Non-aligned Multi-view Multi-label Classification via Learning View-specific Labels

Dawei Zhao, Qingwei Gao, Yixiang Lu, and Dong Sun

***Abstract*—In the multi-view multi-label (MVML) classification problem , multiple views are simultaneously associated with multiple semantic representations. Multi-view multi-label learning inevitably has the problems of consistency, diversity, and non-alignment among views and the correlation among labels. Most of the existing multi-view multi-label methods for non-aligned views assume that each view has a common or shared label set, but because a single view cannot contain the entire label information, they often learn suboptimal results. Based on this, this paper proposes a non-aligned multi-view multi-label classification method that learns view-specific labels (LVSL), aiming to explicitly mine the information of view-specific labels and low-rank label structures in non-aligned views in a unified model framework. Furthermore, to alleviate insufficient available label information, we thoroughly explored the global and local structural information among labels. Specifically, first, we assume that there is structural consistency between the view and the label space and then construct the view-specific label model in turn. Second, to enrich the original label space information, we mine the consistent information of multiple views and the low-rank correlation information hidden among multiple labels. Finally, the contribution weight of each view is combined with learning the complementary information among the views in the decision-making stage, and extend the model to handle nonlinear data. The results of the proposed method compared with existing state-of-the-art algorithms on several datasets validate its effectiveness.**

***Index Terms*—Multi-view multi-label learning, Non-aligned view, View-specific labels learning, Low-rank label correlations, Manifold regularization learning, Kernel extension**

I. Introduction

MVML is used to describe multi-semantic problems of multi-source heterogeneous data objects[1], [2], [3]. In Fig.1, given a natural scene image, it can be represented by multiple view structures (LBP, HOG, HSV) with multiple labels (blue sky, white clouds, desert). Multi-view multi-label is a learning framework for handling high-dimensional heterogeneous multi-semantic data classification problems. Multi-view learning[4], [5], [6], [7] can describe data objects more comprehensively and accurately than single-view learning. For example, video labeled as "Sports", "National Basketball Association", and " Basketball Stars" is represented simultaneously by diverse data forms, such as text, image, and audio. In addition, there are learning paradigms with different perspectives under the same modality. For example, we can use various feature forms to describe image data (texture description, shape description, color, etc.). With the emergence of big data and the rapid development of data collection technology, people are bound to face data classification problems in more complex and changeable real-world scenarios. In the past few decades, multi-view and multi-label learning[8], [9], [10] have been extensively studied as two separate research fields. A fundamental assumption of conventional single-label learning is that the relationships among labels are mutually exclusive. In multi-label learning, the semantic information of the labels is rich, and there is mutual dependence among the labels, which is a theoretical conflict with single-label learning. To solve more complex data classification problems in real-world scenarios, the MVML framework has emerged.

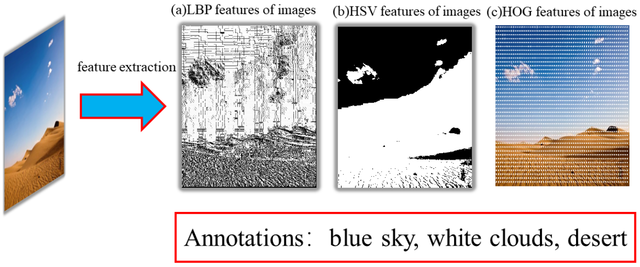

(a)LBP features of images
(b)HSV features of images
(c)HOG features of images
feature extraction
Annotations: blue sky, white clouds, desert

Fig. 1. An example of the multi-view multi-label learning object. (a) LBP features. (b) HSV features. (c) HOG features.

The existing methods have the following problems in the existing methods that urgently need to be solved:

1) There are two major principles in multi-view learning: consistency and diversity in multi-source heterogeneous data[11], [12]. The principle of consistency asserts that it is necessary to keep the consistent information of multiple views as much as possible in multi-view learning. The diversity principle advocates that each view should learn complementary information among views while completing its specific knowledge discovery task.
2) Label correlation learning problem [13], [14]. The correlation among labels in multi-view learning is one of the critical factors for improving multi-label classification performance.

This work is supported by the National Natural Science Foundation of China (No. 61502003) and the Natural Science Research Project of Anhui Province (No. 1608085MF125) (Corresponding author: Qingwei Gao.)

Dawei Zhao is with the School of Electrical Engineering and Automation and with the School of Computer and Technology, Anhui University, Hefei 230601, PR China. E-mail: zhaodwahu@163.com.

Qingwei Gao, Yixiang Lu and Dong Sun are with the School of Electrical Engineering and Automation, Anhui University, Hefei 230601, PR China. (E-mail: qingweigao@ahu.edu.cn; lyxahu@ahu.edu.cn; sundong@ahu.edu.cn)

3) The non-aligned multi-view learning problem[15]. In most multi-view learning methods, it is often explicitly or implicitly assumed that the view samples are uniformly aligned, but in reality, it is often difficult to obtain fully consistent multi-view information. For example: in video recommendations, label data are obtained from different video software, but due to the privacy protection principle of users, we cannot match and align these data with the same user consistently[16]. In the field of face recognition, due to the failure of face landmark detection, multi-view faces cannot be aligned, which harms facial expression recognition[17]. In general, there are many non-aligned multi-view data in the real world, and a single view cannot contain all the label information. Otherwise, multi-view learning will lose its meaning.

Therefore, we naturally face the following challenges: one is how to solve these three problems simultaneously, and the other is to solve the linear inseparability problem of the given data. According to the different solutions, we divide the existing strategies into two types: feature fusion and classification fusion[18], [19]:

The feature fusion strategy usually considers transforming the problem into a multi-view shared subspace information extraction problem and degenerates the multi-view heterogeneous feature information into a multi-label learning problem after fusion[20], [21], [22], [23]. The matrix factorization method[24] is often used to obtain the shared subspace information of the multi-view data and then uses the shared information among the views and the label information of the labeled samples to learn the discriminant predictor. The effectiveness of subspace learning relies on the accurate acquisition of consensus representations, but low-dimensional consensus representation learning becomes more difficult as the number of views increases.

The classification fusion strategy divides the problem into multiple multi-label learning problems and then predicts the unknown example label set by assigning a weight to each view classifier[18], [19], [25], [26]. Because a unified predictor needs to be learned for each view, the classification fusion strategy forces each view to learn common sample label information to learn multiple views and consistent information across multiple labels and assigns different views to each view weight to learn complementary information for this view. Such methods can effectively learn view diversity information, and these individual modes can also improve the robustness of the predictor. Clearly, individual models rely heavily on the performance of each individual classifier. Since it is impossible to label each view separately in reality, the label information learned by this type of method is often the general label information.

Most of the existing methods focus on the first two challenges. For the third problem, the literature [15] gives a mitigation scheme: although the samples among views are not aligned, they can still be implicitly connected through common or shared labels to be learned complementarily. However, this strategy is suboptimal because it assumes that all views have a uniform set of labels. In practice, there is a problem of inconsistent views with their corresponding labels[27]. The intuitive explanation is that each view only observes a part of the corresponding label information, so different views have specific label sets. For example, in Fig. 1, we observe that in subgraphs (a), (b), and (c), all three different views can only obtain a part of the complete label information. Subspace learning can avoid the effect of inconsistent labels for views, it does not focus on the problem of non-aligned multi-views.

With our existing knowledge, it is impossible to learn view-specific features and multi-label structures jointly. Additionally, the data of each view have a complex nonlinear structure, so linear models are no longer sufficient for current needs. This paper proposes an MVML method for jointly learning view-specific labels and multi-label structural information. Specifically, first, a view-specific label matrix is learned based on the structural assumption of similarity between multi-view features and labels. Then, the global label structure and local structure correlation are introduced to enrich view-specific label information. Finally, the joint learning model is extended to nonlinear models.

We designed the model to establish the final optimization goal to study the above problems jointly. Fig.2 illustrates the model framework of the proposed method. The most significant difference between our method and the existing multi-view learning method is that the latter ignores the misalignment of multi-source heterogeneous features and label space. Our experiments prove that this view-specific label learning structure plays an indispensable role. Our main contributions in this paper are as follows:

1) We propose a novel MVML method, that combines view-specific labels and label structure learning.
2) Our method mines view-specific label information for multi-view consistency and complementary information learning.
3) We extend the linear model to the nonlinear model to solve scenarios where the given data are not be linearly separable.

The rest of this article is organized as follows. In Section II, we briefly summarize the related work of multi-view multi-label learning. Section III proposes our method, and Section IV proposes an effective alternative iterative optimization solution method to solve it. A large number of experimental results and analyses are reported in Section V. Section VI summarizes the research directions of this article.

II. Related work

The previous section divided existing approaches into two different strategies, depending on the solution. In this section, we outline the latest research that is closely related to our approach based on the above taxonomy.

