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

Supplementary Material

In this supplementary material, we provide the details of the experiment process, the FS-OCC episode sampling technique, network backbones in Section 1, more experiments of detailed comparisons on loss functions, sensitivity to learning rate, boosting other optimization-based meta-learning methods with TSGA, evaluation on class-imbalanced query sets and cross-domain experiments in Section 2. We also visualize the convergence curves and normal/anomalous samples of the real-world CNC Milling Machine dataset in Section 3.

1. Experimental Details

Experiment Settings. We outline the process for K-shot OCC experiments as follows. The dataset for FS-OCC is divided into three disjoint sets: meta-training set, meta-validation set, and meta-testing set. In the meta-training stage, we sample episodes of K-shot OCC tasks from the meta-training set. For each task, the model learns from support samples (exclusively positive) and classifies the query samples (positive and negative) in the inner loop. In the outer loop, labels for all query samples in an episode are provided, and the model is optimized using prediction and labels. During the meta-testing stage, we randomly sampled a large number of K-shot OCC tasks from the meta-testing set. The pre-trained model is fine-tuned on the support set of each task and evaluated on the query set. The final performance is reported as average accuracy across all tasks.

FS-OCC Task Generation. Following the data sampling technique proposed in OC-MAML [1], an FS-OCC task is sampled as follows: 1) a target class is randomly selected as the positive class, and K non-overlapping positive samples are randomly selected for the support set; 2) Q non-overlapping positive samples are randomly selected from the remaining positive class samples, while Q samples are randomly chosen from each of the other classes to form a class-balanced query set. In the meta-testing stage, each task comprises $Q = 30$ positive and negative samples, resulting in a total of 60 samples ($M = 2Q$). For the real-world CNC Milling Machines dataset, each task contains only $Q = 1$ positive and negative samples due to data limitations. Notably, the query set is kept class-balanced between positive and negative classes during the meta-training stage to capture the meta-knowledge for FS-OCC. In contrast, the query set may not be class-balanced during the meta-testing stage to accommodate diverse task requirements. In our experiments, we maintained a class-balanced query set during meta-testing for consistent evaluation. For more evaluation experiments on different class-imbalance rates please refer to Section 2.

Network Backbone. Conv-4 consists of 4 convolution layers followed by a fully connected layer and a softmax layer for classification. Each block includes a 3×3 convolution with 64 filters, followed by batch normalization, a ReLU activation function, and 2×2 max-pooling. For time-series data, we replace the 2D convolution layer in Conv-4 with the 1D convolution layer. **ResNet-12** [2] comprises 4 residual blocks, followed by a fully connected layer and a softmax layer for classification. Each residual block contains three convolution operations with a 3×3 filter size. Between the convolution operations, batch normalization and ReLU activation are applied. At the end of each residual block, a sequence comprising batch normalization, a skip connection, ReLU activation, and 2×2 max-pooling is implemented. The skip connection itself includes batch normalization and ReLU activation. The first residual block uses 64 filters for each convolution operation, with each successive residual block having double the number of filters from a preceding block.

Loss Function	MIN	OMN	CIFAR-FS
Cross-Entropy Loss	76.2	97.6	79.1
L1 Loss	75.56	97.72	78.38
L2 Loss	76.28	98.06	78.53
Loss Network $\mathcal{N}$	**77.73**	**98.54**	**81.70**

Table 1. Comparison of different loss functions used in the adaptation process for 10-shot OCC tasks on miniImageNet (MIN), Omniglot (OMN), and CIFAR-FS. We highlight the best number in **bold**.

2. More Experiments

Detailed Comparisons of Loss Functions. TSGA employs a learnable loss network during the adaptation process, to address the misalignment between evaluation metrics and the requirements of OCC tasks. To evaluate its effectiveness, we compare it with additional loss functions, *e.g.*, L1-loss and L2-loss, as shown in Table 1. Our results reveal that: 1) In the absence of negative samples, traditional loss functions result in performance degradation. The commonly used loss functions for multi-classification tasks deliver the same bad performance, indicating the misalignment mentioned above exists. 2) The proposed simple yet effective loss network $\mathcal{N}$ achieves significant performance improvements by automatically learning the optimal loss function tailored for one-class samples.

Dataset	Model	$K = 2$	$K = 10$
MIN	MetaSGD	65.0	73.6
	OC-MetaSGD	69.6	75.8
	TSGA (Ours)	**72.22**	**76.84**
CIFAR-FS	MetaSGD	58.4	71.3
	OC-MetaSGD	71.4	77.8
	TSGA (Ours)	**75.98**	**80.48**

Table 2. Comparison of MetaSGD, OC-MetaSGD, and our TSGA for 2-shot and 10-shot OCC tasks on miniImageNet (MIN) and CIFAR-FS. We highlight the best number in **bold**.

