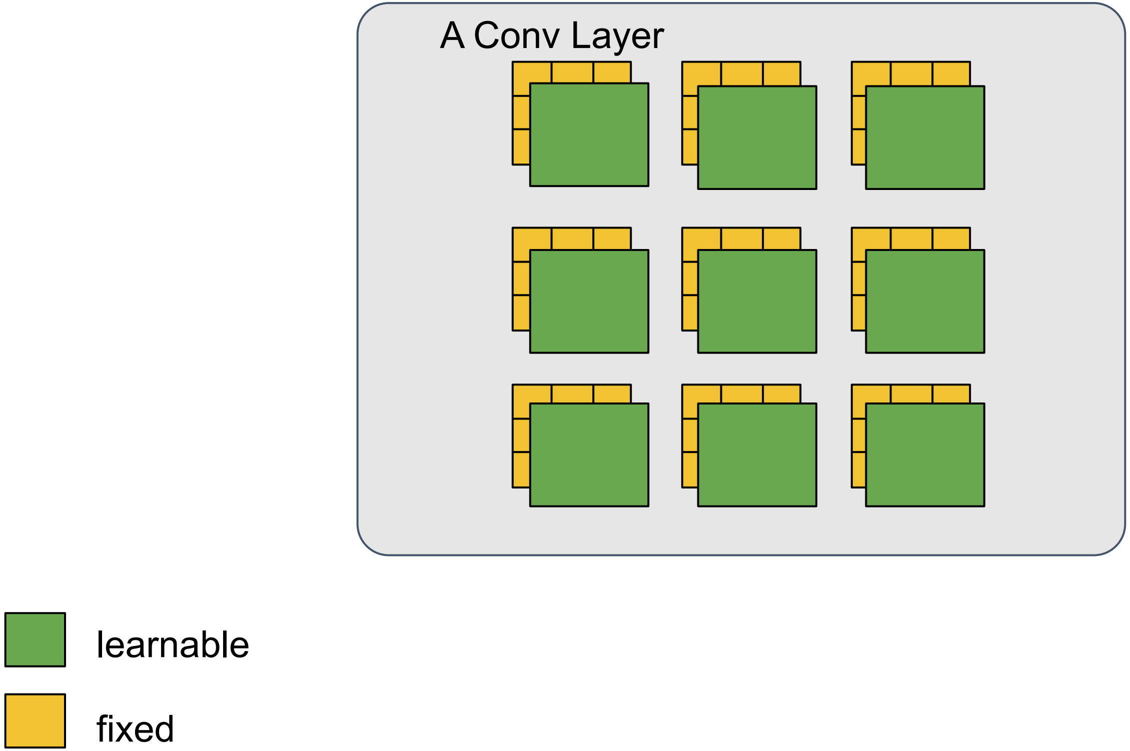
Pre-trained the network with** **many-shot classification task



