MAML is a Noisy Contrastive Learner in Classification

Chia-Hsiang Kao, Wei-Chen Chiu, Pin-Yu Chen
National Yang Ming Chiao Tung University, Taiwan   IBM Research

Contribution

Prove that MAML is a noisy contrastive learning algorithm and propose a zeroing trick to mitigate the noise.

Take Home Message

Q1 Why is MAML effective in learning representations?
A1 Because **MAML implicitly exploits contrastive learning**.

Q2 What is the role of inner loop in MAML?
A2 **In inner loop, classifier memorizes support features.**

Q3 What is the role of support data in MAML?
A3 **The support features act as the prototypes.**

MAML + Our Zeroing Trick

Require inner-/outer-loop learning rate: η / ρ
Require encoder/classifier parameters: θ $\mathbf{w}$
Set $\mathbf{w} \leftarrow 0$ (the zeroing trick)
while not done **do**
    Sample tasks $\{T_1, \ldots T_{N_{batch}}\}$
    for $n = 1, 2, \ldots, N_{batch}$ **do**
        $\{S_n, Q_n\} \leftarrow$ sample from T_n
        $\theta_n = \theta$
        for $i = 1, 2, \ldots, N_{step}$ **do**
            $\theta_n \leftarrow \theta_n - \eta \nabla_{\theta_n} L_{\theta_n, S_n}$
