Improving the Generalization of MAML in Few-Shot Classification via Bi-Level Constraint

Yuanjie Shao, Wenxiao Wu, Xinge You, *Senior Member, IEEE*, Changxin Gao, *Member, IEEE*, and Nong Sang, *Member, IEEE*

***Abstract*— Few-shot classification (FSC), which aims to identify novel classes in the presence of a few labeled samples, has drawn vast attention in recent years. One of the representative few-shot classification methods is model-agnostic meta-learning (MAML), which focuses on learning an initialization that can quickly adapt to novel categories with a few annotated samples. However, due to insufficient samples, MAML can easily fall into the dilemma of overfitting. Most existing MAML-based methods either improve the inner-loop update rule to achieve better generalization or constrain the outer-loop optimization to learn a more desirable initialization, without considering improving the two optimization processes jointly, resulting in unsatisfactory performance. In this paper, we propose a bi-level constrained MAML (BLC-MAML) method for few-shot classification. Specifically, in the inner-loop optimization, we introduce a supervised contrastive loss to constrain the adaptation procedure, which can effectively increase the intra-class aggregation and inter-class separability, thus improving the generalization of the adapted model. In the case of the outer loop, we propose a cross-task metric (CTM) loss to constrain the adapted model to perform well on the different few-shot task. The CTM loss can enforce the adapted model to learn more discriminative and generalized feature representations, further boosting the generalization of the learned initialization. By simultaneously constraining the bi-level optimization procedure, the proposed BLC-MAML can learn an initialization with better generalization. Extensive experiments on several FSC benchmarks show that our method can effectively improve the performance of MAML under both the within-domain and cross-domain settings, and also perform favorably against the state-of-the-art FSC algorithms.**

***Index Terms*— MAML, bi-level constraint, supervised contrastive loss, cross-task metric loss.**

I. Introduction

ALTHOUGH deep learning models have achieved outstanding performance on various computer vision tasks, the generalization ability of deep neural networks heavily relies on a large number of labeled data. However, collecting sufficient amounts of annotated data in newly emerging or rare fields is difficult or expensive [1], [2], [3], which severely precludes the applicability of current supervised learning models to recognize new classes. This issue has inspired research on few-shot learning (FSL) [4], [5], [6], [7], [8], [9], [10], which aims to identify the unseen categories with limited labeled samples.

Due to the low-data properties of few-shot learning, powerful deep models are prone to overfitting, resulting in poor generalization on test data. Various methods have been proposed to solve this problem, and meta-learning [4], [5], [11], [12] become the most prominent methods among them. It can learn task-agnostic meta-knowledge from previous tasks to quickly adapt to novel task with a few data samples. Especially, model-agnostic meta-learning (MAML) is one of the influential optimization-based meta-learning frameworks [5], [13], [14], [15] for its flexible and model independence. MAML aims to learn a good initialization, which can adapt to new tasks with few samples and steps of gradient updates. It is usually formulated as a bi-level optimization problem. In the inner-loop optimization, the shared model initialization is adapted to each task via several gradient descent updates over the support images of that task to learn a task-specific model. In the outer loop, it applies a meta-training objective to evaluate the generalization of the initialization on the query images. Despite its success in various applications, MAML still suffers from poor generalization. Since the number of samples in the support set of each meta-training task is relatively small, the adapted model is still prone to overfitting to the limited support images, resulting in unsatisfactory generalization on the query images.

There have been recent works on improving the generalization of MAML either by improving the inner-loop update rule to obtain a better adaptation process [16], [17] or constraining the outer-loop optimization to learn a more desirable initialization [13], [14] However, these methods do not consider constraining the bi-level optimization procedure simultaneously, which leads to the learned initialization still having the problem of poor generalization. Moreover, existing approaches only use a plain cross-entropy loss to learn a task-specific model in the inner-loop optimization. Due to the few samples available for adaptive learning, just using this supervised loss does not guarantee that the adapted model generalizes well on the query samples.

Manuscript received 26 October 2022; revised 8 December 2022; accepted 17 December 2022. Date of publication 27 December 2022; date of current version 3 July 2023. This work was supported by the National Natural Science Foundation of China under Grant 61901184 and Grant U22B2053. This article was recommended by Associate Editor S. Wang. *(Corresponding author: Nong Sang.)*

Yuanjie Shao and Xinge You are with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: shaoyuanjie@hust.edu.cn; youxg@hust.edu.cn).

Wenxiao Wu, Changxin Gao, and Nong Sang are with the National Key Laboratory of Science and Technology on Multispectral Information Processing, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: wenxiaowu@hust.edu.cn; cgao@hust.edu.cn; nsang@hust.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCSVT.2022.3232717.

Digital Object Identifier 10.1109/TCSVT.2022.3232717

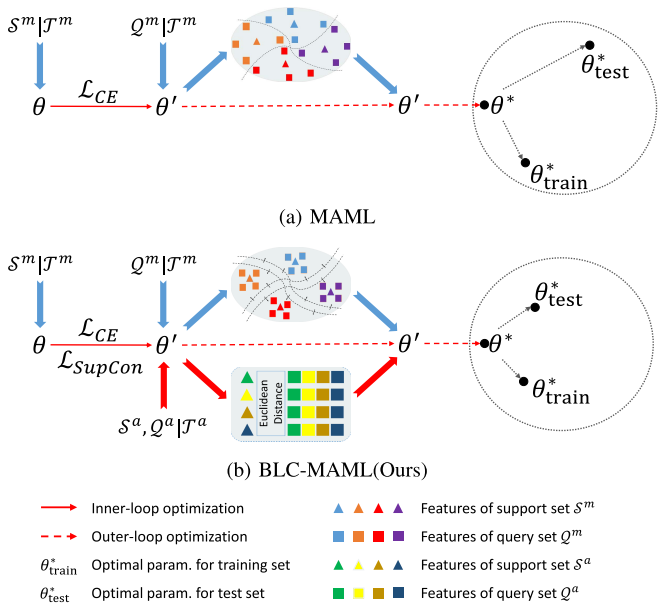


Fig. 1. Overview of both inner-loop and outer-loop optimization of MAML and our BLC-MAML. (a) Pure MAML [5], only leverages cross-entropy loss $\mathcal{L}_{CE}$ to update model in inner-loop, resulting in unsatisfactory generalization. (b) Our proposed method, BLC-MAML, leverages cross-entropy loss $\mathcal{L}_{CE}$ in conjunction with a supervised contrastive loss $\mathcal{L}_{SupCon}$ to update the model in order to obtain a larger margin between classes (inner-loop constraint). Moreover, we introduce an auxiliary task $\mathcal{T}_a$ to constrain the updated model with parameters θ' to obtain a more generalized feature representation (outer-loop constraint). Through the bi-level constraint, we hope to obtain an initial parameter closer to the optimal one.

To this end, in this paper, we propose a bi-level constrained MAML (BLC-MAML) method for few-shot classification. Specifically, inspired by the recent finding [18], robust meta-learning requires features with large inter-class separability and low intra-class variance in addition to vanilla linear separability. In the inner-loop optimization, we introduce a supervised contrastive loss based on cross-entropy loss. We not only hope that each sample is classified correctly but also hope that the categories have good separability from a global perspective, so as to improve the generalization of the adapted model on the query images. In addition, in the outer loop, we propose a cross-task metric (CTM) loss further to improve the generalization performance of the adapted model. We use the CTM loss to explicitly constrain the adapted model not only to perform well on the query image of the current task but also on the other few-shot task with different categories. By doing so, we can enforce the adapted FSC model to pay more attention to the areas related to categories in the images, and learn more discriminative and generalized feature representations, thereby increasing the generalization of the adaptive model. By constraining the bi-level optimization procedure of MAML, we can obtain a metric space with favorable intra-class aggregation and inter-class separability, thus effectively increasing the generalization on the test dataset, as shown in Fig. 1.

We summarize our contributions as following three-folds:

- In addition to adopting the cross-entropy loss, we propose a supervised contrastive loss to constrain the adaptation process in the inner loop, which can reduce overfitting to the support set samples and learn an adapted model with good generalization.
- In the outer loop, we propose a cross-task metric loss to enforce the adapted model to perform well on different few-shot task further to improve the discriminability and generalization of the FSC model.
- We conduct extensive experiments on the standard few-shot learning benchmarks. The experimental results show that our algorithm can effectively improve the performance of MAML and can also perform favorably against the state-of-the-art FSC methods under both within-domain and cross-domain settings.

II. Related Work

Few-shot classification aims to learn to identify novel categories with limited labeled samples. Many efforts have been devoted to solving the above problem. In this section, we first introduce three representative few-shot classification methods: augmentation-based methods, metric-based methods, and optimization-based methods. And then, we briefly discuss the contrastive learning methods, which are related to our work.

