IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 30 January 2024, accepted 12 April 2024, date of publication 16 April 2024, date of current version 1 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3390034

RESEARCH ARTICLE

DBDC-SSL: Deep Brownian Distance Covariance With Self-Supervised Learning for Few-Shot Image Classification

WEI HAN LIU, KIAN MING LIM, (Senior Member, IEEE), THIAN SONG ONG, (Senior Member, IEEE), AND CHIN POO LEE, (Senior Member, IEEE)
Faculty of Information Science and Technology, Multimedia University, Melaka 75450, Malaysia

Corresponding author: Kian Ming Lim (kmlim@mmu.edu.my)

This project is funded by the Malaysian Ministry of Higher Education under the Fundamental Research Grant Scheme (FRGS/1/2021/ICT02/MMU/02/4) which is awarded to the Multimedia University.

ABSTRACT Few-shot image classification remains a persistent challenge due to the intrinsic difficulty faced by visual recognition models in achieving generalization with limited training data. Existing methods primarily focus on exploiting marginal distributions and overlook the disparity between the product of marginals and the joint characteristic functions. This can lead to less robust feature representations. In this paper, we introduce DBDC-SSL, a method that aims to improve few-shot visual recognition models by learning a feature extractor that produces image representations that are more robust. To improve the robustness of the model, we integrate DeepBDC (DBDC) during the training process to learn better feature embeddings by effectively computing the disparity between product of the marginals and joint characteristic functions of the features. To reduce overfitting and improve the generalization of the model, we utilize an auxiliary rotation loss for self-supervised learning (SSL) in the training of the feature extractor. The auxiliary rotation loss is derived from a pretext task, where input images undergo rotation by predefined angles, and the model classifies the rotation angle based on the features it generates. Experimental results demonstrate that DBDC-SSL is able to outperform current state-of-the-art methods on 4 common few-shot image classification benchmark, which are miniImageNet, tieredImageNet, CUB and CIFAR-FS. For 5-way 1-shot and 5-way 5-shot tasks respectively, the proposed DBDC-SSL achieved the accuracy of 68.64±0.43 and 86.02±0.28 on miniImageNet, 73.88±0.48 and 89.03±0.29 on tieredImageNet, 84.67±0.39 and 94.76±0.16 on CUB, and 75.60±0.44 and 88.49±0.31 on CIFAR-FS.

INDEX TERMS Few-shot learning, Brownian distance covariance, metric learning, self-supervised learning, regularization.

I. INTRODUCTION

In recent years, notable progress has been achieved in deep learning within the realm of standard computer vision tasks, particularly in the domain of object recognition. Despite these advancements, a persistent challenge lies in maintaining high accuracy under conditions of limited training data. This has motivated researchers to delve into the domain of few-shot learning. The primary objective of few-shot learning is to identify novel objects using only a small number of training examples per class. This objective closely mirrors real-world scenarios where acquiring labeled data could to be difficult and costly. Similar to human intelligence which exhibits the ability to learn from very few examples, the development of deep learning models capable of learning efficiently from a limited set of training samples across different classes is important in advancing artificial intelligence on a broader scale.

Common approaches for addressing the few-shot learning challenge involve employing the ''learning to learn''

The associate editor coordinating the review of this manuscript and approving it for publication was Hossein Rahmani.

mechanism, commonly known as meta-learning. In meta-learning, the model undergoes training on a series of distinct few-shot classification tasks and is subsequently assessed on test data to acquire parameters that facilitate generalization to new tasks [1], [2], [3]. Recently, metric-based methods have garnered increased attention from researchers due to their better performance compared to other few-shot learning techniques. Typically, many metric-based methods utilize a pre-trained feature extractor on base classes for feature extraction. Subsequently, a classifier is trained based on a chosen metric to compute differences between feature embeddings of test data for classification. Notable examples of metric-based methods include matching networks [4], which employ cosine distance for comparing query features with support features and incorporate a memory mechanism; prototypical networks [5], which employ Euclidean distance to compare query features with the embedding prototype of support features from each class; and relation networks [6], which examine query features with the embedding prototype of support features from each class using a relation module whose parameters are fine-tuned.

Despite many advancements in few-shot learning, researchers continue to explore ways to improve the effectiveness of few-shot learning methods. In situations where the available training data is limited, the training and fine-tuning process of the model becomes unstable and inefficient, primarily due to overfitting. Existing few-shot learning models primarily focus on exploiting marginal distributions of features. However, the disparity between the product of marginals and the joint characteristic functions is often overlooked. This oversight can lead to less robust feature representations because it ignores the deeper statistical relationships between features that arise from their joint distribution. In view of this, we introduce a new few-shot learning framework DBDC-SSL that incorporates deep Brownian Distance Covariance (DBDC) with a self-supervised learning (SSL) loss. Given the support and query images, DBDC effectively measures the discrepancy between the joint distribution of the features based on the images and product of the marginals. This in turn helps the model to learn robust image representations which subsequently improves the performance of the model. On the other hand, we employ a self-supervised learning loss based on an pretext task to classify the degree of rotation of the image when given the embedded features. The aim of this self-supervised learning loss is to reduce overfitting and improve the generalization of the model.

The main contributions of this paper are summarized as follows:

1) To improve the robustness of the model, we incorporate deep Brownian Distance Covariance (DBDC) that effectively measures the discrepancy between the joint distribution of the feature representations and product of the marginals.
2) In addition, to reduce overfitting and improve generalization, a self-supervised learning (SSL) loss based on predicting the rotation of given images is utilized. We then train a new logistic regression classifier to make predictions for the few-shot tasks.
3) Through extensive experiments, we show that the proposed DBDC-SSL is able to achieve higher average accuracy on few-shot recognition datasets.

II. RELATED WORK

Traditionally, few-shot learning predominantly follows an inductive approach. This involves initially training the model using a designated set of training data and subsequently assessing its performance on distinct test data, all without resorting to additional unlabeled data for refinement. We can broadly classify the prevailing methods in few-shot learning into three categories: gradient-based methods, hallucination-based methods, and metric-based methods.

A. GRADIENT-BASED FEW-SHOT CLASSIFICATION

Gradient-based approaches seek to refine the model using a limited set of data samples to address challenges in few-shot learning [1], [2], [3], [7], [8], [9], [10], [11], [12], [13]. These methods fall into two main categories: initialization-based methods [2], [3], [7], [9], [10] and optimization-based methods [1], [8], [11], [12], [13]. Initialization-based techniques aim to acquire an effective starting point for the model's parameters across diverse tasks, enabling proficient performance in new tasks with minimal data samples and parameter updates. For instance, Model-Agnostic Meta-Learning (MAML) [2] strives to optimally initialize parameters based on the loss from a set of tasks in order to improve the fine-tuning process for novel tasks.

Conversely, optimization-based methods aim to acquire an efficient optimizer, facilitating the model's fast adaptation to new tasks with limited data samples and parameter adjustments. These methods often replace the conventional optimizer with an alternative, such as a Long Short-Term Memory-based meta-learner [1] or a mechanism utilizing external memory for parameter updates [8]. Notably, the GCLR-SVM framework [11] was introduced as an end-to-end solution to embed representations into a latent space, augmenting representations through latent code reconstruction with variational information. Furthermore, the A-MET paradigm proposed in [13] adaptively eliminates undesired and incomplete features acquired during pre-training, addressing the objective misalignment between transfer learning and meta-learning. The authors also introduced a GSCM metric, representing samples by jointly re-embedding sample features to yield more consistent prediction results. Additionally, Adaptive Learning Knowledge Networks (ALKN) [14] presented an adaptive learning knowledge module storing learned knowledge memories, coupled with a decoder utilizing query representations and data from the adaptive learning knowledge module for classification.

