Task-Specific Gradient Adaptation for Few-Shot One-Class Classification

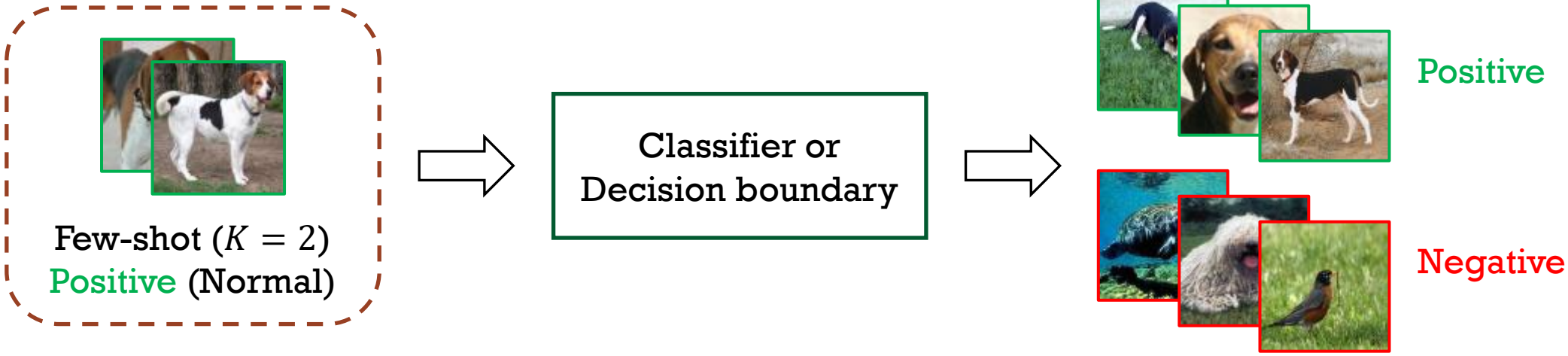
Yunlong Li, Xiabi Liu, Liyuan Pan, Yuchen Ren
Beijing Institute of Technology




Email: {yunlong.li, liuxiabi, liyuan.pan, yuchen.ren}@bit.edu.cn

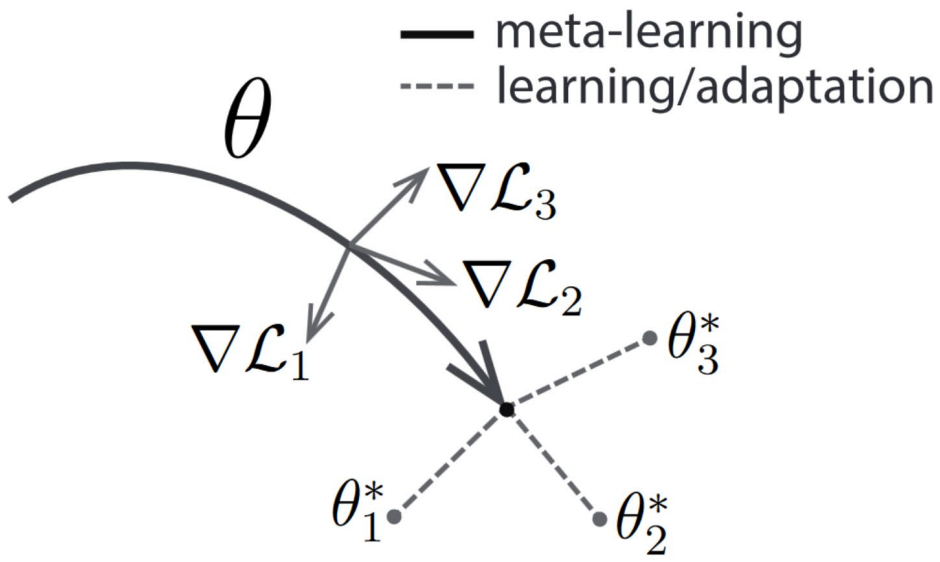
Few-Shot One-Class Classification (FS-OCC)

For each task:



Classify new query samples as either positive or negative when only few-shot positive samples are available during training.