A. Few-Shot Learning

1) Augmentation-Based Methods: For few-shot classification, the intuitive solution is to increase the available samples via data augmentation methods [19], [20], [21], [22]. Wang et al. [19] propose to incorporate random noise into the images to synthesize new data. Schwartz et al. [23] propose a novel auto-encoder, delta-encoder, to synthesize data samples for an unseen class with a few images. Wang et al. [21] utilize a saliency network to obtain the foregrounds and backgrounds of visual images and hallucinate additional data via foreground-background combinations. Yang et al. [24] uses the statistics of rich samples to calibrate the distribution of the few-sample classes, and then a large number of samples can be sampled from these calibrated distributions. Besides, enlightened by Generative Adversarial Networks (GANs) [25], some methods apply GANs to generate images for few-shot learning [26], [27]. Zhang et al. propose a MetaGAN [26] to synthesize images that cannot be distinguished from the real samples sampled from a specific task. Li et al. [27] propose to synthesize fake support features using an adversarial feature hallucination network.

2) Metric-Based Methods: Metric-based methods aim to find a good distance function which can measure data from the same categories closer than different ones [4], [11], [12]. For example, Vinyals et al. [4] apply a recurrent network to measure cosine similarities between query images and each support image. ProtoNet [11] utilizes Euclidean distance as the metric function and applies the distance between a query feature and class-mean representation for classification. RelationNet [12] employs CNN-based relation modules, and GNN [28] uses graph neural networks as the metric function. Besides, some methods [29], [30] propose to improve FSL performance by designing a better feature extractor or a more discriminative

similarity measure. CTM [31] obtains task-relevant features by incorporating the category traversal module into the metric-based few-shot learners. FEAT [29] applies a self-attention-based transformation to yield task-specific and discriminative embeddings. By enforcing equivariance and invariance simultaneously, [30] can learn the features that are independent of the input transformation and encode the structure of geometric transformations. FRN [32] converts the few-shot classification into a reconstruction problem, and predicts the class of a query sample by reconstructing it from the support set of a given class. Hao et al. [6] propose a metric learning framework with global-local interaction, which employs global information to improve local semantic alignment. Zhou et al. [9] present an adaptive deep metric for a new FSL task based on a few labeled images, which overcomes the problem of insufficient discriminative capacity of fixed metrics.

3) Optimization-Based Methods: Optimization-based methods learn to adapt the model parameters or optimizer to the novel tasks with a few example images quickly [5], [15], [16], [17], [33], [34], [35], [36]. Specifically, MAML [5] learns a good model initialization that can rapidly adapt to new tasks via a few steps of gradient updates. Finn et al. [35] also propose a probabilistic version of MAML. ANIL [37] denotes that feature reuse is beneficial to improve the effectiveness of MAML, in which the feature extractor remains frozen and only the classifier of the model is updated in the inner loop. On the contrary, BOIL [15] believes that the dominant factor affecting the performance of MAML is the representation change, which only updates the feature extractor of the model and freezes the classifier in the inner loop. Sharp-MAML [38] leverages the inverse proportional relation between generalization ability and sharpness of local minimizers, aiming at improving the generalization performance of MAML by simultaneously minimizing the loss value and the loss sharpness. To achieve more accurate model adaptation, in addition to learning a good model initialization, Meta-SGD [34] learns a good update direction and learning rate. Ravi et al. [33] propose to use an LSTM-based meta-learner instead of the stochastic gradient descent optimizer to learn the optimizer.

In this work, we focus on model-agnostic meta-learning (MAML), one of the most popular optimization-based meta-learning frameworks. Although MAML has been widely used in various application fields with limited data, it still faces the challenge of poor generalization. To solve this deficiency, some works try to learn a better initialization by adding some regularization to the meta-learning procedure [13], [14]. Yin et al. [14] try to constrain the search space of initialization to prevent rote memorization during meta-training. Jamal et al. [13] propose an entropy-based inequality-minimization regularizer to force the learned initialization to have equal performance on all training tasks. Besides, other studies attempt to obtain a better adaptation process by improving the inner-loop update rule [16], [17]. Baik et al. [16] propose a small meta-network to enhance the fast adaptation process, which can adaptively generate hyper-parameters at each step. Baik et al. [17] utilize meta-learning to learn a task-adaptive loss function in the inner-loop for better generalization. However, these works only use a single cross-entropy loss function during the inner-loop adaptive learning. This will cause the adapted model to easily overfit a few support images and achieve poor generalization on the new images. In addition, the above methods only regularize the single-level optimization procedure to improve the generalization of MAML, and do not consider simultaneously constraining the bi-level optimization process to obtain better adaptation and generalization. In this paper, we propose a bi-level constrained MAML for few-shot learning. In the inner loop, we introduce a supervised contrastive loss function to constrain the adaptation process, thus reducing overfitting to support samples. Meanwhile, we propose a cross-task metric loss in the outer loop to constrain the adapted model to perform well on different few-shot tasks. By simultaneously constraining the bi-level optimization process, we can learn initialization with better generalization.

B. Contrastive Learning

Contrastive learning is a fundamental technique in self-supervised representation learning [39], [40], [41], [42], [43], [44]. The core idea of contrastive learning is to learn an embedding space that an anchor and a single "positive" sample are pulled together. Meanwhile, the anchor is pulled apart from many "negative" samples. Since there are no labels, the positive pair is usually composed of data augmentations of the sample. The negative pairs consist of the anchor and randomly selected small batches. In [41] and [42], they propose a contrastive loss, which maximizes the mutual information between different data views, to achieve the above goal. Inspired by the contrastive self-supervised learning, Khosla et al. [45] propose a supervised contrastive loss for supervised learning by leveraging label information, which pulls the images from the same class closer than those from the different classes. Compared with the cross-entropy loss, it can align the samples of the same class more closely and with better inter-class separability, thus leading to a better generalization performance. In this work, we take advantage of the supervised contrastive loss to constrain the inner-loop optimization procedure in MAML, enabling the adapted model to generalize well on the query images.

III. Method

In this section, we first present the definition of few-shot classification and introduce the conventional Model-agnostic meta-learning (MAML) method. Then we elaborate on our bi-level constrained MAML approach.

A. Problem Definition

Given an N-way K-shot few-shot classification (FSC) task $\mathcal{T}_i$, which is composed of two disjoint datasets: $\mathcal{S}_i$ and $\mathcal{Q}_i$. $\mathcal{S}_i = \{x_s^i, y_s^i\}_{s=1}^{N\times K}$ denotes the support set, which contains N categories with K instances in each category (K is always small, usually set to 1 or 5), and $\mathcal{Q}_i = \{x_q^i, y_q^i\}_{q=1}^{N\times M}$ denotes the query set that shares the same label space with $\mathcal{S}_i$, in which each class has M samples. The goal of few-shot classification is to correctly predict the instances in $\mathcal{Q}_i$ with the support set $\mathcal{S}_i$.

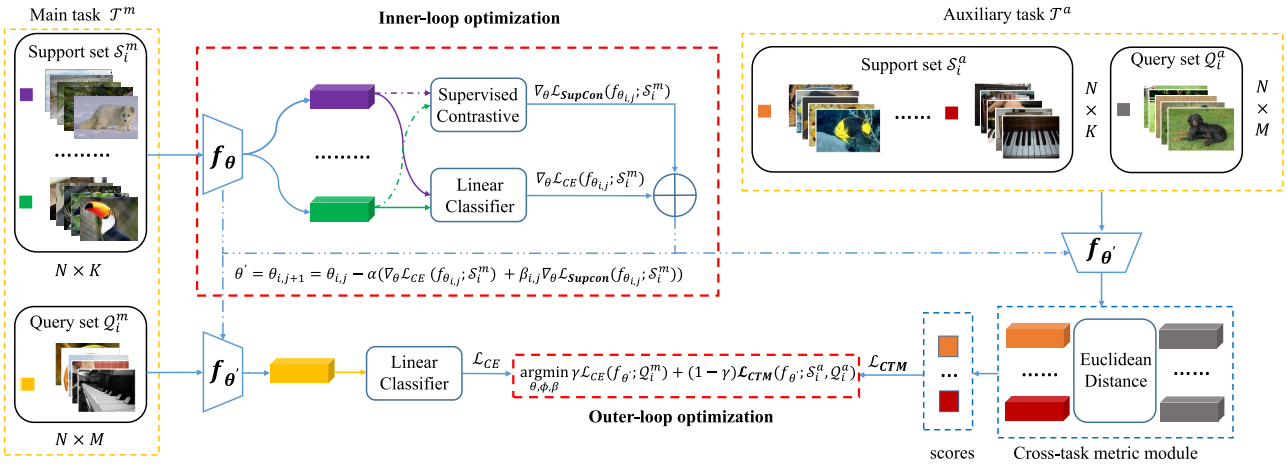


Fig. 2. The framework of our proposed BLC-MAML method. The proposed approach aims to improve the generalization of MAML by simultaneously constraining the inner-loop and outer-loop optimization procedure. In the inner-loop optimization, we utilize CE and SupCon loss to constrain the adaptation procedure jointly to improve the generalization of the adapted model. In the case of the outer loop, in addition to using CE loss, we additionally propose a cross-task metric (CTM) loss to constrain the adapted model $f_{\theta'}$ to perform well on the auxiliary few-shot task. The CTM loss can enforce the adapted model to learn more discriminative and generalized feature representations. The proposed BLC-MAML can learn an initialization with better generalization by constraining the bi-level optimization procedure simultaneously.