Model	OMN		CIFAR-FS	
	pos	neg	pos	neg
OC-MAML †	97.75	98.93	**81.40**	76.16
TSGA (Ours)	**97.91**	**99.04**	80.64	**82.99**

† Our reproduction.

Table 3. Class-imbalanced query sets comparison of OC-MAML and TSGA for 10-shot OCC tasks on Omniglot (OMN) and CIFAR-FS. "pos" and "neg" refer to positive and negative query sets. We highlight the best number in **bold**.

Boost MetaSGD with TSGA. We evaluate the universality of our method by integrating TSGA with other optimization-based meta-learning approaches, *e.g.*, OC-MetaSGD, which achieves performance superior to or comparable with OC-MAML. In this experiment, we only incorporate the loss network, as MetaSGD [3] already learns the learning rate for each parameter, which we retain. As shown in Table 2, the results demonstrate that TSGA consistently enhances OC-MetaSGD by integrating the learnable loss network into the adaptation process, confirming the effectiveness and flexibility of TSGA when applied to optimization-based meta-learning methods.

Evaluation on Class-imbalanced Query Sets. In the fields of novelty detection and anomaly detection, class-balanced query sets are rarely encountered in real-world scenarios. To address this, we evaluate classification accuracy on distinct query sets containing exclusively positive or negative samples, as shown in Table 3. The experimental results demonstrate that TSGA outperforms the OC-MAML baseline. By leveraging the learnable loss network, TSGA enhances its ability to mitigate overfitting to the few available data, which reduces the likelihood of predicting normal samples as anomalies and overfitting to the majority class, which could otherwise result in incorrectly classifying most samples as normal.

Sensitivity to Learning Rate. We perform sensitivity of hyperparameters (*e.g.*, learning rate) experiments in Table 4, demonstrating that TSGA maintains consistent performance across a wide range of learning rates, with minimal degradation even at higher values. This highlights its robustness. The results validate the stability and generalizability of our approach.

Learning rate	1	0.1	0.01	0.001
OC-MAML	66.10	75.64	76.20	76.42
TSGA	**75.70**	**76.34**	**77.73**	**77.98**

Table 4. Comparison of different learning rates used in the adaptation process for 10-shot OCC tasks on miniImageNet (MIN). We highlight the best number in **bold**.

	miniImageNet → CIFAR-FS
OC-MAML	66.82
TSGA	**72.02**

Table 5. Comparison for 10-shot cross-domain OCC tasks. Models are trained on miniImageNet and evaluated on CIFAR-FS. We highlight the best number in **bold**.

Dataset	$K = 2$	$K = 10$
miniImageNet	1.93×10^{-11}	1.96×10^{-5}
Omniglot	1.87×10^{-12}	8.17×10^{-10}
CIFAR-FS	1.26×10^{-12}	1.34×10^{-7}

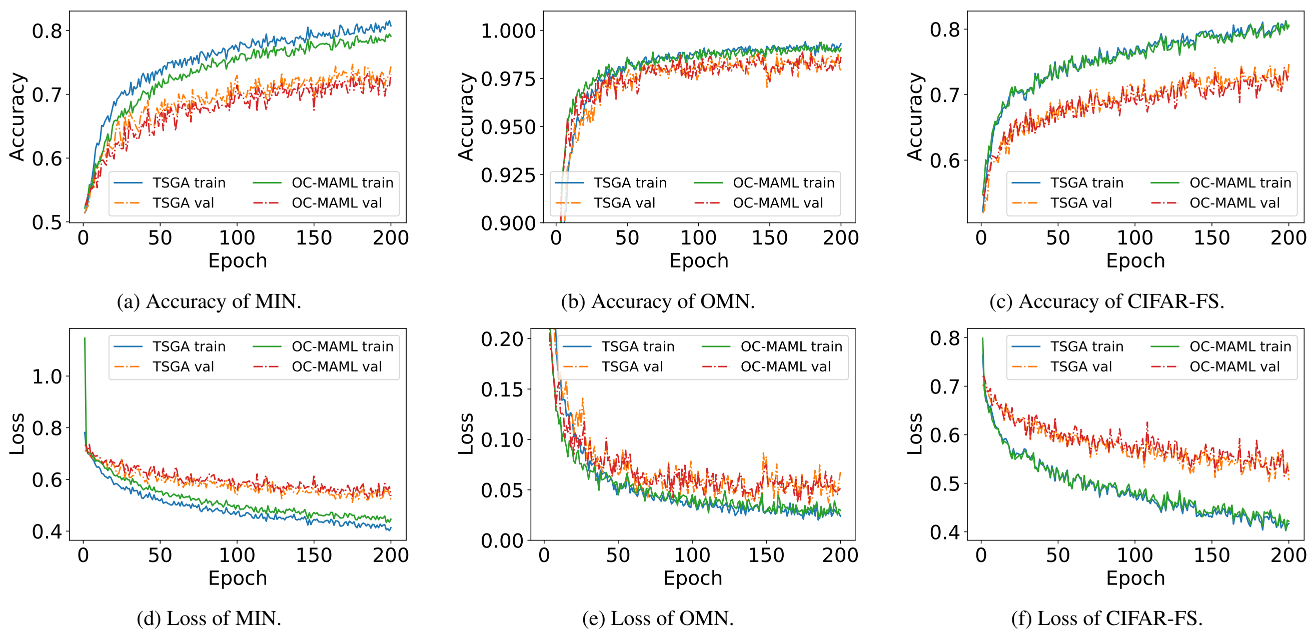
Table 6. Significance tests (One-way ANOVA) results (p-values) for performance comparisons evaluated on multiple datasets and few-shot configurations.

