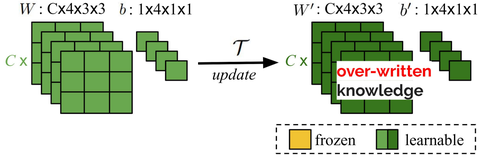
Meta-Transfer Learning for Few-Shot Learning

Qianru Sun[1,3*] Yaoyao Liu[1,2*] Tat-Seng Chua[1] Bernt Schiele[3]

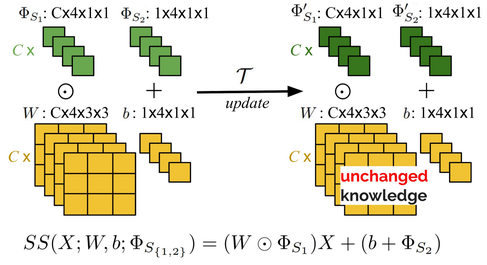
[1]National University of Singapore, [2]Tianjin University, [3]Max Planck Institute for Informatics

Motivation & Contributions

- Few-shot learning is challenging due to the lack of training data.
- Re-thinking promising methods:
  - deep neural networks (DNN)
  - transfer learning (pre-train, fine-tune)
  - meta-learning (meta gradient descent)
- Problem: "catastrophic forgetting"



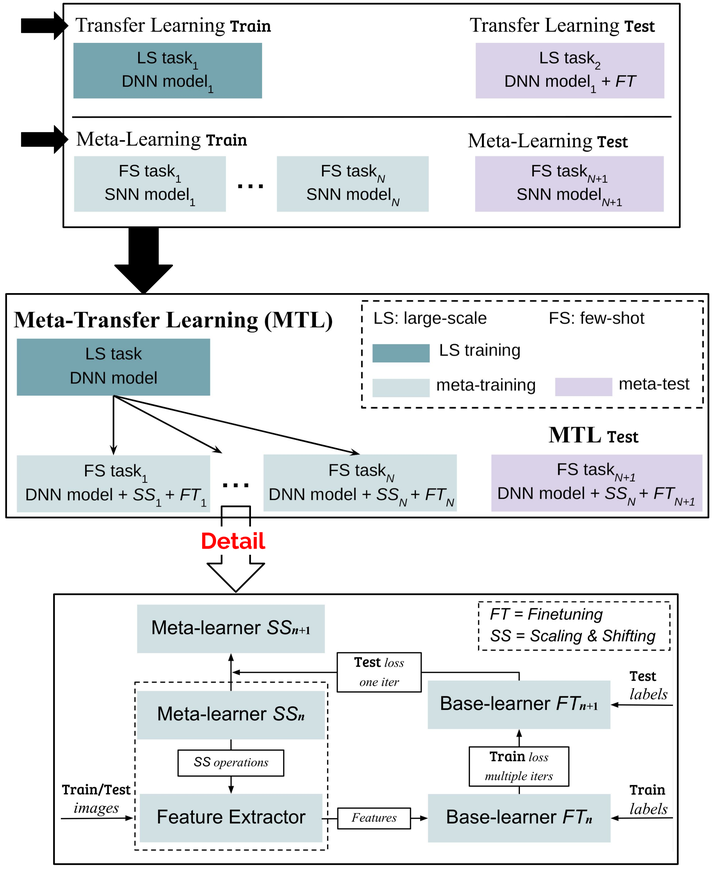
- Our solution: *Scaling & Shifting* (*SS*)



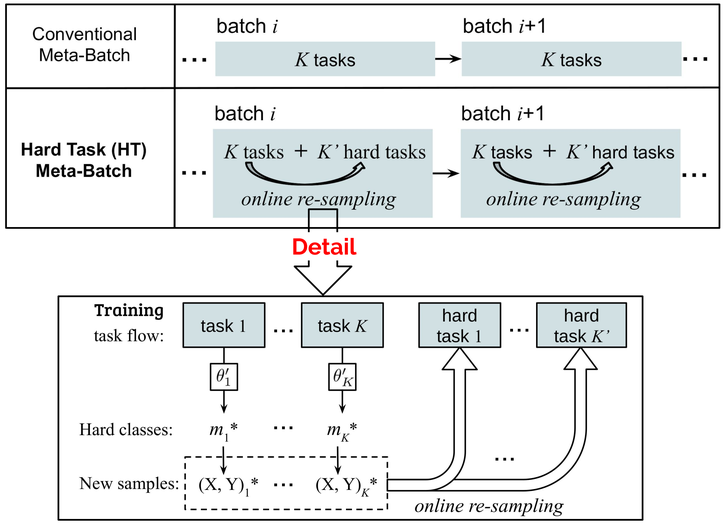
$$SS(X; W, b; \Phi_{S_{\{1,2\}}}) = (W \odot \Phi_{S_1})X + (b + \Phi_{S_2})$$

