Multi-Instance Learning with Emerging Novel Class

Xiu-Shen Wei, *Member, IEEE*, Han-Jia Ye, Xin Mu,
Jianxin Wu, *Member, IEEE*, Chunhua Shen, *Member, IEEE,* and Zhi-Hua Zhou, *Fellow, IEEE*

Abstract—Diverse applications involving complicated data objects such as proteins and images are solved by applying multi-instance learning (MIL) algorithms. However, few MIL algorithms can deal with problems in an open and dynamic environment, where new categories of samples emerge. In this type of emerging novel class setting, algorithms should be able to not only classify the samples from the observed classes accurately, but also recognize the samples from the novel class. In this paper, we focus on the Multi-Instance learning with Emerging Novel class (MIEN) problem, and formulate MIEN from a metric learning perspective. We extract key instances to form the "super-bag" for each observed class, and non-key instances from all the observed classes to form a "meta super-bag". Based on these super-bags, we propose the MIEN-metric method to learn discriminative metrics for classifying MIL bags from the observed classes and recognizing bags from the novel class. Experimental results of diverse domains, *e.g.*, biological function annotation, text categorization and object-centric/scene-centric image classification, show MIEN-metric outperforms other baseline methods significantly when the novel class emerges. Meanwhile, MIEN-metric is comparable with state-of-the-art MIL algorithms for binary classification in the traditional MIL setting.

Index Terms—Multi-instance learning, Emerging novel class, Incremental learning, Metric learning

1 Introduction

During the investigation of drug activity prediction [1], multi-instance learning (MIL) was formally proposed and naturally applied to this problem. In contrast to traditional single-instance learning, MIL receives a set of *bags* that are labeled, rather than receiving a set of labelled instances, which is regarded as a type of weakly supervised learning [2]. Moreover, instances in the MIL bags have no label information. The task of MIL is to train a classifier that labels testing bags, and MIL has already been widely applied in diverse applications [3], [4], such as functional annotation, text categorization, image categorization and so on.

Many existing MIL works [1], [5], [6], [7], [8], [9] solve the binary MIL classification problem. In fact, the multi-class MIL problem is also popular in diverse real-world applications. In knowledge discovery and data mining, especially the medical and biological application fields, MIL was used for detecting lung cancer [10] and protein function annotation (cf. experiments in this paper). Additionally, in computer vision, [11] used MIL for multi-class image categorization, and [12] employed MIL to handle human part detection problems. In the literature, there are two types of multi-class MIL problems: one without background class, and another with background class. A background class in multi-class MIL does not contain any positive instances from any other classes. In this paper, we are dealing with multi-class MIL without background class.

However, both for binary or multi-class classification, it is noteworthy that most previous studies on multi-instance learning were in a stable environment rather than an environment that is open and changes gradually [13], [14]. In this paper, we focus on Multi-instance learning with Emerging Novel class (MIEN). Emerging novel class, so called augmented class recognition, is an important problem in the data mining and machine learning field [15], [16], [17], [18], [19]. In augmented class recognition, after training an effective learner, there will emerge some or usually plenty of samples from a novel class / category which is called the *augmented class*. Because no samples from the augmented class were observed in the training phase, it is a challenging problem. In the MIEN setting, one is required not only to classify the bags from the observed classes accurately, but also to recognize the bags from the novel / augmented class.

For example, in the task of protein function annotation by MIL, multiple distinct functional units (instances) form a protein (bag). We can build a MIL model to automatically annotate the classes of proteins. However, due to the effects of mutation, there may appear some proteins from a novel class. Another example occurs in the case of building a MIL object-centric image classification system: the training data contain three classes, *e.g.*, *elephant*, *fox* and *bird*. But the system has to predict images from more classes in the future. When an image labeled *tiger* comes, a traditional MIL algorithm will predict it in one of the observed classes, such as *fox*, which might make the system unusable. Similar issues take place in the other tasks, such as text categorization, *e.g.*, "deep

• *X.-S. Wei is with Megvii Research Nanjing, Megvii Technology, Nanjing, China. H.-J. Ye, J. Wu and Z.-H. Zhou are with the National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China. X. Mu is with Tecent Inc., Shenzhen, China. C. Shen is with the University of Adelaide, Adelaide, Australia.*
• *The first two authors contributed equally to this work. Corresponding authors: J. Wu and Z.-H. Zhou (Nanjing University).*
• *This work is supported by the National Science Foundation of China (61921006; 61772256), National Key R&D Program of China (2018YFB1004300), the Collaborative Innovation Center of Novel Software Technology and Industrialization, and the program A for Outstanding Ph.D. candidate of Nanjing University (201702A010).*
• *Email: weixs.gm@gmail.com, {yehj, wujx, zhouzh}@lamda.nju.edu.cn, anmarsmu@tencent.com, chunhua.shen@adelaide.edu.au.*

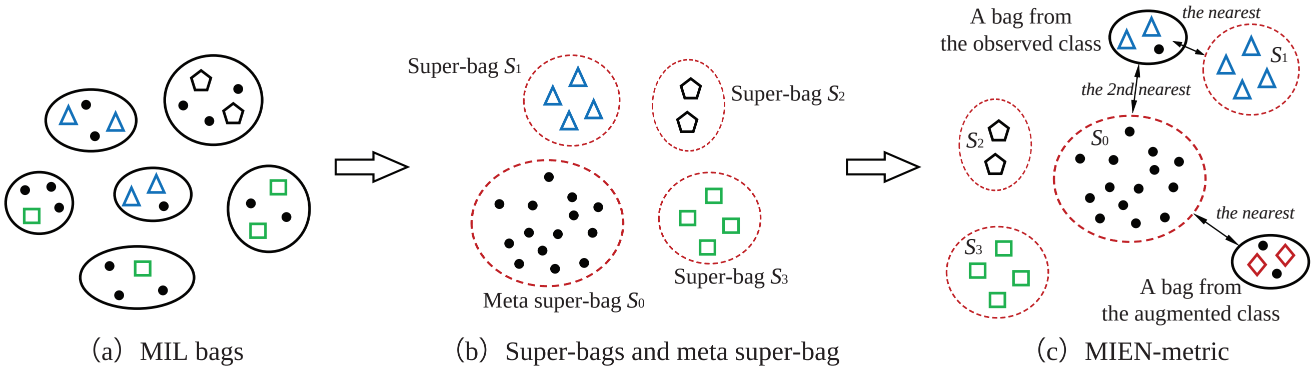

Figure 1. The proposed MIEN-metric method. In (a), black solid circles represent the original MIL bags. Different marks denote the corresponding key instances of different classes, and the black dots are non-key instances. (b) shows the obtained super-bags and the meta super-bag, *i.e.*, red dashed circles. In (c), after learning MIEN-metric, a bag from the augmented class could be recognized: If its nearest super-bag is the meta super-bag, it will be treated as a bag with the novel class.

learning" related topic seldom existed in conferences and journals before 2006, while it became an important novel topic after 2012. When facing these real-life applications, a mature MIL method needs to work in the MIEN setting.

In order to deal with the MIEN problem, we formulate MIEN into a metric learning framework and propose a MIEN-metric method. Fig. 1 gives an illustration of the MIEN-metric procedure. We first extract key instances in each MIL bag X_i, using for example off-the-shelf *key instance detection* methods [20], [21]. The so called key instances are the ones that can determine the bag's label (Several important concepts used in MIEN can be found in Sect. 3.2). Then, the key instances from each observed class merge into the corresponding *super-bag* S_c ($1 \leq c \leq C$, where C is the number of the observed classes). In addition, all the remaining instances which are not considered as key instances combine into one super-bag, which is called the *meta super-bag* S_0. Thus, the super-bags have the discriminative ability to classify the observed classes. The meta super-bag S_0 is used for recognizing bags from the augmented class. After obtaining these super-bags, a *Bag to Super-bag (B2S)* distance $d^2(X_i, S_k)$ ($0 \leq k \leq C$) is proposed to measure the distance between bags and super-bags. In B2S, we also introduce the instance weights to relieve the effect of cases in which key instances might not be detected very accurately.

In the training phase, for a bag X_i from one of the observed classes (*i.e.*, $y_i \in \{1, \ldots, C\}$), there are some key instances and non-key instances existing in S_{y_i} and S_0, respectively. Ideally, there should not exist instances of X_i in the other super-bags $S_{\neg y_i}$. Therefore, the distances $d^2(X_i, S_{y_i})$ and $d^2(X_i, S_0)$ should be small, and be both smaller than $d^2(X_i, S_{\neg y_i})$. Moreover, in order to achieve a better discriminative ability, it is natural to expect the distance $d^2(X_i, S_{y_i})$ should be the smallest, while the distances between that bag and other super-bags should be larger than that by a margin. Consequently, $d^2(X_i, S_0)$ should be the second smallest distance. In MIEN-metric, a metric is learned to make the B2S distance satisfy the aforementioned ranking orders. S_0 is treated as a boundary to separate S_{y_i} from the other super-bags in training. Meanwhile, it is also used for recognizing the bag with a novel class in testing. In the testing phase, for a bag from the observed classes, its nearest super-bag could indicate its class. For a bag from the augmented class, we expect its nearest super-bag to be the meta super-bag. Thus, we can recognize a bag with the novel class in that way.