B. HALLUCINATION-BASED FEW-SHOT CLASSIFICATION

In the realm of few-shot learning, a notable challenge lies in the scarcity of data. Recent efforts to tackle this issue involve the introduction of hallucination-based techniques. These methodologies, detailed in various studies such as [15], [16], [17], and [18], aim to mitigate data limitations by generating additional training samples. Broadly categorized, these techniques fall into two types: the first type transfers appearance variations from the original data categories, as exemplified in [15] and [17], while the second type leverages generative adversarial networks (GANs) to transfer stylistic features, as demonstrated in [16]. In a unique approach, [18] suggests transforming base classes into Gaussian form using power transformation for Maximum A Posteriori (MAP) estimation. Subsequently, the Gaussian mean of novel classes is estimated under the Gaussian prior based on a limited set of samples. This results in each novel class being represented by a distinct Gaussian distribution, from which ample trainable features can be sampled, ultimately enhancing predictive capabilities. It is noteworthy that these techniques are frequently employed in conjunction with other few-shot learning methods, leading to more complexity.

C. METRIC-BASED FEW-SHOT CLASSIFICATION

Recently, metric-based approaches have garnered increasing attention in the literature as highly effective techniques for few-shot learning. These methods excel in discerning between objects with limited examples by exploiting information about the similarity within the available data. Typically, a Convolutional Neural Network (CNN)-based feature extractor is initially trained on a larger dataset. This extractor is then utilized to capture features from the limited data of novel classes. Subsequently, a metric-based classifier is trained to recognize objects based on these features. The employed metric can take various forms, including cosine similarity [4], Euclidean distance [5], a custom convolutional neural network-based distance module [6], [19], [20], or a graph neural network [21], [22], [23].

For instance, the Matching Network [4] employed an end-to-end nearest neighbor classifier with weights and with an attention mechanism based on cosine similarity between two feature embeddings. The Prototypical Network [5] computed the mean of extracted features from support data and compared the Euclidean distance between the class mean and query data for classification. The Relation Network [6] concatenated feature maps of the training set and passed them through a relation module, optimized through mean square error (MSE) to regress the score value to the true label. Task Dependent Adaptive Metric (TADAM) [19] introduced a dynamic task-conditioning module to enhance the feature extractor, incorporating metric scaling and auxiliary task co-training to improve few-shot learning. DeepEMD [20] recently adopted Earth Mover's Distance (EMD) to determine the minimum matching cost between feature vectors of support and query images for few-shot classification. Wang et al. [24] proposed a multi-scale decision network (MSDN) utilizing feature fusion and weighting to enhance the fitting ability of the Relation Network during feature concatenation. Reference [25] proposed to use Brownian Distance Covariance that measures the discrepancy between the joint distribution of the embedded features of the query and support images and the product of the marginals. As a modular layer that can be used in many networks, it can effectively capture the dependency between the two sets of features, which is often neglected by existing methods that only exploit marginal distributions.

On a different note, [21] formulated a Graph Neural Network (GNN) framework for few-shot learning, where extracted features serve as input to a GNN with various layers of nodes and graph convolutional layers. Reference [22] enhanced [21] by introducing the Edge Graph Neural Network (EGNN), predicting edge labels on the graph based on similarity within clusters and dissimilarity between different clusters. Additionally, Distribution Propagation Graph Network (DPGN) [23] introduced a dual complete network comprising a point graph and a distribution graph, with label information propagated from labeled to unlabeled data through multiple updates.

D. TRANSDUCTIVE FEW-SHOT CLASSIFICATION

Transductive few-shot learning, a subset within the metric-based few-shot learning paradigm, has demonstrated improvements compared to other methods like inductive metric-based approaches, gradient-based techniques, and hallucination-based methods, as indicated in recent investigations [26], [27], [28], [29], [30], [31], [32], [33], [34]. In inductive few-shot learning scenarios, models are initially trained on observed and labeled training data and subsequently utilized for predictions on unobserved and unlabeled test data. Conversely, transductive few-shot learning models are trained using both observed and labeled training data and observed but unlabeled test data, and are then employed for classifying the test data.

Transductive Propagation Network (TPN) [28] explicitly addressed transductive inference in few-shot learning settings for the first time. TPN introduced a framework for learning to propagate labels between data instances for unseen classes through episodic meta-learning. In another approach [29], a straightforward method was proposed that minimizes the entropy of model predictions on unlabeled query samples, surprisingly achieving competitive performance compared to more intricate meta-learning methods. A different study [30] suggested using pseudo-labeling and feature shifting in a prototypical network based on cosine similarity. PT-MAP [31] applied Power Transform (PT) to the data to better align it with typical distribution assumptions and utilized Maximum A Posteriori (MAP) for computing class centers during classification. Another investigation [33] derived a regularized manifold by leveraging unlabeled query data and

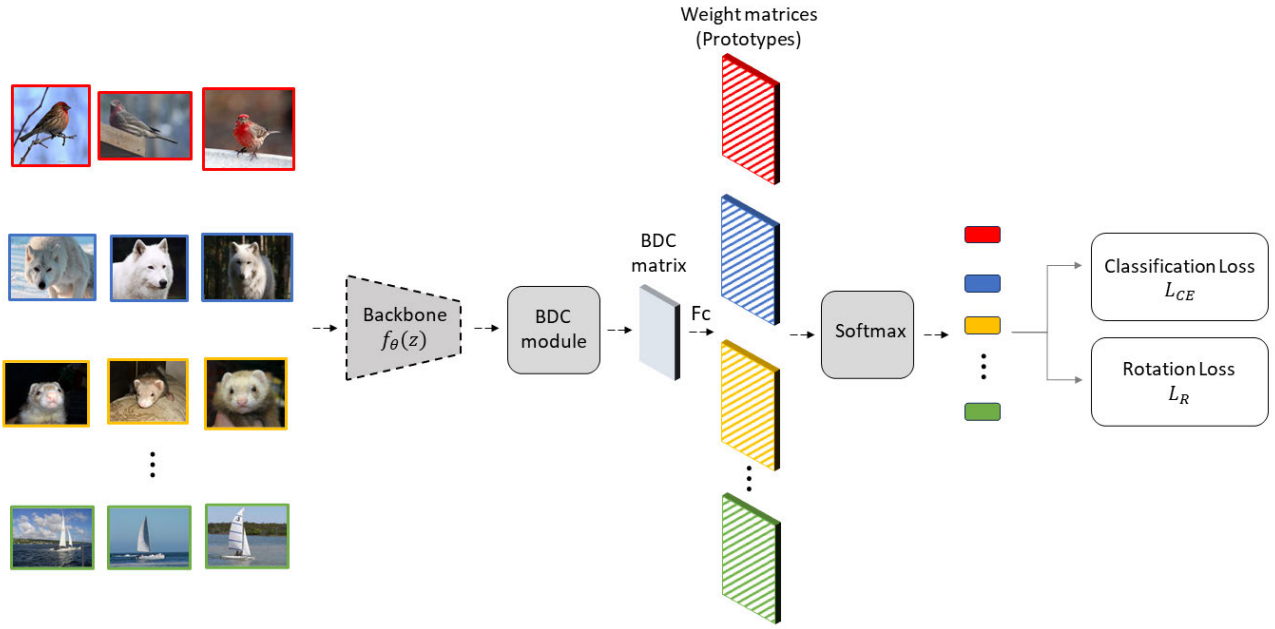

FIGURE 1. The proposed few-shot recognition method DBDC-SSL. Given a set of images, a backbone is used for extracting the feature representations. The feature representations are then passed to BDC module to produce BDC matrix based on the dependency between the features of a query image and the features of support images. The weight matrices are subsequently produced and are used as class prototypes for prediction. The parameters of the model are then updated based on the classification loss L_C and self-supervised learning loss based on rotation L_R.

employed non-parametric embedding propagation to smooth decision boundaries by generating a set of feature interpolations based on a similarity graph. In a subsequent work, [32] proposed minimizing a quadratic binary-assignment function, which achieved competitive performance. This function includes a unary term assigning query samples to the nearest class prototype and a pairwise Laplacian term, encouraging consistent label assignments among nearby query samples. Additionally, [27] introduced a method maximizing mutual information between query features and predictions of a few-shot task while adhering to supervision constraints from the support set. In another study [34], a transductive clustering procedure based on a conditional neural-adaptive feature extractor was developed to yield improved class means for few-shot classification.