- Problem: slow meta-training convergence
- Our solution: hard negative task sampling

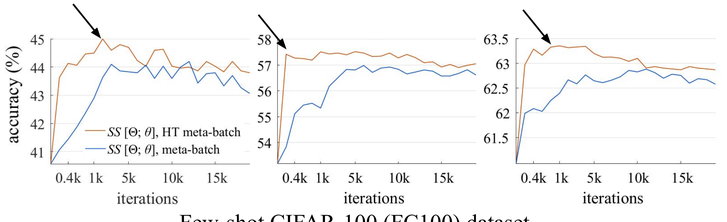
Meta-Transfer Learning



Hard Task Meta-Batch



Faster convergence is achieved!



Few-shot CIFAR-100 (FC100) dataset

Top performance is achieved!

- miniImageNet dataset

method	backbone	1-shot	5-shot
Adv. ResNet, [29]	WRN-40 (pre)	55.2	69.6
Delta-encoder, [44]	VGG-16 (pre)	58.7	73.6
MatchingNets, [53]	4 CONV	43.44 ± 0.77	55.31 ± 0.73
ProtoNets, [48]	4 CONV	49.42 ± 0.78	68.20 ± 0.66
RelationNets, [51]	4 CONV	50.44 ± 0.82	65.32 ± 0.70
MetaNetworks, [31]	5 CONV	49.21 ± 0.96	–
SNAIL, [30]	ResNet-12 (pre)◇	55.71 ± 0.99	68.88 ± 0.92
TADAM, [34]	ResNet-12 (pre)†	58.5 ± 0.3	**76.7 ± 0.3**
MAML, [9]	4 CONV	48.70 ± 1.75	63.11 ± 0.92
Meta-LSTM, [39]	4 CONV	43.56 ± 0.84	60.60 ± 0.71
Hier.Bayes, [13]	4 CONV	49.40 ± 1.830	–
BilevelProgram, [11]	ResNet-12◇	50.54 ± 0.85	64.53 ± 0.68
MetaGAN, [60]	ResNet-12	52.71 ± 0.64	68.63 ± 0.67
adaResNet, [32]	ResNet-12‡	56.88 ± 0.62	71.94 ± 0.57
FT [Θ; θ], HT meta-batch	4 CONV	49.1 ± 1.9	64.1 ± 0.9
FT [Θ; θ], HT meta-batch	ResNet-12 (pre)	59.1 ± 1.9	73.1 ± 0.9
SS [Θ; θ], meta-batch	ResNet-12 (pre)	60.2 ± 1.8	74.3 ± 0.9
SS [Θ; θ], HT meta-batch	ResNet-12 (pre)	**61.2 ± 1.8**	**75.5 ± 0.8**

◇Additional 2 convolutional layers ‡Additional 1 convolutional layer

- Few-shot CIFAR-100 (FC100) dataset

method	backbone	1-shot	5-shot
MAML, [9]‡	4 CONV	38.1 ± 1.7	50.4 ± 1.0
TADAM, [34]	ResNet-12 (pre)†	40.1 ± 0.4	56.1 ± 0.4
FT [Θ; θ], HT meta-batch	4 CONV	39.9 ± 1.8	51.7 ± 0.9
FT [Θ; θ], HT meta-batch	ResNet-12 (pre)	41.8 ± 1.9	55.1 ± 0.9
SS [Θ; θ], meta-batch	ResNet-12 (pre)	43.6 ± 1.8	55.4 ± 0.9
SS [Θ; θ], HT meta-batch	ResNet-12 (pre)	**45.1 ± 1.8**	**57.6 ± 0.9**

†Additional 72 fully connected layers
‡Our implementation using the public code of MAML

★ **Code is available at:** https://github.com/y2l/meta-transfer-learning-tensorflow