With these super-bags and the learned B2S distances, MIEN-metric can handle the multi-instance with augmented class data effectively. In our experiments, we conduct five MIL data sets in diverse applications for MIEN, *i.e.*, biological function annotation, text categorization and object-centric/scene-centric image classification. As shown in our empirical results, existing state-of-the-art augmented class recognition approaches can not work well in the multi-instance setting. By comparing with them, our MIEN-metric achieves the best classification performance when the novel class emerges. Beyond that, we also test MIEN-metric on five benchmark MIL data sets to validate its effectiveness for traditional binary MIL classification (where it has no augmented classes). It shows that the classification accuracy of MIEN-metric is comparable with state-of-the-art MIL classifiers, *e.g.*, miGraph [7] and MIBoosting [22], in a stable MIL environment.

The contributions of this paper are summarized as follows.

- We address a well-recognized practical value problem, *i.e.*, Multi-instance learning with Emerging Novel class (MIEN), and formulate it into a metric learning framework. To our best knowledge, it is the first time to handle emerging novel class in the MIL setting.
- In order to solve MIEN, we develop the MIEN-metric method by learning appropriate metrics on the MIL bag to "super-bag" distance with a ranking restriction, which can jointly classify MIL bags from the observed classes as well as detect the bags from the novel class.
- We conduct experiments on various real-world applications with the novel class emerging to comprehensively evaluate MIEN-metric, *e.g.*, biological function annotation, text categorization, object/scene image classification. Experimental results demonstrate that MIEN-metric consistently achieves superior performance over state-of-the-art methods.

The rest of the paper is organized as follows. In Sect. 2, we introduce some related works about MIL and incremental learning. The proposed MIEN-metric method is presented in

Sect. 3. In Sect. 4, we describe the data sets and empirical settings in experiments. In Sect. 5, we present our experimental results and discussions. Finally, Sect. 6 concludes the paper with future issues.

2 Related work

In this section, we review the related work about incremental learning, anomaly detection and multi-instance learning.

2.1 Incremental learning

Traditional machine learning approaches face many challenges emerged in real-world applications, where the open and dynamic environments [13] break the stationary settings implied in traditional approaches. Incremental learning is one branch of methods dealing with the changing environments, which includes example-incremental learning [23], [24], attribute-incremental learning [25], class-incremental learning [18], [26], etc. The augmented class recognition problem in this paper falls into class-incremental learning. Previous studies on augmented class recognition merely focus on the single-instance setting, rather than MIL.

Moreover, almost all the existing single-instance augmented class recognition algorithms [18], [26] need to leverage the power of unlabeled data or a few data from the augmented classes. A MIL bag can be represented by a single vector (*e.g.*, the mean vector of the instances in a bag) [27], and then employ the single-instance algorithms for recognizing the augmented class. However, because unlabeled or novel-class data is not available in the MIEN problem, these algorithms can not be directly applied to MIEN.

Additionally, zero-shot learning (ZSL) [28], [29], [30], [31] is a popular topic belonging to transfer learning, which is loosely related to augmented class problems in this paper. The similarity between ZSL and MIEN is that they both solve the problems in an open and dynamic environment. But, except for its single-instance setting, the most different point of ZSL from MIEN is that, in ZSL, training and test classes are *disjoint*, *i.e.*, no training examples of the test classes are available [28]. However, in MIEN, the test label space is a superset of the training one. Therefore, MIEN is clearly different from ZSL. Besides, ZSL is commonly studied by the computer vision community, which always depends on some kinds of side information, *e.g.*, attributes (specific patterns corresponding to object “head” or “torso”). While, augmented class recognition belongs to the machine learning area, which considers and focuses on more general learning problems.

2.2 Anomaly detection

Emerging novel class detection (a.k.a. augmented class recognition) is also related to anomaly detection [32]. What distinguishes augmented class recognition with anomaly detection is that: the definition of anomaly detection is the identification of rare items, events or observations which raise suspicious by differing significantly from the majority of the data. While, the items or instances belonging to novel classes are *not* rare, but usually are sizable or even large-scale. Meanwhile, different from relying on the application-specific definition of outliers, the goal of augmented class recognition is to minimize the recognition errors of both observed classes and augmented classes.

In the literature, there are a number of anomaly detection methods. From the technique perspective, these anomaly detection methods can be roughly divided into these following paradigms [32], *i.e.*, classification-based [33], clustering-based [34], nearest neighbor based [35], statistical methods, spectral based methods, etc. In real-world, anomaly detection are used in diverse applications, *e.g.*, fraud detection [36], medical anomaly detection [32], industrial damage detection [32], textual anomaly detection and so on. Several famous and solid anomaly detection methods are LOF [37], iForest [38], etc. Those methods are compared as baselines in our experiments.

2.3 Multi-instance learning

For multi-instance learning, its framework originated from the investigation of drug activity prediction [1]. Since then, many effective multi-instance learning algorithms have been developed in the literature [3], [39], [40], [41], to name a few: EM-DD, miSVM, MIBoosting, MILES, miGraph, miFV, etc. All these MIL algorithms assume that the classes of bags form a closed set and the MIL training and testing data are under the same distribution. Existing MIL algorithms have achieved satisfactory accuracy rates in *the stationary setting* implied in these traditional MIL algorithms.

However, the stationary setting is frequently violated in real applications. Over the last few years, MIL in an open and dynamic environment has attracted more attentions. For example, Zhang and Zhou [42] proposed the MICS approach to deal with MIL in the covariate shift setting [43]. In order to solve image classification problems, Wang et al. [44] used transfer learning to transfer cross-category knowledge from source categories in multi-instance setting for boosting the learning process in the target domain.

In this paper, we focus on the Multi-Instance learning with Emerging Novel class (MIEN) problem. The most related work to ours is multi-instance multi-label learning with novel instances [45]. One difference is that we deal with the multi-class problem in MIL, while they studied the multi-label setting. Multi-class MIL is very popular in diverse applications, *e.g.*, sentiment analysis [46], image classification [12], [47], medical diagnosis [10], biological function annotation (which will be shown in our experiments), etc. In addition, the more important difference is: in [45], the instances of the novel class were observed during training and they were just unlabeled. In contrast, the MIEN problem is more challenging, because *no instance* indicating the augmented classes can be observed in the training set in MIEN. More detailed discussions / results can be found in Sect. 4.

Additionally, our work is also related to key instance detection (KID) [21] which is a classical topic in MIL. KID is more challenging than bag classification, since we can easily label a bag once all the key instances have been detected. In the literature, the early methods of MIL largely focused on the ranking and selection of key instances in positive bags, *e.g.*, DD [48], EM-DD [8], mi-SVM [49]. For the other MIL methods, *e.g.*, MILES [6] and MILIS [50], can also be used for instance selection in an embedded manner.Specifically, MILES learned an ℓ_1 regularized bag-level classifier, and consequently key instances can be selected based on classifier weights but not necessarily one instance from each single bag. While, MILIS selected one instance explicitly from each positive bag, which is the same scenario discussed in the paper.

3 Proposed method

In this section, before presenting the details of the proposed MIEN-metric method, we will first briefly formalize the MIEN problem, and then introduce several crucial concepts of MIEN. The formulation, learning strategies and optimization details of MIEN-metric are followed.

3.1 Problem formalization of MIEN

$D = \{(X_1, y_1), \dots, (X_i, y_i), \dots, (X_N, y_N)\}$ is a MIL training set which contains N "bag-label" pairs. In each bag, $X_i = \{\mathbf{x}_i^1, \dots, \mathbf{x}_i^{n_i}\}$ contains n_i instances, and each instance $\mathbf{x}_i^j \in \mathcal{R}^d$. Instances in a bag X_i correspond to one bag-level class label $y_i \in \mathcal{Y} = \{1, \dots, C\}$. It is supposed that: (I) If a bag X_i is labeled as the c-th class $\Leftrightarrow$ *at least one* instance of X_i belongs to the c-th class; (II) A bag X_i is not assigned to the c-th class $\Leftrightarrow$ *none* instance of X_i belongs to the c-th class. Unlike in classic MIL, during the test phase, we need to predict the classes of bags from an open data set $D_o = \{(X_i, y_i)\}_{i=1}^{\infty}$, where $y_i \in \mathcal{Y}' = \{1, \dots, C, C+1, \dots, K\}$ with $K \geq C$. As there are classes unobservable during the training time, the goal of MIEN is to learn a model $f(X_i) \rightarrow y_i \in \mathcal{Y}' = \{novel, 1, \dots, C\}$, where the option "novel" indicates that X_i belongs to the augmented class. In this paper, without loss of generality, we regard the augmented class as class 0.

3.2 Important concepts of MIEN

We hereby list four important concepts of MIEN for clearly comparisons and understandings.

Concepts	Descriptions
Bags	MIL bags represent samples of MIL which have labels. Each bag contains several instances without label information.
Key instances	The instances in each MIL bag X_i can determine the bag's label ($y_i \in \{1, \dots, C\}$). Additionally, the remaining instances in that bag are non-key instances.
Super-bags	All the key instances from bags belonging to one class ($c \in \{1, \dots, C\}$) merge into the corresponding super-bag S_c.
Meta super-bag	All the non-key instances from all MIL bags combine as the meta super-bag S_0.

3.3 The proposed MIEN-metric method

For classification, an intuitive way is to pull objects from the same class close and to push different-class objects away, thus classification will become easier [51]. In order to handle MIEN, we learn appropriate distance metrics which not only classify bags from the observed classes, but also recognize bags in the novel class.

However, different from the single-instance data, multi-instance data are highly heterogeneous. Hence, a homogeneous distance metric, *e.g.*, the *Bag to Bag (B2B) distance*, may be insufficient to characterize different classes of objects in MIEN. Meanwhile, directly computing the distance between two bags has two drawbacks. (I), it is hard to measure the distance between two instance sets. Existing work [52] used the minimum distance between instances from each bag, which only reflect one side of the characteristic of the bags. (II), pairwise constraints increase the computational cost quadratically.