B. Model-Agnostic Meta-Learning (MAML)

The meta-learning framework [46], known as "learning to learn", provides an excellent solution to solve the above few-shot classification problem. It assumes a series of FSC tasks $\{\mathcal{T}_i\}_{i=1}^T$ drawn from a base data distribution $p(\mathcal{T})$, each of which also contains a support set $\mathcal{S}_i$ and a query set $\mathcal{Q}_i$. The meta-learning framework aims to learn a general learning model (parameterized by θ) through these FSC tasks that can quickly adapt to new tasks. This learning process is called meta-training, and the learned knowledge θ is referred to as meta-knowledge. The objective of the meta-training procedure is defined as follows:

$$\theta^* = \min_\theta \mathop{\mathbb{E}}_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}(\mathcal{T}_i; \theta) \tag{1}$$

where $\mathcal{L}$ measures the performance of a model on a task.

Then, it leverages the learned meta-knowledge θ^* to train a task-specific model ω on a new task $\mathcal{T}_{new}$ via using the support set images $\mathcal{S}_{new}$:

$$\omega_{new}^* = \min_\theta \mathcal{L}(\mathcal{S}_{new}; \omega, \theta^*), \tag{2}$$

By learning how to learn (a.k.a., meta-knowledge), meta-learning can significantly contribute to learning on a new task.

Model-agnostic meta-learning (MAML) [5] is one of the most popular meta-learning frameworks, in which the meta-knowledge denotes the initial parameters of a model. MAML aims to learn a good model initialization θ across tasks, which can quickly adapt to a new task with few images and updates. It formulates the meta-training procedure in Eq. 1 as a bi-level optimization problem: inner-loop optimization and outer-loop optimization. Firstly, a meta-batch consisting of B tasks, each of which is defined as a main task $\mathcal{T}_i^m = \{\mathcal{S}_i^m, \mathcal{Q}_i^m\}$, is sampled in advance. In the course of inner-loop optimization, images from support set $\mathcal{S}_i^m$ of each task $\mathcal{T}_i^m$ are leveraged to fine-tune the model (initialization $\theta_{i,0} = \theta$) via a fixed number of gradient descent. MAML usually adopts cross entropy (CE) loss $\mathcal{L}_{CE}$ to update the weights of the base learner f_θ at j-th step:

$$\theta_{i,j+1} = \theta_{i,j} - \alpha \nabla_\theta \mathcal{L}_{CE}(f_{\theta_{i,j}}; \mathcal{S}_i^m), \tag{3}$$

where α is the step size for inner-loop optimization. Thus, the final weight of base learner after J updates is $\theta' = \theta_{i,J}$. In the outer-loop optimization, it computes the CE loss on the query set $\mathcal{Q}_i^m$ of the same task based on the updated parameter θ', and updates the initial parameter θ as follow:

$$\theta \leftarrow \theta - \eta \sum_{i=1}^{B} \mathcal{L}_{CE}(f_{\theta'}; \mathcal{Q}_i^m). \tag{4}$$

C. Bi-Level Constrained MAML

Traditional MAML-based methods only utilize the CE loss for inner-loop optimization, which only guarantees the correct classification of each sample. However, to achieve better generalization performance, we also need good properties such as slight intra-class variance and large inter-class separability, which are not the direct optimization objectives of the CE loss. To this end, we introduce an additional supervised contrastive (SupCon) loss to constrain the inner-loop optimization process. It can make the samples with the same class close together and push the samples with different classes away, thus can obtain a feature space with a good generalization that different clusters are well separated. In addition, we also consider constraining the outer-loop optimization procedure to obtain better generalization and adaptation. In the outer-loop optimization, we propose a cross-task metric (CTM) loss to constrain the adapted model. We not only hope that the adapted model can perform well on the current task, but also hope it can perform well on the other few-shot classification task with disjoint label space. Therefore, the adapted model can extract discriminant features related to categories and better generalize for different tasks. The whole network architecture of the proposed bi-level constrained MAML is shown in Fig. 2. Next, we will introduce the bi-level constraint loss in detail.

1) Supervised Contrastive Loss for Inner-Loop Optimization: Given each sample $x_s^{m\_i}$ in support set $\mathcal{S}_i^m = \{x_s^{m\_i}, y_s^{m\_i}\}_{s=1}^{N\times K}$, we regard other images belonging to the the same class as positives $\mathcal{P}(s)$, while the remaining images are considered as negatives $\mathcal{N}(s)$. The purpose of supervised contrastive loss is to make the samples of the same category have greater similarity, and those from different classes have less similarity. Its definition is as follows:

$$\mathcal{L}_{SupCon} = -\sum_{s=1}^{NK} \frac{1}{|\mathcal{P}(s)|} \sum_{k\in\mathcal{P}(s)} log \frac{exp(cos(z_s, z_k))}{\sum\limits_{n\in\mathcal{N}(s)} exp(cos(z_s, z_n))}, \tag{5}$$

where z_s denotes the feature of the anchor $x_s^{m\_i}$, z_k and z_n represent the features of the images in the positives and negatives respectively, and $cos(\cdot)$ means the cosine distance between two features.

For the 5-way-1-shot classification problem, since there is only one sample in each class of the support set, we only push the instances from different classes away from each other. Thus Eq. 5 can become the following form:

$$\mathcal{L}_{SupCon} = -\sum_{s=1}^{NK} log \frac{1}{\sum\limits_{n\in\mathcal{N}(s)} exp(cos(z_s, z_n))}, \tag{6}$$

By minimizing the above objective function, we can obtain a feature space with large inter-class margins, thus improving the model's generalization performance. Therefore, we incorporate this SupCon loss into the inner-loop optimization and update the parameters of the base learner f_θ as follows:

$$\begin{aligned}\theta_{i,j+1} = \theta_{i,j} - \alpha(&\nabla_\theta \mathcal{L}_{CE}(f_{\theta_{i,j}}; \mathcal{S}_i^m) \\ &+\beta_{i.j} \nabla_\theta \mathcal{L}_{SupCon}(f_{\theta_{i,j}}; \mathcal{S}_i^m)),\end{aligned} \tag{7}$$

where $\beta_{i,j}$ is a hyperparameter that controls the proportion and effect of SupCon loss.

Task-Adaptive Proportion for Supervised Contrastive Loss: The significant difference between test tasks makes choosing the suitable β for each domain challenge. Therefore, to take full advantage of the SupCon loss, we introduce a task-adaptive module $g(\phi)$ to generate learnable hyperparameter β. As did in ALFA [16], we take layer-wise means of gradients and weights $\tau_{i,j} = [\overline{\nabla}_\theta \mathcal{L}_{SupCon}(f_{\theta_{i,j}}; \mathcal{S}_i^m), \overline{\theta}_{i,j}]$ as input to generate $\beta_{i,j}$.The updated formulation is denoted as:

$$\beta_{i,j} = g_\phi(\tau_{i,j}) \odot \beta_{i,j}^0, \tag{8}$$

where g_ϕ is a 3-layer MLP with ReLU activation between the layers, and $\odot$ denotes the Hadamard product. $\beta_{i,j}^0$ denotes the initial parameter for task $\mathcal{T}_i^m$ at step j.