A. Multi-view Multi-label learning

Direct feature fusion is a method that connects the features of all views in series for classification. For example, RLM-MCML[26] merges multi-view features through a simple concatenation strategy. Meanwhile, the structural relationship among labels is learned based on low-rank labels and sample

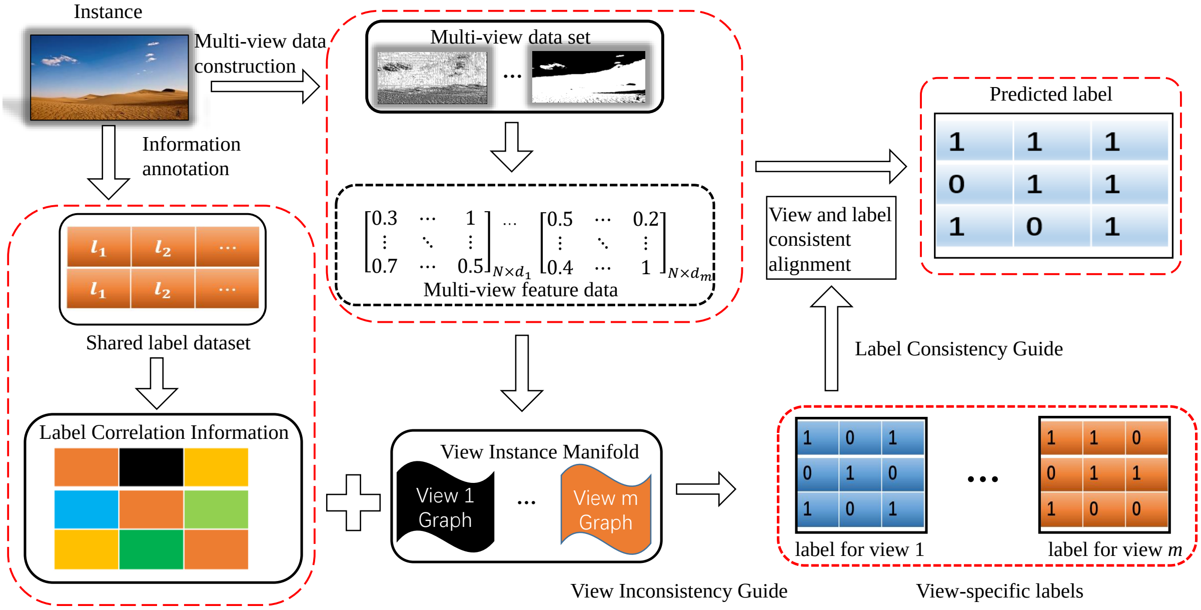


Fig. 2. The framework of the proposed LVSL method. High-order label correlation information is used to augment and complete the shared label set. View inconsistency is guided by view-specific label learning, and label consistency is guided by view-label alignment learning. LVSL combines multi-view feature data with the consistent alignment of views and labels for non-aligned multi-view multi-label classification tasks.

local smoothness assumptions. This degenerate method of merging ignores the unique physical meaning of the view itself. Simultaneously, the high-dimensional heterogeneous features obtained by the merging strategy may lead to the curse of dimensionality and overfitting. The subspace learning method considers that all views have a latent common representation to build a classification model, a feature fusion strategy. For example: in lrMMC[28], the first stage captures the low-dimensional common representation of all views, limits it to a low-rank matrix, and then assigns specific weights to each view to explore the complementarity between different views. In the second stage, the consensus matrix is embedded in the matrix completion for classification. The difference between TMV-LE[22] and lrMMC is that tensor factorization technology is added to learn the high-order relationship between different views when using subspace learning to mine public representations. In addition, the label enhancement method is used when performing multi-label classification. GLMVML[29] learns a consensus multi-view representation through matrix factorization and encodes complementary information from different views. In addition, it also learns global and local label structural information. iMvWL[20] attempts to capture a distinguishable shared subspace from incomplete views through nonnegative matrix factorization and local label structure learning, thereby constructing a robust weak label classifier. LSA-MML[23] uses subspace learning to force the alignment of undiscovered latent patterns to obtain a public representation, revealing the latent semantic patterns in the data. ICM2L[21] utilizes nonnegative matrix factorization to learn the individual and common information of different views, thereby improving the recognition ability of the classifier on rare labels. MLMVL-MM[30] uses multi-label correlation information to merge multiple feature views and maximum margin classification simultaneously. However, with the subspace method, as the number of views increases, it becomes more challenging to learn an effective latent low-dimensional consistency representation, which leads to decrease in the performance of the algorithm.

Classification fusion: Multiple views are fused to perform multi-label classification in the prediction stage. For example, VLSF[31] leverages pairwise label correlations and views contributions to learn view label-specific features in multi-view multi-label learning, addressing the issues of view consistency and complementarity. GRADIS[32] adopts a two-stage label disambiguation method to solve the multi-view partial multi-label problem. First, the candidate labels are disambiguated based on the fusion similarity graph, and the ground-truth labels of the training samples are estimated; then, the disambiguation-guided clustering analysis is used to generate a prediction model for learning label-specific features. $NAIM^3L$[15] uses a classification fusion strategy to describe the global and local structures among labels as high-rank and low-rank, respectively, to alleviate the problem of insufficient available labels, which simultaneously solves the learning problems of missing labels, incomplete views, and non-aligned views. F2L21F[33] proposes a sparse framework for image classification. MLSO[3] builds an SVM classifier based on each data view and jointly learns multi-source multi-label learning tasks under a unified optimization framework. Multi-label classification results are obtained by a weighted combination of decisions from multiple sources. The classification fusion methods generally consider that although the various views are not explicitly aligned, they can still be implicitly connected through public or shared labels[15]. Nevertheless, intuitively, each view has only a subset of the corresponding labels, meaning each view can only catch a subset of common or shared label data. Therefore, there are obvious shortcomings in the premises of the methods mentioned above based on classification fusion.

In addition, the existing multi-view multi-label learning methods have achieved certain results, but most of them are based on linear models. When a given dataset is linearly inseparable, we may not achieve the expected classification effect. For this reason, scholars add nonlinear mapping to the model. For example, TM3L[18] is a two-step learning strategy. The first step is to learn a common representation of multiple views with complementarity and consistency through subspaces, and the second step combines label correlation to build a nonlinear multi-label classifier model. MVLE[34] utilizes the low-dimensional latent semantic space to connect the labels and features of different views and further uses the Hilbert-Schmidt independence criterion (HSIC)[35] to mine the consistency information among different views. SIMM[36] proposes a neural network MVML method, which uses the shared subspace learning and view-specific information identification. On this basis, MML-DAN[37] adopts a self-attention mechanism to model the interaction information of label-specific views to explore consistent label correlations. CDMM[19] utilizes multiple multi-label models to learn view consistency information jointly and introduces HSIC theory to extract the different information among views.

B. Label correlation learning

Different from traditional single-label learning tasks, multi-label learning aims to assign multiple category labels to a sample, which has gained increasing attention in different machine learning tasks. From an intuitive point of view, samples with similar labels are more likely to have strong correlations[38]. Therefore, the existing multi-label methods are divided into three categories according to the different label correlations used[9]. First-order strategies: consider that there is no inherent correlation among labels and that labels are independent of each other[39], [40]. Second-order strategies: consider that the label correlation exists in pairs, and use the distance measurement method to evaluate the correlation of the label pairs[31], [41]. High-order strategies: consider that label correlation in complex scenarios is multifaceted and semantically related [42], [43]. Theoretical research on label propagation dependencies shows that label correlations can reconstruct and enrich original label information[44].

In addition, most of the previous label correlation studies considered the global structural information of labels, but more studies confirmed that the correlation among labels might only be shared with a subset of samples[38]. Therefore, there is a weak correlation or irrelevance among samples with different labels, reflecting the local structural relationship within multiple labels[45]. ML-LRC[46] uses a low-rank structure to capture the complex associations among labels and jointly learns label correlation and multi-label classifiers; GLOCAL[47] builds the global sum of labels by combining multiple regularizers of labels in a multi-label classifier of local structural relationships.

As mentioned above, most of the existing MVML methods consider that all views share a set of labels, but in practical applications, there is a problem of inconsistent view-label information. Moreover, this problem caused by non-aligned view learning has not been directly investigated in previous studies. We propose an MVML method for learning view-specific labels based on the aforementioned issue. First, view-specific label learning addresses the view-label inconsistency of non-aligned views. Then, effective global and local structural regularizers for label correlations are introduced into view-specific label learning. Finally, the complementary information among views is learned by a weighted combination of each view, and the model is extended nonlinearly. The effectiveness of our method is verified on multiple benchmark multi-view multi-label data sets.

III. The proposed method

A. Problem settings

Let $\boldsymbol{X} = \{\boldsymbol{x}^v\}_{v=1}^m$ denote multi-view multi-label data sets with m views, where $\boldsymbol{X}^v = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N]^{\mathrm{T}} \in \mathbb{R}^{N\times d_v}$ is the complete feature space of the v-th view, N represents the number of training samples. $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N] \in \mathbb{R}^{N\times l}$ represents the label space corresponding to the feature set, where $\boldsymbol{y}_i \in \{0,1\}^{N\times l}$ is the label vector of $\boldsymbol{x}_i$, and l represents the number of labels.

B. Problem Formulation

In the initial prediction model of multi-view multi-label classification, label classification learning is a typical regression model problem. The base model advocates different views to predict the same label result to use consistent information between different views. Furthermore, the different contribution weights of each view are considered in the base model to learn the complementary information among views. The objective function can be formally defined as follows:

$$\min_{\boldsymbol{W}^v, \boldsymbol{\theta}^v} \frac{1}{2} \sum_{v=1}^{m} \boldsymbol{\theta}^v \left\{ \left\| \boldsymbol{X}^v \boldsymbol{W}^v - \boldsymbol{Y} \right\|_F^2 \right\} + \frac{\lambda_1}{2} \left\| \boldsymbol{W}^v \right\|_F^2 + \frac{\lambda_2}{2} \left\| \boldsymbol{\theta} \right\|_2^2$$
$$\text{s.t. } \boldsymbol{\theta} \geqslant 0, \ \sum_{v=1}^{m} \boldsymbol{\theta}^v = 1 \tag{1}$$

The variable $\boldsymbol{\theta}^v$ is used to measure the contribution of each view.

There are two main problems currently faced:

1) We need to learn non-aligned views in a common label space.
2) The introduction of multi-label structural learning in multi-label learning helps to improve the classification performance of the algorithm.

Therefore, how to combine these two attributes more effectively and make our model more discriminative is the main issue to be considered below.