Inspired by [53], in this paper, we propose to solve MIEN by taking the class property into consideration, which assesses the relationships between the observed classes and the MIL bags. Different from them, we construct a super-bag S_c for class c which *only* contains *key instances* extracted from relevant bags, rather than a direct combination of all the instances in bags from class c. Beyond constructing super-bags, we also merge the instances which are not key instances into *one* meta super-bag S_0. Thus, instances in S_0 are those dissimilar from patterns in the observed classes. Based on these super-bags, instead of measuring the similarities between bags, we propose a new *Bag to Super-bag (B2S)* distance. Comparing with [53], the proposed super-bag containing key instances are more consistent and have more discriminative ability in metric learning. It is the key of the proposed MIEN-metric method to solve the multi-instance with emerging novel class problem.

In the implementation of MIEN-metric, we use the key instance detection method in [20] to collect key instances. Then, the super-bags S_c and the meta super-bag S_0 are obtained. The B2S distance between bag X_i and super-bag S_k is defined as the Mahalanobis distance between each key instance and its nearest neighbor in a particular bag: $d^2_{M_k}(X_i, S_k) = \sum_{v=1}^{V_k} (\mathbf{s}_k^v - \hat{\mathbf{x}}_i^v)^\top M_k (\mathbf{s}_k^v - \hat{\mathbf{x}}_i^v)$ $(0 \leq k \leq C)$, where $S_k = [\mathbf{s}_k^1, \dots, \mathbf{s}_k^{V_k}]$ is a super-bag with V_k instances. S_k can be the super-bag for an observed class or the meta super-bag. $\hat{\mathbf{x}}_i^v$ is the instance in X_i which is the nearest neighbor to $\mathbf{s}_k^v$ based on the Euclidean distance. M_k is the learned parameters for the k-th metric from X_i to S_k. Here, the impact of the inaccurate Euclidean nearest neighbors can be reduced by the learned metric, cf. [51], [54].

To distinguish the existing C classes and to recognize the augmented class simultaneously, we propose to learn a metric for each super-bag on the aforementioned B2S distance. There are two restrictions on distances. (I) Distance between a bag X_i and its super-bag (*i.e.*, S_{y_i}) should be smaller than the distance between X_i and the other super-bags, including the meta super-bag S_0. (II) Distance between X_i and the meta super-bag S_0 should be smaller than the distances between X_i and other super-bags, *i.e.*, $\{S_c \mid c \geq 1 \text{ and } c \neq y_i\}$.

Therefore, we can formulate MIEN-metric as follows:

$$
\begin{aligned}
\underset{M_0,\ldots,M_C}{\arg\min}\ & \sum\nolimits_{i=1}^{N} d^2_{M_{y_i}}(X_i, S_{y_i}) + \gamma \sum\nolimits_{k=0}^{C} \Omega(M_k) \\
& + \lambda \left(\sum_{i=1}^{N} \sum_{k\geq 0, k\neq y_i} \xi_{ik} + \sum_{i=1}^{N} \sum_{c\geq 1, c\neq y_i} \epsilon_{ic} \right) \\
\text{s.t.}\ & d^2_{M_k}(X_i, S_k) - d^2_{M_{y_i}}(X_i, S_{y_i}) \geq 1 - \xi_{ik} \\
& d^2_{M_c}(X_i, S_c) - d^2_{M_0}(X_i, S_0) \geq 1 - \epsilon_{ic} \\
& \xi_{ik} \geq 0 \ , \ \epsilon_{ic} \geq 0 \ , \ M_k \succeq 0 \ (0 \leq k \leq C) \\
& 1 \leq i \leq N \ , \ 0 \leq k \leq C \,\text{and}\, k \neq y_i \\
& 1 \leq c \leq C \,\text{and}\, c \neq y_i \ , \qquad (1)
\end{aligned}
$$

where ξ_{ik} and ϵ_{ic} are non-negative slack variables. $\Omega(\cdot)$ is a convex regularization on the metric M_k, and λ and γ are two non-negative trade-off parameters. "$\succeq$" presents positive semidefinite.

The first constraint makes the distance to distinguish the observed classes. The two constraints work as a ranking restriction, *i.e.*, it makes the distance $d^2_{M_{y_i}}(X_i, S_{y_i})$ the smallest one, and the distance $d^2_{M_0}(X_i, S_0)$ the second smallest. Both are smaller than the distance between X_i and other existing but irrelevant super-bags. Using the distance between X_i and the meta super-bag S_0 as an intermediate level keeps a margin between relevant and irrelevant classes, which will benefit the class discrimination process. More importantly, this ranking constraint makes S_0 a reject option. When a bag does not have an obvious pattern with the observed classes (the existing super-bags), it will be close to S_0, which will be regarded as a bag from augmented classes.

3.4 Learning strategies of MIEN-metric

MIEN-metric can be directly solved with semi-definite programming, which has a high computational cost. To improve its effectiveness and efficiency, two modifications are made to the computations of the B2S distance.

First, we introduce the instance weights $\mathbf{w}_k \in \mathcal{R}_+^{V_k}$ into the B2S distance. Therefore, the instances in the super-bag could have different impact and capacity, which also relieves the effect of detecting key instance inaccurately. The B2S distance is expressed in the following form:

$$
d_k^2(X_i, S_k) = \sum\nolimits_{v=1}^{V_k} w_k^v (\mathbf{s}_k^v - \hat{\mathbf{x}}_i^v)^\top M_k (\mathbf{s}_k^v - \hat{\mathbf{x}}_i^v) \ ,
$$

where w_k^v is the v-th element of $\mathbf{w}_k$, which corresponds to the v-th instance weight.

Second, to accelerate distance computation and reduce the training cost, we form the metric M_k as a combination of T bases [54], *i.e.*, $M_k = \sum_{t=1}^{T} \alpha_k^t \mathbf{b}_t \mathbf{b}_t^\top$. $\mathbf{b}_t \in \mathcal{R}^d$ is the basis vector generated the same as [54] before training, and $\boldsymbol{\alpha}_k \in \mathcal{R}_+^T$ is the basis coefficient vector for the super-bag S_k, which have non-negative elements to ensure the positive semi-definite property of the metric M_k. Hence,

$$
\begin{aligned}
d_k^2(X_i, S_k) &= \sum_{t=1}^{T} \alpha_k^t \sum_{v=1}^{V_k} w_k^v (\mathbf{s}_k^v - \hat{\mathbf{x}}_i^v)^\top \mathbf{b}_t \mathbf{b}_t^\top (\mathbf{s}_k^v - \hat{\mathbf{x}}_i^v) \\
&= \boldsymbol{\alpha}_k{}^\top D_{ik} \mathbf{w}_k \ ,
\end{aligned}
$$

where α_k^t is the t-th basis coefficient. $D_{ik} \in \mathcal{R}^{T \times V_k}$ contains the intermediate distance values between the bag X_i and each metric basis, which can be pre-computed. Its entry $D_{ik}^{tv} = (\mathbf{s}_k^v - \hat{\mathbf{x}}_i^v)^\top \mathbf{b}_t \mathbf{b}_t^\top (\mathbf{s}_k^v - \hat{\mathbf{x}}_i^v)$ is the distance between the super-bag instance $\mathbf{s}_k^v$ and its corresponding pair $\hat{\mathbf{x}}_i^v$ in X_i based on the t-th basis $\mathbf{b}_t$. Consequently, we can reformulate the MIEN objective function in Eq. (1) to an optimization problem with the basis combination coefficient $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_0; \boldsymbol{\alpha}_1; \ldots; \boldsymbol{\alpha}_C] \in \mathcal{R}_+^{T \times (C+1)}$ and the super-bag instance weights $\mathbf{w} = [\mathbf{w}_0; \mathbf{w}_1; \ldots; \mathbf{w}_C] \in \mathcal{R}_+^{V_0 + \cdots + V_C}$:

$$
\begin{aligned}
\underset{\boldsymbol{\alpha}, \mathbf{w}}{\arg\min}\ & \sum\nolimits_{i=1}^{N} \boldsymbol{\alpha}_{y_i}{}^\top D_{iy_i} \mathbf{w}_{y_i} + \gamma(\Omega_1(\boldsymbol{\alpha}) + \Omega_2(\mathbf{w})) \\
& + \lambda \left(\sum_{i=1}^{N} \sum_{k\geq 0, k\neq y_i} \xi_{ik} + \sum_{i=1}^{N} \sum_{c\geq 1, c\neq y_i} \epsilon_{ic} \right) \\
\text{s.t.}\ & \boldsymbol{\alpha}_k{}^\top D_{ik} \mathbf{w}_k - \boldsymbol{\alpha}_{y_i}{}^\top D_{iy_i} \mathbf{w}_{y_i} \geq 1 - \xi_{ik} \\
& \boldsymbol{\alpha}_c{}^\top D_{ic} \mathbf{w}_c - \boldsymbol{\alpha}_0{}^\top D_{i0} \mathbf{w}_0 \geq 1 - \epsilon_{ic} \\
& \xi_{ik} \geq 0 \ , \ \epsilon_{ic} \geq 0 \ , \ \boldsymbol{\alpha} \geq \mathbf{0} \ , \ \mathbf{w} \geq \mathbf{0} \\
& 1 \leq i \leq N \ , \ 0 \leq k \leq C \,\text{and}\, k \neq y_i \\
& 1 \leq c \leq C \,\text{and}\, c \neq y_i \ . \qquad (2)
\end{aligned}
$$

In Eq. (2), it is obvious that both the observed classes classification and the augmented class recognition are based on non-negative coefficients $\boldsymbol{\alpha}$ and $\mathbf{w}$. $\mathbf{w}$ is the weights for all instances in the super-bags. Since all instances in the super-bags initially have the same weights, we set the regularizer as $\Omega_2(\mathbf{w}) = \|\mathbf{w} - \mathbf{1}\|_2^2$. $\mathbf{1}$ is a vector with all values equal to 1 and it has the same length as $\mathbf{w}$. For metric construction, we expect that each metric is a *sparse* combination of the same basis set built from all instances, which has a *global* view of data. The selection of bases on the one hand reduces the computational burden in the training phase. On the other hand, it makes different super-bag metrics have different *local* constructions, which can reflect the local pattern of each class (super-bag) and the meta super-bag. Thus, we configure $\Omega_1(\boldsymbol{\alpha}) = \|\boldsymbol{\alpha}\|_1$, and because of its non-negative property, the ℓ_1-norm equals to the sum of elements in $\boldsymbol{\alpha}$.