2) Cross-Task Metric Loss for Outer-Loop Optimization: To further increase the adapted model's generalization, we propose a cross-task metric (CTM) loss to constrain the outer-loop optimization. In each meta-episode training procedure, in addition to sampling a main FSC task $\mathcal{T}_i^m$, we also select an auxiliary FSC task $\mathcal{T}_i^a = \{\mathcal{S}_i^a, \mathcal{Q}_i^a\}$ with different classes to assist in training, where $\mathcal{S}_i^a = \{x_s^{a\_i}, y_s^{a\_i}\}_{s=1}^{N\times K}$ and $\mathcal{Q}_i^a = \{x_q^{a\_i}, y_q^{a\_i}\}_{q=1}^{N\times M}$ represent support set and query set respectively. $\mathcal{T}_i^a$ can come from the same domain as the main task $\mathcal{T}_i^m$, and also can be drawn from other domains. The CTM loss is defined as follows:

$$\begin{aligned}&\mathcal{L}_{CTM}(f_{\theta_e'}; \mathcal{S}_i^a, \mathcal{Q}_i^a) \\ &= -\frac{1}{NM}\sum_{q=1}^{NM}\sum_{c=1}^{N} \mathbb{I}\{y_q^{a\_i} == c\} \log p(y_q^{a\_i} = c | x_q^{a\_i}),\end{aligned} \tag{9}$$

where $\mathbb{I}\{\cdot\}$ means the 0-1 indicator function and $p(y_q^{a\_i} = c|x_q^{a\_i})$ represents the probability that the query sample $x_q^{a\_i}$ belongs to class c:

$$p(y_q^{a\_i} = c | x_q^{a\_i}) = \frac{\exp(-\left\| f_{\theta_e'}(x_q^{a\_i}) - \mu_c \right\|_2^2)}{\sum\limits_j \exp(-\left\| f_{\theta_e'}(x_q^{a\_i}) - \mu_j \right\|_2^2)}, \tag{10}$$

Here, $f_{\theta_e'}$ is the feature embedding module of the adapted model $f_{\theta'}$, $\mu_c = \frac{1}{|(\mathcal{S}_i^a)_c|} \sum\limits_{(x_s^{a\_i}, y_s^{a\_i})\in(\mathcal{S}_i^a)_c} f_{\theta_e'}(x_s^{a\_i})$ denotes the feature prototype of the c-th class in support set $\mathcal{S}_i^a$, and $\left\| f_{\theta_e'}(x_q^{a\_i}) - \mu_c \right\|_2$ represents the Euclidean distance between $f_{\theta_e'}(x_q^{a\_i})$ and μ_c.

Algorithm 1 Bi-Level Constrained MAML

Require: Task distribution $p(\mathcal{T})$ and $p(\mathcal{T}')$, initial parameters θ and ϕ, learning rate λ, initial proportioan β, balance hyperparameter γ

while training **do**
  Sample a batch of B task pairs $\{\mathcal{T}_i^m, \mathcal{T}_i^a\}$, where $\mathcal{T}_i^m = \{\mathcal{S}_i^m, \mathcal{Q}_i^m\} \sim p(\mathcal{T})$ and $\mathcal{T}_i^a = \{\mathcal{S}_i^a, \mathcal{Q}_i^a\} \sim p(\mathcal{T}')$
  for each task pair $\{\mathcal{T}_i^m, \mathcal{T}_i^a\}$ **do**
    Initialize $\theta_{i,0} = \theta, \beta_{i,0} = \beta$
    for inner-loop step $j \in [0, J-1]$ **do**
      Compute the CE loss of $\mathcal{S}_i^m$: $\mathcal{L}_{CE}(f_{\theta_{i,j}}; \mathcal{S}_i^m)$
      Compute the supervised contrastive loss of $\mathcal{S}_i^m$: $\mathcal{L}_{SupCon}(f_{\theta_{i,j}}; \mathcal{S}_i^m)$
      Compute the layer-wise means of gradients and weights: $\tau_{i,j} = [\overline{\nabla}_\theta \mathcal{L}_{SupCon}(f_{\theta_{i,j}}; \mathcal{S}_i^m), \overline{\theta}_{i,j}]$
      Generate the task-adaptive hyperparameter: $\beta_{i,j} = g_\phi(\tau_{i,j}) \odot \beta_{i,j-1}$
      Compute adapted weights through gradient desent: $\theta_{i,j+1} = \theta_{i,j} - \alpha(\nabla_\theta \mathcal{L}_{CE}(f_{\theta_{i,j}}; \mathcal{S}_i^m) + \beta_{i.j} \nabla_\theta \mathcal{L}_{SupCon}(f_{\theta_{i,j}}; \mathcal{S}_i^m))$
    end for
  end for
  Compute query loss on all the B task pairs: $\mathcal{L}_{outer} = \gamma \sum_{i=1}^{B} \mathcal{L}_{CE}(f_{\theta'}; \mathcal{Q}_i^m) + (1-\gamma) \sum_{i=1}^{B} \mathcal{L}_{CTM}(f_{\theta_e'}; \mathcal{S}_i^a, \mathcal{Q}_i^a)$
  Perform gradient descent to update parameters: $(\theta, \phi, \beta) \leftarrow (\theta, \phi, \beta) - \eta \nabla_{(\theta,\phi,\beta)} \mathcal{L}_{outer}$
end while

Ensure: The final parameters θ, ϕ and β.

When optimizing the outer loop, we not only hope that the adapted model $f_{\theta'}$ performs well on the query images of the main task $\mathcal{T}_i^m$, but also hope it can classify the query instances of the auxiliary task $\mathcal{T}_i^a$ correctly. To this end, for outer-loop optimization, we use the following loss function to update initial parameters θ:

$$\mathcal{L}_{outer} = \gamma \sum_{i=1}^{B} \mathcal{L}_{CE}(f_{\theta'}; \mathcal{Q}_i^m) + (1-\gamma) \sum_{i=1}^{B} \mathcal{L}_{CTM}(f_{\theta'_e}; \mathcal{S}_i^a, \mathcal{Q}_i^a). \quad (11)$$

where γ balances the main and auxiliary task.

Since the CTM loss emphasizes that the adapted model should be suitable for different FSC tasks, the learned feature extraction model thus has better generalization. We summarize the complete procedure of our proposed method in Algorithm 1.

IV. Experiments

In this section, we first elaborate on the implementation details of our experiments. Then, we compare our approach with some competitive methods under the within-domain and cross-domain few-shot classification settings. Finally, we conduct ablation studies to analyze the effectiveness of the proposed bi-level constraint loss.

A. Implementation Details

For all experiments, we use ResNet-12 network architecture as the backbone of the base learner f_θ and implement our method within PyTorch. ResNet-12 contains four residual blocks, each consisting of 3 modules of convolution, batch normalization, and ReLU with a 2×2 max-pool layer applied at the end. The convolution filters in each residual block are 64, 128, 256, and 512. Following the protocol in [47], we also pretrain the feature extractor by minimizing the standard cross-entropy classification loss on the corresponding training dataset. We pretrain the feature embedding model on the mini-ImageNet [4] for cross-domain few-shot classification. We also note that the training set's label space is entirely disjoint from the validation and test set.

In the training stage, we train our framework in an episodic manner and use the validation set to select the best model. In each training episode, we follow the 5-way K-shot setting (K is set as 1 or 5) and pick 16 instances randomly for the query set of each task. During the inner-loop optimization, the base learner f_θ is updated for 5 times with step size α, which is set to 0.2 for mini-ImageNet and 0.01 for CUB. The initial proportion for β is set to 4 and 0.01 for mini-ImageNet and CUB, respectively. In the case of outer-loop optimization, all the experiments use the Adam optimizer, and the learning rate λ and meta-batch size B are set to 10^{-4} and 3, respectively. The value of balance hyperparameter γ is 0.5 for all of our experiments.

In the evaluation stage, the number of model updates is the same as the training stage, which is five times. For each dataset, we sample 2000 independent 5-way 1/5-shot classification tasks for evaluation, in which 15 samples per class are selected for the query set of each task, and report the average classification with 95% confidence interval.

TABLE I

Few-Shot Classification Accuracy (%) With 95% Confidence Interval on Mini-ImageNet Under 5-Way 1/5-Shot Scenarios. The Best Results Are in Bold

Methods	Backbone	Reference	mini-Imagenet 1-shot	mini-Imagenet 5-shot
MAML†[5]	ResNet-12	ICML'17	63.70±0.34	74.19±0.54
ALFA†[16]	ResNet-12	NIPS'20	63.66±0.44	78.73±0.32
GNN+FT[47]	ResNet-10	ICLR'20	66.32±0.80	81.98±0.55
MeTAL[17]	ResNet-12	ICCV'21	59.64±0.38	76.20±0.19
ALFA+MeTAL[17]	ResNet-12	ICCV'21	66.61±0.28	81.43±0.25
ProtoNet[11]	ResNet-12	NIPS'17	62.11±0.44	80.77±0.30
DeepEMD[48]	ResNet-12	CVPR'20	65.91±0.82	82.41±0.56
AMD[49]	ResNet-12	IJCAI'20	65.87±0.43	82.05±0.29
GLIML[6]	ResNet-12	T-CSVT'21	66.23±0.20	81.63±0.14
WS(SVM)[8]	ResNet-18	T-CSVT'21	67.53±0.24	80.90±0.21
Zhang et al.[7]	ResNet-12	T-CSVT'21	65.91±0.53	82.91±0.49
DAM[9]	ResNet-12	T-CSVT'22	60.39±0.21	73.84±0.16
FRN[32]	ResNet-12	CVPR'21	66.45±0.19	82.83±0.13
Cheng et al.[50]	ResNet-12	T-IP'22	66.88±0.20	82.22±0.14
Cao et al.[51]	ResNet-12	T-IP'22	63.47±0.20	81.27±0.15
Sum-min[52]	SF-12	CVPR'22	**68.32±0.62**	82.71±0.46
BLC-MAML(Ours)	ResNet-12	–	66.26±0.32	**82.92±0.30**

† Reproduced under our settings.