One-Class Model-Agnostic Meta-Learning (OC-MAML)



One-Class Support Set

Cross Entropy Loss

$$\boldsymbol{\theta}_i' = \boldsymbol{\theta} - \alpha\nabla_{\boldsymbol{\theta}}\ell(\mathcal{S};\boldsymbol{\theta}),$$

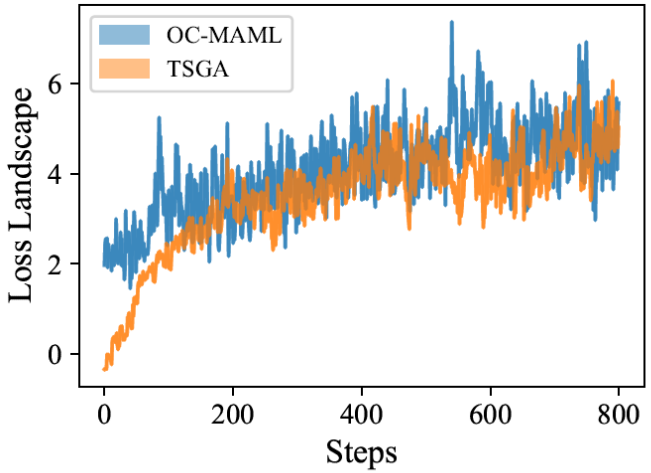
$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \eta\nabla_{\boldsymbol{\theta}}\sum_{\mathcal{T}_i}\ell(\mathcal{Q};\boldsymbol{\theta}_i')$$

Class-Balanced Query Set

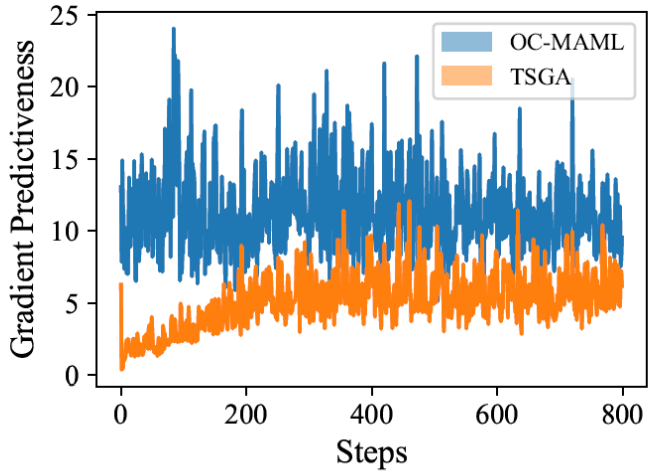
Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks[C]//International conference on machine learning. PMLR, 2017: 1126-1135.
Frikha A, Krompaß D, Köpken H G, et al. Few-shot one-class classification via meta-learning[C]//Proceedings of the AAAI conference on artificial intelligence. 2021, 35(8): 7448-7456

Misalignment

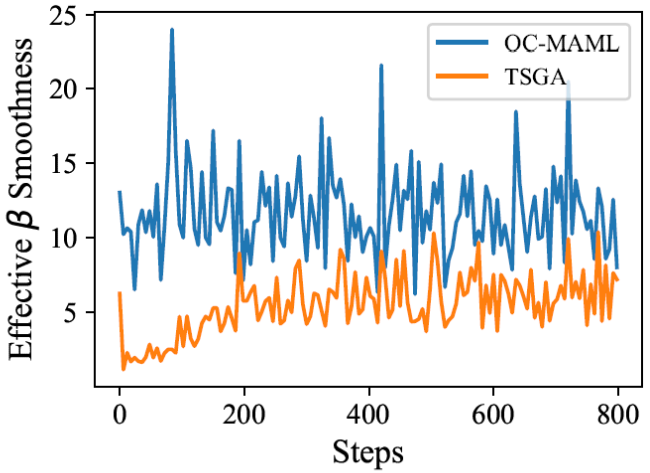
between Cross Entropy Loss and OCC Tasks during adaptation



