Common-Individual Semantic Fusion for Multi-View Multi-Label Learning

Gengyu Lyu[1] , **Weiqi Kang**[1] , **Haobo Wang**[2] , **Zheng Li**[1] , **Zhen Yang**[1*] , **Songhe Feng**[3*]
[1]Faculty of Information Technology, Beijing University of Technology
[2]School of Software Technology, Zhejiang University
[3]School of Computer Science and Technology, Beijing Jiaotong University
{lyugengyu, lizhengcn, yangzhen}@bjut.edu.cn, weiqikang@emails.bjut.edu.cn,
wanghaobo@zju.edu.cn, shfeng@bjtu.edu.cn

Abstract

In Multi-View Multi-Label Learning, each instance is described by several heterogeneous features and associated with multiple valid labels simultaneously. Existing methods mainly focus on leveraging feature-level view fusion to capture a common representation for multi-label classifier induction. In this paper, we take a new perspective and propose a new semantic-level fusion model named Common-Individual Semantic Fusion Multi-View Multi-Label Learning Method (CISF). Different from previous feature-level fusion model, our proposed method directly focuses on semantic-level view fusion and simultaneously take both the common semantic across different views and the individual semantic of each specific view into consideration. Specifically, we first assume each view involves some common semantic labels while owns a few exclusive semantic labels. Then, the common and exclusive semantic labels are separately forced to be consensus and diverse to excavate the consistences and complementarities among different views. Afterwards, we introduce the low-rank and sparse constraint to highlight the label co-occurrence relationship of common semantics and the view-specific expression of individual semantics. We provide theoretical guarantee for the strict convexity of our method by properly setting parameters. Extensive experiments on various data sets have verified the superiority of our method.

1 Introduction

Multi-View Multi-Label Learning (MVML) learns from the training data, where each object is represented by several heterogeneous feature representations and associated with multiple class labels simultaneously [Bickel and Scheffer, 2004; Wu *et al.*, 2019; Wu *et al.*, 2020; Lyu *et al.*, 2022a]. Recently, such learning paradigm has been widely used in many real-world applications. For example, in the task of news webpage classification, one news webpage can be represented by diverse channel information including *video*, *image* and *text*, while at the same time it is annotated with multiple class labels such as *FIFA World Cup*, *Sports* and *Lionel Messi*. Multi-view multi-label learning provides an effective framework to learn a desired multi-label classifier from such heterogeneous data and predict proper labels for unseen examples.

Figure 1: An Example of Multi-View Multi-Label Learning. The news webpage is represented by several different channels including *video*, *image* and *text*, and associated with multiple class labels such as *FIFA World Cup*, *Sports* and *Lionel Messi* simultaneously.

The key to deal with multi-view multi-label data lies in how to effectively integrate these heterogenous features while ensuring all relevant labels can be characterized comprehensively. A general strategy is to learn a latent multi-view subspace representation, which can characterize multiple semantic labels as much intact as possible. [Zhang *et al.*, 2018a] proposes a matrix factorization based shared subspace representation method, which employs Hilbert-Schmidt independence criterion to strengthen its ability of consensus semantic characterization. [Zhao *et al.*, 2023] also seeks for a latent low-dimensional representation, while it focuses on each specific view and employs the structural view-label consistency information to enhance the expressions of view-specific semantics. [Lu *et al.*, 2023] proposes a bipartite graph based multi-view embedding representation method, and it imposes a joint low-rank constraint on both the embedding representation matrix and multi-label classifier matrix to enhance its robustness toward the labels with dependencies. Obviously, the above MVML methods just leverage the multi-view consensus and complementary relationship in the feature space and they formulate multi-view subspace as implicit semantic-aware representation.

*Zhen Yang and Songhe Feng are the corresponding authors.