In the test phase, we compute the distance between a test bag X_* and each super-bag. Distance to each super-bag is comparable since in the training phase they have the ranking relationships. According to the principles in the training phase, if the distance to one of the C observed super-bags is the smallest, we believe the bag's class is the same as the one of it. Otherwise, as aforementioned, the meta super-bag S_0 plays a role as the reject option: If S_0 is the closest one to X_*, which indicates X_* is distant from other observed super-bags. Therefore, it is a bag from the augmented class.

3.5 Optimization for MIEN-metric

In this section, we present the optimization details of MIEN-metric. MIEN-metric is solved in an alternating style, *i.e.*, we fix $\mathbf{w}$ and solve $\boldsymbol{\alpha}$, and vice versa.

When $\mathbf{w}$ is fixed, the B2S distance can be simplified as $d^2(X_i, S_k) = \boldsymbol{\alpha}_k{}^\top G_{ik}$, where $G_{ik} = D_{ik} \mathbf{w}_k \in \mathcal{R}^T$. Because of the non-negative property of $\boldsymbol{\alpha}$, the ℓ_1-norm regularizer can be transformed to $\mathbf{1}^\top \boldsymbol{\alpha}$, and the objective function is

Algorithm 1 Optimizing the sub-problem with $\boldsymbol{\alpha}$

```
Require: Training data $(X_i, y_i), i = 1, \dots, N$, current solution of $w$, inter-mediate distance matrix $D$, parameters $\lambda$ and $\gamma$;
Compute matrix $G$ with $G_{ik} = D_{ik}\mathbf{w}_k \in \mathcal{R}^T$;
Set $t_{-1} = 0$, $t_0 = 1$ and $L_0 = \frac{1}{N}$;
Initialize $\boldsymbol{\alpha} = \mathbf{1}$;
for $s = 0, 1, \cdots$ do
    Set $\omega_s = \frac{t_{s-2}-1}{t_{s-1}}$ and $L_s = L_{s-1}$;
    Compute auxiliary solution $\hat{\boldsymbol{\alpha}}_s = \boldsymbol{\alpha}_s + \omega_s(\boldsymbol{\alpha}_s - \boldsymbol{\alpha}_{s-1})$
    Compute gradient at $\hat{\boldsymbol{\alpha}}_s$: $\nabla_\alpha = \frac{\partial \mathcal{F}(\hat{\boldsymbol{\alpha}}_s)}{\partial \boldsymbol{\alpha}}$
    while True do
        Compute $\boldsymbol{\alpha}' = \hat{\boldsymbol{\alpha}}_s - \frac{1}{L_s}\nabla_\alpha$
        Truncate negative values for current solution $\boldsymbol{\alpha}'$
        if $\mathcal{F}(\boldsymbol{\alpha}') \leq \mathcal{F}(\hat{\boldsymbol{\alpha}}_s) + \frac{\|\nabla_\alpha\|^2}{2L_s}$ then
            Break;
        end if
        $L_s = 2L_s$
    end while
    update $t_s = (1 + \sqrt{1 + 4t_{s-1}^2})/2$
    if Objective $\mathcal{F}(\boldsymbol{\alpha})$ changes smaller than a criterion then
        Break
    end if
end for
return $\boldsymbol{\alpha}_{s+1}$.
```

transformed to:

$$\begin{aligned}\arg\min_{\boldsymbol{\alpha} \geq 0} \; & \sum\nolimits_{i=1}^N {\boldsymbol{\alpha}_{y_i}}^\top G_{iy_i} + \gamma \boldsymbol{\alpha}^\top \mathbf{1} \\ & + \lambda \sum\nolimits_{i=1}^N \sum\nolimits_{k \geq 0, k \neq y_i} \ell(\mathcal{G}(k, y_i)) \\ & + \lambda \sum\nolimits_{i=1}^N \sum\nolimits_{c \geq 1, c \neq y_i} \ell(\mathcal{G}(c, 0)) \, , \end{aligned} \tag{3}$$

where $\mathcal{G}(a, b) = {\boldsymbol{\alpha}_a}^\top G_{ia} - {\boldsymbol{\alpha}_b}^\top G_{ib}$, and the loss function $\ell(\cdot)$ is the hinge loss, *i.e.*, $\ell(x) = \max(0, 1 - x)$. This sub-problem is linear on each super-bag's metric combination weights, which is a convex problem. When $\boldsymbol{\alpha}$ is fixed, the B2S distance can be simplified as $d_k^2(X_i, S_k) = E_{ik}\mathbf{w}_k$ with $E_{ik} = {\boldsymbol{\alpha}_k}^\top D_{ik} \in \mathcal{R}^{1 \times V_k}$. The objective can be also transformed into a convex sub-problem. To accelerate both two sub-problems, we substitute the hinge loss to the smoothed hinge loss, and use the accelerated projected gradient descent method to get global solutions of each sub-problem iteratively.

Specifically, we substitute the hinge loss to its smoothed version:

$$\ell_s(x) = \begin{cases} 0 & \text{if} \quad x \geq 1 \\ \frac{1}{2} - x & \text{if} \quad x \leq 0 \\ \frac{1}{2}(1 - x)^2 & \text{otherwise} \end{cases} .$$

We denote the smoothed objective as $\mathcal{F}(\boldsymbol{\alpha})$. The gradient w.r.t. $\boldsymbol{\alpha}$ can be calculated as $\frac{\partial \mathcal{F}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = [\frac{\partial \mathcal{F}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}_0}; \frac{\partial \mathcal{F}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}_1}; \dots; \frac{\partial \mathcal{F}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}_C}]$.

To clearly present the gradient computations and demonstrate different scenarios of the smoothed hinge loss, we define two types of indicator vectors for each bag, *i.e.*, $\mathcal{I}_i^1 \in \mathcal{R}^{C+1}$ and $\mathcal{I}_i^2 \in \mathcal{R}^{C+1}$ with $i = 1, \dots, N$. $\mathcal{I}^1$ and $\mathcal{I}^2$ correspond to two smoothed hinge losses in Eq. (3), respectively. The elements in $\mathcal{I}_i^1$ and $\mathcal{I}_i^2$ can be computed:

$$\mathcal{I}_{ik}^1 = \begin{cases} 0 & \text{if} \quad \mathcal{G}(k, y_i) \geq 1 \\ -1 & \text{if} \quad \mathcal{G}(k, y_i) \leq 0 \\ \mathcal{G}(k, y_i) - 1 & \quad \text{otherwise} \end{cases}$$

with $0 \leq k \leq C$, and

$$\mathcal{I}_{ic}^2 = \begin{cases} 0 & \text{if} \quad \mathcal{G}(c, 0) \geq 1 \\ -1 & \text{if} \quad \mathcal{G}(c, 0) \leq 0 \\ \mathcal{G}(c, 0) - 1 & \quad \text{otherwise} \end{cases}$$

with $1 \leq c \leq C$.

Then, $\frac{\partial \overline{\mathcal{F}}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}_0}$ is a combination of three parts:

$$\frac{\partial \mathcal{F}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}_0} = \gamma \mathbf{1} + \lambda \sum_{i=1}^N \mathcal{I}_{i0}^1 G_{i0} - \lambda \sum_{i=1}^N \sum_{c \geq 1, c \neq y_i} \mathcal{I}_{ic}^2 G_{i0} \, . \tag{4}$$

For $1 \leq c^* \leq C$, another indicator $\mathcal{I}_i^3 \in \mathcal{R}^{C+1}$ should be defined, which denotes the true class for each training bag. If $k = y_i$, $\mathcal{I}_{ik}^3 = 1$; otherwise, $\mathcal{I}_{ik}^3 = 0$. With these definitions, $\frac{\partial \mathcal{F}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}_{c^*}}$ can be computed:

$$\begin{aligned} \frac{\partial \mathcal{F}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}_{c^*}} = & \sum_{i=1}^N \mathcal{I}_{ic^*}^3 G_{iy_i} + \gamma \mathbf{1} \\ & + \lambda \sum_{i=1}^N (1 - \mathcal{I}_{ic^*}^3)(\mathcal{I}_{ic^*}^1 G_{ic^*}) - \mathcal{I}_{ic^*}^3 (\sum_{k \geq 0, k \neq y_i} \mathcal{I}_{ic^*}^1 G_{ic^*}) \\ & + \lambda \sum_{i=1}^N (1 - \mathcal{I}_{ic^*}^3)(\mathcal{I}_{ic^*}^2 G_{ic^*}) \, . \end{aligned}$$

Given the concrete formulation of gradient w.r.t. $\boldsymbol{\alpha}$, the sub-problem can be optimized with Accelerated Projected Gradient Descent [55]. The whole optimization process is summarized in Algo. 1.