Eq.1 assumes that the samples among views share a common label set, which is an implicit solution to view alignment consistency. However, there is no such explicit or implicit alignment view sample in a large amount of data in reality because the labels that each view in the real world can observe may only be part of the entire information, so it is necessary to learn a particular non-aligned multi-view method that solves the inconsistency of observable information in each view. For the first question, we propose a display view non-alignment method, introducing the concept of view-specific labels. Then, we have the following equation:

$$\begin{aligned}&\min_{\boldsymbol{W}^v,\boldsymbol{P}^v,\boldsymbol{\theta^v}} \frac{1}{2}\sum_{v=1}^{m}\left\{F\left(\boldsymbol{W}^v\right)+\lambda_1\left\|\boldsymbol{W}^v\right\|_F^2+\lambda_4\left\|\boldsymbol{P}^v-\boldsymbol{Y}\right\|_F^2\right\}\\&+\frac{\lambda_2}{2}\left\|\boldsymbol{\theta}\right\|_2^2\\&\text{s.t. }\ \boldsymbol{\theta}\geqslant 0,\ \sum_{v=1}^{m}\boldsymbol{\theta}^v=1\end{aligned} \tag{2}$$

$$F\left(\boldsymbol{W}^v\right)=\boldsymbol{\theta}^v\left(\left\|\boldsymbol{X}^v\boldsymbol{W}^v-\boldsymbol{P}^v\right\|_F^2+\lambda_3 Tr\left(\left(\boldsymbol{P}^v\right)^{\mathrm{T}}\boldsymbol{L}^v\boldsymbol{P}^v\right)\right) \tag{3}$$

Where $\boldsymbol{P}^v$ represents the view-specific label matrix, the second term of Eq.3 represents the introduction of the topological structure of each view in the feature space, which ensures that the local geometric structure between the feature space and the semantic matrix of different views is consistent. $\boldsymbol{L}^v = \boldsymbol{D}^v - \boldsymbol{S}^v$ is the graph Laplacian matrix. $\boldsymbol{S}_{ij}$ measures the similarity between instances $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$. The local geometric structure is constructed from the nearest neighbor graph on the feature space $\boldsymbol{X}^v$ in our work. In addition, the calculation of the similarity between the two instances of the v-th view is as follows:

$$\boldsymbol{S}_{i,j}^v=\begin{cases} e^{-\frac{\left\|\boldsymbol{x}_i^v-\boldsymbol{x}_j^v\right\|^2}{2\sigma^2}}, & if\ \boldsymbol{x}_j^v\in N_p\left(\boldsymbol{x}_i^v\right)\ \ or\ \boldsymbol{x}_i^v\in N_p\left(\boldsymbol{x}_j^v\right)\\ 0, & otherwise\end{cases} \tag{4}$$

where $N_p\left(\boldsymbol{x}\right)$ is the set of p nearest neighbors of instance $\boldsymbol{X}^v$.

For the second problem, we introduce a structural learning method of label correlation. We know that most existing multi-label label correlation learning methods have two limitations:

1) Label correlation is usually regarded as prior knowledge and cannot correctly describe the true dependency relationship among labels;
2) The consideration of the local structure of the label relationship in the label space is ignored.

For the first limitation, we use the idea of label propagation to build a joint learning model of view-specific labels and label correlations to solve them. Specifically, we believe that in addition to keeping the structure consistent with different view features, the view-specific labels should also consider the impact of label correlation on the information supplement of the original label space. Therefore, we introduce label correlation to supplement the original label matrix:

$$\begin{aligned}&\min_{\boldsymbol{W}^v,\boldsymbol{P}^v,\boldsymbol{C},\boldsymbol{\theta^v}} \frac{1}{2}\sum_{v=1}^{m}\left\{F\left(\boldsymbol{W}^v\right)+\lambda_1\left\|\boldsymbol{W}^v\right\|_F^2+\lambda_4\left\|\boldsymbol{P}^v-\boldsymbol{YC}\right\|_F^2\right\}\\&+\frac{\lambda_2}{2}\left\|\boldsymbol{\theta}\right\|_2^2\\&\text{s.t. }\ \boldsymbol{\theta}\geqslant 0,\ \sum_{v=1}^{m}\boldsymbol{\theta}^v=1\end{aligned} \tag{5}$$

Regarding the second limitation, we believe that in addition to focusing on the global features of multi-labels, we also need to capture some local structural information. For example, there is usually a group of labels so that the labels in a group have a strong correlation with each other and are independent of different labels. Therefore, we use $\left\|\cdot\right\|_*$ to represent the nuclear norm to limit the label correlation matrix $\boldsymbol{C}$ to a low-rank structure. Finally, obtain the objective function as follows:

$$\begin{aligned}&\min_{\boldsymbol{W}^v,\boldsymbol{P}^v,\boldsymbol{C},\boldsymbol{\theta^v}} \frac{1}{2}\sum_{v=1}^{m}\left\{F\left(\boldsymbol{W}^v\right)+\lambda_1\left\|\boldsymbol{W}^v\right\|_F^2+\lambda_4\left\|\boldsymbol{P}^v-\boldsymbol{YC}\right\|_F^2\right\}\\&+\frac{\lambda_5}{2}\left\|\boldsymbol{C}\right\|_*+\frac{\lambda_2}{2}\left\|\boldsymbol{\theta}\right\|_2^2\\&\text{s.t. }\ \boldsymbol{\theta}\geqslant 0,\ \sum_{v=1}^{m}\boldsymbol{\theta}^v=1\end{aligned} \tag{6}$$

Based on the above problems, we jointly learn non-aligned multi-view and multi-label semantic structures. Furthermore, because Eq.6 is a linear model, it cannot solve the inseparable linearity of given data. At present, some existing multi-label learning algorithms (such as [14], [34], and [48]) use nonlinear models to achieve good performance. We use the feature map $\phi\left(\cdot\right)$ to map the feature space $\boldsymbol{X}$ to a higher-dimensional (possibly infinite-dimensional) Hilbert space $\phi\left(\cdot\right)$. According to the expression theorem, we rerepresent the linear combination of input variables $\boldsymbol{W}$ as $\boldsymbol{W}=\phi(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{A}$, according to the expression theorem [40]. Suppose $\boldsymbol{K}$ is the kernel matrix $\boldsymbol{K}_{ij}=\boldsymbol{\kappa}\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)=\phi\left(\boldsymbol{x}\right)\phi(\boldsymbol{x})^{\mathrm{T}}$, where $\boldsymbol{\kappa}\left(\cdot,\cdot\right)$ is the kernel function used (the Gaussian kernel is used in this paper). Then, Eq.3 and Eq.6 can be rewritten as:

$$F\left(\boldsymbol{A}^v\right)=\boldsymbol{\theta}^v\left(\left\|\boldsymbol{K}^v\boldsymbol{A}^v-\boldsymbol{P}^v\right\|_F^2+\lambda_3 Tr\left(\left(\boldsymbol{P}^v\right)^{\mathrm{T}}\boldsymbol{L}^v\boldsymbol{P}^v\right)\right) \tag{7}$$

$$\begin{aligned}&\min_{\boldsymbol{A}^v,\boldsymbol{P}^v,\boldsymbol{C},\boldsymbol{\theta}} \frac{1}{2}\sum_{v=1}^{m}\left\{F\left(\boldsymbol{A}^v\right)+\lambda_1 Tr\left(\left(\boldsymbol{A}^v\right)^{\mathrm{T}}\boldsymbol{K}^v\boldsymbol{A}^v\right)+\lambda_4\left\|\boldsymbol{P}^v-\boldsymbol{YC}\right\|_F^2\right\}\\&+\frac{\lambda_5}{2}\left\|\boldsymbol{C}\right\|_*+\frac{\lambda_2}{2}\left\|\boldsymbol{\theta}\right\|_2^2\\&\text{s.t. }\ \boldsymbol{\theta}\geqslant 0,\ \sum_{v=1}^{m}\boldsymbol{\theta}^v=1\end{aligned} \tag{8}$$

In the next section, we will solve problem 8 with alternate iterative optimization.

IV. Optimization

A. Model optimization

The optimization problem in Eq.8 is convex, and the resulting problem can be solved by following the alternate optimization procedure.

Fix $\boldsymbol{P}^v$, $\boldsymbol{C}$ and $\boldsymbol{\theta}$, Optimize $\boldsymbol{A}^v$.

$$L\left(\boldsymbol{A}^v\right) = \\ \sum_{v=1}^{m}\left\{\boldsymbol{\theta}^v\left(\frac{1}{2}\left\|\boldsymbol{K}^v\boldsymbol{A}^v-\boldsymbol{P}^v\right\|_F^2\right)+\frac{\lambda_1}{2}Tr\left(\left(\boldsymbol{A}^v\right)^{\mathrm{T}}\boldsymbol{K}^v\boldsymbol{A}^v\right)\right\} \tag{9}$$

Taking the derivative of $L\left(\boldsymbol{A}^v\right)$ w.r.t $\boldsymbol{A}^v$ and setting the derivative to 0 can obtain a closed solution w.r.t. $\boldsymbol{A}^v$:

$$\boldsymbol{A}^v=\left(\boldsymbol{\theta}^v\boldsymbol{K}^v+\lambda_1\boldsymbol{I}\right)^{-1}\left(\boldsymbol{\theta}^v\boldsymbol{P}^v\right) \tag{10}$$

Fix $\boldsymbol{A}^v$, $\boldsymbol{C}$ and $\boldsymbol{\theta}$, Optimize $\boldsymbol{P}^v$.

$$\begin{aligned} L\left(\boldsymbol{P}^v\right)&=\min_{\boldsymbol{P}^v}\sum_{v=1}^{m}\left\{\boldsymbol{\theta}^v\boldsymbol{T}^v+\frac{\lambda_4}{2}\left\|\boldsymbol{P}^v-\boldsymbol{Y}\boldsymbol{C}\right\|_F^2\right\} \\ \boldsymbol{T}^v&=\left(\frac{1}{2}\left\|\boldsymbol{K}^v\boldsymbol{A}^v-\boldsymbol{P}^v\right\|_F^2+\frac{\lambda_3}{2}Tr\left(\left(\boldsymbol{P}^v\right)^{\mathrm{T}}\boldsymbol{L}^v\boldsymbol{P}^v\right)\right)\end{aligned} \tag{11}$$

Taking the derivative of $L\left(\boldsymbol{P}^v\right)$ w.r.t. $L\left(\boldsymbol{P}^v\right)$ and setting the derivative to 0 can obtain a closed solution w.r.t. $L\left(\boldsymbol{P}^v\right)$.

$$\boldsymbol{P}^v=\left(\lambda_3\boldsymbol{\theta}^v\boldsymbol{L}^v+\left(\boldsymbol{\theta}^v+\lambda_4\right)\boldsymbol{I}\right)^{-1}\left(\boldsymbol{\theta}^v\boldsymbol{K}^v\boldsymbol{A}^v+\lambda_4\boldsymbol{Y}\boldsymbol{C}\right) \tag{12}$$

Fix $\boldsymbol{A}^v$, $\boldsymbol{P}^v$ and $\boldsymbol{\theta}$, Optimize $\boldsymbol{C}$.

