Learning to Learn via Self-Critique

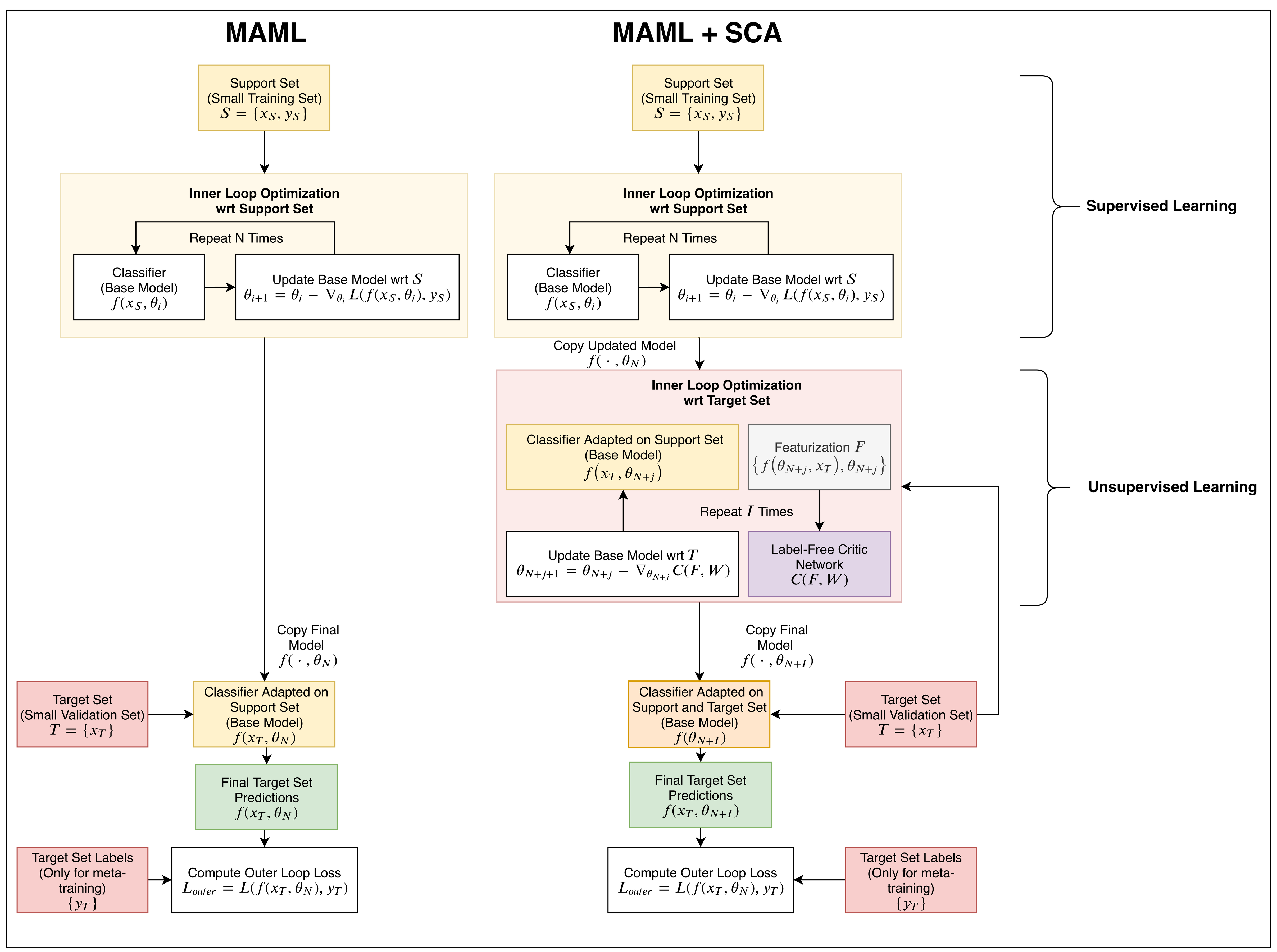
Antreas Antoniou[1] and Amos Storkey[1]

[1]University of Edinburgh

Overview

1. One of the most succesful methods for tackling few-shot learning is via meta-learning.
2. A few-shot learning task is composed of a small training set (i.e. a support set), and a small validation set (i.e. a validation set).
3. All of current few-shot learning methods, have, to-date, been using information only from the support set, to learn about the task at hand.
4. However, the target-set, which also contains useful task-related information, contains no label information. Thus, it's impossible to use discriminative loss functions to learn from it.
5. Instead, we propose, learning an unsupervised loss function, parameterized as a neural network, which can inspect the base-model's target-set features, and compute a loss value, which can be used with gradient descent to update our base-model wrt target-set information.
6. We refer to the resulting method as Self-Critique and Adapt (SCA), which can be added to any existing meta-learning few-shot model to allow extraction of information from a target-set.