When $\boldsymbol{\alpha}$ is fixed, the B2S distance can be simplified as $d_k^2(X_i, S_k) = E_{ik}\mathbf{w}_k$ with $E_{ik} = {\boldsymbol{\alpha}_k}^\top D_{ik} \in \mathcal{R}^{1 \times V_k}$. The sub-problem objective function can be formulated into:

$$\begin{aligned} \min_{\mathbf{w}} \; & \sum_{i=1}^N E_{iy_i}\mathbf{w}_{y_i} + \gamma \|\mathbf{w} - \mathbf{1}\|_2^2 \\ & + \lambda \sum_{i=1}^N \sum_{k \geq 0, k \neq y_i} \ell(E_{ik}\mathbf{w}_k - E_{iy_i}\mathbf{w}_{y_i}) \\ & + \lambda \sum_{i=1}^N \sum_{c \geq 1, c \neq y_i} \ell(E_{ic}\mathbf{w}_c - E_{i0}\mathbf{w}_0) \, . \end{aligned} \tag{5}$$

Similarly, we use the smoothed hinge loss here for accelerating. Eq. (5) can be also solved with Accelerated Gradient Descent, which has a similar procedure as that for solving $\boldsymbol{\alpha}$.

Table 1
Detailed characteristics of the MIL data sets.

Dataset	Style	♯ attribute	♯ bag	♯ instance	♯ class
Protein1	Biological	216	816	1,234	3
Protein2	Biological	216	544	805	4
Text	Text	300	2,493	46,531	12
COREL	Object	200	1,000	6,000	10
NYU-v1	Scene	200	2,284	22,505	7

Table 2
Average F1 on the biological MIL data sets.

Data set	LOF	iForest	I-MLOF	OC-MIL	OVR-MIL	MIMLNC	MIEN-metric
Protein1	41.14±1.89	47.89±1.06	51.61±1.12	48.76±1.05	58.47±1.40	53.03±1.14	**63.61±1.24**
Protein2	45.60±1.00	51.08±1.74	54.76±1.61	54.48±2.88	62.27±1.98	58.56±1.09	**68.82±1.68**

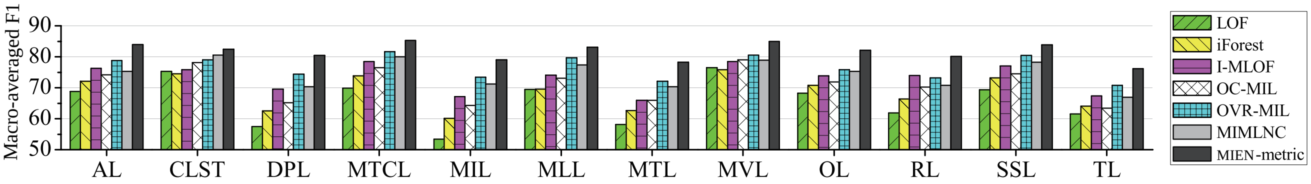

Figure 2. Comparison results on the *Text* data set. The horizontal axis presents the augmented class of the twelve text classes, and the vertical axis is the mean macro-averaged F1. (Best if viewed in color.)

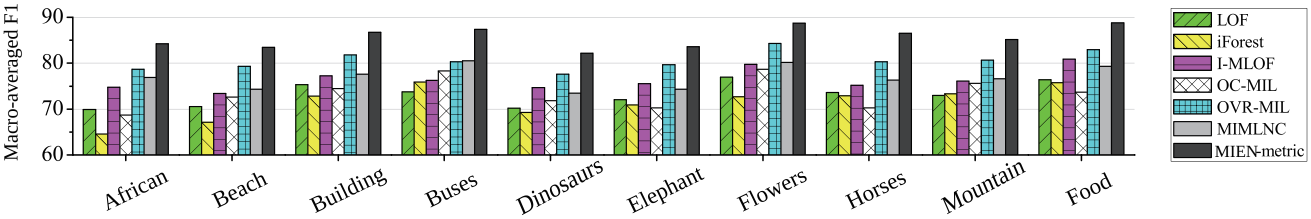

Figure 3. Comparison results for object classification. The horizontal axis presents the augmented class of the ten object-centric image classes, and the vertical axis is the mean macro-averaged F1. (Best if viewed in color.)

4 Experiments

In this section, we describe the empirical setup and the data sets used in the experiments. Then, we introduce the comparative methods in experiments.

4.1 Data sets and experimental setup

We conduct experiments on five data sets from diverse domains, *i.e.*, biological function annotation (on *Protein1* and *Protein2*), text categorization (on *Text*), object-centric image classification (on *COREL*) and scene-centric image classification (on *NYU-v1*). Detailed characteristics of these data sets are summarized in Table 1.

Similar to the *Musk* data set of drug activity prediction [1], biological protein function annotation is another typical and natural multi-instance application. In protein function, "domains" (instances) can be thought as distinct functional and structural units of a protein (bag). A protein is linked with one Gene Ontology (GO) term (class). For each protein (bag), multiple frequency vectors of 216-d represents its domains (instances). To our best knowledge, the data sets are the first biological multiple class MIL data sets, which are available at http://www.lamda.nju.edu.cn/code_MIEN.ashx.

For text categorization, one academic paper is treated as a multi-instance bag, its abstract and its references' corresponding abstracts are instances. The *Text* data set [56] selects 12 topics/classes in machine learning fields, *i.e.*, active learning (AL), clustering (CLST), deep learning (DPL), metric learning (MTCL), multi-instance learning (MIL), multi-label learning (MLL), multi-task learning (MTL), multi-view learning (MVL), online learning (OL), reinforcement learning (RL), semi-supervised learning (SSL) and transfer learning (TL). The TF-IDF feature is extracted to present an instance. After that, we use Principal Component Analysis (PCA) to reduce noise, and also reduce dimension of these instances into 300-d.

We also conduct experiments on both object-centric and scene-centric styles image classification. For object-centric image classification, we employ the popular used *COREL* image MIL data set [57], which contains 10 categories, *i.e.*, African people, beach, historical building, buses, dinosaurs, elephants, flowers, horses, mountains and food. Each category has 100 images of 384×256 image resolution. According to the suggestions of [4], we treat each image as a bag, and divide one image into six equal sized image patches (instances). For each image patch, the pre-trained AlexNet [58] model is employed to extract the 4096-d deep features. PCA is used to compress these features into 200-d.

For scene-centric classification, the *NYU-v1* data set [59] is used in experiments. *NYU-v1* has 2,284 images of 640 × 480 image resolution, and the images in the data set belong to 7 scene-level classes, *i.e.*, bathroom, bedroom, bookstore, cafe, kitchen, living room and office. Except for that, these images also supply pixel-level semantic labels. These semantic labels are regarded as the guide for separating image patches (instances). Each semantic label corresponds to one image patch, and each image patch is treated as an instance in that MIL bag (the whole image). For each image patch (instance), we use the pre-trained Places205-AlexNet [60] to extract the 4096-d deep features. There are totally 22,505 instances in the *NYU-v1* data set. Similar to *COREL*, we also employ PCA to compress the features into 200-d.

4.2 Comparative approaches

On these data sets, we compare MIEN-metric with state-of-the-art approaches including:

- **LOF**: Local Outlier Factor [37] is a single-instance outlier detector. We first represent one bag with the mean vector of all the instances in that bag, and then employ LOF to recognize the augmented class. Finally, for the rest MIL bags which are not annotated as the augmented class, we get their predictions by an one-vs-rest MIL (OVR-MIL) classifier built by miGraph [7].

- **iForest**: Isolation Forest [38] is a state-of-the-art single-instance outlier detector. For iForest, the mean vector of one bag is also used as the representation. Then, the rest procedures are similar to those in LOF.
- **I-MLOF**: Instance-neighborhood based multi-instance outlier detector [61] is a representative outlier detection algorithm in the multi-instance setting. We employ I-MLOF for recognizing the MIL augmented class, and then classify the rest bags by OVR-MIL.
- **OC-MIL**: One-class multi-instance learning [62] learns a MIL decision function for multi-instance outlier detection. In our experiments, OC-MIL is used in the training set to learn an one-class MIL classifier, and recognize the augmented class emerging in the test set. The rest bags are predicted by one-vs-rest MIL.
- **OVR-MIL**: One-vs-rest is a powerful scheme for multi-class MIL classification. As aforementioned, we build OVR-MIL on the training set based on a state-of-the-art MIL algorithm miGraph. In the original OVR-MIL, a test bag X is predicted as class y if $y = \mathrm{argmax}_{1 \le c \le C} f_c(X)$, where f_c is the binary MIL classifier for the class c. To adapt OVR-MIL for recognizing the augmented class, we let the model return the class y only when $\max_c f_c(X) > 0$, otherwise return the augmented class.
- **MIMLNC** [45] is a method for solving the multi-instance multi-label with novel classes problem. As aforementioned in related work, it has a strong assumption that instances of the novel class were indeed observed in training, just without labels. Moreover, it requires each instance has the corresponding instance-level label, which is harsh to MIL and is also impractical in MIEN. But, we still compare it with MIEN-metric as a baseline. As it is a multi-label method, we take the labels with the highest scores as the multi-class predictions.

5 Results and ablation studies

In this section, we first present the empirical results of our MIEN-metric, and compare with several state-of-the-art approaches. Then, we conduct the ablation experiments to show the robustness of our MIEN-metric. Finally, we also report the classification accuracy of five MIL benchmark data sets for traditional MIL classification.