Compared with variables $\boldsymbol{A}^v$ and $\boldsymbol{P}^v$ that can directly obtain closed solutions, it is difficult to directly optimize $\boldsymbol{C}$ because of the nonsmooth regularization term in Eq.8. To make the objective function Eq.8 separable, we introduced the auxiliary variable $\boldsymbol{Z}$ to replace $\boldsymbol{C}$, and then an equivalent objective function can be expressed as:

$$\begin{aligned}&\min_{\boldsymbol{C},\boldsymbol{Z}}\sum_{v=1}^{m}\left\{\frac{\lambda_4}{2}\left\|\boldsymbol{P}^v-\boldsymbol{Y}\boldsymbol{C}\right\|_F^2\right\}+\frac{\lambda_5}{2}\|\boldsymbol{Z}\|_*\\ &\text{s.t. } \boldsymbol{C}=\boldsymbol{Z}\end{aligned} \tag{13}$$

We use augmented Lagrangian multipliers (ALMs) to solve this problem and reformulate the objective function 13 as:

$$\begin{aligned}&\min_{\boldsymbol{C},\boldsymbol{Z},\boldsymbol{\Lambda}}\sum_{v=1}^{m}\left\{\frac{\lambda_4}{2}\left\|\boldsymbol{P}^v-\boldsymbol{Y}\boldsymbol{C}\right\|_F^2\right\}+\frac{\lambda_5}{2}\|\boldsymbol{Z}\|_*+\\ &\frac{\mu}{2}\left\|\boldsymbol{C}-\boldsymbol{Z}+\frac{\boldsymbol{\Lambda}}{\mu}\right\|_F^2-\frac{1}{2\mu}\left\|\boldsymbol{\Lambda}\right\|_F^2\end{aligned} \tag{14}$$

Then, the inexact ALM (IALM) method is used to iteratively solve each variable in 14 by the block coordinate descent method. μ and Λ are expressed as nonnegative penalty factors and Lagrangian multipliers, respectively. According to the optimization strategy of IALM[49], we divide Eq.14 into the following subproblems:

$\boldsymbol{C}$-subproblem.

$$\boldsymbol{C}=\left(\sum_{v=1}^{m}\boldsymbol{Y}^T\boldsymbol{Y}+\frac{\mu}{\lambda_4}\boldsymbol{I}\right)^{-1}\left(\sum_{v=1}^{m}\left(\boldsymbol{P}^v\right)^{\mathrm{T}}\boldsymbol{Y}+\frac{\mu}{\lambda_4}\boldsymbol{Z}-\frac{\boldsymbol{\Lambda}}{\lambda_4}\right) \tag{15}$$

$\boldsymbol{Z}$-subproblem.

$$\boldsymbol{Z}^*=\arg\min_{\boldsymbol{Z}}\frac{\lambda_5}{\mu}\|\boldsymbol{Z}\|_*+\left\|\boldsymbol{Z}-\boldsymbol{C}-\frac{\boldsymbol{\Lambda}}{\mu}\right\|_F^2 \tag{16}$$

Update multiplier $\boldsymbol{\Lambda}$.

$$\boldsymbol{\Lambda}=\boldsymbol{\Lambda}+\mu\left(\boldsymbol{C}-\boldsymbol{Z}\right) \tag{17}$$

The subproblems of $\boldsymbol{Z}$ can be solved by the singular value threshold [50] method.

Fix $\boldsymbol{A}^v$, $\boldsymbol{P}^v$ and $\boldsymbol{C}$, Optimize $\boldsymbol{\theta}$.

$$\begin{aligned}&\min_{\boldsymbol{\theta}^v}\sum_{v=1}^{m}\left\{\boldsymbol{\theta}^v\boldsymbol{T}^v\right\}+\frac{\lambda_2}{2}\left\|\boldsymbol{\theta}\right\|_2^2\\ &\text{s.t. } \boldsymbol{\theta}\geqslant 0,\ \sum_{v=1}^{m}\boldsymbol{\theta}^v=1\end{aligned} \tag{18}$$

In summary, we introduce a kernel model to generate the predicted label vector $\boldsymbol{Y}_t$:

$$\boldsymbol{Y}_t\!=\operatorname{sign}\left(\boldsymbol{P}_t-\eta\right) \tag{19}$$

where $\boldsymbol{P}_t=\sum\limits_{v=1}^{m}\boldsymbol{\theta}^v\boldsymbol{K}_{test}^v\boldsymbol{A}^{v*}$, and η is the given threshold obtained by cross-validation.

Algorithm 1 Non-aligned Multi-view Multi-label Classification via **L**earning **V**iew-**S**pecific **L**abels (LVSL).

```
Require:
    The training data set: $\{\boldsymbol{X}^v\}_{v=1}^m \in \mathbb{R}^{N\times d}$;
    The label dataset: $\boldsymbol{Y} \in \mathbb{R}^{N\times l}$;
    The trade-off parameters: $\lambda_1$, $\lambda_2$, $\lambda_3$, $\lambda_4$, $\lambda_5$, $\mu = 10^{-1}$, $\rho = 1.5$, $\max_\mu = 10^6$;
    Randomly initialize $\boldsymbol{P}^v$, $\boldsymbol{A}^v$, $\boldsymbol{C}$ and $\boldsymbol{\theta}^v$;
Ensure:
    Final prediction objective function: $\boldsymbol{Y}_t$;
 1: Let $iter = 0$
 2: for Train Data $v = 1$ to $m$ do
 3:     $iter = iter + 1$;
 4:     Update variables $\boldsymbol{A}^v$, $\boldsymbol{P}^v$, $\boldsymbol{C}$, and $\boldsymbol{Z}$ by Eq.10, Eq.12, Eq.15, and Eq.16, respectively;
 5:     Update multipliers $\boldsymbol{\Lambda}$ by Eq.17;
 6:     Update the penalty parameter $\mu$ by $\mu = \min(\rho\mu, \max_\mu)$;
 7:     Update weight $\boldsymbol{\theta}^v$ by Eq.18;
 8: end for
 9: Calculate the prediction function of the test set by Eq.19
10: return $\boldsymbol{Y}_t$.
```

B. Complexity analysis

In this section, we mainly analyze the complexity of the optimization parts listed in Algorithm 1. The time complexity of LVSL is mainly controlled by step 4. The complexity of updating $\boldsymbol{A}^v$ in each iteration is $\mathcal{O}\left(N^3+N^2l\right)$, and the complexity of updating $\boldsymbol{P}^v$ is $\mathcal{O}\left(N^3+N^2l+Nl^2\right)$. The update of $\boldsymbol{C}$ costs $\mathcal{O}\left(mNl^2+l^3\right)$. The update of $\boldsymbol{Z}$ costs $\mathcal{O}\left(l^3\right)$. The time complexity of constructing $\boldsymbol{L}^v$ for each iteration is $\mathcal{O}\left(N^2d_{\max}\right)$. In summary, the total time complexity of LVSL is $\mathcal{O}\left(t\left(N^3+N^2l+Nl^2+N^2d_{\max}+l^3\right)\right)$, where t is the number of iterations. Typically, the model reaches its optimum after ten iterations converge quickly.

V. Evaluation and discussion

A. Experimental settings

We performed experiments on 7 benchmark multi-view multi-label data sets, which can be downloaded from Mulan[51][1]. Pascal07, Corel5k, ESPgame, Iaprtc12, and Mirflickr are the five widely used image datasets[2] from [52], [53]. The details of the datasets are summarized in Table I.

To verify the effectiveness of the proposed method, we compare our method with the following seven competing methods. Two of these methods use a concatenation strategy, which builds a multi-label learning model based on each data view and combines the weights of the output results to make the final prediction. Other methods are multi-view multi-label learning methods.

- MLkNN[40]: A lazy learning algorithm for multi-label learning. The k-nearest neighbor parameter is set to 10.
- LSML[54]: Multi-label classification method for joint learning of missing labels and label features. The parameters are set according to the given recommendations $\lambda_1 = 10^2$, $\lambda_2 = 10^{-2}$, $\lambda_3 = 10^{-2}$, $\lambda_4 = 10^{-3}$, and $\lambda_5 = 10^{-5}$.
- ICM2L[21]: Individual-view and commonality-view mining MVML classification method. Parameter configurations are implemented according to the suggestions given in the paper.
- iMvWL[20]: Incomplete multi-view weak label learning. In the experiment, the complete view information is available. Parameter configurations are implemented according to the suggestions given in the paper.
- TM3L[18]: Two-step multi-view multi-label classification method with missing labels. Specific parameters are selected according to the given optimal configuration.
- CDMM[19]: A neural network multi-view multi-label classification learning method based on view consistency and diversity. Specific parameters are selected according to the given optimal configuration.
- SIMM[36]: The multi-view multi-label classification method for subspace learning based on view-specific information mining. Specific parameters are selected according to the given optimal configuration.
- LVSL[3]: The non-aligned multi-view multi-label method by learning view-specific labels. The parameters λ_1, λ_3, and λ_5 are searched in the range of $\left\{10^{-5}, 10^{-4}, \cdots, 10^{-1}\right\}$, the parameter λ_2 is searched in $\left\{10^3, 10^4, \cdots, 10^6\right\}$, and the parameter λ_4 is searched in $\left\{10^{-3}, 10^{-4}, \cdots, 10^3\right\}$.

For all the above methods, the parameters are tuned to achieve the best performance by grid search.

B. Evaluation metrics

We use five evaluation metrics that are widely used in multi-label learning to measure the performance of each algorithm. The specific evaluation metrics are average precision (AP), coverage (CV), Hamming loss (HL), one error (OE), and ranking loss (RL). The larger the value of AP is, the better. The smaller the other evaluation metrics values are, the better. The detailed metric definitions can be found in[9], [10].

C. Experimental results

We performed fivefold cross-validation on each dataset, and each algorithm repeated the experiment 5 times. The average and standard deviation of each metric value under each dataset are reported in Tables II to VI. We show the best results in red and the second-best results in blue.