III. METHODOLOGY

In this section, the common few-shot setting is first introduced. After that, the details of DBDC-SSL are described. A summary of DBDC-SSL is shown in Figure 1.

A. FEW-SHOT SETTING

We examine few-shot learning within the framework of a labeled training set denoted as $D_{base} = \{\mathbf{z}_j, \mathbf{y}_j\}_{j=1}^{N_{base}}$, where each sample is represented by its raw images $\mathbf{z}_j$ and its corresponding one-hot encoded label $\mathbf{y}_j$. The set of classes for this base dataset is represented by Y_{base}. In few-shot scenarios, there is a distinct test dataset $X_{test} = \{\mathbf{z}_j, \mathbf{y}_j\}_{j=1}^{N_{test}}$ with a set of classes Y_{test}, ensuring $Y_{base} \cap Y_{test} = \emptyset$. In the context of few-shot classification tasks, the labeled data samples are randomly sampled from the test dataset. Each task involves N distinct classes, with K_{sup} labeled samples from each class, resulting in an N-way K_{sup}-shot task. The set of these labeled samples is denoted as the support set *sup*, with the size $|s| = K_{sup} \cdot N$. In addition, each task has an unlabeled query set *que* comprising K_{que} examples from each of the N classes, resulting in a query set size $|que| = K_{que} \cdot N$, typically consisting of unseen examples.

After training the models on the base classes, few-shot learning methods employ the labeled support sets to adapt to new tasks, conducting evaluations on the unlabeled query sets. In the mean time, the raw images from the support set *sup* and query set *que* are denoted as Z_{sup} and Z_{que} respectively, with their actual labels Y_{sup} and Y_{que}. The predicted labels of the support set are represented as $\hat{Y_{sup}}$, while the predicted labels of the query set are denoted as $\hat{Y_{que}}$.

B. DEEP BROWNIAN DISTANCE COVARIANCE

In this work, Brownian Distance Covariance (BDC) is used to measure the dependency between the features of a query image and the features of support images, which is often neglected by existing methods that only exploit marginal distributions [25]. The foundation of the Brownian Distance Covariance (BDC) theory is initially laid out in [35] and [36]. Given two random vectors, it is a method that measures

the dependency between them By considering their joint characteristic function.

Consider random vectors X and Y with dimensions p and q in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively. Let $f_{XY}(\mathbf{x}, \mathbf{y})$ represent their joint probability density function. The following equation defines the joint characteristic function of X and Y:

$$\phi_{XY}(\mathbf{t}, \mathbf{s}) = \int_{\mathbb{R}^p} \int_{\mathbb{R}^q} \exp(i(\mathbf{t}^T\mathbf{x} + \mathbf{s}^T\mathbf{y}))f_{XY}(\mathbf{x}, \mathbf{y})d\mathbf{x}d\mathbf{y} \quad (1)$$

where i is the imaginary unit, and $\mathbf{t}$ and $\mathbf{s}$ act as the parameter vector for the characteristic function associated with X and Y respectively.

Assuming finite first moments for random vectors X and Y, the BDC metric is expressed as follows:

$$\rho(X, Y) = \int_{\mathbb{R}^p} \int_{\mathbb{R}^q} \frac{|\phi_{XY}(\mathbf{t}, \mathbf{s}) - \phi_X(\mathbf{t})\phi_Y(\mathbf{s})|^2}{c_p c_q \|\mathbf{t}\|^{1+p}\|\mathbf{s}\|^{1+q}} d\mathbf{t}d\mathbf{s} \quad (2)$$

where $\|\cdot\|$ represents Euclidean norm, $c_p = \pi^{(1+p)/2}/\Gamma((1+p)/2)$ and Γ denote the complete gamma function.

For a set of m independent and identically distributed (i.i.d.) observations $(\mathbf{z}_1, \mathbf{y}_1), \ldots, (\mathbf{z}_m, \mathbf{y}_m)$, one intuitive method is to establish the BDC metric by utilizing the observed characteristic functions:

$$\phi_{XY}(\mathbf{t}, \mathbf{s}) = \frac{1}{m}\sum_{k=1}^{m} \exp(i(\mathbf{t}^T\mathbf{x}_k + \mathbf{s}^T\mathbf{y}_k)) \quad (3)$$

Let $\widehat{\mathbf{A}} = (\widehat{a}_{kl}) \in \mathbb{R}^{m\times m}$ denote the matrix of Euclidean distances calculated between pairs of observations in the set X, with $\widehat{a}_{kl} = |\mathbf{x}_k - \mathbf{x}_l|$. Similarly, we establish the matrix $\widehat{\mathbf{B}} = (\widehat{b}_{kl}) \in \mathbb{R}^{m\times m}$ representing Euclidean distances, with $\widehat{b}_{kl} = |\mathbf{y}_k - \mathbf{y}_l|$. Following this, the BDC metric can be expressed as:

$$\rho(X, Y) = \mathrm{tr}\big(\mathbf{A}^T\mathbf{B}\big) \quad (4)$$

Note that tr$(\cdot)$ represents the trace of a matrix, T represents matrix transpose, and $\mathbf{A} = (a_{kl})$ as the *BDC matrix*. In this context, a_{kl} is defined as $\widehat{a}_{kl} - \frac{1}{m}\sum_{l=1}^{m}\widehat{a}_{kl} - \frac{1}{m}\sum_{k=1}^{m}\widehat{a}_{kl} - \frac{1}{m^2}\sum_{k=1}^{m}\sum_{l=1}^{m}\widehat{a}_{kl}$, where $\frac{1}{m}\sum_{l=1}^{m}\widehat{a}_{kl}$ represents the means of the k-th row, $\frac{1}{m}\sum_{k=1}^{m}\widehat{a}_{kl}$ represents the means of the l-th column, and $\frac{1}{m^2}\sum_{k=1}^{m}\sum_{l=1}^{m}\widehat{a}_{kl}$ represents the means of all entries of the matrix $\widehat{\mathbf{A}}$. The computation of the matrix $\mathbf{B}$ mirrors that of $\widehat{\mathbf{B}}$. Because of the symmetry inherent in a BDC matrix, $\rho(X, Y)$ can also be formulated as the inner product of two BDC vectors, designated as $\mathbf{a}$ and $\mathbf{b}$:

$$\rho(X, Y) = \langle \mathbf{a}, \mathbf{b} \rangle = \mathbf{a}^T\mathbf{b} \quad (5)$$

In this context, $\mathbf{a}$ (and similarly, $\mathbf{b}$) is obtained by extracting the upper triangular portion of $\mathbf{A}$ (and $\mathbf{B}$, respectively) and subsequently undergoing vectorization.