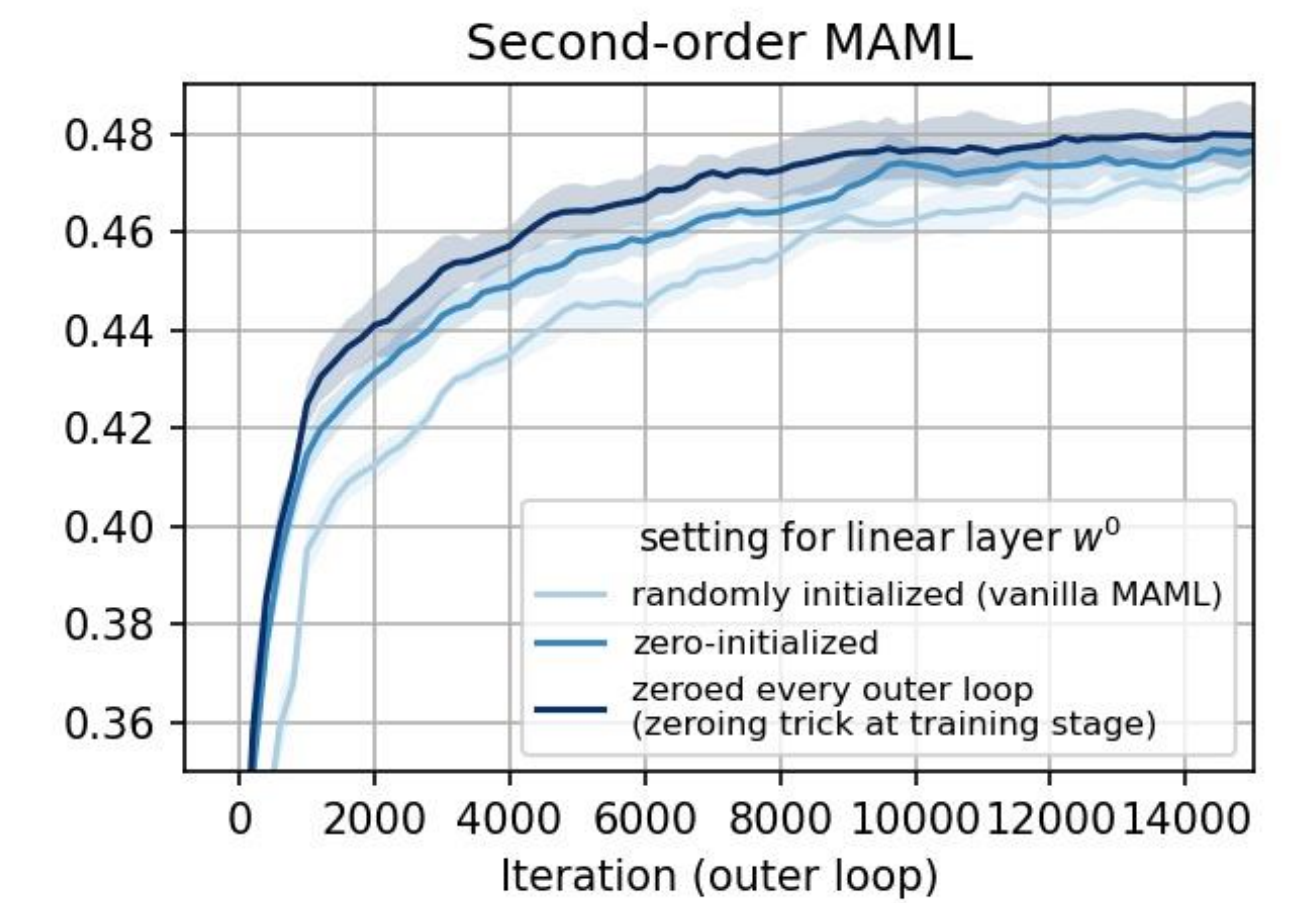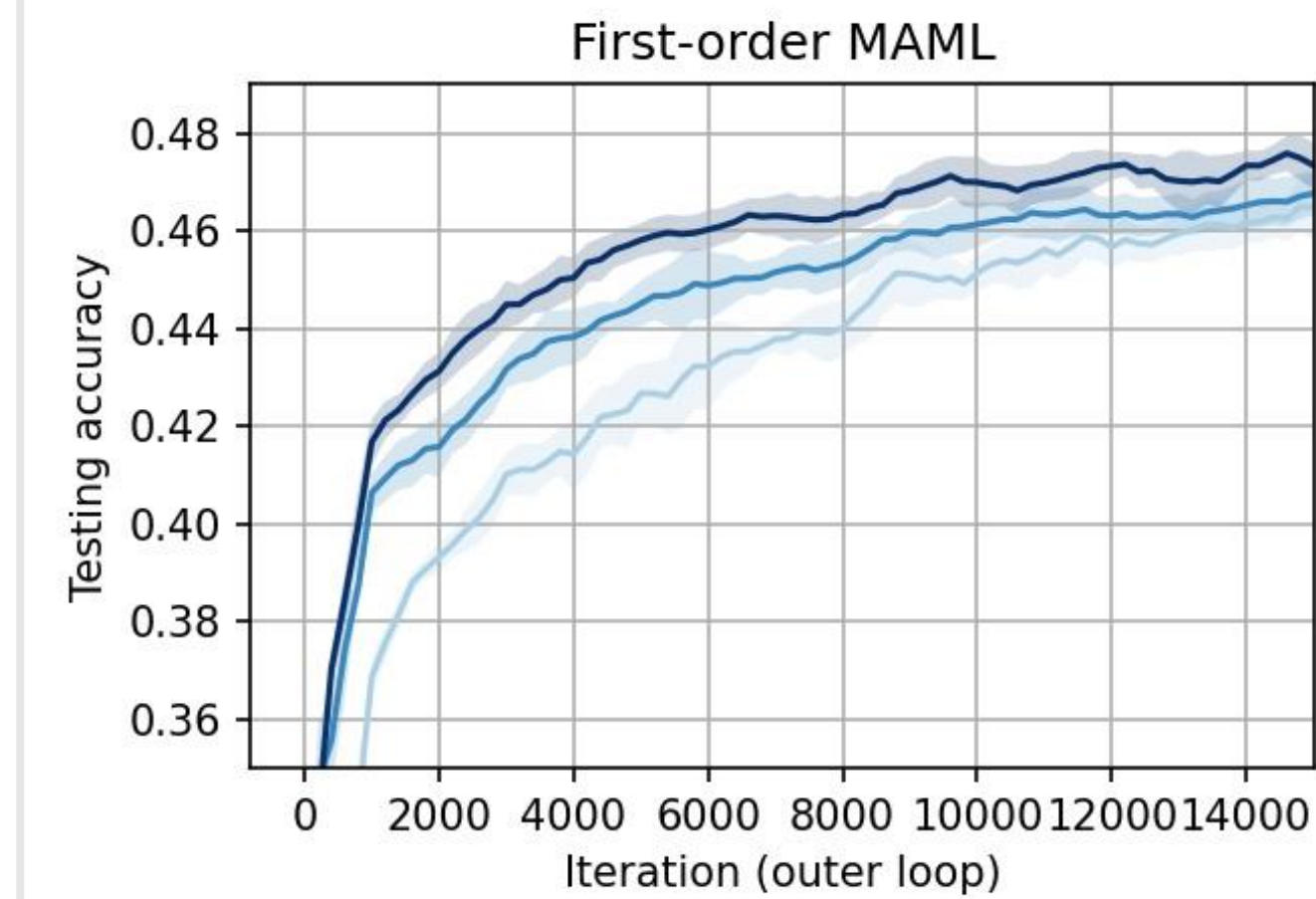   **end for**
    end for
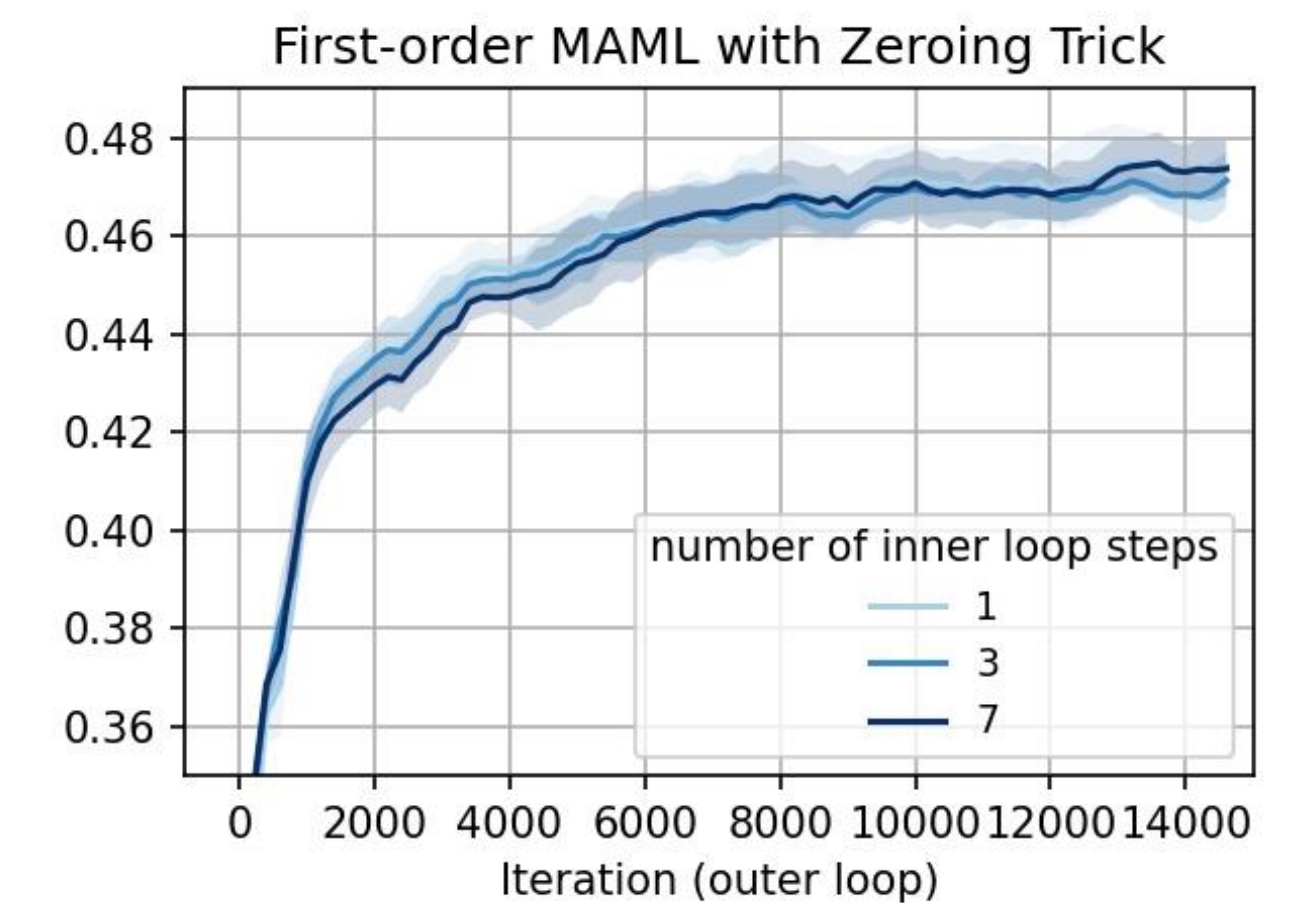   Update $\theta \leftarrow \theta - \rho \sum_{n=1}^{N_{batch}} \nabla_\theta L_{\theta_n, Q_n}$
    Set $\mathbf{w} \leftarrow 0$ (the zeroing trick)
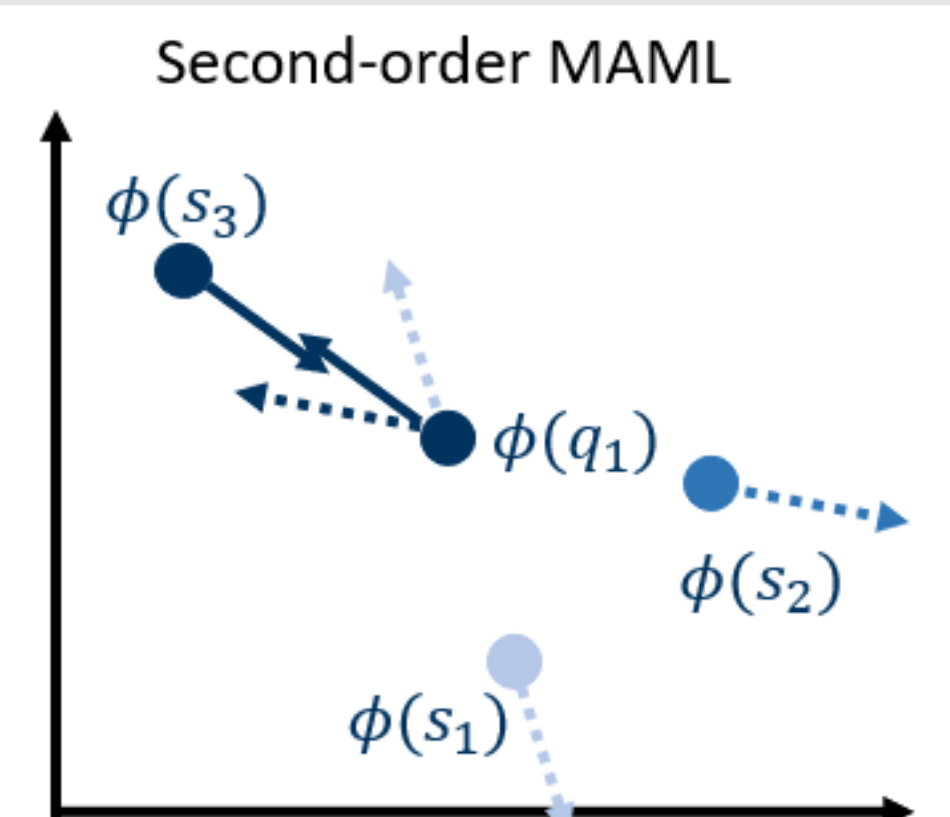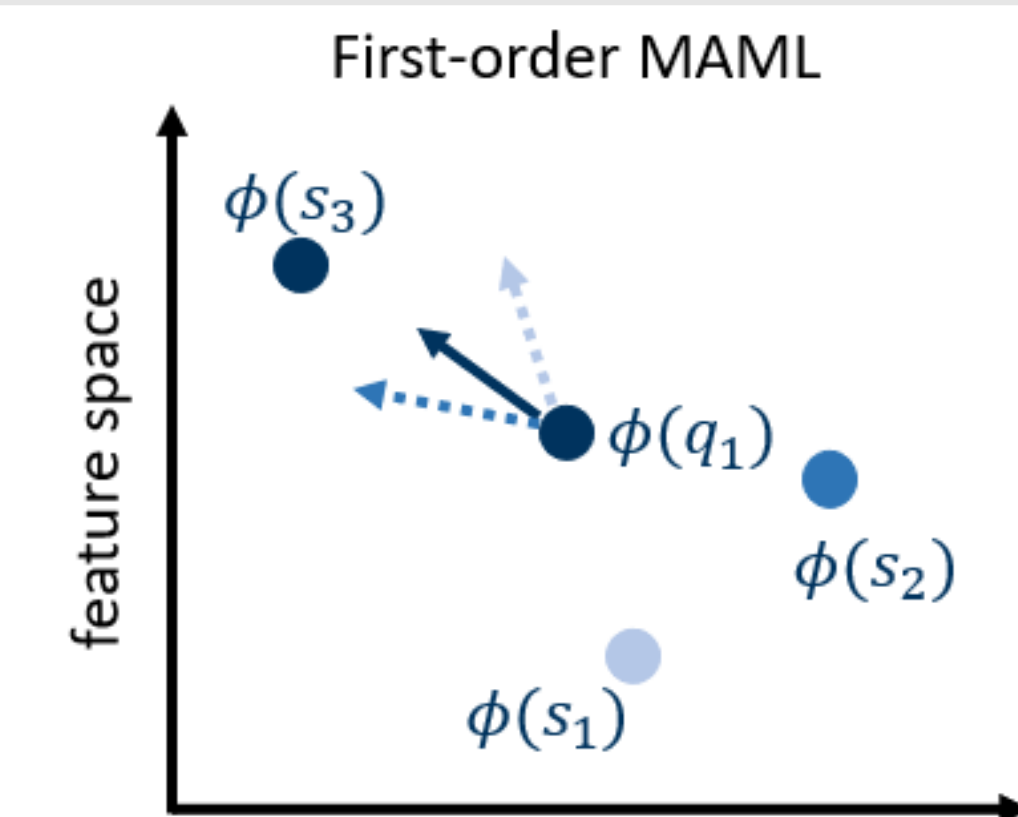**end while**