(a) Loss variation



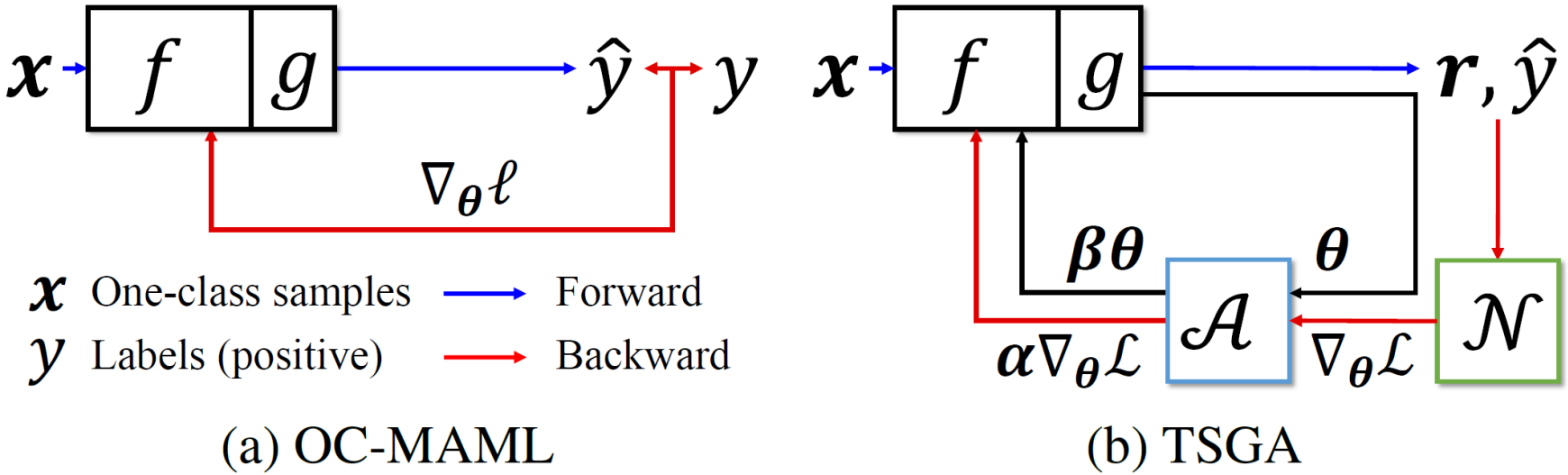
(b) Changes in gradients



(c) Maximum difference in gradients

Task-Agnostic Adaptation

disregards task-specific meta-knowledge



(a) OC-MAML (b) TSGA

Task-Specific Gradient Adaptation (TSGA)

$$\boldsymbol{\theta}_i' = \boldsymbol{\theta} - \alpha \nabla_{\boldsymbol{\theta}} \ell(\mathcal{S}; \boldsymbol{\theta}),$$

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \eta \nabla_{\boldsymbol{\theta}} \sum_{\mathcal{T}_i} \ell(\mathcal{Q}; \boldsymbol{\theta}_i')$$

Inner loop

Outer loop

$$\ell(\mathcal{Q}; \boldsymbol{\theta}_i') = \mathbb{E}_\ell \left[\sum_{t=1}^{T} w_t \ell(\boldsymbol{\theta}_i^t) \right],$$

$$\ell(\boldsymbol{\theta}_i^t) = \frac{1}{M} \sum_{(\boldsymbol{x}, y \in \mathcal{Q})} \ell(\hat{y}, y; \boldsymbol{\theta}_i^t), t \in [1, T]$$

$$(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\varphi}) \leftarrow (\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\varphi}) - \eta \nabla_{(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\varphi})} \sum_{\mathcal{T}_i} \ell(\mathcal{Q}; \boldsymbol{\theta}_i')$$