Meta-transfer Learning



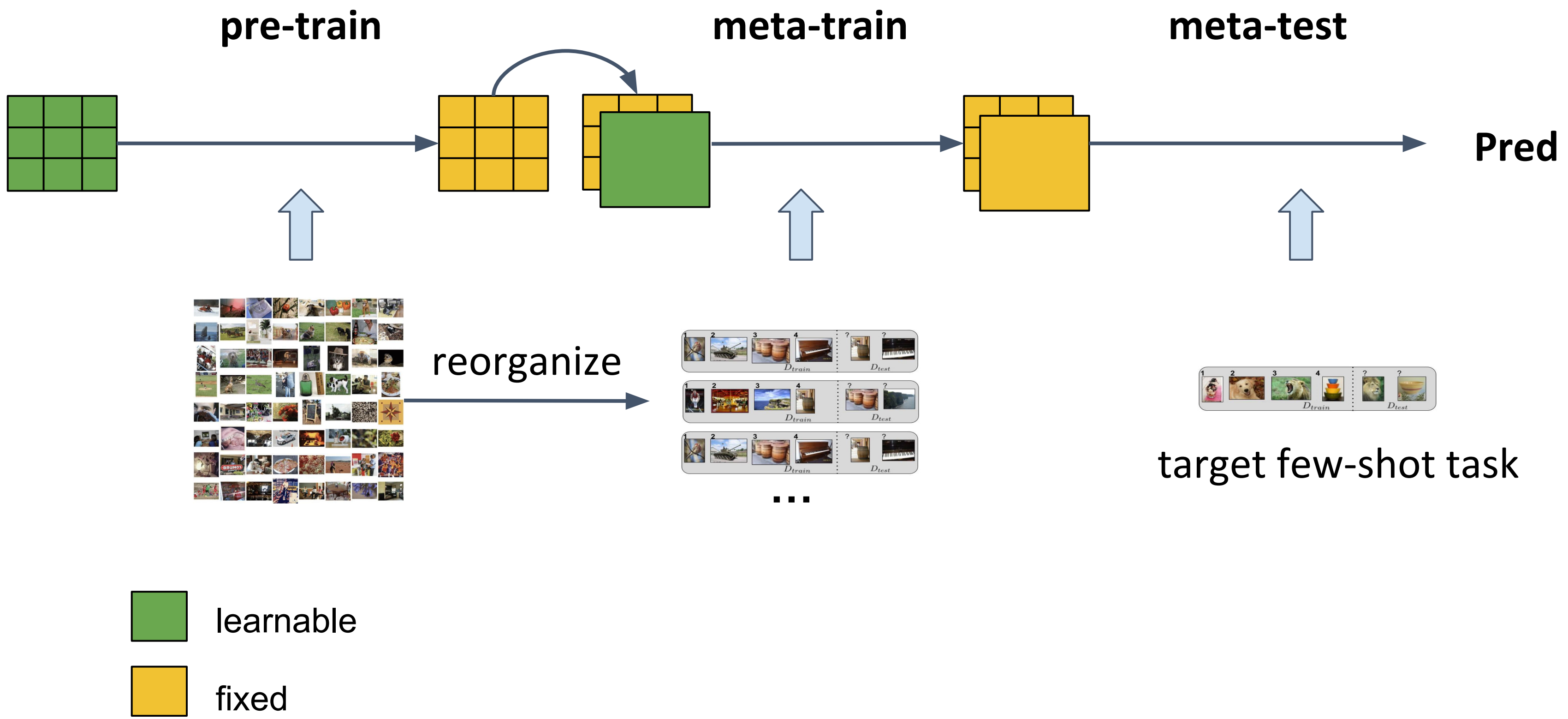
Meta-transfer Learning



Applying the scaling weights for each filter

Parameter number is reduced to approximately 1/9

The Pipeline

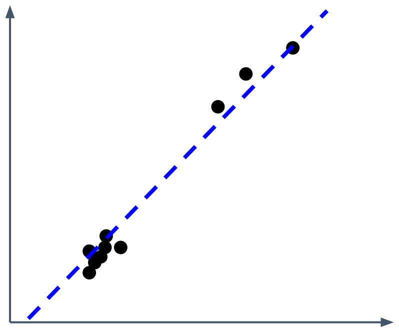


Overview of the Methods

- *Meta-transfer Learning*

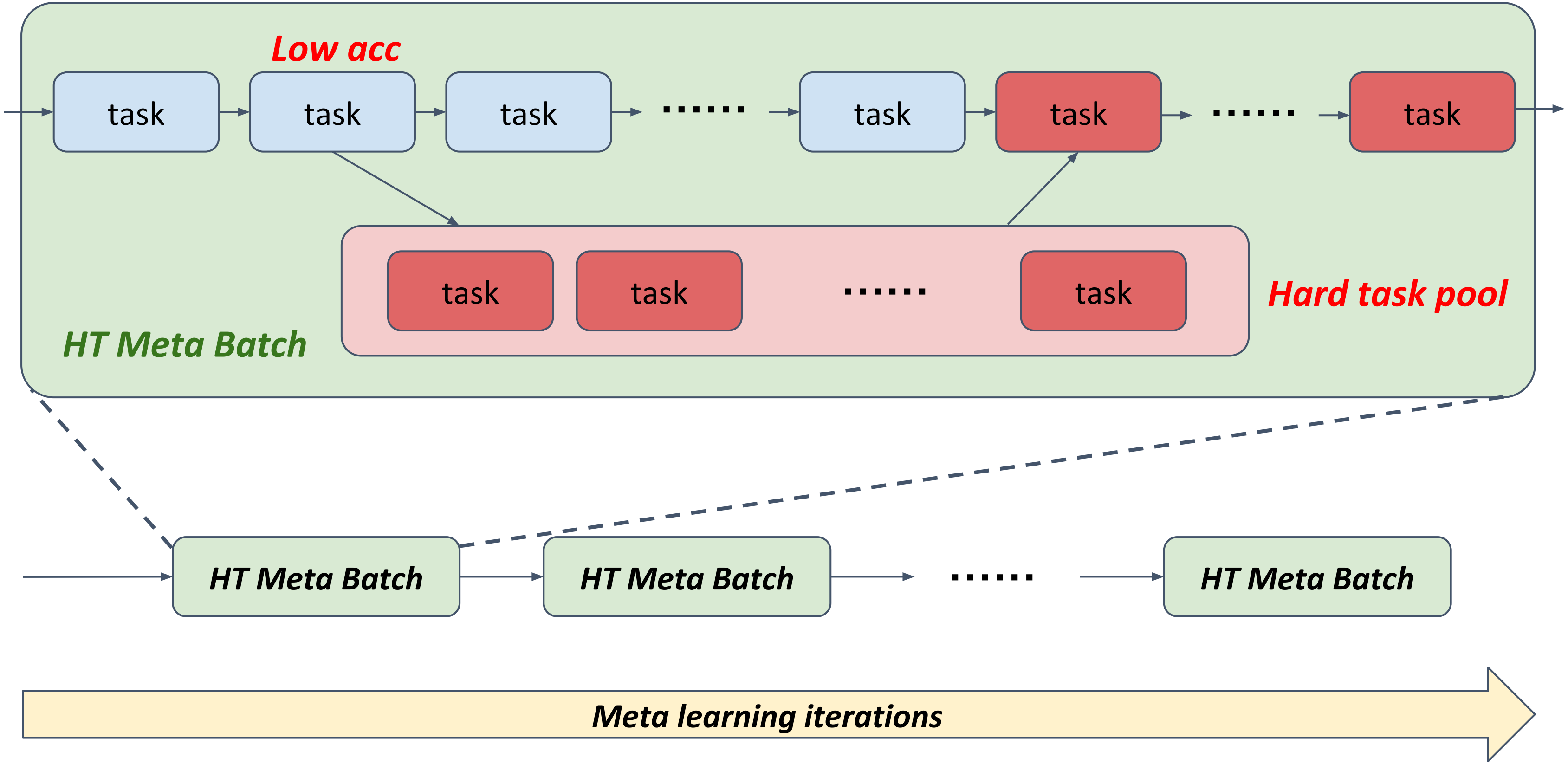
- ***Hard Task Meta Batch***

  The idea is from hard example mining[1]
  Hard example -> hard task



[1] Shrivastava et al. "Training region-based object detectors with online hard example mining." CVPR 2016.

Hard Task Meta Batch



OUTLINE

- Research Background
- Method
  - Meta-transfer Learning
  - Hard-task Meta Batch
- Experiments and Conclusions

Datasets

- miniImageNet
  - Reorganized from ImageNet
  - Vinyals et al.[1] first devised the dataset, and it is widely used in evaluating few-shot learning methods
  - 100 classes (64 meta-train, 16 meta-val, 20 meta-test)

- Fewshot-CIFAR100 (FC100)
  - Reorganized from CIFAR100
  - Splitted by Oreshkin et al.[2]
  - 100 classes (60 meta-train, 20 meta-val 20 meta-test)
  - 20 super-classes (12 meta-train, 4 meta-val 4 meta-test)

[1] Vinyals et al. "Matching networks for one shot learning." NIPS 2016;
[2] Oreshkin et al. "TADAM: Task dependent adaptive metric for improved few-shot learning." NIPS 2018.

Evaluation

- Image Classification Accuracy
  - 600 testing tasks randomly sampled from the meta-test set
  - 5-class
  - 1-shot and 5-shot on *mini*ImageNet
  - 1-shot, 5-shot and 10-shot on FC100

* The same evaluation protocol with MAML[1]

[1] Finn et al. "Model-agnostic meta-learning for fast adaptation of deep networks." ICML 2017.