It becomes evident that the BDC metric is disentangled based on Eq. (4) and Eq. (5). When provided with a set of input images, the matrix of BDC for each set of features can be independently calculated. Utilizing the given set of features involves employing a two-layer module designed for dimension reduction and BDC matrix computation. To achieve this, a convolutional layer with a 1×1 filter for reducing dimensions is incorporated directly following the final convolutional layer of the backbone.

For feature extraction, a network parameterized by θ is used to extract features from a color image $\mathbf{z}$. The feature embedding from the network based on the image is represented as a tensor of dimensions $h \times w \times d$. Here, h and w correspond to spatial height and width, respectively, while d denotes the number of channels. This tensor is reshaped into a matrix $\mathbf{X} \in \mathbb{R}^{hw\times d}$, and each column $\boldsymbol{\chi}_k \in \mathbb{R}^{hw}$ or each row (upon transposition) $\mathbf{x}_j \in \mathbb{R}^d$ can be considered as an observation of the random vector X.

Next, we consecutively calculate the matrix of squared Euclidean distances denoted as $\widetilde{\mathbf{A}} = (\tilde{a}_{kl})$, with $\tilde{a}_{kl}$ signifying the squared Euclidean distance of $\mathbf{X}$'s k-th column and l-th column. Following this, we derive the Euclidean distance matrix $\widehat{\mathbf{A}} = (\sqrt{\tilde{a}_{kl}})$. Finally, we subtract the mean of row, the mean of column, as well as the mean of the elements based on $\widehat{\mathbf{A}}$ obtain the BDC matrix $\mathbf{A}$:

$$\begin{aligned}
\widetilde{\mathbf{A}} &= 2\big(\mathbf{1}(\mathbf{X}^T\mathbf{X} \circ \mathbf{I})\big)_{\mathrm{sym}} - 2\mathbf{X}^T\mathbf{X} \\
\widehat{\mathbf{A}} &= \big(\sqrt{\tilde{a}_{kl}}\big) \\
\mathbf{A} &= \widehat{\mathbf{A}} - \frac{2}{d}\big(\mathbf{1}\widehat{\mathbf{A}}\big)_{\mathrm{sym}} + \frac{1}{d^2}\mathbf{1}\widehat{\mathbf{A}}\mathbf{1}
\end{aligned} \quad (6)$$

In this context, $\mathbf{1} \in \mathbb{R}^{d\times d}$ represents a matrix where each element is assigned a value of 1, $\mathbf{I}$ denotes the identity matrix, and the symbol $\circ$ signifies the Hadamard product. The notation $(\mathbf{U})\mathrm{sym} = \frac{1}{2}(\mathbf{U} + \mathbf{U}^T)$ is employed to indicate the symmetric component of the matrix $\mathbf{U}$. Consequently, it is evident that DeepBDC serves as a parameter-free spatial pooling layer. Its high modularity renders it adaptable to various network architectures in the context of few-shot classification. It is important to note that we use the notation $\mathbf{A}_\theta(\mathbf{z})$ to express that the we derive the BDC matrix based on the parameterized network f_θ and takes an input image $\mathbf{z}$.

C. SELF-SUPERVISED LEARNING WITH ROTATION LOSS

In this work, we utilize one pretext task for self-supervised learning. The chosen pretext task is classifying the rotation angle. First, the input image is rotated based on a set of angles. The auxiliary objective of the model involves categorizing the rotational degree undergone by the image. We employ a 4-way linear classifier, denoted as c_{W_r}, applied to the feature representation $\mathbf{A}_\theta(\mathbf{z}^r)$, where z^r represents the image z rotated by r degrees, and $r \in C_R = \{0°, 90°, 180°, 270°\}$. The goal of this linear classifier is to predict 1 class among the 4 classes within C_R. The self-supervision loss is defined as the following:

$$L_R = \frac{1}{|C_R|} * \sum_{z\in\mathcal{D}_{base}} \sum_{r\in C_R} L(c_{W_r}(\mathbf{A}_\theta(\mathbf{z})), r) \quad (7)$$

where $|C_R|$ denotes the cardinality of C_R and L represents the cross-entropy loss.

With this self-supervised learning protocol based on rotation loss, the training improves the learned backbone

TABLE 1. Average accuracy on miniImageNet and tieredImageNet. † denotes our implementation using their publicly released code.

		miniImageNet		tieredImageNet	
Methods	**Network**	**1-shot**	**5-shot**	**1-shot**	**5-shot**
TADAM [19]	ResNet-12	58.50±0.30	76.70±0.38	–	–
Baseline++ [37] †	ResNet-12	60.56±0.45	77.40±0.34	–	–
ProtoNet [5] †	ResNet-12	62.11±0.44	80.77±0.30	68.31±0.51	83.85±0.36
MetaOptNet [38]	ResNet-12	62.64±0.44	78.63±0.46	65.99±0.72	81.56±0.63
SimpleShot [39]	ResNet-18	62.85±0.20	80.02±0.14	69.09±0.22	84.58±0.16
Meta-Baseline [40]	ResNet-12	63.17±0.23	79.26± 0.17	68.62±0.27	83.29±0.18
$S2M2_R$ [41]	ResNet-18	64.06±0.18	80.58±0.12	–	–
CTM [42]	ResNet-18	64.12±0.82	80.51±0.13	68.41±0.39	84.28±1.73
CovNet [43] †	ResNet-12	64.59±0.45	82.02±0.29	69.75±0.52	84.21±0.26
DN4 [44] †	ResNet-12	64.73±0.44	79.85±0.31	–	–
Good-Embed [45]	ResNet-12	64.82±0.60	82.14±0.43	71.52±0.69	86.03±0.58
ADM [46] †	ResNet-12	65.87±0.43	82.05±0.29	70.78±0.52	85.70±0.43
DeepEMD [20]	ResNet-12	65.91±0.82	82.41±0.56	71.16±0.87	86.03±0.58
FRN [47]	ResNet-12	66.45±0.19	82.83±0.13	71.16±0.22	86.01±0.15
FEAT [48]	ResNet-12	66.78±0.20	82.05±0.14	70.80±0.23	84.79±0.16
BML [49]	ResNet-12	67.04±0.63	83.63±0.29	68.99±0.50	85.49±0.34
IEPT [50]	ResNet-12	67.05±0.44	82.90±0.30	72.24±0.50	86.73±0.34
MELR [51]	ResNet-12	67.40±0.43	83.40±0.28	72.14±0.51	87.01±0.35
STL DeepBDC [25]	ResNet-12	67.83±0.43	85.45±0.29	73.82±0.47	89.00±0.30
DBDC-SSL	**ResNet-12**	**68.64±0.43**	**86.02±0.28**	**73.88±0.48**	**89.03±0.29**