Cross-Domain Generalization. The distribution shift between training and test tasks is indeed a significant challenge. Our additional results on cross-domain tasks (e.g., miniImageNet to CIFAR-FS) in Table 5 showcase the adaptability of TSGA, achieving a notable accuracy of 72.02% compared to OC-MAML's 66.82%, demonstrating its robustness under significant distribution shifts.

Significance Tests. The analysis of variance (ANOVA) revealed statistically significant performance differences across all datasets and few-shot configurations in Table 6. All p-values are below the 0.05 threshold, indicating statistically significant differences between models.

3. Visualization

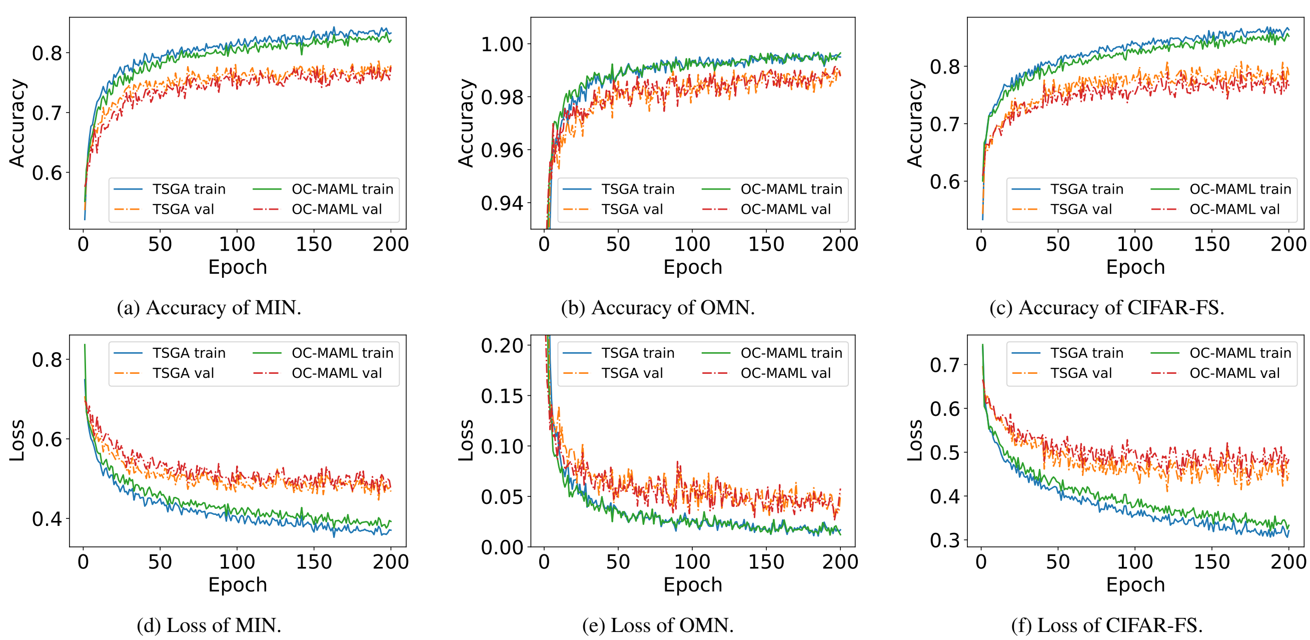
Convergence Curves. To validate the convergence results, we plot the changing tendency curves of training and validation accuracy and loss on the query set with the number of epochs in our experiments, as shown in Figure 1 and Figure 2. The convergence tendency can be easily observed in the figures, which show significantly more stable and faster convergence during the optimization, substantiating the ef-



(a) Accuracy of MIN. (b) Accuracy of OMN. (c) Accuracy of CIFAR-FS.