The $Friedman$ test[55], as a common strategy for comparing whether multiple algorithms have the same performance. Table VII summarizes the Friedman statistical F_F value of each evaluation metric and the critical value at the 0.05 significance level. Observing Table VII, we know that the F_F statistics of all metrics are greater than the critical value. Obviously, all metrics negate the null hypothesis, so we need to use a post-hoc test method to illustrate the significant differences among the approaches. In this article, we choose the $Nemenyi$ test[39], [56], [57] as the post-hoc test method. In Fig.3, the algorithm performance is sorted from left to right, and the best algorithm is ranked on the far right.Specifically, if the average ranking difference among the comparison algorithms is within a CD value, they are connected with a red solid line. From the reports in Tables II to VI and Figures 3(a) to 3(e), the following conclusions can be drawn:

- Among 35 configurations (7 datasets and 5 evaluation metrics), ours ranked first and second at 71.4% and 14.3%, respectively.
- Fig. 3 shows that LVSL is significantly better than other methods in 40% of cases, followed by CDMM and SIMM in 20% of cases. It is worth noting that our method is always better than CDMM.
- Encouragingly, by observing Tables II to VI, we find that our method achieves better performance on all metrics of $Emotions$ and $Yeast$. The overall CV metric performance of LVSL is not as good as SIMM, but it is not much different from the better results.

The analysis in addition to the experimental results is as follows:

- Compared with LSML and MLkNN, it can be seen that the performance of the traditional multi-label method connected to the multi-view multi-label learning approaches is flawed, mainly because they ignore the consistency and complementary information mining of multi-view and the physical interpretation of the characteristics of different views.
- The comparison among LVSL and iMvWL, ICM2L, and TM3L shows that our view-specific label learning method has better performance in mining the information among non-aligned multi-view labels and features. iMvWL ignores the diversity of views and has limitations in view information extraction.
- LVSL, SIMM, TM3L, and CDMM use nonlinear mapping to solve the linear inseparability problem. In general,

[1] datasets: http://mulan.sourceforge.net/datasets-mlc.html.

[2] datasets: http://lear.inrialpes.fr/people/guillaumin/data.php

[3] code: https://github.com/zhaodwahu/LVSL.

TABLE I
MULTI-VIEW MULTI-LABEL DATA SETS

Views	Emotions	Yeast	Pascal07	Corel5k	ESPgame	Iaprtc12	Mirflickr
1	rhythmic attributes (8)	Genetic Expression (79)	DenseSift (1000)	DenseHue (100)	DenseHue (100)	DenseHue (100)	DenseHue (100)
2	timbre attributes (64)	phylogenetic profile (24)	HarrisSift (1000)	DenseSift (1000)	DenseSift (1000)	DenseSift (1000)	DenseSift (1000)
3	-	-	Gist(512)	Gist(512)	Gist(512)	Gist(512)	Gist(512)
4	-	-	HSV(4096)	HSV(4096)	HSV(4096)	HSV(4096)	HSV(4096)
5	-	-	RGB(4096)	Lab(4096)	Lab(4096)	Lab(4096)	Lab(4096)
6	-	-	Tags(804)	RGB(4096)	RGB(4096)	RGB(4096)	RGB(4096)
Domain	music	biology	image	image	image	image	image
The number of labels	6	14	20	260	268	291	457
The number of samples	593	2417	9963	4999	20770	19627	25000

TABLE II
EXPERIMENTAL RESULTS (MEAN ± STD) ON AVERAGE PRECISION (↑).

Dataset	AP(↑)							
	ML*k*NN	LSML	ICM2L	iMvWL	CDMM	TM3L	SIMM	LVSL
Emotions	0.718±0.021	0.785±0.011	0.578±0.022	0.584±0.015	0.790±0.019	0.781±0.016	0.780±0.027	0.801±0.028
Yeast	0.762±0.009	0.610±0.008	0.708±0.014	0.704±0.011	0.781±0.009	0.772±0.010	0.765±0.016	0.785±0.006
Pascal07	0.464±0.007	0.663±0.009	0.460±0.025	0.660±0.013	0.759±0.006	0.781±0.005	0.786±0.005	0.769±0.005
Corel5k	0.349±0.008	0.418±0.008	0.258±0.004	0.274±0.003	0.545±0.007	0.516±0.008	0.534±0.006	0.546±0.007
ESPgame	0.259±0.003	0.319±0.002	0.219±0.013	0.237±0.002	0.400±0.003	0.383±0.002	0.378±0.013	0.402±0.003
Iaprtc12	0.340±0.004	0.328±0.007	0.204±0.000	0.242±0.001	0.432±0.004	0.421±0.004	0.401±0.017	0.434±0.003
Mirflickr	0.076±0.001	0.096±0.002	0.093±0.001	0.094±0.005	0.102±0.002	0.099±0.000	0.142±0.009	0.128±0.002

TABLE III
EXPERIMENTAL RESULTS (MEAN ± STD) ON COVERAGE (↓).

Dataset	CV(↓)							
	ML*k*NN	LSML	ICM2L	iMvWL	CDMM	TM3L	SIMM	LVSL
Emotions	0.376±0.020	0.307±0.017	0.530±0.036	0.506±0.014	0.304±0.018	0.309±0.030	0.307±0.014	0.298±0.014
Yeast	0.452±0.006	0.623±0.013	0.503±0.006	0.494±0.009	0.426±0.008	0.446±0.010	0.450±0.004	0.424±0.008
Pascal07	0.319±0.003	0.124±0.003	0.308±0.048	0.189±0.015	0.111±0.003	0.109±0.002	0.106±0.002	0.110±0.002
Corel5k	0.290±0.004	0.184±0.006	0.334±0.000	0.286±0.004	0.179±0.011	0.187±0.009	0.148±0.006	0.175±0.003
ESPgame	0.437±0.003	0.342±0.002	0.479±0.001	0.447±0.004	0.337±0.005	0.367±0.004	0.308±0.012	0.320±0.003
Iaprtc12	0.376±0.005	0.307±0.003	0.497±0.001	0.435±0.003	0.284±0.006	0.335±0.003	0.270±0.019	0.266±0.006
Mirflickr	0.386±0.004	0.329±0.005	0.499±0.002	0.492±0.007	0.332±0.002	0.307±0.001	0.306±0.016	0.316±0.004

TABLE IV
EXPERIMENTAL RESULTS (MEAN ± STD) ON HAMMING LOSS (↓).

Dataset	HL(↓)							
	ML*k*NN	LSML	ICML	iMvWL	CDMM	TM3L	SIMM	LVSL
Emotions	0.262±0.010	0.224±0.012	0.375±0.015	0.395±0.011	0.207±0.014	0.218±0.014	0.246±0.008	0.205±0.011
Yeast	0.196±0.005	0.261±0.008	0.278±0.008	0.269±0.005	0.189±0.006	0.196±0.006	0.207±0.005	0.182±0.009
Pascal07	0.072±0.001	0.066±0.002	0.115±0.003	0.086±0.002	0.049±0.001	0.050±0.000	0.046±0.001	0.046±0.001
Corel5k	0.013±0.000	0.013±0.000	0.022±0.000	0.022±0.000	0.011±0.000	0.012±0.000	0.011±0.000	0.011±0.000
ESPgame	0.017±0.000	0.017±0.000	0.029±0.000	0.028±0.000	0.018±0.000	0.017±0.000	0.017±0.000	0.017±0.000
Iaprtc12	0.019±0.000	0.019±0.000	0.032±0.000	0.031±0.000	0.019±0.000	0.019±0.000	0.019±0.000	0.018±0.000
Mirflickr	0.006±0.000	0.006±0.000	0.013±0.000	0.013±0.000	0.006±0.000	0.006±0.000	0.006±0.000	0.006±0.000

TABLE V
EXPERIMENTAL RESULTS (MEAN ± STD) ON ONE ERROR (↓).

Dataset	OE(↓)							
	ML*k*NN	LSML	ICML	iMvWL	CDMM	TM3L	SIMM	LVSL
Emotions	0.359±0.041	0.295±0.016	0.530±0.030	0.521±0.021	0.304±0.034	0.314±0.019	0.310±0.056	0.268±0.039
Yeast	0.233±0.017	0.358±0.019	0.235±0.024	0.292±0.020	0.211±0.013	0.217±0.022	0.225±0.028	0.210±0.015
Pascal07	0.585±0.012	0.474±0.016	0.589±0.002	0.397±0.021	0.308±0.011	0.267±0.007	0.255±0.008	0.284±0.009
Corel5k	0.602±0.017	0.522±0.012	0.697±0.007	0.687±0.003	0.362±0.006	0.395±0.008	0.363±0.011	0.356±0.008
ESPgame	0.650±0.008	0.559±0.005	0.713±0.030	0.674±0.000	0.465±0.009	0.482±0.007	0.476±0.022	0.465±0.009
Iaprtc12	0.535±0.006	0.541±0.007	0.720±0.005	0.624±0.002	0.439±0.007	0.443±0.008	0.447±0.021	0.444±0.005
Mirflickr	0.905±0.003	0.877±0.004	0.908±0.004	0.887±0.003	0.869±0.004	0.878±0.003	0.865±0.006	0.841±0.003

TABLE VI
EXPERIMENTAL RESULTS (MEAN ± STD) ON RANKING LOSS (↓).

Dataset	RL(↓)							
	MLkNN	LSML	ICML	iMvWL	CDMM	TM3L	SIMM	LVSL
Emotions	0.255±0.020	0.177±0.012	0.443±0.013	0.414±0.012	0.174±0.013	0.184±0.025	0.178±0.024	0.165±0.024
Yeast	0.170±0.006	0.346±0.012	0.215±0.011	0.214±0.008	0.151±0.008	0.161±0.012	0.165±0.008	0.150±0.006
Pascal07	0.256±0.004	0.084±0.003	0.241±0.040	0.138±0.011	0.070±0.002	0.068±0.002	0.066±0.002	0.066±0.002
Corel5k	0.127±0.003	0.076±0.002	0.149±0.002	0.130±0.003	0.069±0.005	0.072±0.003	0.059±0.002	0.069±0.002
ESPgame	0.181±0.002	0.134±0.001	0.203±0.002	0.190±0.002	0.124±0.002	0.135±0.002	0.120±0.006	0.118±0.002
Iaprtc12	0.135±0.002	0.105±0.001	0.189±0.001	0.165±0.002	0.089±0.002	0.106±0.002	0.089±0.007	0.085±0.001
Mirflickr	0.224±0.003	0.183±0.004	0.288±0.001	0.285±0.009	0.184±0.003	0.169±0.000	0.222±0.013	0.179±0.003

TABLE VII
THE CORRESPONDING STATISTICAL F_F VALUE OF EACH EVALUATION METRIC AND CRITICAL VALUE UNDER THE $Friedman$ TEST.