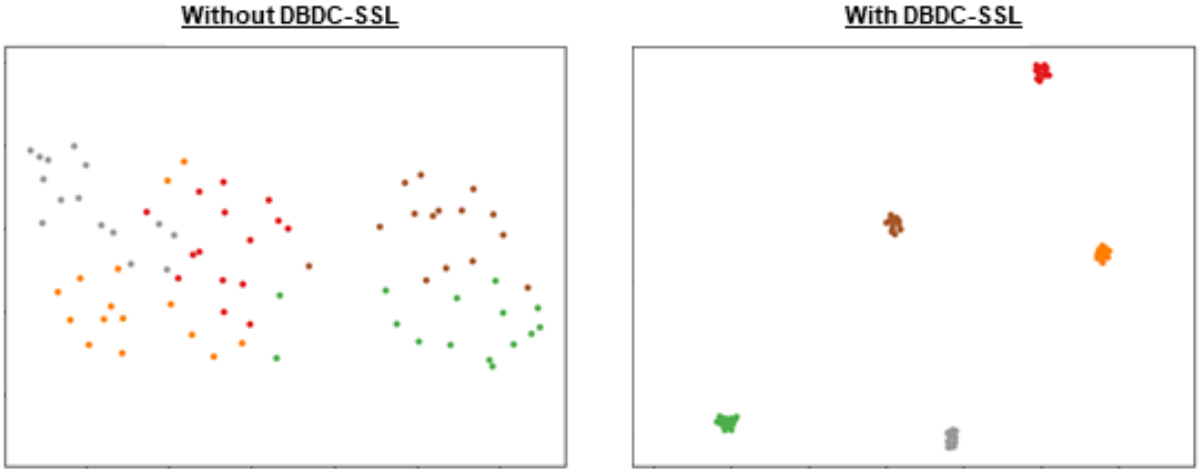

FIGURE 2. UMAP 2-dimensional visualisation [62] of the features of 75 query images based on a randomly sampled 5-way 1-shot few-shot classification task from miniImageNet without DBDC-SSL and with DBDC-SSL.

model such that given a set of input images, the backbone model can generate feature vectors with better decision boundaries between the set of classes. This extends the efficiency of the model to classify new unseen novel classes based on limited novel data.

D. DEEP BROWNIAN DISTANCE COVARIANCE WITH SELF-SUPERVISED LEARNING

We conduct training for a standard image classification task based on the entire meta-training dataset $\mathcal{D}_{\text{base}}$. During this training, a learner is built from scratch using both the cross-entropy loss L_{CE}, which measures the disparity between predictions and actual labels, and the auxiliary rotation loss L_R.

$$L_{CE}(\mathbf{y}, \hat{\mathbf{y}}) = -\frac{1}{n}\sum_{j=1}^{n} \mathbf{y}_j \log(\hat{\mathbf{y}}_j) \tag{8}$$

Overall, the final loss is defined as:

$$\mathcal{L} = \alpha \cdot L_{CE} + \beta \cdot L_R \tag{9}$$

where α and β are both set to 0.5 in all experiments.

Based on the meta-testing dataset $\mathcal{D}_{\text{test}}$, we randomly select a set of tasks $(\mathcal{C}_{\text{sup}}, \mathcal{C}_{\text{que}})$. For every task, there are K amount of classes. For the classification task, we construct and train a linear classifier based on the instances from $\mathcal{C}_{\text{sup}}$, with the pre-trained model serving as a feature extractor. In this work, we employ the logistic regression model for the final classification task.

E. SUMMARY

In summary, the proposed DBDC-SSL incorporates deep Brownian Distance Covariance (DBDC) that effectively measures the discrepancy between the product of the marginals and the joint distribution of the feature representations, which

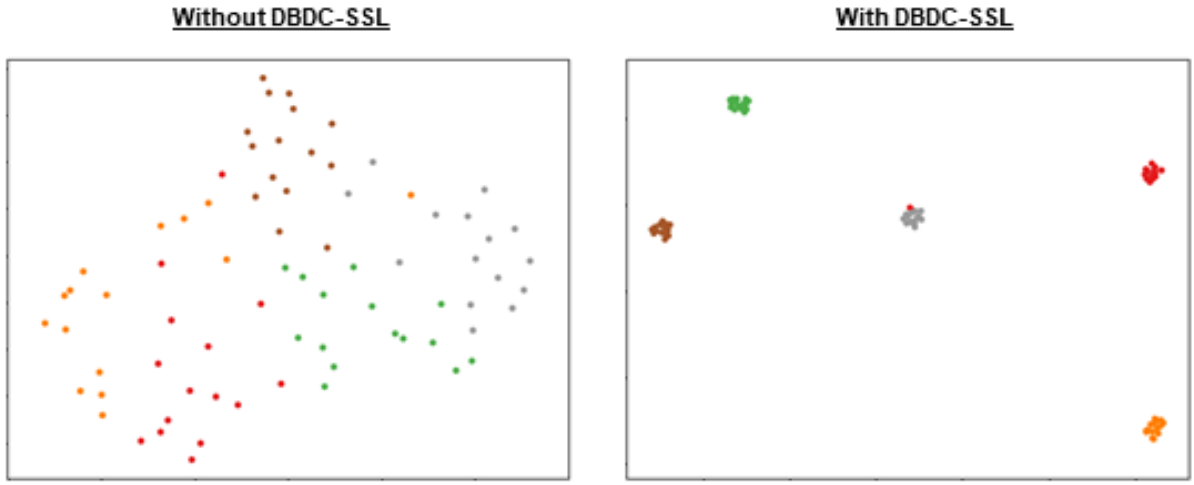

FIGURE 3. UMAP 2-dimensional visualisation [62] of the features of 75 query images based on a randomly sampled 5-way 1-shot few-shot classification task from tieredImageNet without DBDC-SSL and with DBDC-SSL.

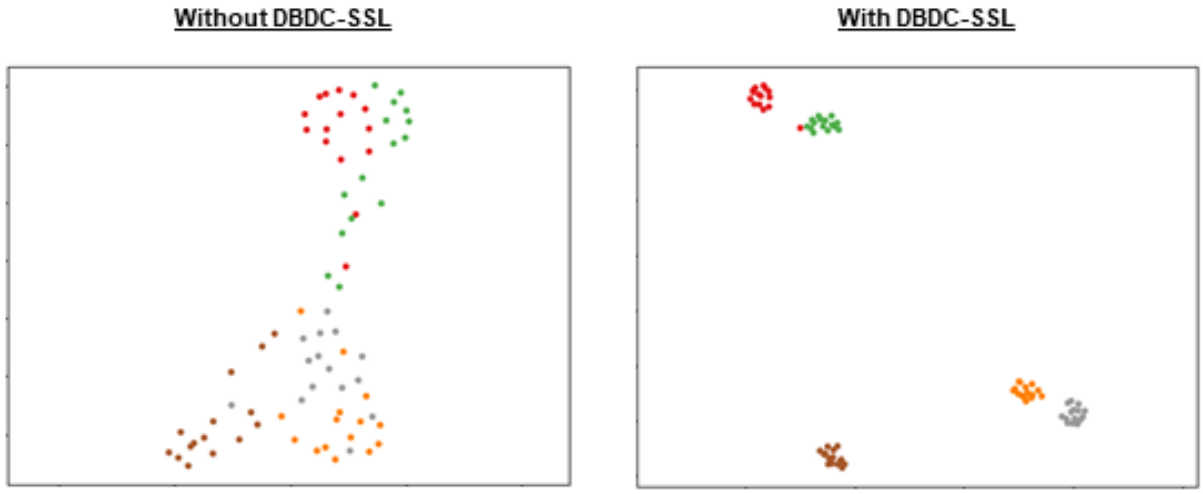

FIGURE 4. UMAP 2-dimensional visualisation [62] of the features of 75 query images based on a randomly sampled 5-way 1-shot few-shot classification task from CUB without DBDC-SSL and with DBDC-SSL.

helps to improve the robustness of the model. In addition, auxiliary loss for self-supervised learning based on predicting the rotation of given images is utilized to reduce overfitting and improve generalization of the model. As a result, due to better feature representations, the classifier is able to make predictions with a lower error rate, which in turn boosts the performance of the model.