TABLE II

Few-Shot Classification Accuracy (%) With 95% Confidence Interval on CUB Under 5-Way 1/5-Shot Scenarios. The Best Results Are in Bold

Methods	Backbone	Reference	CUB 1-shot	CUB 5-shot
MAML†[5]	ResNet-12	ICML'17	71.27±0.56	80.14±0.39
MatchinigNet[4]	ResNet-18	NIPS'16	72.36±0.90	83.64±0.60
ProtoNet[11]	ResNet-18	NIPS'17	71.88±0.91	87.42±0.48
RelationNet[12]	ResNet-18	CVPR'18	67.59±1.02	82.75±0.58
DeepEMD [48]	ResNet-12	CVPR'20	75.65±0.82	88.69±0.50
IEPT [53]	Conv4-64	ICLR'21	69.97±0.49	84.33±0.33
WS(SVM)[8]	ResNet-18	T-CSVT'21	73.20±0.17	86.05±0.09
GLIML [6]	ResNet-12	T-CSVT'21	75.43±0.82	87.71±0.56
HGNN [54]	ResNet-12	T-CSVT'21	69.43±0.49	87.67±0.45
TEDC [10]	ResNet-12	T-CSVT'22	76.11±0.21	89.54±0.12
MAP-Net [55]	ResNet-12	T-IP'22	**82.45±0.23**	88.30±0.17
Cheng et al. [50]	WRN-28-10	T-IP'22	73.89±0.69	89.69±0.40
Sum-min [52]	SF-12	CVPR'22	79.60±0.80	90.48±0.44
BLC-MAML(Ours)	ResNet-12	–	**82.21±0.42**	**90.99±0.23**

† Reproduced under our settings.

B. Evaluation for Within-Domain Few-Shot Classfication

To verify the effectiveness of the proposed method, we first compare our algorithm with some state-of-the-art (SOTA) few-shot classification methods under the within-domain setting, where the training dataset and the test dataset are sampled from the same domain. We evaluate the performance on the following two standard benchmarks.

TABLE III

CROSS-DOMAIN FEW-SHOT CLASSIFICATION ACCURACY (%) ON CUB, PLACES, CARS AND PLANTAE WITH 95% CONFIDENCE INTERVAL UNDER 5-WAY 1/5 SHOT SCENARIOS. WE APPLY OUR METHOD TO MAML AND COMPARE WITH OTHER COMPETITIVE METHODS. THE BEST RESULTS ARE IN BOLD

Methods	Backbone	Reference	CUB		Places	
			1-shot	5-shot	1-shot	5-shot
MAML†♭[5]	ResNet-10	ICML'17	45.29±0.44	59.09±0.39	53.26±0.50	67.16±0.40
GNN♭[28]	ResNet-10	ICLR'18	45.69±0.68	62.25±0.65	53.10±0.80	70.84±0.65
GNN+FT♭[47]	ResNet-10	ICLR'20	47.47±0.75	66.98±0.68	55.77±0.79	73.94±0.67
GNN+FT♯[47]	ResNet-10	ICLR'20	48.24±0.75	70.37±0.68	54.81±0.81	74.48±0.70
Fine-tuning♭[56]	ResNet-10	ECCV'20	43.53±0.40	63.76±0.40	50.57±0.40	70.68±0.40
NSAE(CE+CE)♭[57]	ResNet-10	ICCV'21	–	69.96±0.80	–	71.86±0.72
ConFT♭[58]	ResNet-10	ICCV'21	45.57±0.76	70.53±0.75	49.97±0.86	72.09±0.68
ALFA+MeTal♭[17]	ResNet-12	ICCV'21	–	70.22±0.14	–	–
TPN+ATA♭[59]	ResNet-10	IJCAI'21	50.26±0.50	65.31±0.40	**57.03±0.50**	72.12±0.40
BLC-MAML(Ours)♭	ResNet-10	–	49.50±0.44	70.34±0.38	54.52±0.33	**73.98±0.38**
BLC-MAML(Ours)♭	ResNet-12	–	**51.00±0.30**	**70.57±0.40**	55.78±0.34	**75.17±0.39**
BLC-MAML(Ours)♯	ResNet-12	–	49.46±0.31	**71.42±0.40**	53.39±0.33	73.64±0.38

Methods	Backbone	Reference	Cars		Plantae	
			1-shot	5-shot	1-shot	5-shot
MAML†♭[5]	ResNet-10	ICML'17	35.47±0.39	46.97±0.41	36.76±0.38	50.08±0.38
GNN♭[28]	ResNet-10	ICLR'18	31.79±0.51	44.28±0.63	35.60±0.56	52.53±0.59
GNN+FT♭[47]	ResNet-10	ICLR'20	31.61±0.53	44.90±0.64	35.95±0.58	53.85±0.62
GNN+FT♯[47]	ResNet-10	ICLR'20	33.26±0.56	47.68±0.63	37.54±0.62	57.85±0.68
Fine-tuning♭[56]	ResNet-10	ECCV'20	35.12±0.40	51.21±0.40	38.77±0.40	56.45±0.40
NSAE(CE+CE)♭[57]	ResNet-10	ICCV'21	–	54.91±0.70	–	59.55±0.80
ConFT♭[58]	ResNet-10	ICCV'21	**39.11±0.70**	**61.53±0.70**	**43.09±0.80**	62.54±0.70
BOIL♭[15]	ResNet-12	ICLR'21	–	49.71±0.28	–	–
TPN+ATA♭[59]	ResNet-10	IJCAI'21	34.18±0.40	46.95±0.40	39.83±0.40	55.08±0.40
BLC-MAML(Ours)♭	ResNet-10	–	35.85±0.37	53.20±0.43	40.11±0.28	60.24±0.38
BLC-MAML(Ours)♭	ResNet-12	–	**37.21±0.26**	53.61±0.43	**41.87±0.29**	62.19±0.39
BLC-MAML(Ours)♯	ResNet-12	–	36.71±0.26	**56.03±0.43**	41.08±0.29	**63.42±0.39**

† Reproduced under our settings.
♭ Trained on single domain.
♯ Trained on multi domains.

1) Mini-ImageNet: The mini-ImageNet [4] consists of a subset of 100 classes in the ImageNet dataset [60], each class has 600 images. Following [33], we split these classes into 64-16-20 for training, validation, and testing, respectively.

2) CUB: Caltech-UCSD Birds-200-2011 [61] (CUB) dataset contains 11,788 images in 200 bird classes. We follow the split proposed by [62], where the ratio of train, validation, and test is set as 2:1:1, i.e., 100-50-50 classes for training-validation-testing.

5-way 5-shot and 5-way 1-shot classification results on mini-ImageNet and CUB can be found in Table I and Table II. These tables show that our method achieves significant performance gains on both datasets compared to the baseline method MAML. For example, on the mini-ImageNet dataset, our method improves the performance of the 5-shot classification from 74.19% to 82.92%, and the 1-shot classification from 63.70% to 66.26%. For another benchmark, CUB, the performance improvement is more significant. We achieve more than 10% improvement for both the 5-shot and 1-shot classification. These results show that our proposed bi-level optimization loss can effectively reduce the overfitting of the model to the limited support images and improve the generalization performance on the query samples.

Besides, the proposed BLC-MAML approach also achieves competitive results compared to the SOTA metric-based few-shot classification methods. In particular, compared to the SOTA methods TEDC [10] and MAP-Net [55], our method outperforms it by 1.4% and 2.7% on the 5-shot classification of the CUB dataset, respectively. Compared with Sum-min [52], our method also achieves comparable performance on both mini-ImageNet and CUB datasets.

C. Evaluation for Cross-Domain Few-Shot Classfication

In addition to experimental verification on the within-domain FSC, we also evaluate the performance of our method under the cross-domain setting on the following four benchmarks: *CUB*, *Cars*, *Places* and *Plantae*. Unlike the within-domain setting, cross-domain few-shot classification (CD-FSC) [62] introduces a more challenging problem in which tasks for training and testing are sampled from significantly different domains. This cross-domain setting shares a similar philosophy with the topic of domain generalization (DG) [63], [64], they both want to learn a model that can be generalized well to the unseen domains. However, in the DG task, although the source and target datasets come from

different domains, they share the same label space. While in the cross-domain few-shot setting, it not only emphasizes that source and target dataset should come from different domains, but also that their categories can not overlap. Therefore, the CD-FSC is a more challenging cross-domain task than DG. In the experiments, we present two cross-domain settings. One is to follow the setting of [62], which only uses mini-Imagenet as the base dataset to train the model and treats one of the above four datasets as the test domain. The other is to train our base learner on the multiple datasets and test only on the remaining one, which follows the protocol of [47]. Below we will detail the training and testing datasets selection in both settings.