Metric	F_F	Critical Value($\alpha = 0.05$)
AP	27.273	2.2371
CV	25.825	
HL	9.760	
OE	22.484	
RL	24.007	

LVSL is always better than the other three methods. View-specific labels and multi-label structural learning can effectively improve classification performance. In addition, SIMM also ignores the impact of label correlation, which leads to its poor overall performance.

- LVSL performs worse than SIMM on the AP and CV metrics on the Pascal07 and Mirflickr datasets for two main reasons. (1) LVSL uses a single kernel function for kernel mapping of multiple views, but it is undeniable that the performance of the kernel method often depends on the choice of the kernel function. Because the nonlinear relationship among the data of each view may be different, the optimal kernel function for one view may not be suitable for another view[58], which provides a new direction for our future research work. SIMM does not need to consider this problem. (2) SIMM develops the shared subspace based on the information among each view. In our work, considering the problem of the non-aligned view, the information among views cannot be directly communicated, which affects the performance of the LVSL to a certain extent.

Additionally, there are two main reasons for the advantage of our method over deep learning methods:

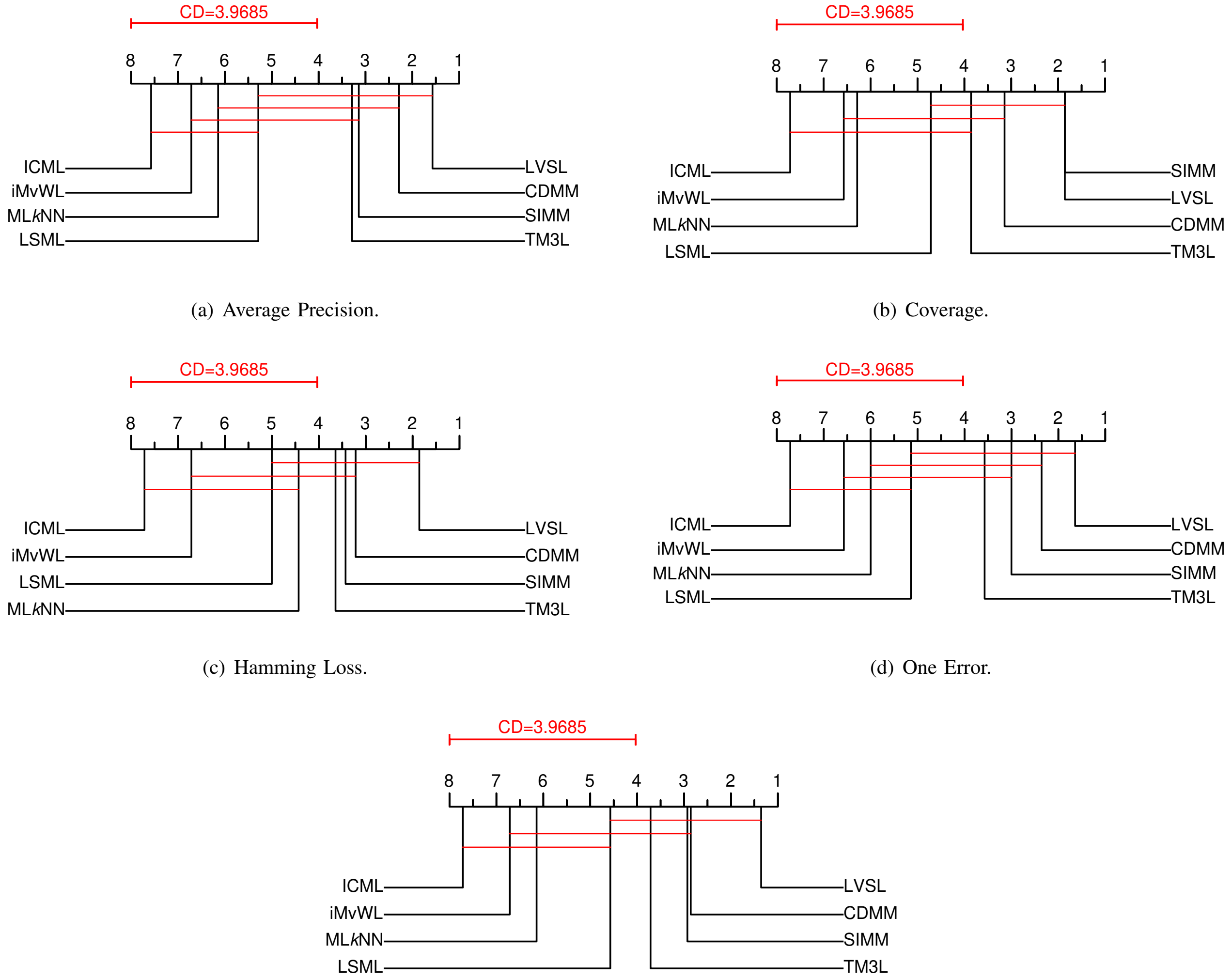
- The current multi-view multi-label learning tasks cannot directly perform end-to-end training through deep learning and require solutions that benefit from some traditional feature extraction techniques. Therefore, the feature representation capability of deep learning is limited in this task, and due to its powerful nonlinear data processing capability, our method using kernel tricks can also achieve this purpose[48].
- The training data in this paper are relatively limited, and deep learning may overfit the training data, resulting in insufficient model generalization ability. The traditional method has good generalization ability, interpretability, sufficient transparency, and universality[59]. Therefore, to some extent, traditional methods are more suitable for solving the complex tasks proposed in this paper.

D. Ablation Analysis

In this section, to further verify the effectiveness of each component in LVSL, we conducted additional ablation analysis experiments and reported the values on the five evaluation metrics in Table VIII. LVSL-I, LVSL-II, and LVSL-III are variants of LVSL, which exclude the influence of view-specific labels, label correlations, and view contributions, respectively. Comparing the results of LVSL-I and LVSL on Table VIII, it can be found that the overall performance is significantly improved after adding view-specific labels, which confirmed our clear motivation to use view-specific label learning to solve the problem of the non-aligned view. Comparing LVSL-II and LVSL, it is found that LVSL is better than LVSL-II in most cases, which proves the necessity of capturing label structure information and verifies the effectiveness of using the label association matrix $\boldsymbol{C}$ to complement the original label matrix $\boldsymbol{Y}$. In some cases, LVSL-III and LVSL have the same performance, showing that our contribution measurement method has room for further improvement.

E. Sensitivity Analysis

LVSL has five important hyperparameters λ_1, λ_2, λ_3, λ_4 and λ_5. We separately tested the sensitivity of LVSL to five hyperparameters on the $corel5k$ dataset, and we fixed four of the parameters as the best (for example, $\lambda_1 = 10^{-1}$, $\lambda_2 = 10^5$, $\lambda_3 = 10^{-1}$, $\lambda_4 = 10^2$ and $\lambda_5 = 10^{-1}$) and then changed the



(a) Average Precision.

(b) Coverage.

(c) Hamming Loss.

(d) One Error.

(e) Ranking Loss.

Fig. 3. The performance comparison results of LVSL and other comparison methods using the $Nemenyi$ test (CD = 3.9685 at the 0.05 significance level) under five evaluation metrics.

value of one of the parameters within the given range. Fig. 4 shows the changes in LVSL on the AP and RL evaluation metrics.

The hyperparameter λ_1 controls the complexity of the model coefficients and adjusts the balance between overfitting and underfitting. When λ_1 is too small, it will cause overfitting problems in the model, and underfitting problems will occur when λ_1 is too large. The hyperparameter λ_2 controls the contribution of different views. The hyperparameter λ_3 controls the structural diversity among different views. The hyperparameter λ_4 controls the global consistency of information between the view-specific label and the real label. The hyperparameter λ_5 controls the effect of local label correlation.

Fig. 4 shows that the parameter λ_1 has a better effect in taking the intermediate value, and intuitively, the intermediate value ensures the balance of the model fitting. When the parameter λ_2 achieves 10^5, the effect is better. A larger value means that the influence of the contribution weight of each view is ignored, and a smaller value will be too sensitive to the contribution of view parameters and ignore the complementary information between views. The parameter λ_3 and λ_5 values tend to take smaller values, but values that are too small will ignore the contribution of the corresponding regularization term, so we generally choose the median value. The performance is better when the parameter λ_4 takes a larger value. A larger value can fully learn the view consistency information of multiple views, but an excessively large value will also lead to insufficient complementary learning of view-specific labels. Our parameter sensitivity analysis results on other datasets are similar, and similar conclusions can be drawn.

F. Further Analysis

We report the algorithm efficiency analysis of LVSL in this section. Fig. 5 shows the iterative trend of our method on two datasets. Fig. 5 shows that the value of the objective function is significantly reduced during the initial iteration, and as the optimization process proceeds, the value of the objective

TABLE VIII
COMPARISON RESULTS OF LVSL-I, LVSL-III, LVSL-III AND LVSL. LVSL-I WITHOUT VIEW-SPECIFIC LABEL STRUCTURE, LVSL-III WITHOUT LABEL CORRELATION, AND LVSL-III WITH THE SAME CONTRIBUTION WEIGHT FOR ALL VIEWS.