IV. EXPERIMENTS

In this section, the datasets, evaluation protocols, implementation details, and results obtained by the proposed DBDC-SSL are described. Three common benchmarks are used to evaluate the performance, which are miniImageNet, tieredImageNet, CIFAR-FS and CUB.

A. DATASETS

1) miniImageNet

This dataset [1], [4] is widely used in few-shot image classification. It is a subset of ILSVRC-12 [57] that contains 60,000 randomly selected images from 100 classes with the size of 84 × 84 pixels. It consists of 64 training classes, 16 validation classes, and 20 test classes.

2) tieredImageNet

This dataset is a larger subset of ILSVRC-12 [57] that is made of a total of 34 high categories (608 classes), which are partitioned into 20 training categories (351 classes), 6 categories for validation (97 classes) and 8 categories (160 classes) for testing [58]. Similar to miniImageNet, 600 random images with the size of 84 × 84 pixels are sampled for each class.

3) CUB

This fine-grained classification dataset [59] is made of 200 classes and 6,033 images. Following the protocol of [37], it is split into 100 training classes, 50 validation classes, and 50 test classes with the images resized to 84 × 84 pixels.

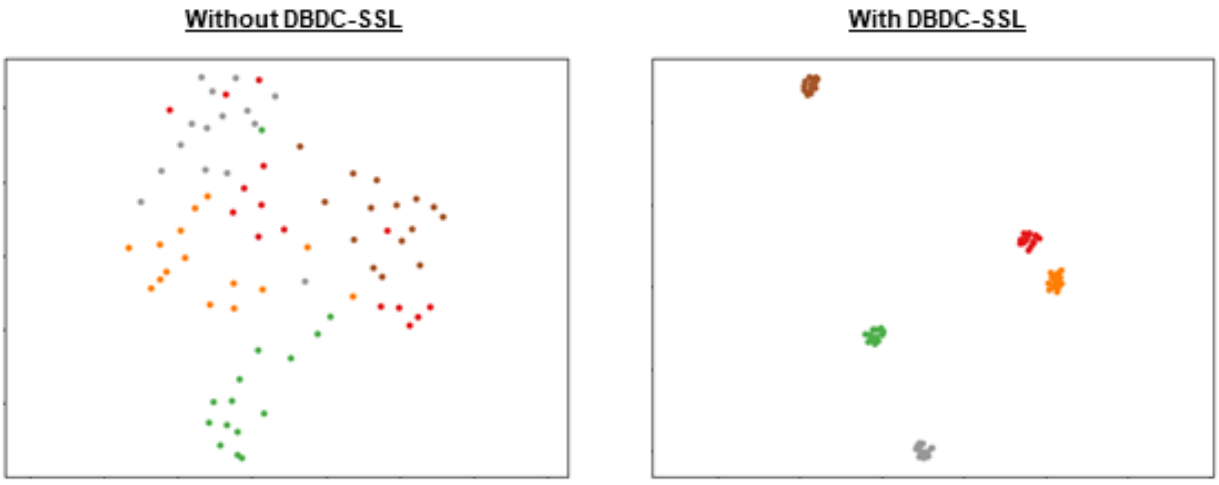

FIGURE 5. UMAP 2-dimensional visualisation [62] of the features of 75 query images based on a randomly sampled 5-way 1-shot few-shot classification task from CIFAR-FS without DBDC-SSL and with DBDC-SSL.

TABLE 2. Average accuracy on CUB. † denotes our implementation using their publicly released code.

Method	Network	1-shot	5-shot
Baseline++ [37]	ResNet-18	67.02±0.90	83.58±0.54
MAML [2]	ResNet-18	68.42±1.07	83.47±0.62
$S2M2_R$ [41]	ResNet-18	71.43±0.28	85.55 ± 0.52
MatchNet [4]	ResNet-12	71.87±0.85	85.08±0.57
Neg-Cosine [52]	ResNet-18	72.66±0.85	89.40±0.43
ProtoNet [5]	ResNet-18	72.99±0.88	86.64±0.51
AA [53]	ResNet-18	74.22±1.09	88.65±0.55
LR+ICI [54]	ResNet-12	76.16	90.32
Good-Embed [45] †	ResNet-18	77.92±0.46	89.94±0.26
ADM [46] †	ResNet-18	79.31±0.43	90.69±0.21
CovNet [43] †	ResNet-18	80.76±0.42	92.05±0.20
ProtoNet [5] †	ResNet-18	80.90±0.43	89.81±0.23
LaplacianShot [32]	ResNet-18	80.96	88.68
FRN [47] †	ResNet-18	82.55±0.19	92.98±0.10
STL DeepBDC [25]	ResNet-18	84.01±0.42	94.02±0.24
DBDC-SSL	**ResNet-18**	**84.67±0.39**	**94.76±0.16**

4) CIFAR-FS

This dataset is a randomly sampled subset of CIFAR-100 [60]. It is split into 64 base, 16 validation and 20 novel classes. For every class, there are 600 random images with the size of 32 × 32 pixels.

B. EVALUATION PROTOCOLS

The experiments are evaluated based on the standard few-shot classification settings, which are 5-way 1-shot and 5-way 5-shot tasks. The training data consist of 1 or 5 labelled data from each of the 5 classes, while the test data consist of 15 instances randomly selected from the same classes. The experimental results are obtained by averaging the accuracy with 95% confidence interval scores across 2000 randomly generated tasks.

C. TRAINING PROCEDURE AND HYPERPARAMETERS

During the pre-training phase of feature extractor, two backbones, ResNet-12 [38], [45] and ResNet-18 [6], [52], [53] are used for fair comparisons with previous methods.

TABLE 3. Average accuracy on CIFAR-FS. † denotes our implementation using their publicly released code.

Method	Network	1-shot	5-shot
$S2M2_R$ [41]	ResNet-18	63.66±0.17	76.07±0.19
MetaOptNet [38]	ResNet-12	72.00±0.70	84.20±0.50
NCA nearest centroid [55]	ResNet-12	72.49±0.12	85.15±0.09
STL DeepBDC [25] †	ResNet-12	73.07±0.46	87.69±0.31
BML [49]	ResNet-12	73.45±0.47	88.04±0.33
RENet [56]	ResNet-12	74.51±0.46	86.60±0.32
DBDC-SSL	**ResNet-12**	**75.60±0.44**	**88.49±0.31**

In the training phase for all the datasets, we utilize conventional techniques for data augmentation following [20], [37], and [47]. The data augmentation methods include andom horizontal flip, color jittering, and random resized crop. We use the SGD algorithm as the optimizer for our method. The momentum and weight decay of the optimizer used to train our proposed DBDC-SSL are set to 0.9 and 5e-4 respectively. For ResNet-12, we apply DropBlock regularization [61] during training following [38], [40], and [48].

D. COMPARISON WITH THE STATE-OF-THE-ART METHODS

Experiments of standard 5-way 1-shot and 5-way 5-shot classification tasks are carried out on three datasets: mini-ImageNet, CUB and CIFAR-FS. For a fair comparison, only the results based on the feature extractor backbone ResNet-12 and ResNet-18 are compared. As additional data is required at test time, results from a transductive setting are excluded in this work. We observed that the proposed DBDC-SSL has consistent accuracy gains over the existing methods as shown in Table 1, Table 2 and Table 3.