Method



Results

Model	Test Accuracy			
	Mini-Imagenet		CUB	
	1-shot	5-shot	1-shot	5-shot
MAML++ (Low-End)	$52.15 \pm 0.26\%$	$68.32 \pm 0.44\%$	-	-
MAML++ (Low-End) with (preds)	$52.52 \pm 1.13\%$	$70.84 \pm 0.34\%$	-	-
MAML++ (Low-End) with (preds, params)	$52.68 \pm 0.93\%$	$69.83 \pm 1.18\%$	-	-
MAML++ (Low-End) with (preds, task-embedding)	$\mathbf{54.84 \pm 1.24}\%$	$70.95 \pm 0.17\%$	-	-
MAML++ (Low-End) with (preds, task-embedding, params)	$54.24 \pm 0.99\%$	$\mathbf{71.85 \pm 0.53}\%$	-	-
MAML++ (High-End)	$58.37 \pm 0.27\%$	$75.50 \pm 0.19\%$	$67.48 \pm 1.44\%$	$83.80 \pm 0.35\%$
MAML++ (High-End) with (preds)	$\mathbf{62.86 \pm 0.70}\%$	$77.07 \pm 0.19\%$	$70.33 \pm 0.78\%$	$85.47 \pm 0.40\%$
MAML++ (High-End) with (preds, task-embedding)	$62.29 \pm 0.38\%$	$\mathbf{77.64 \pm 0.40}\%$	$\mathbf{70.46 \pm 1.18}\%$	$\mathbf{85.63 \pm 0.66}\%$

Table: Ablation Studies on Mini-ImageNet and CUB

Model	Test Accuracy			
	Mini-ImageNet		CUB	
	1-shot	5-shot	1-shot	5-shot
Matching networks	$43.56 \pm 0.84\%$	$55.31 \pm 0.73\%$	$61.16 \pm 0.89\%$	$72.86 \pm 0.70\%$
Meta-learner LSTM	$43.44 \pm 0.77\%$	$60.60 \pm 0.71\%$	-	-
MAML	$48.70 \pm 1.84\%$	$63.11 \pm 0.92\%$	$55.92 \pm 0.95\%$	$72.09 \pm 0.76\%$
LLAMA	$49.40 \pm 1.83\%$	-	-	-
REPTILE	$49.97 \pm 0.32\%$	$65.99 \pm 0.58\%$	-	-
PLATIPUS	$50.13 \pm 1.86\%$	-	-	-
Meta-SGD (our features)	$54.24 \pm 0.03\%$	$70.86 \pm 0.04\%$	-	-
SNAIL	$55.71 \pm 0.99\%$	$68.88 \pm 0.92\%$	-	-
gidaris2018dynamic	$56.20 \pm 0.86\%$	$73.00 \pm 0.64\%$	-	-
munkhdalai2017meta	$57.10 \pm 0.70\%$	$70.04 \pm 0.63\%$	-	-
TADAM	$58.50 \pm 0.30\%$	$76.70 \pm 0.30\%$	-	-
qiao2018few	$59.60 \pm 0.41\%$	$73.74 \pm 0.19\%$	-	-
LEO	$61.76 \pm 0.08\%$	$77.59 \pm 0.12\%$	-	-
Baseline	-	-	$47.12 \pm 0.74\%$	$64.16 \pm 0.71\%$
Baseline ++	-	-	$60.53 \pm 0.83\%$	$79.34 \pm 0.61\%$
MAML (Local Replication)	$48.25 \pm 0.62\%$	$64.39 \pm 0.31\%$	-	-
MAML++ (Low-End)	$52.15 \pm 0.26\%$	$68.32 \pm 0.44\%$	-	-
MAML++ (Low-End) +	$54.84 \pm 0.99\%$	$71.85 \pm 0.53\%$	-	-
MAML++ (High-End)	$58.37 \pm 0.27\%$	$75.50 \pm 0.19\%$	$67.48 \pm 1.44\%$	$83.80 \pm 0.35\%$
MAML++ (High-End) +	$\mathbf{62.86 \pm 0.79}\%$	$\mathbf{77.64 \pm 0.40}\%$	$\mathbf{70.46 \pm 1.18}\%$	$\mathbf{85.63 \pm 0.66}\%$

Table: Comparative Results on Mini-ImageNet and CUB

Conclusion

We have demonstrated that learning a neural network-based loss function to extract information from an unsupervised dataset, can be very useful in improving performance on a given few-shot learning task. This fact, showcases that fully learnable losses can, in fact, help improve both the effectiveness and efficiency of existing machine learning models.