1) Mini-ImageNet→CUB: This cross-domain task was first introduced by [62], which further explores the issue of domain difference in few-shot classification. We perform meta-training on the mini-ImageNet base classes and test on the test set of CUB.

2) Mini-ImageNet→Cars: Cars [65] is composed of 196 classes with 16,185 images in total. Following the split in [47], we sample 49 classes for both validation and test set.

3) Mini-ImageNet→Places: Places365-standard [66] (Places) contains 1,803,460 images with the image number per class varying from 3,068 to 5,000. Under this setting, all 64 classes of mini-ImageNet are used for meta-training, and the test set (91 classes) of Places is adopted for evaluation.

4) Mini-ImageNet→Plantae: The Plantae dataset is a subset of iNat2017 [67]. Like CUB, we randomly sample 200 classes from the Plantae super-class of iNat2017, 100 for training, 50 for validation, and 50 for testing. We only utilize the test data in this setting to validate the model performance.

5) Multi-Domains→CUB/Cars/Places/Plantae: Under this cross-domain setting, we use the leave-one-out setting to select the test domain. For example, for "**multi-domains→Cars**", we choose the cars dataset as the test set, mini-Imagenet, and the remaining datasets as training datasets.

A comparison of the cross-domain few-shot classification results on the above four datasets is shown in Table III. From the table, we can observe that our method can also effectively improve the performance of MAML on cross-domain tasks for both 1-shot and 5-shot classification. Especially in the 5-shot setting, our method achieves an average performance improvement of about 9% on the four datasets compared to MAML. Of course, we can also see that the performance improvement in the 1-shot setting is not as significant as that in the 5-shot setting, which may be the fact that in the 1-shot setting, there is only one labeled sample per class. The supervised contrastive loss has no available positive sample pairs, resulting in minor performance improvement. In addition, compared with the metric-based meta-learning methods GNN+FT [47] and TPN+ATA [59], our method also achieves superior performance on all four datasets. The above experimental results also demonstrate that our proposed bi-level optimization loss can significantly improve the generalization of the initial parameters on the cross-domain datasets.

It is well known that in the current cross-domain few-shot classification tasks, the methods that achieve the best performance are based on transfer learning, such as NSAE [57] and ConFT [58]. Nonetheless, our method can also achieve competitive results compared with these SOTA cross-domain few-shot classification methods. For example, compared with the Fine-tuning [56] method, our method outperforms it on both 5-shot and 1-shot classification on all four datasets. Even compared to NSAE and ConFT, our method is not inferior on some datasets, e.g., on the 5-shot classification of Places dataset, our method can outperform them by 2% and 1%, respectively.

At the same time, we can also notice in Table III that training with multiple datasets can further improve the model's performance. When training on a single dataset, we can only sample different few-shot classification tasks of the same domain when optimizing the inner loop. However, with multiple datasets, we can select FSC tasks in different domains to constrain the outer-loop optimization and increase the training data, further improving improve the model's generalization performance.

TABLE IV
EFFECT OF SUPERVISED CONTRASTIVE LOSS AND CROSS-TASK MERTIC LOSS

	Losses		miniImageNet	CUB
	$\mathcal{L}_{SupCon}$	$\mathcal{L}_{CTM}$		
a)			74.19±0.54	80.14±0.39
b)	✓		77.29±0.40	87.55±0.32
c)		✓	80.19±0.32	88.79±0.27
d)	✓	✓	**82.92±0.30**	**90.99±0.23**

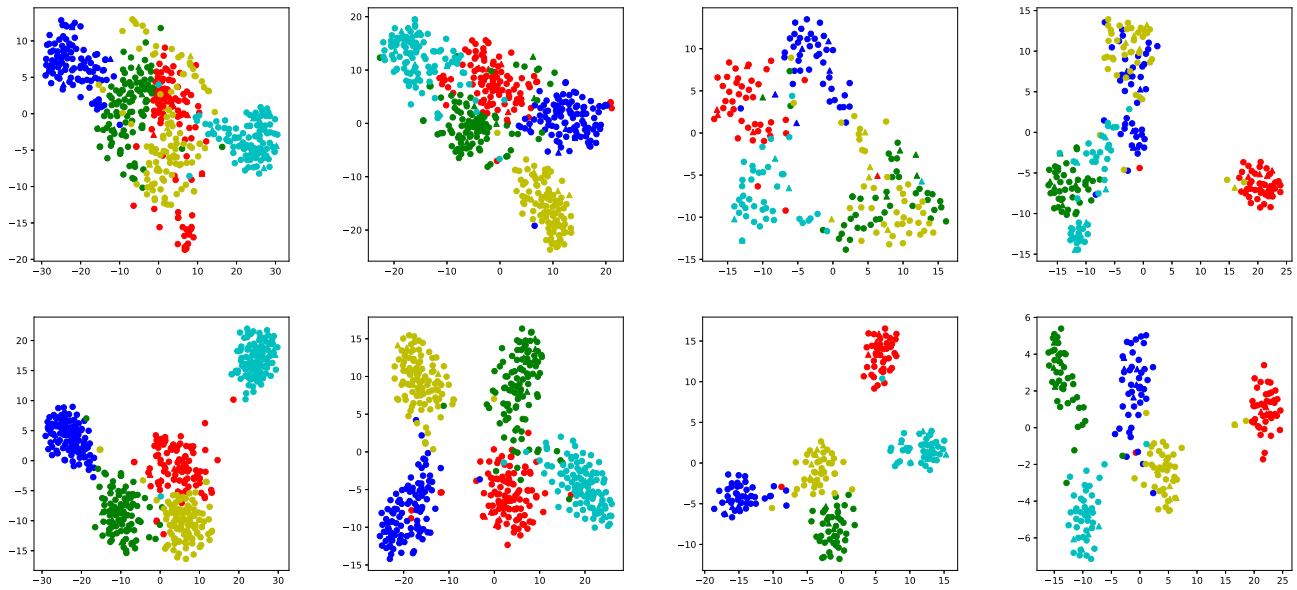
TABLE V
THE COMPARISON OF EUCLIDEAN AND COSINE DISTANCE IN THE 5-SHOT SCENARIOS ON THE MINI-IMAGENET

Distance function		miniImageNet
Euclidean distance	Cosine diatance	
✓		**82.92±0.30**
	✓	77.46±0.28

D. Ablation Study

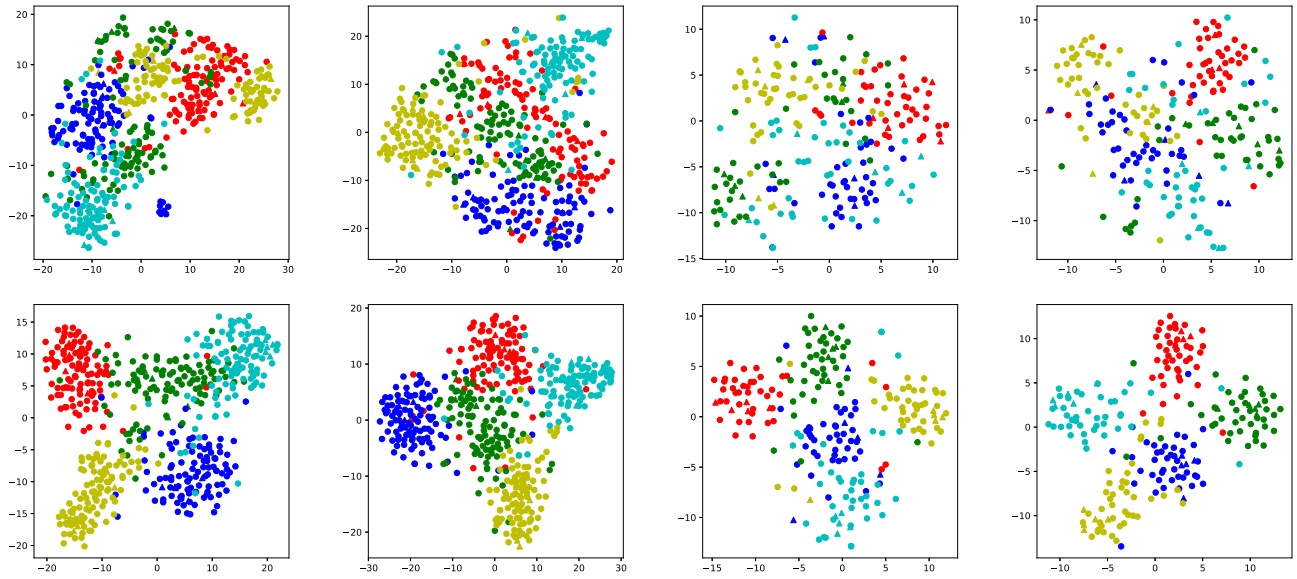
To verify the effectiveness of the proposed bi-level optimization loss: supervised contrastive (SupCon) loss, and the cross-task metric (CTM) loss, we conduct a series of ablation studies to analyze the impact of the above two losses. We construct the following MAML methods for comparison: a) **MAML**: f_θ is trained by pure MAML; b) **MAML+**$\mathcal{L}_{SupCon}$: f_θ is trained by MAML with inner-loop constraint (SupCon loss); c) **MAML+**$\mathcal{L}_{CTM}$: f_θ is trained by MAML with outer-loop constraint (CTM loss); d) **Bi-level constrained MAML (BLC-MAML)**: MAML with both inner-loop and outer-loop constraint.