On miniImageNet, DBDC-SSL outperforms other methods with the highest accuracy of 68.64±0.43 on miniImageNet 5-way 1-shot task and 86.02±0.28 on miniImageNet 5-way 5-shot task. Likewise, in comparison with other methods on tieredImageNet, DBDC-SSL obtains higher accuracy

TABLE 4. Results of ablation studies on 5-way 1-shot and 5-way 5-shot tasks on miniImageNet.

Method		miniImageNet	
DBDC	SSL	1-shot	5-shot
✗	✗	57.64±0.45	73.00±0.31
✗	✓	63.90±0.44	81.03±0.31
✓	✗	67.83±0.43	85.45±0.29
✓	✓	**68.64±0.43**	**86.02±0.28**

TABLE 5. Results of ablation studies on 5-way 1-shot and 5-way 5-shot tasks on tieredImageNet.

Method		tieredImageNet	
DBDC	SSL	1-shot	5-shot
✗	✗	61.12±0.52	76.20±0.34
✗	✓	73.71±0.50	88.59±0.32
✓	✗	73.82±0.47	89.00±0.30
✓	✓	**73.88±0.48**	**89.03±0.29**

TABLE 6. Results of ablation studies on 5-way 1-shot and 5-way 5-shot tasks on CUB.

Method		CUB	
DBDC	SSL	1-shot	5-shot
✗	✗	70.40±0.48	82.98±0.34
✗	✓	80.68±0.43	90.85±0.26
✓	✗	84.01±0.42	94.02±0.24
✓	✓	**84.67±0.39**	**94.76±0.16**

TABLE 7. Results of ablation studies on 5-way 1-shot and 5-way 5-shot tasks on CIFAR-FS.

Method		CIFAR-FS	
DBDC	SSL	1-shot	5-shot
✗	✗	67.60±0.50	80.34±0.38
✗	✓	74.81±0.47	87.47±0.34
✓	✗	74.51±0.46	86.60±0.32
✓	✓	**75.60±0.44**	**88.49±0.31**

at 73.88±0.48 on 5-way 1-shot tasks and 89.03±0.29 on 5-way 5-shot tasks. For the fine-grained dataset CUB, DBDC-SSL is able to outperform other methods with 84.67±0.39 on 5-way 1-shot tasks and 94.76±0.16 on 5-way 5-shot tasks. In addition, the proposed DBDC-SSL has the highest accuracy among other methods on CIFAR-FS with 75.60±0.44 on 5-way 1-shot tasks and 88.49±0.31 on 5-way 5-shot tasks. These comparisons demonstrate that our models have better robustness and generalization when intergrating DBDC with SSL.

E. ABLATION STUDIES AND DISCUSSIONS

To investigate the effects of the major components of the proposed DBDC-SSL, an ablation study is conducted on mini-ImageNet to study the effects of DBDC and SSL. Based on Table 4, Table 5, Table 6, and Table 7, it is consistently shown that both DBDC and SSL are crucial to improve the mean accuracy of the model. With either DBDC or SSL, the model shows a substantial improvement in accuracy compared to the setting where DBDC and SSL are not utilized. This supports the hypothesis that DBDC and SSL improves the robustness of the feature representations from the model by effectively measuring the discrepancy between the product of the marginals and the joint distribution of the feature representations, as well as reducing overfitting. This causes the model to generalize better and in turn have higher accuracy.

In addition, we utilized 2-dimensional UMAP [62] for feature visualization. The UMAP graph of the feature representations from novel images based on a randomly sampled 5-way 1-shot task from miniImageNet, tieredImageNet, CUB, and CIFAR-FS respectively is shown in Figure 2, Figure 3, Figure 4, and Figure 5. The visualization shows that without DBDC-SSL, the points of each cluster are more sparse, showing higher variance. In contrast, when DBDC-SSL is integrated, it can be observed that the segregated clusters are with less variance. This indicates that DBDC and SSL contribute to the generation of features with shorter inter-class distances and higher intra-class distances, which in turn improves the performance of the model.

V. CONCLUSION

In this paper, DBDC-SSL is proposed for few-shot learning. The proposed DBDC-SSL utilizes deep Brownian Distance Covariance that effectively measures the discrepancy between the product of the marginals and the joint distribution of the feature representations. This in turn helps the model to learn robust image representations which subsequently improves the performance of the model. In addition, to reduce overfitting and improve the generalization of the model, we incorporate a self-supervised learning loss based on an auxiliary task to classify the degree of rotation of the image when given the embedded features. By doing so, the learned representations become more robust, which allows the few-shot recognition model to achieve good performance in mean accuracy. Through extensive experiments, the performance of the proposed DBDC-SSL is shown to be able to outperform many state-of-the-art methods in few-shot learning in both mean accuracy. Thus, the proposed framework in this work is applicable to many practical problems.

REFERENCES

[1] S. Ravi and H. Larochelle, "Optimization as a model for few-shot learning," in *Proc. Int. Conf. Learn. Represent.*, 2017.
[2] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, Aug. 2017, pp. 1126–1135.
[3] A. Nichol, J. Achiam, and J. Schulman, "On first-order meta-learning algorithms," 2018, *arXiv:1803.02999*.
[4] O. Vinyals, C. Blundell, T. Lillicrap, and D. Wierstra, "Matching networks for one shot learning," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2016, pp. 3630–3638.
[5] Jake Snell, Kevin Swersky, and Richard Zemel, "Prototypical networks for few-shot learning," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 4080–4090.
[6] 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.
[7] A. A. Rusu, "Meta-learning with latent embedding optimization," in *Proc. Int. Conf. Learn. Represent.*, 2018.
[8] T. Munkhdalai and H. Yu, "Meta networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2554–2563.

[9] H. Xu, J. Wang, H. Li, D. Ouyang, and J. Shao, "Unsupervised meta-learning for few-shot learning," *Pattern Recognit.*, vol. 116, Aug. 2021, Art. no. 107951.

[10] B. Zhang, K.-C. Leung, X. Li, and Y. Ye, "Learn to abstract via concept graph for weakly-supervised few-shot learning," *Pattern Recognit.*, vol. 117, Sep. 2021, Art. no. 107946.

[11] X. Zhong, C. Gu, M. Ye, W. Huang, and C.-W. Lin, "Graph complemented latent representation for few-shot image classification," *IEEE Trans. Multimedia*, vol. 25, pp. 1979–1990, 2022.

[12] H. Zhang, H. Li, and P. Koniusz, "Multi-level second-order few-shot learning," *IEEE Trans. Multimedia*, vol. 25, pp. 2111–2126, 2022.

[13] Y. Zheng, X. Zhang, Z. Tian, W. Zeng, and S. Du, "Detach and unite: A simple meta-transfer for few-shot learning," *Knowl.-Based Syst.*, vol. 277, Oct. 2023, Art. no. 110798.

[14] M. Yan, "Adaptive learning knowledge networks for few-shot learning," *IEEE Access*, vol. 7, pp. 119041–119051, 2019.

[15] B. Hariharan and R. Girshick, "Low-shot visual recognition by shrinking and hallucinating features," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3037–3046.

[16] A. Antoniou, A. Storkey, and H. Edwards, "Data augmentation generative adversarial networks," in *Proc. Int. Conf. Learn. Represent. Workshops*, 2018.

[17] Q. Luo, L. Wang, J. Lv, S. Xiang, and C. Pan, "Few-shot learning via feature hallucination with variational inference," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 3962–3971.

[18] J. Wu, N. Dong, F. Liu, S. Yang, and J. Hu, "Feature hallucination via maximum a posteriori for few-shot learning," *Knowl.-Based Syst.*, vol. 225, Aug. 2021, Art. no. 107129.