Datasets	Methods	Evaluation Metrics				
		AP	CV	HL	OE	RL
Emotions	LVSL-I	0.794± 0.021	0.304±0.014	0.213±0.018	0.282±0.047	0.171±0.018
	LVSL-II	0.789± 0.026	0.304±0.016	0.208±0.016	0.297±0.054	0.172±0.028
	LVSL-III	0.798± 0.018	0.300±0.010	0.212±0.011	0.277±0.041	0.166±0.013
	LVSL	**0.801± 0.028**	**0.298±0.014**	**0.205±0.011**	**0.268±0.039**	**0.165±0.024**
Yeast	LVSL-I	0.777± 0.030	0.435±0.008	0.185±0.002	0.224±0.008	0.158±0.003
	LVSL-II	0.779± 0.009	0.432±0.015	0.184±0.004	0.220±0.012	0.155±0.006
	LVSL-III	**0.785± 0.006**	0.426±0.003	0.185±0.004	0.211±0.015	0.151±0.005
	LVSL	**0.785± 0.006**	**0.424±0.008**	**0.182±0.009**	**0.210±0.015**	**0.150±0.006**
Pascal07	LVSL-I	0.747± 0.009	0.112±0.004	0.049±0.001	0.326±0.013	0.071±0.003
	LVSL-II	0.762± 0.004	0.112±0.002	0.049±0.001	0.301±0.009	0.070±0.002
	LVSL-III	0.759± 0.004	**0.110±0.002**	0.048±0.001	0.310±0.002	0.074±0.003
	LVSL	**0.769± 0.005**	**0.110±0.002**	**0.046±0.001**	**0.284±0.009**	**0.066±0.002**
Corel5k	LVSL-I	0.543± 0.007	**0.171±0.002**	0.012±0.000	0.361±0.009	0.070±0.002
	LVSL-II	0.542± 0.008	0.186±0.011	**0.011±0.000**	0.367±0.014	0.071±0.005
	LVSL-III	0.546± 0.008	0.178±0.006	**0.011±0.000**	0.364±0.008	**0.068±0.002**
	LVSL	**0.546± 0.007**	0.175±0.003	**0.011±0.000**	**0.356±0.008**	0.069±0.002
ESPgame	LVSL-I	0.403± 0.005	0.322±0.005	**0.017±0.000**	0.466±0.006	0.119±0.001
	LVSL-II	**0.404± 0.006**	0.321±0.002	**0.017±0.000**	0.467±0.006	0.119±0.001
	LVSL-III	0.403± 0.001	0.323±0.002	**0.017±0.000**	0.466±0.005	0.119±0.001
	LVSL	0.402± 0.003	**0.320±0.003**	**0.017±0.000**	**0.465±0.009**	**0.118±0.002**
Iaprtc12	LVSL-I	0.424± 0.006	0.277±0.004	0.019±0.000	0.470±0.006	0.089±0.001
	LVSL-II	0.425± 0.005	0.275±0.003	0.019±0.000	0.472±0.007	0.089±0.001
	LVSL-III	0.420± 0.002	0.276±0.003	0.019±0.000	0.478±0.003	0.089±0.001
	LVSL	**0.434±0.003**	**0.266±0.005**	**0.018±0.000**	**0.444±0.005**	**0.085±0.001**
Mirflickr	LVSL-I	0.116±0.002	0.321±0.004	**0.006±0.000**	0.895±0.006	0.242±0.002
	LVSL-II	0.118±0.002	0.320±0.003	**0.006±0.000**	0.864±0.002	0.180±0.002
	LVSL-III	0.120±0.002	0.319±0.002	**0.006±0.000**	0.867±0.002	0.180±0.001
	LVSL	**0.128±0.002**	**0.316±0.004**	**0.006±0.000**	**0.841±0.003**	**0.179±0.003**

function gradually converges. LVSL tends to converge for 10 iterations on both datasets, proving that it can converge faster. Our convergence results on other datasets are similar.

VI. Conclusion

This paper proposes a novel multi-view multi-label classification method that jointly learns view-specific labels and label structures. LVSL differs from existing work on multi-view multi-label classification by implicitly concatenating common or shared labels in that it assigns a specific label to each view to solve the problem of inconsistent labels for views in non-aligned views. When constructing view-specific labels, the consistency and diversity information among the views in multi-view learning are learned, and the label correlation information in multi-label learning is also combined. A large number of experiments show that the proposed non-aligned view learning method is a promising solution for multi-view multi-label classification based on view-specific labels.

This method is of great significance for future research on the feasibility of the multi-view multi-label classification of non-aligned views. Future work will be devoted to proposing more new methods to study view-specific label learning problems via multi-kernel learning.

References

[1] Yong Luo, Tongliang Liu, Dacheng Tao, and Chao Xu. Multiview matrix completion for multilabel image classification. IEEE Transactions on Image Processing, 24(8):2355–2368, 2015.

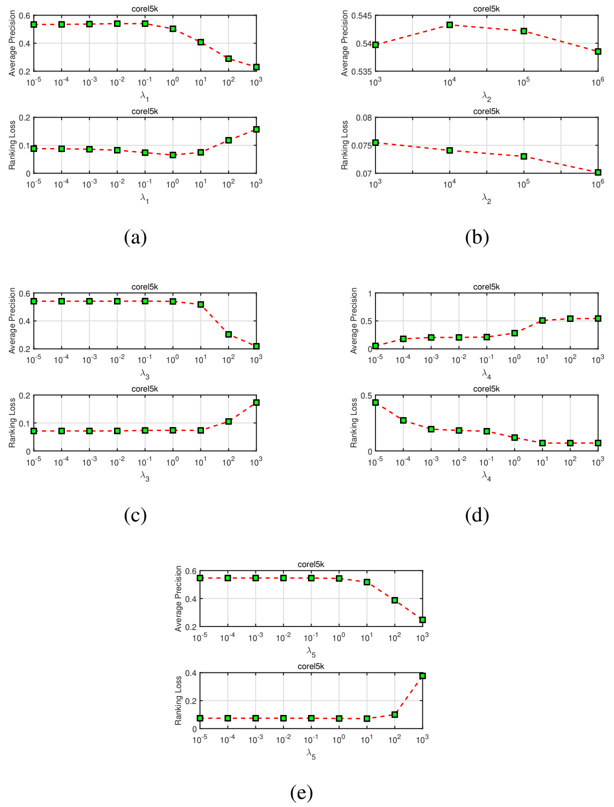
[2] Shiliang Sun and Daoming Zong. Lcbm: A multi-view probabilistic model for multi-label classification. IEEE transactions on pattern analysis and machine intelligence, 2020.

[3] Jia Zhang, Candong Li, Zhenqiang Sun, Zhiming Luo, Changen Zhou, and Shaozi Li. Towards a unified multi-source-based optimization framework for multi-label learning. Applied Soft Computing, 76:425–435, 2019.

[4] Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun. Multi-view learning overview: Recent progress and new challenges. Information Fusion, 38:43–54, 2017.

[5] Hongchang Gao, Feiping Nie, Xuelong Li, and Heng Huang. Multi-view subspace clustering. In Proceedings of the IEEE international conference on computer vision, pages 4238–4246, 2015.

[6] Yingming Li, Ming Yang, and Zhongfei Zhang. A survey of multi-view representation learning. IEEE transactions on knowledge and data engineering, 31(10):1863–1883, 2018.

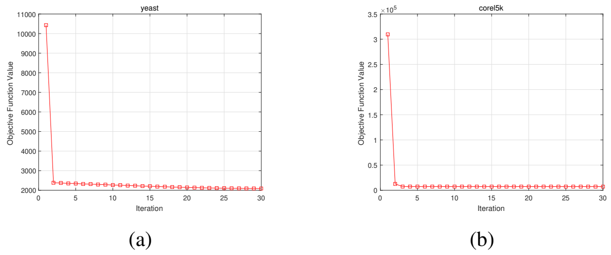


(a) (b)

(c) (d)

(e)

Fig. 4. Parameter sensitivity analysis of the LVSL algorithm on the Corel5k dataset. (a) Effect of λ_1 with other fixed parameters. (b) Effect of λ_2 with other fixed parameters. (c) Effect of λ_3 with other fixed parameters. (d) Effect of λ_4 with other fixed parameters. (e) Effect of λ_5 with other fixed parameters.



(a) (b)

Fig. 5. Convergence analysis of LVSL on Yeast and Corel5k. (a) Convergence trend on Yeast. (b) Convergence trend on Corel5k.

[7] Jie Wen, Ke Yan, Zheng Zhang, Yong Xu, Junqian Wang, Lunke Fei, and Bob Zhang. Adaptive graph completion based incomplete multi-view clustering. IEEE Transactions on Multimedia, 23:2493–2504, 2020.

[8] Grigorios Tsoumakas, Ioannis Katakis, and Ioannis Vlahavas. Mining multi-label data. In Data mining and knowledge discovery handbook, pages 667–685. Springer, 2009.

[9] Minling Zhang and Zhihua Zhou. A review on multi-label learning algorithms. IEEE Transactions on Knowledge and Data Engineering, 26(8):1819–1837, 2014.

[10] Eva Gibaja and Sebastián Ventura. A tutorial on multilabel learning. ACM Computing Surveys (CSUR), 47(3):1–38, 2015.

[11] Xiaochun Cao, Changqing Zhang, Huazhu Fu, Si Liu, and Hua Zhang. Diversity-induced multi-view subspace clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 586–594, 2015.

[12] Youwei Liang, Dong Huang, and Chang-Dong Wang. Consistency meets inconsistency: A unified graph learning framework for multi-view clustering. In 2019 IEEE International Conference on Data Mining (ICDM), pages 1204–1209. IEEE, 2019.

[13] Grigorios Tsoumakas and Ioannis Katakis. Multi-label classification: An overview. International Journal of Data Warehousing and Mining (IJDWM), 3(3):1–13, 2007.

[14] Lei Feng, Jun Huang, Senlin Shu, and Bo An. Regularized matrix factorization for multilabel learning with missing labels. IEEE Transactions on Cybernetics, 2020.

[15] X Li and S Chen. A concise yet effective model for non-aligned incomplete multi-view and missing multi-label learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

[16] Aminu Da¡¯u and Naomie Salim. Recommendation system based on deep learning methods: a systematic review and new directions. Artificial Intelligence Review, 53(4):2709–2748, 2020.

[17] Bo-Kyeong Kim, Suh-Yeon Dong, Jihyeon Roh, Geonmin Kim, and Soo-Young Lee. Fusing aligned and non-aligned face information for automatic affect recognition in the wild: a deep learning approach. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 48–57, 2016.