5.1 Empirical results of MIEN

In our experiments, for *Protein1* and *Protein2*, we randomly select one class as the augmented class, and the others are regarded as the observed classes. Furthermore, to get a reliable estimate, we use a leave-one-class-out strategy on *Text*, *COREL* and *NYU-v1*, considering in turn each class as the augmented class, and the others as the observed classes. We randomly split the observed classes into two equal parts: One part is for training, and the other part plus the augmented class are for testing. The optimal parameters of these methods are obtained by cross-validations on the training sets. The tuning parameters, *i.e.*, λ and γ, are selected from a set of $\left[10^{-3}, 10^{2}, \ldots, 10^{2}, 10^{3}\right]$. The corresponding selected parameters are directly used for test sets. Experiments are repeated ten times with different training/test data splittings on *Protein1* and *Protein2*, and three times on *Text*, *COREL* and *NYU-v1*. For evaluation, we focus on the macro-averaged F1 by treating the augmented class as the $(C+1)$-th class to eliminate the influence of unbalanced data [63]. The average results of F1 are reported.

The comparison results on two protein data sets are reported in Table 2. On both protein function annotation data sets, MIEN-metric achieves the best classification performance. Because LOF and iForest lose the characteristics of MIL data, it is not a surprise that their F1 scores are the lowest ones among all the methods. In addition, because the number of bags from the augmented class is large (*i.e.*, they are not outliers which are rare and dispersed), the outlier based augmented recognition method I-MLOF can not work well. For MIMLNC, MIEN apparently does not fit its assumption. Thus, its performance is not satisfactory. Even in most cases of experiments, it can not predict the bags with novel classes. Moreover, MIEN-metric is significantly better than the one based on one-class MIL, and also outperforms OVR-MIL (which is a strong baseline method in augmented class recognition, cf. [26]) by a large margin, *i.e.*, 5∼6% F1 score. Similar trends are shown on the text categorization and both object-centric and scene-centric image classification data sets. As presented in Fig. 2, Fig. 3 and Fig. 4, the proposed MIEN-metric wins all the configurations compared with the other six augmented class recognition approaches.

5.2 Robustness experiments

In MIEN-metric, we firstly use an off-the-shelf key instance detection (KID) algorithm in [20] to extract key instances and construct the super-bags and the meta super-bag. As aforementioned in Sect. 3, the instance weights are introduced into MIEN-metric to relieve the effect of cases in which key instances might not be detected very accurately. In this section, we conduct the ablation experiments to show the robustness of MIEN-metric, *i.e.*, the classification performance of MIEN-metric is not dependent on which KID algorithm is used.

miSVM [49] is an instance-level MIL classification algorithm by explicitly treating instance-level labels as unobserved integer variables, subjected to constraints defined by their bag-level labels. Therefore, miSVM can return the bag-level predictions, and meanwhile the instance-level predictions can be also obtained. In our robustness experiments, we employ miSVM to predict the positive instances of each observed class by building one-vs-rest MIL classifiers on the training set. These positive instances are treated as the detected key instances for each observed class. As discussed in [49], miSVM was designed as a MIL classification algorithm, rather than a KID algorithm. Its key instance detection results are not very accurate. We denote this KID method as "miSVM-based". In addition, EM-DD [8] is another KID alternative, which is weaker key instance selection method and does not leverage the power of a learned classifier for key instance selection, unlike miSVM. Thus, we further employ "EM-DD based" to demonstrate the robustness of MIEN-metric.

After obtaining key instances, the rest training and testing procedures are the same as those in MIEN-metric. Table 3,

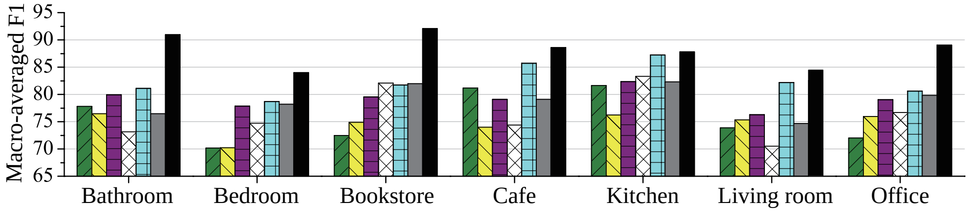

Figure 4. Comparison results for scene classification. The horizontal axis presents the augmented class of the seven scene-centric image classes, and the vertical axis is the mean macro-averaged F1. The legend is as the same as the legends in Fig. 2.

Table 3
Average F1 on the biological MIL data sets.

Method	*Protein1*	*Protein2*
miSVM-based	62.16	68.01
EM-DD based	62.07	67.89
MIEN-metric	63.61	68.82

Table 4
Average F1 on *Text*. The first row presents the augmented class of the twelve classes.

Method	*AL*	*CLST*	*DPL*	*MTCL*	*MIL*	*MLL*
miSVM-based	83.29	82.51	80.09	84.54	78.56	82.83
EM-DD based	82.91	82.22	80.01	84.22	78.12	82.56
MIEN-metric	83.94	82.46	80.49	85.27	79.02	83.14
Method	*MTL*	*MVL*	*OL*	*RL*	*SSL*	*TL*
miSVM-based	77.82	84.53	81.73	79.87	83.51	75.89
EM-DD based	77.56	84.18	81.26	79.54	83.23	75.55
MIEN-metric	78.26	84.98	82.09	80.11	83.82	76.17

Table 5
Average F1 on *COREL*. The first row presents the augmented class of the ten classes.

Method	*Afri.*	*Beach*	*Build.*	*Buses*	*Dinos.*
miSVM-based	84.13	83.24	86.48	87.11	82.12
EM-DD based	84.05	83.09	86.27	86.91	82.01
MIEN-metric	84.22	83.46	86.73	87.38	82.15
Method	*Eleph.*	*Flow.*	*Horses*	*Moun.*	*Food*
miSVM-based	83.58	88.63	86.31	85.01	88.77
EM-DD based	83.44	88.39	86.16	84.82	88.66
MIEN-metric	83.58	88.76	86.55	85.17	88.79

Table 4, Table 5 and Table 6 report the mean macro-averaged F1 scores of miSVM-based, EM-DD based and MIEN-metric on the biological, text, object-centric image and scene-centric image MIL data sets, respectively. As the results shown in these tables, for the image data set, the classification performance of miSVM-based and MIEN-metric is very similar. For the biological and text data sets, the mean macro-averaged F1 scores of miSVM-based are just slightly lower than MIEN-metric. We also do the pairwise t test between the results of miSVM-based and the ones of MIEN-metric. In the pairwise t test, we get $h = 0$ ($p = 0.8636$), which indicates that the null hypothesis cannot be rejected at the 5% significance level. That is to say that, $h = 0$ also shows miSVM-based and MIEN-metric are not significantly different from each other. For EM-DD based results, they are merely slightly lower than the results of mi-SVM based. Thus, the proposed MIEN-metric method is robust, and is agnostic to which key instance detection algorithm is used.

Table 6
Average F1 on *NYU-v1*. The first row presents the augmented class of the seven classes.

Method	*Bath.*	*Bed.*	*Book.*	*Cafe*	*Kitch.*	*Liv.*	*Offi.*
miSVM-based	90.32	83.11	92.01	87.20	87.22	81.89	89.01
EM-DD based	90.10	82.97	91.84	86.92	86.86	81.20	88.89
MIEN-metric	90.97	83.98	92.10	88.57	87.81	82.55	89.06

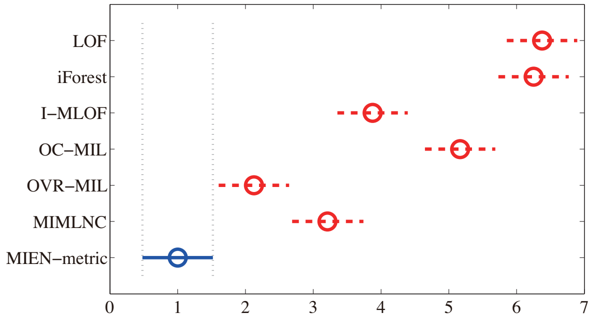

Figure 5. Friedman test for comparing MIEN-metric with other methods. The best result (*i.e.*, MIEN-metric) is presented in solid bars (colored in blue) on the leftmost side in the diagram.

5.3 Friedman test

To have an overall evaluation of the performances in the MIEN setting of the methods over these MIL data sets, we perform the Friedman test. In Fig. 5, the horizontal axis indicates the rank values. The circle is the average rank for each algorithm and the bar indicates the critical values for a two-tailed test at 90% significance level. As we can see, the average rank of MIEN-metric is the 1-st, which performs significantly better than other baseline methods.

5.4 Results on benchmark MIL data sets

To further investigate the effectiveness of MIEN-metric in a stable environment, we test it on five benchmark MIL data sets (*Musk1*, *Musk2*, *Elephant*, *Fox* and *Tiger*) for the traditional binary classification task. On each of five, we run ten times 10-fold cross validation and report the average results in Table 7. For MIEN-metric, it is allowed to return the augmented class predictions. However, once the bag is predicted as the one from the augmented class, it will be treated as a wrong prediction. For the observed classes of MIEN-metric's predictions, traditional classification accuracy is reported. Comparing with the state-of-the-art MIL algorithms, *i.e.*, miSVM [49], MIBoosting [22], miGraph [7] and miFV [64], MIEN-metric achieves comparable accuracy on these data sets. In addition, we also report the prediction ratio of novel class bags, which equals to the number of bags predicted

Table 7
Classification accuracy (mean±std.) on five MIL benchmark data sets. Note that, MIEN-metric is still allowed to return augmented class predictions in this traditional binary classification MIL setting. The prediction ratio is also reported below. ("↓": The lower, the better.) *The results show that our MIEN-metric can achieve good binary classification accuracy (in the traditional classification setting) beyond handling augmented classes well.*

Algorithm	*Musk1*	*Musk2*	*Elephant*	*Fox*	*Tiger*
miSVM	.874±.029	.836±.025	.822±.028	.582±.012	.789±.029
MIBoosting	.837±.023	.790±.023	.827±.025	.638±.022	.784±.023
miFV	.909±.042	.884±.039	.852±.027	.621±.031	.806±.054
miGraph	.889±.024	.903±.019	.869±.020	.616±.018	.801±.034
MIEN-metric	.883±.025	.881±.025	.849±.024	.610±.019	.800±.033
Prediction ratio of novel class bags (↓)	0	0.002	0.007	0.161	0.009

as from the augmented class dividing the whole number of testing MIL bags. As presented in Table 7, on *Musk1*, MIEN-metric returns *no* bags with a novel class, and the ratio of other benchmark datasets are almost less than 1% (except for *Fox*). The results prove that MIEN-metric can also overcome traditional MIL problems in a stable environment.