(d) Loss of MIN. (e) Loss of OMN. (f) Loss of CIFAR-FS.

Figure 1. Training and validation accuracy and loss tendency curves of query sets for 2-shot OCC tasks on miniImageNet (MIN), Omniglot (OMN), and CIFAR-FS. “train” and “val” represent meta-training and meta-validation set, respectively.



(a) Accuracy of MIN. (b) Accuracy of OMN. (c) Accuracy of CIFAR-FS.

(d) Loss of MIN. (e) Loss of OMN. (f) Loss of CIFAR-FS.

Figure 2. Training and validation accuracy and loss tendency curves of query sets for 10-shot OCC tasks on miniImageNet (MIN), Omniglot (OMN), and CIFAR-FS. “train” and “val” represent meta-training and meta-validation set, respectively.

fectiveness of our TSGA over OC-MAML.

CNC Milling Machine Dataset. Figure 3 shows exemplary normal and anomalous samples belonging to OP00 from the CNC Milling Machine dataset. Due to the extreme scarcity of anomalous time-series data (including both quantity and duration of series), it is difficult to train an OCC model. In addition, the real-world dataset also faces various challenges, *e.g.*, cross-machine and cross-time generalization. Please refer to [4] for more details of the CNC Milling Machine dataset.

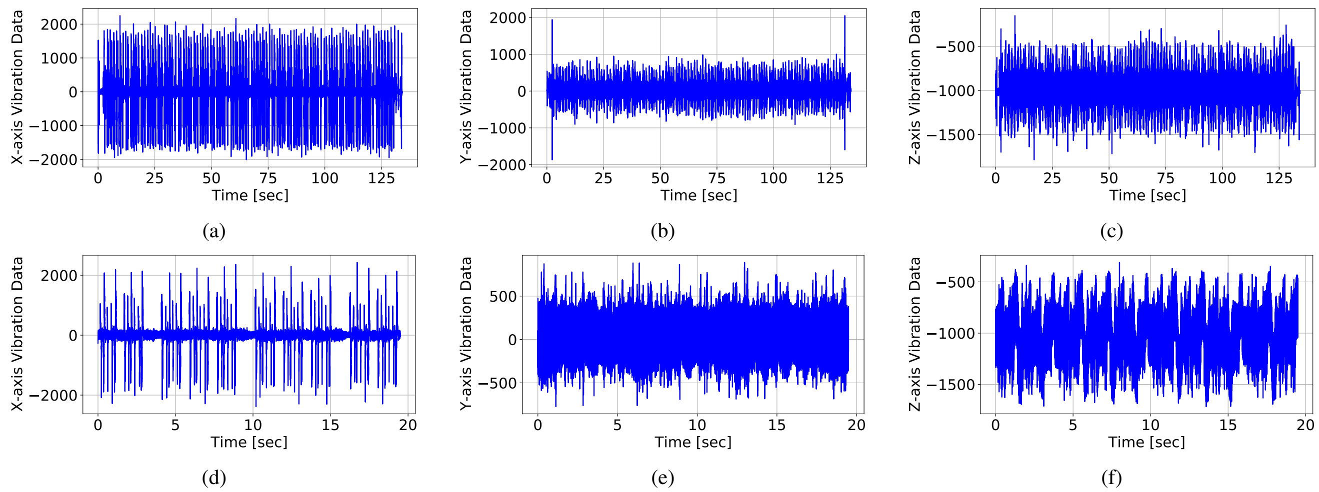


Figure 3. Exemplary normal (top) and anomalous (bottom) samples belonging to OP00 from the CNC Milling Machine dataset. (a)-(c): x, y, z-axis of a normal sample. (d)-(f): x, y, z-axis of a anomalous sample.

References

[1] Ahmed Frikha, Denis Krompaß, Hans-Georg Köpken, and Volker Tresp. Few-shot one-class classification via meta-learning. In *Proceedings of the AAAI conference on artificial intelligence*, pages 7448–7456, 2021. 1

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1

[3] Zhenguo Li, Fengwei Zhou, Fei Chen, and Hang Li. Meta-sgd: Learning to learn quickly for few-shot learning. *arXiv preprint arXiv:1707.09835*, 2017. 2

[4] Mohamed-Ali Tnani, Michael Feil, and Klaus Diepold. Smart data collection system for brownfield cnc milling machines: A new benchmark dataset for data-driven machine monitoring. *Procedia CIRP*, 107:131–136, 2022. 3