[18] Dawei Zhao, Qingwei Gao, Yixiang Lu, and Dong Sun. Two-step multi-view and multi-label learning with missing label via subspace learning. Applied Soft Computing, 102:107120, 2021.

[19] Dawei Zhao, Qingwei Gao, Yixiang Lu, Dong Sun, and Yusheng Cheng. Consistency and diversity neural network multi-view multi-label learning. Knowledge-Based Systems, 218:106841, 2021.

[20] Qiaoyu Tan, Guoxian Yu, Carlotta Domeniconi, Jun Wang, and Zili Zhang. Incomplete multi-view weak-label learning. In IJCAI, pages 2703–2709, 2018.

[21] Qiaoyu Tan, Guoxian Yu, Jun Wang, Carlotta Domeniconi, and Xiangliang Zhang. Individuality-and commonality-based multiview multi-label learning. IEEE transactions on cybernetics, 51(3):1716–1727, 2019.

[22] Fangwen Zhang, Xiuyi Jia, and Weiwei Li. Tensor-based multi-view label enhancement for multi-label learning. In IJCAI, pages 2369–2375, 2020.

[23] Changqing Zhang, Ziwei Yu, Qinghua Hu, Pengfei Zhu, Xinwang Liu, and Xiaobo Wang. Latent semantic aware multi-view multi-label classification. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 32, 2018.

[24] Shiping Wang, Witold Pedrycz, Qingxin Zhu, and William Zhu. Subspace learning for unsupervised feature selection via matrix factorization. Pattern Recognition, 48(1):10–19, 2015.

[25] Yongshan Zhang, Jia Wu, Zhihua Cai, and S Yu Philip. Multi-view multi-label learning with sparse feature selection for image annotation. IEEE Transactions on Multimedia, 22(11):2844–2857, 2020.

[26] Weijieying Ren, Lei Zhang, Bo Jiang, Zhefeng Wang, Guangming Guo, and Guiquan Liu. Robust mapping learning for multi-view multi-label classification with missing labels. In International Conference on Knowledge Science, Engineering and Management, pages 543–551. Springer, 2017.

[27] Xiaoyu Zhang, Jian Cheng, Changsheng Xu, Hanqing Lu, and Songde Ma. Multi-view multi-label active learning for image classification. In 2009 IEEE International Conference on Multimedia and Expo, pages 258–261. IEEE, 2009.

[28] Meng Liu, Yong Luo, Dacheng Tao, Chao Xu, and Yonggang Wen. Low-rank multi-view learning in matrix completion for multi-label image classification. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 29, 2015.

[29] Changming Zhu, Duoqian Miao, Zhe Wang, Rigui Zhou, Lai Wei, and Xiafen Zhang. Global and local multi-view multi-label learning. Neurocomputing, 371:67–77, 2020.

[30] Zheng Fang and Zhongfei Zhang. Simultaneously combining multi-view multi-label learning with maximum margin classification. In 2012 IEEE 12th International Conference on Data Mining, pages 864–869. IEEE, 2012.

[31] Jun Huang, Xiwen Qu, Guorong Li, Feng Qin, Xiao Zheng, and Qingming Huang. Multi-view multi-label learning with view-label-specific features. IEEE Access, 7:100979–100992, 2019.

[32] Ze-Sen Chen, Xuan Wu, Qing-Guo Chen, Yao Hu, and Min-Ling Zhang. Multi-view partial multi-label learning with graph-based disambiguation. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 3553–3560, 2020.

[33] Xiaofeng Zhu, Xuelong Li, and Shichao Zhang. Block-row sparse multiview multilabel learning for image classification. IEEE transactions on cybernetics, 46(2):450–461, 2015.

[34] Pengfei Zhu, Qi Hu, Qinghua Hu, Changqing Zhang, and Zhizhao Feng. Multi-view label embedding. Pattern Recognition, 84:126–135, 2018.

[35] Arthur Gretton, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf. Measuring statistical dependence with hilbert-schmidt norms. In

International conference on algorithmic learning theory, pages 63–77. Springer, 2005.

[36] Xuan Wu, Qing-Guo Chen, Yao Hu, Dengbao Wang, Xiaodong Chang, Xiaobo Wang, and Min-Ling Zhang. Multi-view multi-label learning with view-specific information extraction. In IJCAI, pages 3884–3890, 2019.

[37] Jundong Shen, Yi Zhang, Cheng Yu, and Chongjun Wang. Multi-view multi-label learning with dual-attention networks for stroke screen. In 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pages 1124–1128. IEEE, 2020.

[38] Sheng-Jun Huang and Zhi-Hua Zhou. Multi-label learning by exploiting label correlations locally. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 26, 2012.

[39] Min-Ling Zhang and Lei Wu. Lift: Multi-label learning with label-specific features. IEEE transactions on pattern analysis and machine intelligence, 37(1):107–120, 2014.

[40] Min-Ling Zhang and Zhi-Hua Zhou. Ml-knn: A lazy learning approach to multi-label learning. Pattern recognition, 40(7):2038–2048, 2007.

[41] Ze-Bang Yu and Min-Ling Zhang. Multi-label classification with label-specific feature generation: A wrapped approach. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

[42] Jun Huang, Guorong Li, Shuhui Wang, Weigang Zhang, and Qingming Huang. Group sensitive classifier chains for multi-label classification. In 2015 IEEE international conference on multimedia and expo (ICME), pages 1–6. IEEE, 2015.

[43] Jesse Read, Bernhard Pfahringer, Geoff Holmes, and Eibe Frank. Classifier chains for multi-label classification. Machine learning, 85(3):333–359, 2011.

[44] Clara Pizzuti. A multi-objective genetic algorithm for community detection in networks. In 2009 21st IEEE International Conference on Tools with Artificial Intelligence, pages 379–386. IEEE, 2009.

[45] Baolin Guo, Chenping Hou, Jincheng Shan, and Dongyun Yi. Low rank multi-label classification with missing labels. In 2018 24th International Conference on Pattern Recognition (ICPR), pages 417–422. IEEE, 2018.

[46] Linli Xu, Zhen Wang, Zefan Shen, Yubo Wang, and Enhong Chen. Learning low-rank label correlations for multi-label classification with missing labels. In 2014 IEEE international conference on data mining, pages 1067–1072. IEEE, 2014.

[47] Yue Zhu, James T Kwok, and Zhi-Hua Zhou. Multi-label learning with global and local label correlation. IEEE Transactions on Knowledge and Data Engineering, 30(6):1081–1094, 2017.

[48] Zhongchen Ma and Songcan Chen. Expand globally, shrink locally: Discriminant multi-label learning with missing labels. Pattern Recognition, 111:107675, 2021.

[49] Zhouchen Lin, Risheng Liu, and Zhixun Su. Linearized alternating direction method with adaptive penalty for low-rank representation. In Proceedings of the 24th International Conference on Neural Information Processing Systems, pages 612–620, 2011.

[50] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. SIAM Journal on optimization, 20(4):1956–1982, 2010.

[51] Grigorios Tsoumakas, Eleftherios Spyromitros-Xioufis, Jozef Vilcek, and Ioannis Vlahavas. Mulan: A java library for multi-label learning. Journal of Machine Learning Research, 12(Jul):2411–2414, 2011.

[52] Matthieu Guillaumin, Jakob Verbeek, and Cordelia Schmid. Multimodal semi-supervised learning for image classification. In 2010 IEEE Computer society conference on computer vision and pattern recognition, pages 902–909. IEEE, 2010.

[53] Xuanwu Liu, Guoxian Yu, Carlotta Domeniconi, Jun Wang, Yazhou Ren, and Maozu Guo. Ranking-based deep cross-modal hashing. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, pages 4400–4407, 2019.

[54] Jun Huang, Feng Qin, Xiao Zheng, Zekai Cheng, Zhixiang Yuan, Weigang Zhang, and Qingming Huang. Improving multi-label classification with missing labels by learning label-specific features. Information Sciences, 492:124–146, 2019.

[55] Junjie Zhang, Qi Wu, Chunhua Shen, Jian Zhang, and Jianfeng Lu. Multilabel image classification with regional latent semantic dependencies. IEEE Transactions on Multimedia, 20(10):2801–2813, 2018.

[56] Zhi-Fen He, Ming Yang, Yang Gao, Hui-Dong Liu, and Yilong Yin. Joint multi-label classification and label correlations with missing labels and feature selection. Knowledge-Based Systems, 163:145–158, 2019.

[57] Janez Demšar. Statistical comparisons of classifiers over multiple data sets. Journal of Machine learning research, 7(Jan):1–30, 2006.

[58] Tiejian Zhang, Xinwang Liu, Lei Gong, Siwei Wang, Xin Niu, and Li Shen. Late fusion multiple kernel clustering with local kernel alignment maximization. IEEE Transactions on Multimedia, 2021.

[59] Niall O¡¯Mahony, Sean Campbell, Anderson Carvalho, Suman Harapanahalli, Gustavo Velasco Hernandez, Lenka Krpalkova, Daniel Riordan, and Joseph Walsh. Deep learning vs. traditional computer vision. In Science and information conference, pages 128–144. Springer, 2019.

Dawei Zhao received the M.S. degree from the Anqing Normal University and the Ph.D. degree at the School of Computers Science and Technology, Anhui University, Hefei, China. His current research interests mainly focus on multi-view learning, pattern recognition and image processing.

Qingwei Gao received the Ph.D. degree at the School of Information and Communication Engineering, University of science and technology of China, Hefei, China. He is currently a professor with the School of electrical engineering and automation, Anhui University. His research interests include pattern recognition, wavelet analysis, image processing and fractal signal processing.

Yixiang Lu received the B.S. degree in measurement control and instrumentation and the M.S. degree in detection technology and automatic equipment from Anhui University, Hefei, China, in 2004 and 2007, respectively. He is currently pursuing the Ph.D. degree in circuit and system at the Anhui University. His research interests include wavelet analysis, image processing and synthetic aperture radar.

Dong Sun received the B.S. degree from the automation department of Anhui University in 2003, the M.S. degree in pattern recognition from the Anhui University in 2006, and the Ph.D. degree in computer science and application from Anhui University in 2016. His current research interest is primarily in the area of sparse representation, fractal, and image restoration.