

Meta-Learning Experiment Report

Pang Liu
jeff.pang.liu@gmail.com

December 15, 2025

1 Introduction and Data Preparation

This experiment is based on the Mini-ImageNet dataset and explores **meta-learning** and the **MAML algorithm**. It compares meta-learning with a baseline method, and investigates the effects of different N-way, K-shot, and inner loop update steps. **Tasks are generated episodically by sampling N classes into Support and Query sets.** To match the fixed classifier output size, disjoint original labels are mapped to a consistent index $0, \dots, N - 1$, ensuring valid loss calculation for each task. The results demonstrate that MAML exhibits significantly superior rapid adaptation capabilities in few-shot image classification tasks.

2 Comparisons and Analysis

2.1 MAML Model vs. Fine-tuning Baseline

Configuration: 5-way 5-shot.

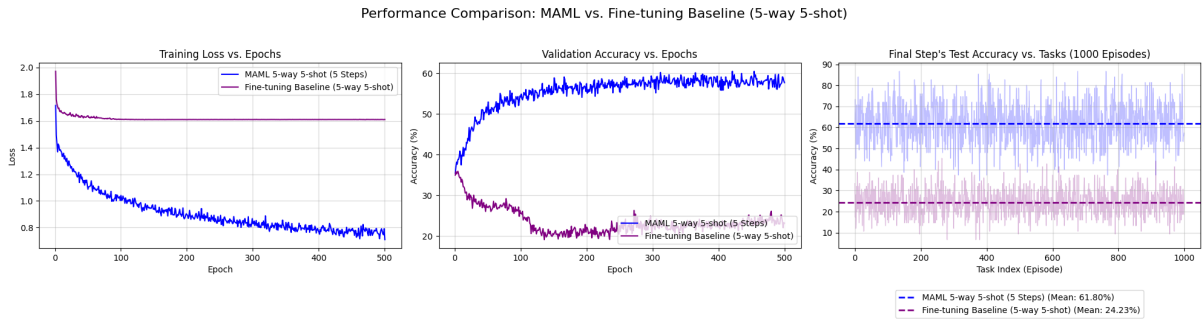


Figure 1: MAML vs. Fine-tuning Baseline: Training Loss, Validation Accuracy, Final Step Test

From the experiment result, we can see that the **MAML model significantly outperforms the Fine-tuning Baseline**. The MAML model achieved a final test accuracy of **61.80%**, whereas the baseline only reached **24.23%**.

From the plots, we observe that the training loss for the Baseline (purple line) remained high (≈ 1.61) and flat throughout the 500 epochs, and the validation accuracy hovered around 20-25%. Since 5-way classification has a random guess probability of 20%, the Baseline effectively failed to learn how to generalize to new tasks. In contrast, MAML's loss consistently decreased, and accuracy increased.

This result aligns with expectations and what we learned. The Fine-tuning Baseline trains a single network across randomly sampled tasks without a specific meta-objective. It struggles to find a single set of weights that works for all tasks simultaneously. MAML uses the inner/outer loop to explicitly learn an initialization that is sensitive to task-specific gradients. This allows MAML to adapt rapidly to unseen tasks.

2.2 Effect of N in N-Way Classification

Configuration: 5-way 5-shot vs. 10-way 5-shot.

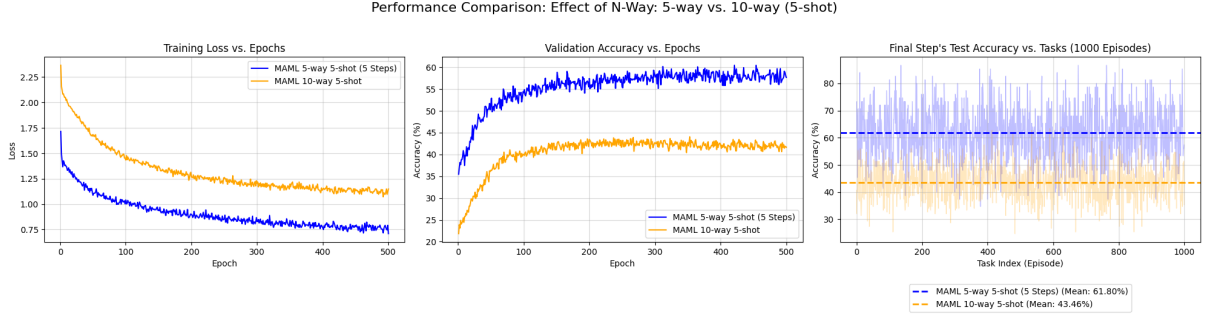


Figure 2: Performance comparison between 5-way and 10-way classification

The **5-way configuration performs significantly better** than the 10-way configuration (61.80% vs. 43.46%).

This result is expected because increasing N increases the difficulty of the classification problem. First, the random baseline drops from 20% ($1/5$) for 5-way to 10% ($1/10$) for 10-way. Second, with 10 classes, the model must carve out more complex decision boundaries in the embedding space to distinguish between twice as many categories using the same amount of support data per class. The higher initial loss observed in the 10-way experiment reflects this higher entropy and task difficulty.