6 Conclusion

In this paper, we attempted to deal with a practical but challenging task, *i.e.*, the Multi-instance with Emerging Novel class (MIEN) problem. To solve MIEN, we formulated as a metric learning problem and proposed the MIEN-metric method. In MIEN-metric, we first obtain several super-bags and one meta super-bag by extracting key instances in MIL bags. After learning appropriate metrics on the Bag to Super-bag distance with a ranking restriction, MIEN-metric not only effectively handles the MIEN problem in a dynamic environment, but also achieves satisfactory results on MIL binary classification tasks in a stable environment.

An interesting future issue is to develop the advanced MIEN-metric method for multiple augmented classes in the online setting, which can incorporate the augmented classes for life-long learning. Another interesting future issue is to incorporate a MIEN-like approach into the recently proposed *abductive learning* [65], a new paradigm which encompasses machine learning and logical reasoning, to enable it handle data with inner-structure and emerging variables.

Acknowledgment

The authors would like to thank Jian-Sheng Wu for supplying raw data for the *Protein* data sets, and thank Yang Yu, De-Chuan Zhan and Sheng-Jun Huang for valuable discussions. The authors also thank the anonymous reviewers, whose comments have helped improving this paper.

References

[1] T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, "Solving the multiple instance problem with axis-parallel rectangles," *Artificial Intelligence*, vol. 89, no. 1-2, pp. 31–71, 1997.

[2] Z.-H. Zhou, "A brief introduction to weakly supervised learning," *National Science Review*, vol. 1, no. 5, pp. 44–53, 2018.

[3] J. Amores, "Multiple instance classification: Review, taxonomy and comparative study," *Artificial Intelligence*, vol. 201, pp. 81–105, 2013.

[4] X.-S. Wei and Z.-H. Zhou, "An empirical study on image bag generators for multi-instance learning," *Machine Learning*, vol. 105, no. 2, pp. 155–198, 2016.

[5] V. Cheplygina, D. M. J. Tax, and M. Loog, "Dissimilarity-based ensembles for multiple instance learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1379–1391, 2016.

[6] Y. Chen, J. Bi, and J.-Z. Wang, "MILES: Multiple-instance learning via embedded instance selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 1931–1947, 2006.

[7] Z.-H. Zhou, Y.-Y. Sun, and Y.-F. Li, "Multi-instance learning by treating instances as non-I.I.D. samples," in *Proceedings of International Conference on Machine Learning*, 2009, pp. 1249–1256.

[8] Q. Zhang and S. A. Goldman, "EM-DD: An Improved Multiple-Instance Learning Technique," in *Proceedings of IEEE International Conference on Data Engineering*, 2000, pp. 233–243.

[9] D. Zhang, J. He, and R. Lawrence, "MI2LS: Multi-instance learning from multiple informationsources," in *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2013, pp. 149–157.

[10] L. Zhu, B. Zhao, and Y. Gao, "Multi-class multi-instance learning for lung cancer image classification based on bag feature selection," in *Fuzzy Systems and Knowledge Discovery*, 2008, pp. 487–492.

[11] S. Vijayanarasimhan and K. Grauman, "Keywords to visual categories: multiple-instance learning for weakly supervised object categorization," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[12] Y.-T. Chen, C.-S. Chen, Y.-P. Hung, and K.-Y. Chang, "Multi-class multi-instance boosting for part-based human detection," in *Proceedings of IEEE International Conference on Computer Vision*, 2009, pp. 1177–1184.

[13] Z.-H. Zhou, "Learnware: On the future of machine learning," *Frontiers of Computer Science*, vol. 10, no. 4, pp. 589–590, 2016.

[14] T. G. Dietterich, "Robust artificial intelligence and robust human organizations," *Frontiers of Computer Science*, vol. 13, no. 1, pp. 1–3, 2019.

[15] X. Mu, K. M. Ting, and Z.-H. Zhou, "Classification under streaming emerging new classes: A solution using completely-random trees." *IEEE IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 8, pp. 1605–1618, 2017.

[16] Y. Zhu, K. M. Ting, and Z.-H. Zhou, "Multi-label learning with emerging new labels," in *Proceedings of IEEE International Conference on Data Mining*, 2016, pp. 1371–1376.

[17] M. Fink, S. Shalev-Shwartz, Y. Singer, and S. Ullman, "Online multiclass learning by interclass hypothesis sharing," in *Proceedings of International Conference on Machine Learning*, 2006, pp. 313–320.

[18] I. Kuzborskij, F. Orabona, and B. Caputo, "From N to N+1: Multiclass transfer incremental learning," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3358–3365.

[19] W. J. Scheirer, A. R. Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 1757–1772, 2013.

[20] X. Wang, B. Wang, X. Bai, W. Liu, and Z. Tu, "Max-margin multiple-instance dictionary learning," in *Proceedings of International Conference on Machine Learning*, 2013, pp. 846–854.

[21] G. Liu, J. Wu, and Z.-H. Zhou, "Key instance detection in multi-instance learning," in *Proceedings of Asian Conference on Machine Learning*, 2012.

[22] X. Xu and E. Frank, "Logistic regression and boosting for labeled bags of instances," in *Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2004, pp. 272–281.

[23] A. Fern and R. Givan, "Online ensemble learning: An empirical study," *Machine Learning*, vol. 53, pp. 71–109, 2003.

[24] R. Polikar, L. Upda, S. Upda, and V. Honavar, "Learn++: An incremental learning algorithm for supervised neural networks," *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. 31, pp. 497–508, 2001.

[25] L. Zhu, B. Zhao, and Y. Gao, "Learning using hidden information (learning with teacher)," in *Proceedings of International Joint Conference on Neural Networks*, 2009, pp. 3188–3195.

[26] Q. Da, Y. Yu, and Z.-H. Zhou, "Learning with augmented class by exploiting unlabeled data," in *Proceedings of AAAI Conference on Artificial Intelligence*, 2014, pp. 1760–1766.

[27] Z.-H. Zhou and M.-L. Zhang, "Solving multi-instance problems with classifier ensemble based on constructive clustering," *Knowledge and Information Systems*, vol. 11, no. 2, pp. 155–170, 2007.

[28] C. H. Lampert, H. Nickisch, and S. Harmeling, "Learning to detect unseen object classes by between-class attribute transfer," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 951–958.

[29] R. Socher, M. Ganjoo, C. D. Manning, and A. Y. Ng, "Zero-shot learning through cross-modal transfer," in *Advances in Neural Information Processing Systems*, 2013, pp. 935–943.

[30] E. Kodirov, T. Xiang, Z. Fu, and S. Gong, "Unsupervised domain adaptation for zero-shot learning," in *Proceedings of IEEE International Conference on Computer Vision*, 2015, pp. 2452–2460.

[31] X. Li, Y. Guo, and D. Schuurmans, "Semi-supervised zero-shot classification with label representation learning," in *Proceedings of IEEE International Conference on Computer Vision*, 2015, pp. 4211–4219.

[32] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection : A survey," *ACM Computing Surveys*, vol. 41, no. 15, pp. 1–72, 2009.

[33] K. Das and J. Schneider, "Detecting anomalous records in categorical datasets," in *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2007, pp. 220–229.

[34] P. Sun and S. Chawla, "On local spatial outliers," in *Proceedings of IEEE International Conference on Data Mining*, 2004, pp. 209–216.

[35] M. Wu and C. Jermaine, "Outlier detection by sampling with accuracy guarantees," in *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2006, pp. 767–772.

[36] Y.-J. Lee, Y.-R. Yeh, and Y.-C. F. Wang, "Anomaly detection via online oversampling principal component analysis," *IEEE IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 7, pp. 1460–1470, 2013.

[37] M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in *Proceedings of ACM SIGMOD Conference on Management Of Data*, 2000, pp. 93–104.

[38] F. T. Liu, K.-M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proceedings of IEEE International Conference on Data Mining*, 2008, pp. 413–422.

[39] W.-J. Li and D. Y. Yeung, "MILD: Multiple-instance learning via disambiguation," *IEEE IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 1, pp. 76–89, 2010.

[40] J. Wu, S. Pan, X. Zhu, C. Zhang, and X. Wu, "Multi-instance learning with discriminative bag mapping," *IEEE IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 6, pp. 1065–1080, 2018.

[41] H. Yuan, M. Fang, and X. Zhu, "Hierarchical sampling for multi-instance ensemble learning," *IEEE IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 12, pp. 2900–2905, 2013.

[42] W.-J. Zhang and Z.-H. Zhou, "Multi-instance learning with distribution change," in *Proceedings of AAAI Conference on Artificial Intelligence*, 2014, pp. 2184–2190.