Image Classification Accuracy

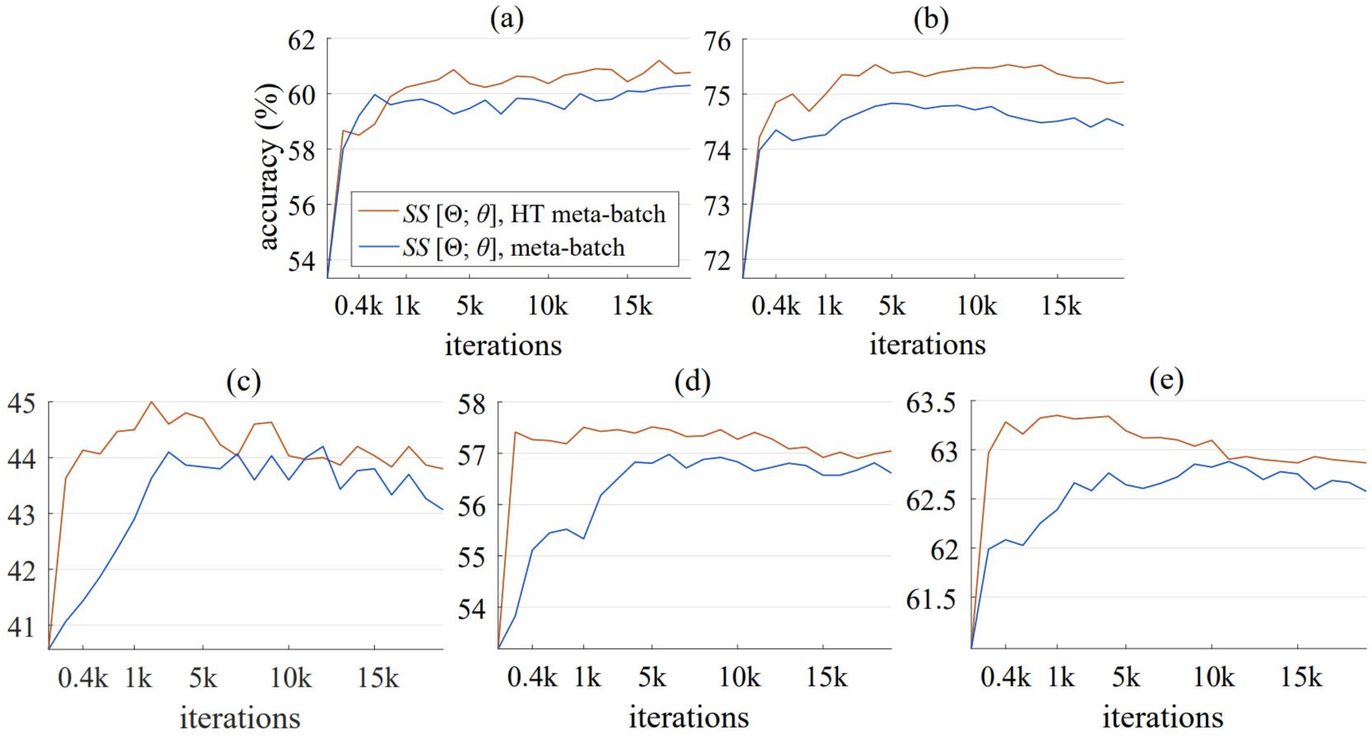
Methods	*mini*ImageNet (5-class)		FC100 (5-class)		
	1-shot	5-shot	1-shot	5-shot	10-shot
MatchingNets [1]	43.4 ± 0.8 %	55.3 ± 0.7 %			
Meta-LSTM [2]	43.6 ± 0.8 %	60.6 ± 0.7 %			
MAML [3]	48.7 ± 1.8 %	63.1 ± 0.9 %			
ProtoNets [4]	49.4 ± 0.8 %	68.2 ± 0.7 %			
TADAM [5]	58.5 ± 0.3 %	**76.7 ± 0.3 %**	40.1 ± 0.4 %	56.1 ± 0.4 %	61.6 ± 0.5 %
Ours (MTL + HT)	**61.2 ± 1.8 %**	75.5 ± 0.8 %	**45.8 ± 1.9 %**	**57.0 ± 1.0 %**	**63.4 ± 0.8 %**

[1] Vinyals et al. "Matching networks for one shot learning." NIPS 2016;
[2] Sachin et al. "Optimization as a model for few-shot learning." ICLR 2017;
[3] Chelsea et al. "Model-agnostic meta-learning for fast adaptation of deep networks." ICML 2017;
[4] Snell et al. "Prototypical networks for few-shot learning." NIPS 2017;
[5] Oreshkin et al. "TADAM: Task dependent adaptive metric for improved few-shot learning." NIPS 2018.

Ablation Study

Method	*mini*ImageNet (5-class)		FC100 (5-class)		
	1-shot	5-shot	1-shot	5-shot	10-shot
Train from scratch	45.3	64.6	38.4	52.6	58.6
Finetune on pre-train model	55.9	71.4	41.6	54.9	61.6
Ours (MTL)	60.2	74.3	43.6	55.4	62.4
Ours (MTL + HT)	**61.2**	**75.5**	**45.1**	**57.6**	**63.4**

Validation Accuracy



(a) (b) *mini*Imagenet, 1-shot and 5-shot   (c) (d) (e) FC100, 1-shot, 5-shot, and 10-shot

Conclusions

- A novel ***MTL*** method that learns to transfer large-scale pre-trained DNN weights for solving few-shot learning tasks.

- A novel ***HT meta-batch*** learning strategy that forces meta-transfer to "grow faster and stronger through hardship".

- Extensive experiments on ***miniImageNet*** and ***FC100***, and achieving the state-of-the-art performance.

Paper and Code

This work:

Meta-transfer Learning for Few-shot Learning. In *CVPR 2019.*

arXiv preprint: https://arxiv.org/pdf/1812.02391.pdf

GitHub repo: https://github.com/y2l/meta-transfer-learning-tensorflow

Thank you!
Any questions?

Email: yaoyao.liu@u.nus.edu