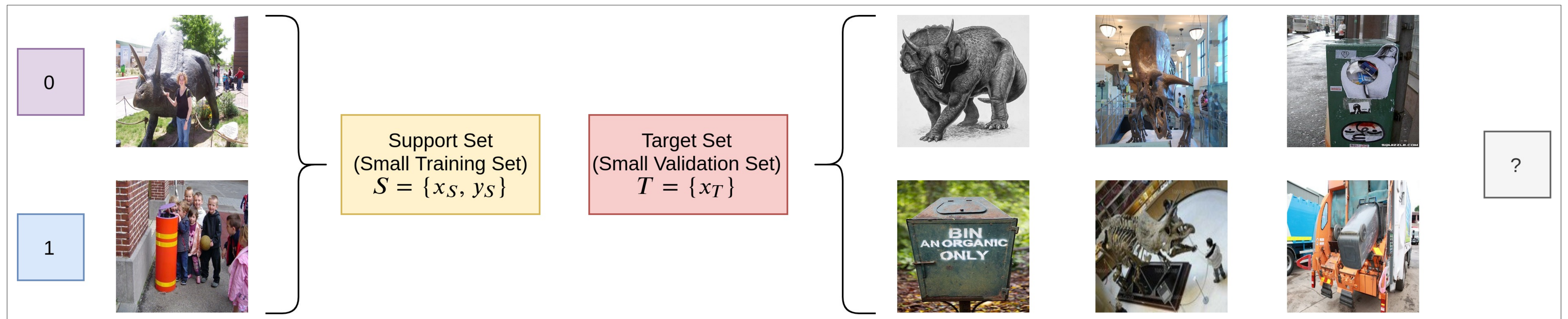
Learning to Learn via Self-Critique

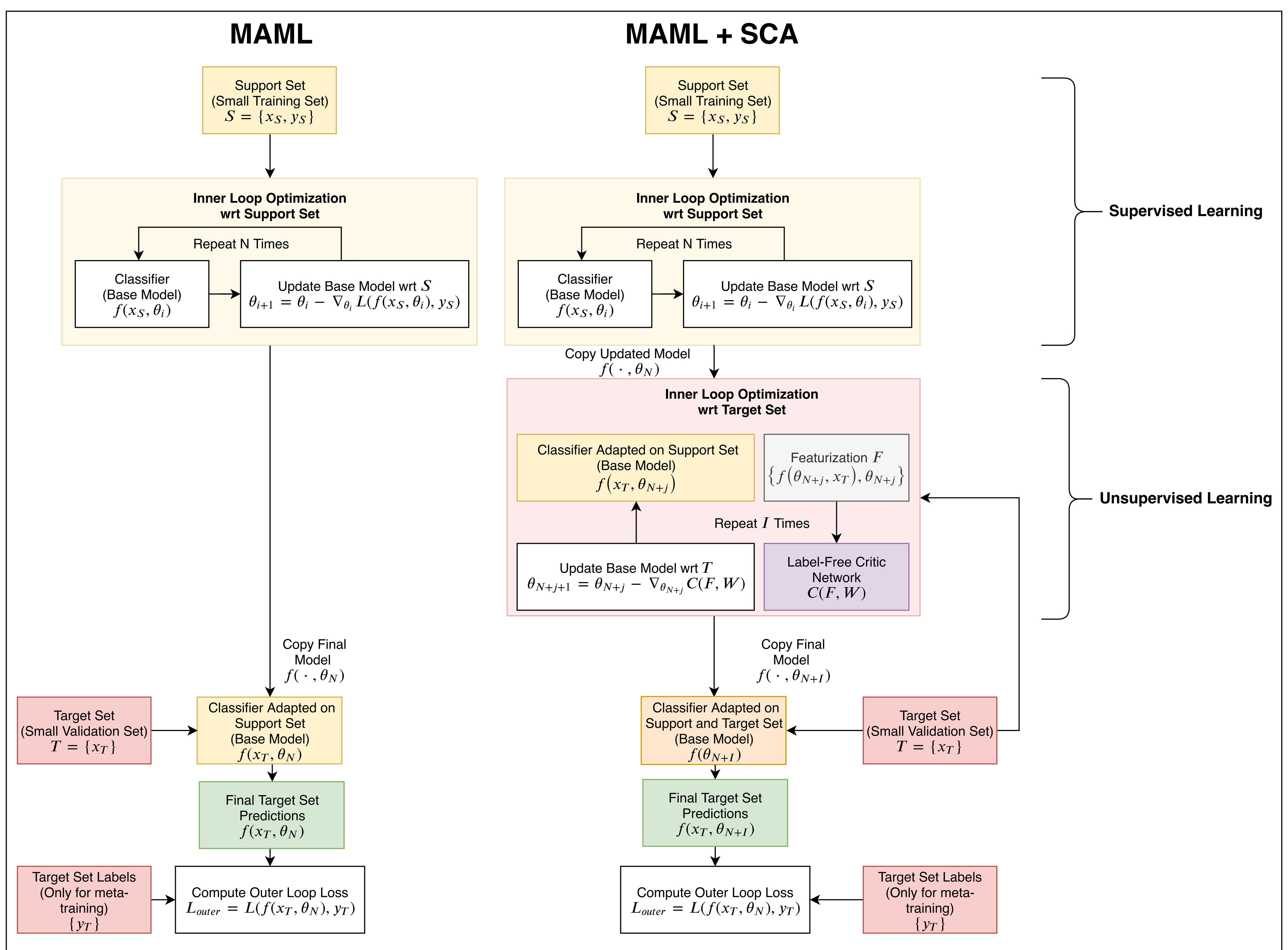
Antreas Antoniou[1] and Amos Storkey[1]