2.3 Effect of K in K-Shot Configuration

Configuration: 5-way 5-shot vs. 5-way 1-shot.

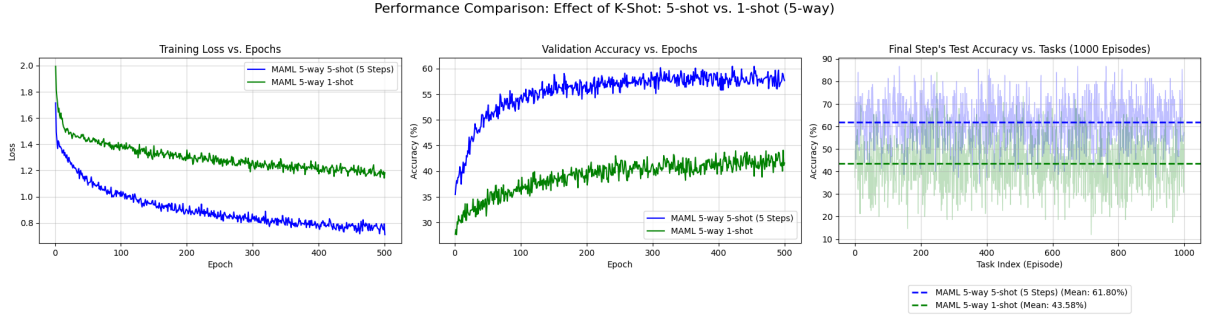


Figure 3: Performance comparison between 5-shot and 1-shot learning

From above results, we can see that the **5-shot configuration performs significantly better** than the 1-shot configuration (61.80% vs. 43.58%).

The shot influence the number of labeled examples per class available for the inner-loop adaptation. In 1-shot learning, the model relies on a single image to calculate the gradient and update its weights. This gradient is highly noisy and may not represent the true distribution of the class. With only 1 sample, the model is highly prone to overfitting the support set during the inner loop updates. With 5 examples (5-shot), the inner-loop optimization has a more representative sample, allowing MAML to compute a more stable gradient direction.

2.4 Effect of the Number of Inner-Loop Update Steps

Configuration: 5 steps vs. 2 steps (during training).

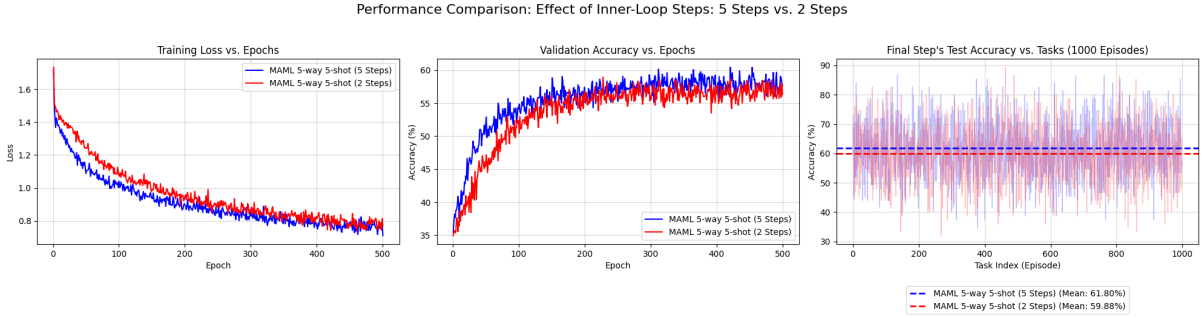


Figure 4: Performance comparison between 5 inner-loop steps and 2 steps

Two results are very close due to the experiment result. But we still can see that **5 update steps performs slightly better** than the configuration with 2 update steps (61.80% vs. 59.88%).

Increasing the number of inner-loop steps allows the task-specific model to descend further down the loss surface for the specific task. Two steps might be insufficient to fully adapt the initial weights to the new task’s optimal state. Furthermore, by training with 5 steps, the meta-model learns an initialization that is robust enough to remain stable and improve over a longer trajectory of gradient updates.

3 Conclusion and Ablation Study

The table below summarizes the final averaged test accuracy across all conducted experiments (computed over 1000 test episodes with a 95% confidence interval).

Table 1: Final Test Accuracy Summary for All MAML Experiments

Experiment	Configuration	Test Accuracy
MAML	5-way 5-shot (5 Steps)	61.80% \pm 0.60%
Effect of N	10-way 5-shot	43.46% \pm 0.39%
Effect of K	5-way 1-shot	43.58% \pm 0.69%
Effect of Steps	5-way 5-shot (2 Steps)	59.88% \pm 0.58%
Fine-tuning (Baseline)	5-way 5-shot	24.23% \pm 0.39%

Table 1 comprehensively shows this experiment’s result and show the advantage of MAML over baseline, and show the importance of proper settings of N-way and N-shot. Specifically, MAML outperforms standard fine-tuning baselines with same N-way and N-shot; 5-way performs better than 10-way, 5-shot performs better than 1-shot, 5 Steps performs better than 2 Steps.

In the future work, we can explore more details of configuration settings, to explore the best through different settings and find the pattern or tradeoffs between different combinations.