$$\mathcal{L}(\boldsymbol{\theta}, \mathcal{S}; \boldsymbol{\phi}) = \frac{1}{K} \sum_{(\boldsymbol{x}, y \in \mathcal{S})} \mathcal{N}(\hat{y}, \boldsymbol{r}, y; \boldsymbol{\phi})$$

$$\boldsymbol{\theta}_i' = \boldsymbol{\beta}\boldsymbol{\theta} - \boldsymbol{\alpha} \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, \mathcal{S}; \boldsymbol{\phi})$$

$$= [\boldsymbol{\beta}, \boldsymbol{\alpha}] \begin{bmatrix} \boldsymbol{\theta} \\ -\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, \mathcal{S}; \boldsymbol{\phi}) \end{bmatrix}$$

$$[\boldsymbol{\beta}, \boldsymbol{\alpha}] = \mathcal{A}\left(\begin{bmatrix} \boldsymbol{\theta} \\ -\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, \mathcal{S}; \boldsymbol{\phi}) \end{bmatrix}; \boldsymbol{\varphi} \right)$$

Ablation Study

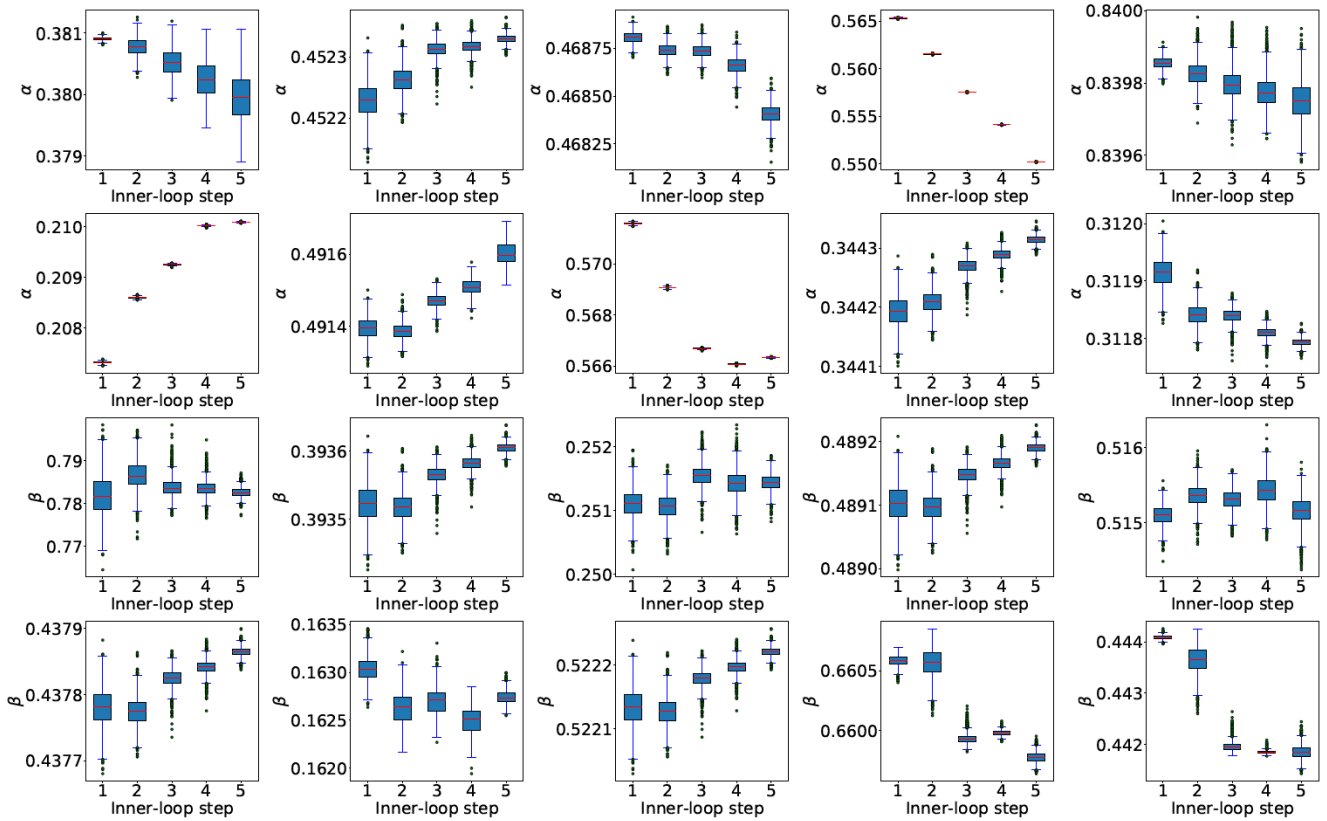
ID	Setting	$K = 2$			$K = 10$		
		MIN	OMN	CIFAR-FS	MIN	OMN	CIFAR-FS
$\mathbb{A}$	Baseline	69.1	96.6	70.0	76.2	97.6	79.1
$\mathbb{B}$	$+\mathcal{A}$	72.13	97.89	75.91	76.46	98.31	81.11
$\mathbb{C}$	$+\mathcal{N}$	72.09	97.99	75.59	76.66	98.45	81.13
$\mathbb{D}$	$+\mathcal{A}+\mathcal{N}$ (TSGA)	**73.41** (+4.31)	**98.11** (+1.51)	**76.84** (+6.84)	**77.73** (+1.53)	**98.54** (+0.94)	**81.70** (+2.60)