All the ablation studies are conducted on the mini-ImageNet and the CUB datasets for 5-way 5-shot classification. The comparison results are shown in Table IV. As shown in Table IV, compare with **MAML**, the **MAML+**$\mathcal{L}_{SupCon}$ achieves higher performance on both datasets. This demonstrates that the proposed SupCon loss can effectively reduce the overfitting to the support images in the inner-loop optimization, and



(a) mini-ImageNet (b) CUB

Fig. 3. The t-SNE visualization of four within-domain 5-way 5-shot classification tasks, the first two columns are from mini-ImageNet and the last two from CUB. The first row illustrates the feature visualization results of the pure MAML, while the second row represents the results of our BLC-MAML. It can be clearly observed that our method not only increases the similarity within classes but also obtains a more significant inter-class margin, thus having better generalization.



(a) Places (b) CUB

Fig. 4. The t-SNE visualization of four cross-domain 5-way 5-shot classification tasks, the first two columns are from Places and the last two from CUB. The first and second rows illustrate the feature visualization results of MAML and our BLC-MAML, respectively. All the experiments are trained on the base classes of mini-ImageNet.

increase the generalization of the adapted model on the query samples. We can also see that only adding the CTM loss to the outer-loop optimization can also effectively improve the performance of the model. Finally, we can observe that the MAML with both inner-loop and outer-loop constraints achieves the best classification results. This suggests that the two losses are complementary, and considering both inner-loop and outer-loop constraints can obtain initial parameters with better generalization and adaptation.

In addition, we also compare the effects of using different distance measures in CTM loss. The comparison of Euclidean and cosine distance in the 5-shot scenarios on the mini-ImageNet is shown in Table V. From this table, we can observe that using Euclidean distance has better performance than cosine distance. Actually, for the distance measure in Cross-Task Metric (CTM) Loss, we follow the setting in ProtoNet [11] and use the Euclidean distance to calculate the similarity between the features of query image and the class

prototype. And in [11], the authors also analyse the effect of these two distance metric on ProtoNet. The experimental results show that using Euclidean distance improves performance substantially over cosine distance. The possible reason is that in the ProtoNet computing the class prototype as the mean of embedded support points is more naturally suited to Euclidean distances since cosine distance is not a Bregman divergence. Since the Euclidean distance is more suitable for calculating the distance between the mean prototype and the feature, we apply this distance measure for the proposed CTM loss.

E. T-SNE Visualization

In this part, we demonstrate the effectiveness of our method by visualizing the feature distribution of the test tasks. In Fig. 3, we show the t-SNE representation of four within-domain 5-way 5-shot classification tasks from mini-ImageNet and CUB. We can observe that our proposed method reduces the variation of features within each class while enabling features of different classes to be separated from each other. Besides, we also present the feature visualization results on cross-domain tasks drawn from Places and CUB, as shown in Fig. 4. The model is trained on the mini-ImageNet. From the figure, we can also see that the features extracted by our method on cross-domain tasks have good intra-class aggregation and inter-class separability, which indicates that the initialization learned by our method has good generalization performance.

V. Conclusion

This paper proposes a bi-level constrained MAML (BLC-MAML) method for few-shot classification. Specifically, in the inner loop of MAML, we introduce a supervised contrastive (SupCon) loss to constrain the adaptation procedure. The proposed SupCon loss can learn a feature space with large inter-class separability and low intra-class variance, thus improving the generalization ability of the adaptive model on query images. Besides, in the outer loop, we propose a cross-task metric loss to constrain the optimization process, which constrains the adaptive model to perform well on different FSC tasks. The CTM loss can force the adapted FSC model to pay more attention to category-relevant regions in the image and learn more discriminative and generalized feature representations, thereby improving the generalization ability of the adaptive model. By constraining the bi-level optimization process of MAML, the proposed method can effectively reduce the overfitting of deep models to the few labeled samples and obtain desirable generalization on the test data. Extensive experimental results on some FSC benchmarks show that our method can effectively improve the performance of MAML under both the within-domain and the cross-domain settings.

References

[1] W. Zhang, K. Ma, G. Zhai, and X. Yang, "Uncertainty-aware blind image quality assessment in the laboratory and wild," *IEEE Trans. Image Process.*, vol. 30, pp. 3474–3486, 2021.

[2] W. Zhang, C. Ma, Q. Wu, and X. Yang, "Language-guided navigation via cross-modal grounding and alternate adversarial learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 9, pp. 3469–3481, Sep. 2021.

[3] B. Li, W. Zhang, M. Tian, G. Zhai, and X. Wang, "Blindly assess quality of in-the-wild videos via quality-aware pre-training and motion perception," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 9, pp. 5944–5958, Sep. 2022.

[4] O. Vinyals et al., "Matching networks for one shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.

[5] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1126–1135.

[6] F. Hao, F. He, J. Cheng, and D. Tao, "Global-local interplay in semantic alignment for few-shot learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 7, pp. 4351–4363, Jul. 2022.

[7] L. Zhang, L. Zuo, Y. Du, and X. Zhen, "Learning to adapt with memory for probabilistic few-shot learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 11, pp. 4283–4292, Nov. 2021.

[8] Z. Chi, Z. Wang, M. Yang, D. Li, and W. Du, "Learning to capture the query distribution for few-shot learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 7, pp. 4163–4173, Jul. 2022.

[9] F. Zhou, L. Zhang, and W. Wei, "Meta-generating deep attentive metric for few-shot classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 10, pp. 6863–6873, Oct. 2022.

[10] J. Zhang, X. Zhang, and Z. Wang, "Task encoding with distribution calibration for few-shot learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 9, pp. 6240–6252, Sep. 2022.

[11] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Proc. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[12] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1199–1208.

[13] M. A. Jamal and G.-J. Qi, "Task agnostic meta-learning for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11719–11727.

[14] M. Yin, G. Tucker, M. Zhou, S. Levine, and C. Finn, "Meta-learning without memorization," in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–21.

[15] J. Oh, H. Yoo, C. Kim, and S.-Y. Yun, "BOIL: Towards representation change for few-shot learning," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–24.

[16] S. Baik, M. Choi, J. Choi, H. Kim, and K. M. Lee, "Meta-learning with adaptive hyperparameters," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 20755–20765.

[17] S. Baik, J. Choi, H. Kim, D. Cho, J. Min, and K. M. Lee, "Meta-learning with task-adaptive loss function for few-shot learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9465–9474.

[18] M. Agarwal, M. Yurochkin, and Y. Sun, "On sensitivity of meta-learning to support data," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 20447–20460.

[19] Y.-X. Wang, R. Girshick, M. Hebert, and B. Hariharan, "Low-shot learning from imaginary data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7278–7286.

[20] A. Khoreva, R. Benenson, E. Ilg, T. Brox, and B. Schiele, "Lucid data dreaming for video object segmentation," *Proc. Int. J. Comput. Vis.*, vol. 127, no. 9, pp. 1175–1197, 2019.

[21] H. Zhang, J. Zhang, and P. Koniusz, "Few-shot learning via saliency-guided hallucination of samples," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2770–2779.

[22] V. Sushko, J. Gall, and A. Khoreva, "One-shot GAN: Learning to generate samples from single images and videos," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 2596–2600.

[23] E. Schwartz et al., "Delta-encoder: An effective sample synthesis method for few-shot object recognition," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–11.

[24] S. Yang, L. Liu, and M. Xu, "Free lunch for few-shot learning: Distribution calibration," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–13.