[43] Y.-L. Zhang and Z.-H. Zhou, "Multi-instance learning with key instance shift," in *Proceedings of International Joint Conference on Artificial Intelligence*, 2017, pp. 3441–3447.

[44] Q. Wang, L. Ruan, and L. Si, "Adaptive knowledge transfer for multiple instance learning in image classification," in *Proceedings of AAAI Conference on Artificial Intelligence*, 2014, pp. 1334–1340.

[45] A. T. Pham, R. Raich, X. Z. Fern, and J. R. Arriaga, "Multi-instance multi-label learning in the presence of novel class instances," in *Proceedings of International Conference on Machine Learning*, 2015, pp. 2427–2435.

[46] D. Kotzias, M. Denil, N. de Freitas, and P. Smyth, "MI2LS: Multi-instance learning from multiple informationsources," in *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2013, pp. 149–157.

[47] X. Xu and B. Li, "Multiple class multiple-instance learning and its application to image categorization," *Intenational Journal of Image and Graphics*, vol. 7, pp. 427–444, 2007.

[48] O. Maron and T. Lozano-Pérez, "A framework for multiple-instance learning," in *Advances in Neural Information Processing Systems*, 1998, pp. 570–576.

[49] S. Andrews, I. Tsochantaridis, and T. Hofmann, "Support vector machines for multiple-instance learning," in *Advances in Neural Information Processing Systems*, 2002, pp. 561–568.

[50] Z. Fu, A. Robles-Kelly, and J. Zhou, "MILIS: Multiple instance learning with instance selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 958–977, 2011.

[51] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, pp. 207–244, 2009.

[52] H. Wang, F. Nie, and H. Huang, "Learning instance specific distance for multi-instance classification," in *Proceedings of AAAI Conference on Artificial Intelligence*, 2011, pp. 507–512.

[53] H. Wang, H. Huang, F. Kamangar, F. Nie, and C. Ding, "Maximum margin multi-instance learning," in *Advances in Neural Information Processing Systems*, 2011, pp. 1–9.

[54] Y. Shi, A. Bellet, and F. Sha, "Sparse compositional metric learning," in *Proceedings of AAAI Conference on Artificial Intelligence*, 2014, pp. 2078–2084.

[55] N. Li, R. Jin, and Z.-H. Zhou, "Top rank optimization in linear time," in *Advances in Neural Information Processing Systems*, 2014, pp. 1502–1510.

[56] Y. Zhu, J. Wu, Y. Jiang, and Z.-H. Zhou, "Learning with augmented multi-instance view," in *Proceedings of Asian Conference on Machine Learning*, 2014, pp. 234–249.

[57] Y. Chen and J. Wang, "Image categorization by learning and reasoning with regions," *Journal of Machine Learning Research*, vol. 5, pp. 913–939, 2004.

[58] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[59] N. Silberman and R. Fergus, "Indoor scene segmentation using a structured light sensor," in *Proceedings of Proceedings of IEEE International Conference on Computer Vision Workshops*, 2011, pp. 601–608.

[60] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in *Advances in Neural Information Processing Systems*, 2014, pp. 487–495.

[61] O. Wu, J. Gao, W. Hu, B. Li, and M. Zhu, "Identifying multi-instance outliers," in *Proceedings of SIAM International Conference on Data Mining*, 2010, pp. 430–441.

[62] X. Wang, Z. Zhang, Y. Ma, X. Bai, W. Liu, and Z. Tu, "One-class multiple instance learning via robust PCA for common object discovery," in *Proceedings of Asian Conference on Computer Vision*, 2014, pp. 246–258.

[63] X.-Y. Liu, S.-T. Wang, and M.-L. Zhang, "Transfer synthetic over-sampling for class-imbalance learning with limited minority class data," *Frontiers of Computer Science*, vol. 13, no. 5, pp. 996–1009, 2019.

[64] X.-S. Wei, J. Wu, and Z.-H. Zhou, "Scalable algorithms for multi-instance learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 4, pp. 975–987, 2017.

[65] Z.-H. Zhou, "Abductive learning: Towards bridging machine learning and logical reasoning," *Science China Information Sciences*, vol. 62, no. 076101, 2019.

Xiu-Shen Wei (M'18) received his BS degree in computer science, and his Ph.D. degree in computer science and technology from Nanjing University. He is now the Research Director of Megvii Research Nanjing, Megvii Technology, China. He has published about twenty academic papers on the top-tier international journals and conferences, such as IEEE TPAMI, IEEE TIP, IEEE TNNLS, Machine Learning Journal, CVPR, ICCV, IJCAI, ICDM, ACCV, etc. He achieved the first place in the iNaturalist competition (in association with CVPR 2019), the first place in the Apparent Personality Analysis competition (in association with ECCV 2016) and the first runner-up in the Cultural Event Recognition competition (in association with ICCV 2015) as the team director. He also received the Presidential Special Scholarship (the highest honor for Ph.D. students) in Nanjing University, and received the Outstanding Reviewer Award in CVPR 2017. His research interests are computer vision and machine learning. He has served as a PC member of ICCV, CVPR, ECCV, NIPS, IJCAI, AAAI, etc. He is a member of the IEEE.

Han-Jia Ye received his Ph.D. degree in computer science, Nanjing University, China in 2019. In the same year, he became an assistant professor in the National Key Lab for Novel Software Technology, the School of Artificial Intelligence at Nanjing University, China. His research interests lie primarily in machine learning, including distance metric learning, multi-modal/multi-task learning, few-shot learning, meta-learning, and semantic mining.

Xin Mu is currently working as a senior researcher at Advertising and Marketing Services, Tencent Inc. He got his Ph.D. degree in the LAMDA Group at Nanjing University, advised by Prof. Zhi-Hua Zhou. He visited at Singapore Management University in the LARC lab as the research assistant, supervised by Prof. Ee-Peng Lim and Prof. Feida Zhu from May 2015 to May 2018. His research interest is mainly on machine learning and data mining. He is currently working on ensemble methods, stream mining and creating deep learning models for user profile modeling in advertising recommendation systems. He has won the second place in the 5th China Conference on Data Mining (Data Mining Competition).

Jianxin Wu (M'09) received his BS and MS degrees in computer science from Nanjing University, and his PhD degree in computer science from the Georgia Institute of Technology. He is currently a professor in the Department of Computer Science and Technology at Nanjing University, China, and is associated with the National Key Laboratory for Novel Software Technology, China. He has served as an area chair for CVPR, ICCV and AAAI. His research interests are computer vision and machine learning.

Chunhua Shen is a Professor at School of Computer Science, University of Adelaide. He is a Project Leader and Chief Investigator at the Australian Research Council Centre of Excellence for Robotic Vision (ACRV), for which he leads the project on machine learning for robotic vision. Before he moved to Adelaide as a Senior Lecturer, he was with the computer vision program at NICTA (National ICT Australia), Canberra Research Laboratory for about six years. His research interests are in the intersection of computer vision and statistical machine learning. He studied at Nanjing University, at Australian National University, and received his PhD degree from the University of Adelaide. From 2012 to 2016 he held an Australian Research Council Future Fellowship.

Zhi-Hua Zhou (S'00-M'01-SM'06-F'13) received the BSc, MSc and PhD degrees in computer science from Nanjing University, China, in 1996, 1998 and 2000, respectively, all with the highest honors. He joined the Department of Computer Science & Technology at Nanjing University as an Assistant Professor in 2001, and is currently Professor, Head of the Department of Computer Science and Technology, and Dean of the School of Artificial Intelligence; he is also the Founding Director of the LAMDA group. His research interests are mainly in artificial intelligence, machine learning and data mining. He has authored the books *Ensemble Methods: Foundations and Algorithms*, *Evolutionary Learning: Advances in Theories and Algorithms*, *Machine Learning* (in Chinese), and published more than 150 papers in top-tier international journals or conference proceedings. He has received various awards/honors including the National Natural Science Award of China, the IEEE Computer Society Edward J. McCluskey Technical Achievement Award, the PAKDD Distinguished Contribution Award, the IEEE ICDM Outstanding Service Award, the Microsoft Professorship Award, etc. He also holds 24 patents. He is the Editor-in-Chief of the *Frontiers of Computer Science*, Associate Editor-in-Chief of the *Science China Information Sciences*, Action or Associate Editor of the *Machine Learning*, *IEEE Transactions on Pattern Analysis and Machine Intelligence* , *ACM Transactions on Knowledge Discovery from Data*, etc. He served as Associate Editor-in-Chief for *Chinese Science Bulletin* (2008-2014), Associate Editor for *IEEE Transactions on Knowledge and Data Engineering* (2008-2012), *IEEE Transactions on Neural Networks and Learning Systems* (2014-2017), *ACM Transactions on Intelligent Systems and Technology* (2009-2017), *Neural Networks* (2014-2016), etc. He founded ACML (Asian Conference on Machine Learning), served as Advisory Committee member for IJCAI (2015-2016), Steering Committee member for ICDM, PAKDD and PRICAI, and Chair of various conferences such as Program co-chair of AAAI 2019, General co-chair of ICDM 2016, and Area chair of NeurIPS, ICML, AAAI, IJCAI, KDD, etc. He was the Chair of the IEEE CIS Data Mining Technical Committee (2015-2016), the Chair of the CCF-AI (2012-2019), and the Chair of the CAAI Machine Learning Technical Committee (2006-2015). He is a foreign member of the Academy of Europe, and a Fellow of the ACM, AAAI, AAAS, IEEE, IAPR, IET/IEE, CCF, and CAAI.