[1]University of Edinburgh

- **Observation:** Existing SOTA few-shot learning systems learn only from the labelled support set without leveraging the unlabelled validation (target) sets.
- **Question:** How can we extract information from unlabelled target sets, to enhance few-shot learning systems?
- **Problem:** No supervised labels means that we can't use discriminative training to learn.
- **Solution:** Meta-learn an unsupervised loss function that can extract such information, targeted towards performing better on a task, called ***Self-Critique and Adapt (SCA)***.
- **Demonstration:** State-of-the-art, currently best-in-class few-shot learning results.



Unsupervised information in the target set can be used to disambiguate the semantics of the task classes.



Model	Test Accuracy			
	Mini-Imagenet		CUB	
	1-shot	5-shot	1-shot	5-shot
MAML++ (Low-End)	$52.15 \pm 0.26\%$	$68.32 \pm 0.44\%$	$62.19 \pm 0.53\%$	$76.08 \pm 0.51\%$
MAML++ (Low-End) with (preds)	$52.52 \pm 1.13\%$	$70.84 \pm 0.34\%$	$66.13 \pm 0.97\%$	$77.62 \pm 0.77\%$
MAML++ (Low-End) with (preds, params)	$52.68 \pm 0.93\%$	$69.83 \pm 1.18\%$	-	-
MAML++ (Low-End) with (preds, task-embedding)	$\mathbf{54.84 \pm 1.24}\%$	$70.95 \pm 0.17\%$	$65.56 \pm 0.48\%$	$77.69 \pm 0.47\%$
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 the conditioning information of the critique network. The combination of the task-embedding and the predictions appear to produce the best results.

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	$48.70 \pm 1.84\%$	$63.11 \pm 0.92\%$	$55.92 \pm 0.95\%$	$72.09 \pm 0.76\%$
SNAIL	$55.71 \pm 0.99\%$	$68.88 \pm 0.92\%$	-	-
Qiao et al 2018	$59.60 \pm 0.41\%$	$73.74 \pm 0.19\%$	-	-
Baseline	-	-	$47.12 \pm 0.74\%$	$64.16 \pm 0.71\%$
Baseline ++	-	-	$60.53 \pm 0.83\%$	$79.34 \pm 0.61\%$
Latent Embedding Optimization	$61.76 \pm 0.08\%$	$77.59 \pm 0.12\%$	-	-
MAML (Local Replication)	$48.25 \pm 0.62\%$	$64.39 \pm 0.31\%$	-	-
MAML++ (Low-End - Original)	$52.15 \pm 0.26\%$	$68.32 \pm 0.44\%$	$62.19 \pm 0.53\%$	$76.08 \pm 0.51\%$
MAML++ (Low-End - Original) + SCA (Ours)	$54.84 \pm 0.99\%$	$71.85 \pm 0.53\%$	$66.13 \pm 0.97\%$	$77.62 \pm 0.77\%$
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) + SCA (Ours)	$\mathbf{62.86 \pm 0.79}\%$	$\mathbf{77.64 \pm 0.40}\%$	$\mathbf{70.46 \pm 1.18}\%$	$\mathbf{85.63 \pm 0.66}\%$

Table: Test accuracy comparison with legacy and other SOTA methods. Our methods produce the top performance across the board.

[1] Jürgen Schmidhuber.
Learning to control fast-weight memories: An alternative to dynamic recurrent networks.
Neural Computation, 4(1):131–139, 1992.

[2] Yoshua Bengio, Samy Bengio, and Jocelyn Cloutier.
Learning a synaptic learning rule.
Université de Montréal, Département d'informatique et de recherche opérationnelle, 1990.

[3] Geoffrey E Hinton and David C Plaut.
Using fast weights to deblur old memories.
1987.

[4] Chelsea Finn, Pieter Abbeel, and Sergey Levine.
Model-agnostic meta-learning for fast adaptation of deep networks.
In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1126–1135. JMLR. org, 2017.

[5] Oriol Vinyals, Charles Blundell, Tim Lillicrap, Daan Wierstra, et al.
Matching networks for one shot learning.
In *Advances in Neural Information Processing Systems*, pages 3630–3638, 2016.

[6] Sachin Ravi and Hugo Larochelle.
Optimization as a model for few-shot learning.
Proceedings of the 4th International Conference on Learning Representations, 2016.

[7] Siyuan Qiao, Chenxi Liu, Wei Shen, and Alan L Yuille.
Few-shot image recognition by predicting parameters from activations.
In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7229–7238, 2018.

[8] Yanbin Liu, Juho Lee, Minseop Park, Saehoon Kim, and Yi Yang.
Transductive propagation network for few-shot learning.
arXiv preprint arXiv:1805.10002, 2018.

[9] Andrei A Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, and Raia Hadsell.
Meta-learning with latent embedding optimization.
arXiv preprint arXiv:1807.05960, 2018.