A Motivating Example

Setting:
5-way 1-shot using MAML with **one** inner-loop iteration under **MSE loss**.

Model:
Image → ϕ → w → Output
         Encoder Classifier

Value:
0   1

Inner Loop (1 step)

1 Inner loop update using
$w^1 = w^0 - \eta \sum_{k=1}^{5} -\phi(s_k) t_k^{\top}$

2 The ith column of classifier is updated by adding support feature of the ith class.

Outer Loop

Class 3 q_1

$\eta \phi(q_1)^{\top} \phi(s_1)$ 0
$\eta \phi(q_1)^{\top} \phi(s_3)$ 1
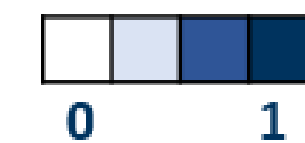
3 Loss from negative sample
q_1 and s_1 have different labels. The loss happens to ask their inner product of features to be zero.

4 Loss from positive sample
q_1 and s_3 have same labels. The loss happens to ask their inner product of features to be one.

1 Using Zeroing Trick Mitigates Inherent Noise in MAML

setting for linear layer w^0
randomly initialized (vanilla MAML)
zero-initialized
zeroed every outer loop (zeroing trick at training stage)

2 Without Inherent Noise, a Larger Number of Inner Loop Update Steps Is Not Necessary

number of inner loop steps
1
3
7

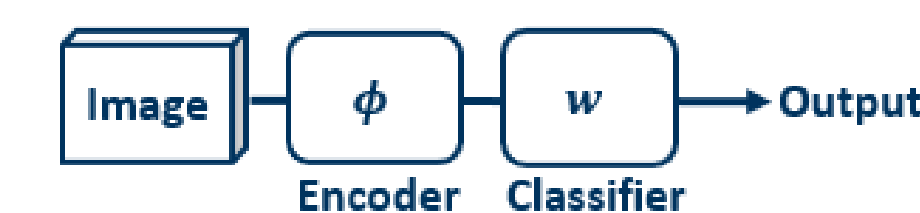## 3 We Identify the Difference Between FOMAML and SOMAML From a Contrastive Learning Perspective.