[19] B. Oreshkin, P. R. López, and A. Lacoste, "TADAM: Task dependent adaptive metric for improved few-shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 1–13.

[20] 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. 12200–12210.

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

[22] J. Kim, T. Kim, S. Kim, and C. D. Yoo, "Edge-labeling graph neural network for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11–20.

[23] L. Yang, L. Li, Z. Zhang, X. Zhou, E. Zhou, and Y. Liu, "DPGN: Distribution propagation graph network for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 13387–13396.

[24] X. Wang, B. Ma, Z. Yu, F. Li, and Y. Cai, "Multi-scale decision network with feature fusion and weighting for few-shot learning," *IEEE Access*, vol. 8, pp. 92172–92181, 2020.

[25] J. Xie, F. Long, J. Lv, Q. Wang, and P. Li, "Joint distribution matters: Deep Brownian distance covariance for few-shot classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 7962–7971.

[26] W. Cui and Y. Guo, "Parameterless transductive feature re-representation for few-shot learning," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 2212–2221.

[27] M. Boudiaf, I. Ziko, J. Rony, J. Dolz, P. Piantanida, and I. B. Ayed, "Information maximization for few-shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 2445–2457.

[28] Y. Liu, J. Lee, M. Park, S. Kim, E. Yang, S. J. Hwang, and Y. Yang, "Learning to propagate labels: Transductive propagation network for few-shot learning," in *Proc. Int. Conf. Learn. Represent.*, 2018.

[29] G. S. Dhillon, P. Chaudhari, A. Ravichandran, and S. Soatto, "A baseline for few-shot image classification," in *Proc. Int. Conf. Learn. Represent.*, 2019.

[30] J. Liu, L. Song, and Y. Qin, "Prototype rectification for few-shot learning," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 741–756.

[31] Y. Hu, V. Gripon, and S. Pateux, "Leveraging the feature distribution in transfer-based few-shot learning," in *Proc. Int. Conf. Artif. Neural Netw. (ICANN)*, 2021, pp. 487–499.

[32] I. Ziko, J. Dolz, E. Granger, and I. B. Ayed, "Laplacian regularized few-shot learning," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 11660–11670.

[33] P. Rodríguez, I. Laradji, A. Drouin, and A. Lacoste, "Embedding propagation: Smoother manifold for few-shot classification," in *Proc. Eur. Conf. Comput. Vis.*, Aug. 2020, pp. 121–138.

[34] C. Chen, X. Yang, C. Xu, X. Huang, and Z. Ma, "ECKPN: Explicit class knowledge propagation network for transductive few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 6592–6601.

[35] G. J. Székely and M. L. Rizzo, "Brownian distance covariance," *Ann. Appl. Statist.*, vol. 3, no. 4, pp. 1236–1265, Dec. 2009.

[36] G. J. Székely, M. L. Rizzo, and N. K. Bakirov, "Measuring and testing dependence by correlation of distances," *Ann. Statist.*, vol. 35, pp. 2769–2794, 2007.

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

[38] K. Lee, S. Maji, A. Ravichandran, and S. Soatto, "Meta-learning with differentiable convex optimization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 10649–10657.

[39] Y. Wang, W.-L. Chao, K. Q. Weinberger, and L. van der Maaten, "SimpleShot: Revisiting nearest-neighbor classification for few-shot learning," 2019, *arXiv:1911.04623*.

[40] Y. Chen, Z. Liu, H. Xu, T. Darrell, and X. Wang, "Meta-baseline: Exploring simple meta-learning for few-shot learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9042–9051.

[41] P. Mangla, M. Singh, A. Sinha, N. Kumari, V. N. Balasubramanian, and B. Krishnamurthy, "Charting the right manifold: Manifold mixup for few-shot learning," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 2207–2216.

[42] 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.

[43] D. Wertheimer and B. Hariharan, "Few-shot learning with localization in realistic settings," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6551–6560.

[44] W. Li, L. Wang, J. Xu, J. Huo, Y. Gao, and J. Luo, "Revisiting local descriptor based image-to-class measure for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7253–7260.

[45] Y. Tian, Y. Wang, D. Krishnan, J. B. Tenenbaum, and P. Isola, "Rethinking few-shot image classification: A good embedding is all you need?" 2020, *arXiv:2003.11539*.

[46] 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.

[47] 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. 8008–8017.

[48] 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. 8805–8814.

[49] Z. Zhou, X. Qiu, J. Xie, J. Wu, and C. Zhang, "Binocular mutual learning for improving few-shot classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 8382–8391.

[50] 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.*, 2021.

[51] N. Fei, Z. Lu, T. Xiang, and S. Huang, "MELR: Meta-learning via modeling episode-level relationships for few-shot learning," in *Proc. Int. Conf. Learn. Represent.*, 2021, pp. 1–20.

[52] B. Liu, "Negative margin matters: Understanding margin in few-shot classification," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 438–455.

[53] A. Afrasiyabi, J.-F. Lalonde, and C. Gagné, "Associative alignment for few-shot image classification," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 18–35.

[54] Y. Wang, C. Xu, C. Liu, L. Zhang, and Y. Fu, "Instance credibility inference for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12833–12842.

[55] S. Laenen and L. Bertinetto, "On episodes, prototypical networks, and few-shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 24581–24592.

[56] D. Kang, H. Kwon, J. Min, and M. Cho, "Relational embedding for few-shot classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 8802–8813.
[57] 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.
[58] M. Ren, E. Triantafillou, S. Ravi, J. Snell, K. Swersky, J. B. Tenenbaum, H. Larochelle, and R. S. Zemel, "Meta-learning for semi-supervised few-shot classification," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–7.
[59] P. Welinder, S. Branson, T. Mita, C. Wah, F. Schroff, S. Belongie, and P. Perona, "'Caltech-UCSD Birds 200," California Inst. Technol., Pasadena, CA, USA, Tech. Rep. CNS-TR-2010-001, 2010.
[60] L. Bertinetto, J. F. Henriques, P. Torr, and A. Vedaldi, "Meta-learning with differentiable closed-form solvers," in *Proc. Int. Conf. Learn. Represent.*, 2018.
[61] G. Ghiasi, T.-Y. Lin, and Q. V. Le, "Dropblock: A regularization method for convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 31, 2018, pp. 1–5.
[62] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," 2018, *arXiv:1802.03426*.

WEI HAN LIU received the bachelor's degree from Multimedia University, Malaysia, where he is currently pursuing the Ph.D. degree with the Faculty of Information Science and Technology. His research interests include deep learning, computer vision, and few-shot learning.

KIAN MING LIM (Senior Member, IEEE) received the B.IT. degree (Hons.) in information systems engineering and the Master of Engineering Science and Ph.D. (I.T.) degrees from Multimedia University. He is currently a Senior Lecturer with the Faculty of Information Science and Technology, Multimedia University. His research interests include machine learning, computer vision, and pattern recognition.

THIAN SONG ONG (Senior Member, IEEE) is currently a Professor with the Faculty of Information Science and Technology, Multimedia University, Malaysia. His research interests include biometric security and machine learning. He has published more than 60 international refereed journals and conference papers in the related fields.

CHIN POO LEE (Senior Member, IEEE) received the Master of Science and Ph.D. degrees in abnormal behavior detection and gait recognition. She is currently a Senior Lecturer with the Faculty of Information Science and Technology, Multimedia University, Malaysia. Her research interests include action recognition, computer vision, gait recognition, and deep learning.