Results

Model	Backbone	$K = 2$			$K = 10$		
		MIN	OMN	CIFAR-FS	MIN	OMN	CIFAR-FS
OC-ProtoNet [3]	Conv-4	64.4	88.8	64.5	66.4	91.6	67.6
OW-ProtoNet [4]	Conv-4	67.0	–	70.9	74.4	–	76.7
AHDD [2]	Conv-4	71.93	96.81	75.34	77.00	97.87	80.95
OC-MetaOptNet [1]	ResNet-12	51.8	–	56.3	67.4	–	75.5
OC-MetaSGD [1]	Conv-4	69.6	–	71.4	75.8	–	77.8
OC-MAML [1]	Conv-4	69.1	96.6	70.0	76.2	97.6	79.1
TSGA (Ours)	Conv-4	**73.41** $\pm$ 0.4	**98.11** $\pm$ 0.1	**76.84** $\pm$ 0.4	**77.73** $\pm$ 0.4	**98.54** $\pm$ 0.1	**81.70** $\pm$ 0.4
OC-MAML [1]	ResNet-12	76.60 $\pm$ 0.4	98.61 $\pm$ 0.1	74.26 $\pm$ 0.4	77.27 $\pm$ 0.4	99.18 $\pm$ 0.1	82.05 $\pm$ 0.4
TSGA (Ours)	ResNet-12	**78.04** $\pm$ 0.4	**98.97** $\pm$ 0.1	**79.53** $\pm$ 0.4	**79.24** $\pm$ 0.4	**99.29** $\pm$ 0.1	**84.05** $\pm$ 0.4

[1] Frikha A, Krompaß D, Köpken H G, et al. Few-shot one-class classification via meta-learning[C]//Proceedings of the AAAI conference on artificial intelligence. 2021, 35(8): 7448-7456.
[2] Ren Y, Liu X, Pan L, et al. Adaptive Hypersphere Data Description for few-shot one-class classification[J]. Applied Intelligence, 2024, 54(24): 12885-12897.
[3] Dahia G, Pamplona Segundo M. Meta learning for few-shot one-class classification[J]. AI, 2021, 2(2): 195-208.
[4] Kruspe A. One-way prototypical networks[J]. arXiv preprint arXiv:1906.00820, 2019.

Visualization



Thank you!

Project Website:
https://github.com/noragami2333/TSGA