[25] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[26] R. Zhang, T. Che, Z. Ghahramani, Y. Bengio, and Y. Song, "MetaGAN: An adversarial approach to few-shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–10.

[27] K. Li, Y. Zhang, K. Li, and Y. Fu, "Adversarial feature hallucination networks for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 13470–13479.

[28] V. G. Satorras and J. B. Estrach, "Few-shot learning with graph neural networks," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–13.

[29] H.-J. Ye, H. Hu, D.-C. Zhan, and F. Sha, "Few-shot learning via embedding adaptation with set-to-set functions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8808–8817.

[30] M. N. Rizve, S. Khan, F. S. Khan, and M. Shah, "Exploring complementary strengths of invariant and equivariant representations for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10836–10846.

[31] H. Li, D. Eigen, S. Dodge, M. Zeiler, and X. Wang, "Finding task-relevant features for few-shot learning by category traversal," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1–10.

[32] D. Wertheimer, L. Tang, and B. Hariharan, "Few-shot classification with feature map reconstruction networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 8012–8021.

[33] A. Srinivasan, A. Bharadwaj, M. Sathyan, and S. Natarajan, "Optimization of image embeddings for few shot learning," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–6.

[34] Z. Li, F. Zhou, F. Chen, and H. Li, "Meta-SGD: Learning to learn quickly for few-shot learning," 2017, *arXiv:1707.09835*.

[35] C. Finn, K. Xu, and S. Levine, "Probabilistic model-agnostic meta-learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 1–12.

[36] A. Antoniou, H. Edwards, and A. Storkey, "How to train your MAML," in *Proc. Int. Conf. Learn. Represent.*, 2019, pp. 1–11.

[37] A. Raghu, M. Raghu, S. Bengio, and O. Vinyals, "Rapid learning or feature reuse? Towards understanding the effectiveness of MAML," in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–21.

[38] M. Abbas, Q. Xiao, L. Chen, P.-Y. Chen, and T. Chen, "Sharp-MAML: Sharpness-aware model-agnostic meta learning," in *Proc. Int. Conf. Mach. Learn.*, 2022, pp. 1–23.

[39] Z. Wu, Y. Xiong, S. X. Yu, and D. Lin, "Unsupervised feature learning via non-parametric instance discrimination," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3733–3742.

[40] O. Henaff, "Data-efficient image recognition with contrastive predictive coding," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 4182–4192.

[41] A. Van Den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," 2018, *arXiv:1807.03748*.

[42] Y. Tian, D. Krishnan, and P. Isola, "Contrastive multiview coding," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 776–794.

[43] B. Xu, X. Shu, and Y. Song, "X-invariant contrastive augmentation and representation learning for semi-supervised skeleton-based action recognition," *IEEE Trans. Image Process.*, vol. 31, pp. 3852–3867, 2022.

[44] R. Yan, P. Huang, X. Shu, J. Zhang, Y. Pan, and J. Tang, "Look less think more: Rethinking compositional action recognition," in *Proc. 30th ACM Int. Conf. Multimedia*, Oct. 2022, pp. 3666–3675.

[45] P. Khosla et al., "Supervised contrastive learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 18661–18673.

[46] T. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, "Meta-learning in neural networks: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 5149–5169, Sep. 2022.

[47] H.-Y. Tseng, H.-Y. Lee, J.-B. Huang, and M.-H. Yang, "Cross-domain few-shot classification via learned feature-wise transformation," in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–18.

[48] C. Zhang, Y. Cai, G. Lin, and C. Shen, "DeepEMD: Few-shot image classification with differentiable earth Mover's distance and structured classifiers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12203–12213.

[49] W. Li, L. Wang, J. Huo, Y. Shi, Y. Gao, and J. Luo, "Asymmetric distribution measure for few-shot learning," in *Proc. 29th Int. Joint Conf. Artif. Intell.*, Jul. 2020, pp. 2957–2963.

[50] J. Cheng, F. Hao, L. Liu, and D. Tao, "Imposing semantic consistency of local descriptors for few-shot learning," *IEEE Trans. Image Process.*, vol. 31, pp. 1587–1600, 2022.

[51] C. Cao and Y. Zhang, "Learning to compare relation: Semantic alignment for few-shot learning," *IEEE Trans. Image Process.*, vol. 31, pp. 1462–1474, 2022.

[52] A. Afrasiyabi, H. Larochelle, J.-F. Lalonde, and C. Gagne, "Matching feature sets for few-shot image classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 9014–9024.

[53] M. Zhang, J. Zhang, Z. Lu, T. Xiang, M. Ding, and S. Huang, "IEPT: Instance-level and episode-level pretext tasks for few-shot learning," in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–16.

[54] C. Chen, K. Li, W. Wei, J. T. Zhou, and Z. Zeng, "Hierarchical graph neural networks for few-shot learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 1, pp. 240–252, Feb. 2021.

[55] Z. Ji, Z. Hou, X. Liu, Y. Pang, and J. Han, "Information symmetry matters: A modal-alternating propagation network for few-shot learning," *IEEE Trans. Image Process.*, vol. 31, pp. 1520–1531, 2022.

[56] Y. Guo et al., "A broader study of cross-domain few-shot learning," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 124–141.

[57] H. Liang, Q. Zhang, P. Dai, and J. Lu, "Boosting the generalization capability in cross-domain few-shot learning via noise-enhanced supervised autoencoder," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9424–9434.

[58] R. Das, Y.-X. Wang, and J. M. F. Moura, "On the importance of distractors for few-shot classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9030–9040.

[59] H. Wang and Z.-H. Deng, "Cross-domain few-shot classification via adversarial task augmentation," in *Proc. 13th Int. Joint Conf. Artif. Intell.*, Aug. 2021, pp. 1–7.

[60] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[61] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, "The Caltech-UCSD birds-200–2011 dataset," California Inst. Technol., Pasadena, CA, USA, Tech. Rep. 2010-001, 2011.

[62] W.-Y. Chen, Y.-C. Liu, Z. Kira, Y.-C. F. Wang, and J.-B. Huang, "A closer look at few-shot classification," in *Proc. Int. Conf. Learn. Represent.*, 2019, pp. 1–17.

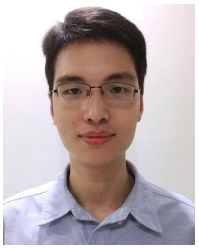
[63] Y. Li, Y. Yang, W. Zhou, and T. Hospedales, "Feature-critic networks for heterogeneous domain generalization," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 3915–3924.

[64] J. Wang et al., "Generalizing to unseen domains: A survey on domain generalization," *IEEE Trans. Knowl. Data Eng.*, early access, May 26, 2022, doi: 10.1109/TKDE.2022.3178128.

[65] J. Krause, M. Stark, J. Deng, and L. Fei-Fei, "3D object representations for fine-grained categorization," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2013, pp. 554–561.

[66] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1452–1464, Jul. 2018.

[67] G. Van Horn et al., "The iNaturalist species classification and detection dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8769–8778.



Yuanjie Shao received the B.S. and M.S. degrees from the College of Mechanical and Electronic Information, China University of Geosciences, Wuhan, China, in 2010 and 2013, respectively, and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, in 2018. He is currently a Lecturer with the School of Electronic Information and Communications, Huazhong University of Science and Technology. His research interests include pattern recognition and computer vision.

Wenxiao Wu received the B.E. degree from the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China, in 2021, where he is currently pursuing the M.E. degree. His research interests include deep learning and computer vision.

Xinge You (Senior Member, IEEE) received the B.S. and M.S. degrees in mathematics from Hubei University, Wuhan, China, in 1990 and 2000, respectively, and the Ph.D. degree from the Department of Computer Science, Hong Kong Baptist University, Hong Kong, in 2004. He is currently a Professor with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan. His research results have expounded in more than 60 publications at prestigious journals and prominent conferences, such as IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, NeurIPS, CVPR, ICCV, and ECCV. His current research interests include image processing, wavelet analysis and its applications, pattern recognition, machine earning, and computer vision. He served/serves as an Associate Editor for the IEEE TRANSACTIONS ON CYBERNETICS and IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS.

Changxin Gao (Member, IEEE) received the Ph.D. degree in pattern recognition and intelligent systems from the Huazhong University of Science and Technology in 2010. He is currently an Associate Professor with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology. His research interests include pattern recognition and surveillance video analysis.

Nong Sang (Member, IEEE) received the B.E. degree in computer science and engineering, the M.S. degree in pattern recognition and intelligent control, and the Ph.D. degree in pattern recognition and intelligent systems from the Huazhong University of Science and Technology in 1990, 1993, and 2000, respectively. He is currently a Professor with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology. His research interests include object detection, object tracking, image/video segmentation, and analysis of surveillance videos.