In this paper, different from traditional feature-level multi-view fusion MVML methods, we take the first attempt to conduct multi-view fusion under the guidance of semantic fusion, and propose a Common-Individual Semantic Fusion Multi-View Multi-Label Learning Method (CISF). Specifically, we first assume each view corresponds to one view-specific label set and each of them contains two different kinds of semantics - Common Semantics & Individual Semantics. The common semantics refers to the core semantic labels shared by multiple views, which reflects the consensus information across different views. The individual semantics refers to the exclusive semantic labels owned by each specific view, which characterizes the complementary information of diverse views. Afterwards, we separately introduce the low-rank and sparse constraint to highlight the label co-occurrence relationship of the common semantics and the exclusive semantic representation of the individual semantics. Finally, we embed an adaptive global label correlations to enhance the semantic integrity for improving the performance of the final multi-label model.

In summary, the main contribution of our paper lie in the following aspects:

- We propose a new Common-Individual Semantic Fusion Multi-View Multi-Label Learning Method (CISF). To the best of our knowledge, it is the first attempt to directly leverage multi-view fusion on the semantic space.
- Considering that single common semantics cannot characterize all relevant labels, we simultaneously consider the commonality and individuality of multi-view data, and introduce an adaptive global label correlations to enhance the semantic integrity of the final model.
- We provide theoretical guarantee for the strict convexity of CISF by properly setting parameters and develop an alternative optimization algorithm to solve it. Extensive results have verified the superiority of our method.

2 Related Work

2.1 Multi-View Learning (MVL)

Multi-View Learning aims to learn a desired multi-view representation from different views by leveraging the consensus and complementary information across heterogeneous features [Bickel and Scheffer, 2004; Wang *et al.*, 2016; Zhang *et al.*, 2017; Gu *et al.*, 2023]. A core purpose of multi-view learning is to encapsulate multi-view information of different views to learn a share or common representation for clustering. Based on the way to exploit multi-view information, existing multi-view learning methods can be roughly divided into the following categories: canonical correlation analysis [Andrew *et al.*, 2013; Wang *et al.*, 2015], multi-view subspace clustering [Gao *et al.*, 2015; Cao *et al.*, 2015; Kang *et al.*, 2020; Wang *et al.*, 2021], multi-view matrix factorization [Liu *et al.*, 2013; Zhao *et al.*, 2017], and deep multi-view clustering [Li *et al.*, 2019; Xu *et al.*, 2023]. Besides, there are also many other multi-view learning methods for different tasks, such as retrieval [Yan *et al.*, 2020], recommendation [Flanagan *et al.*, 2021] and classification [Han *et al.*, 2022; Lyu *et al.*, 2022a], etc.

2.2 Multi-Label Learning (MLL)

In Multi-Label Learning, each instance is represented by a single feature vector and annotated with multiple valid labels [Wen *et al.*, 2022]. Label correlation is a fundamental challenge to be utilized for improving the performance of multi-label learning. Based on the order of label correlations being exploited for model training, existing MLL methods can be roughly grouped into three categories: first-order correlations [Zhang *et al.*, 2018b], second-order correlations [Madjarov *et al.*, 2010; Li *et al.*, 2017] and high-order correlations [Burkhardt and Kramer, 2018]. The above methods mainly are formulated under full supervised settings while such phenomenon may not hold in real-world scenarios due to expensive annotation efforts. Recently, some weakly supervised MLL frameworks are proposed and have been widely used in many applications, such as semi-supervised MLL [Wang *et al.*, 2020], MLL with missing labels [Zhu *et al.*, 2018], partial multi-label learning [Lyu *et al.*, 2020; Li *et al.*, 2021; Lyu *et al.*, 2022b; Wang *et al.*, 2023], etc.

2.3 Multi-View Multi-Label Learning (MVML)

In Multi-View Multi-Label Learning, each instance is represented by several heterogeneous features and associated with multiple valid labels [Lyu *et al.*, 2024]. Obviously, such paradigm can be regard as an integration of multi-view learning and multi-label learning, and the key to deal with MVML data lies in how to integrate heterogeneous features effectively while realize multi-label classification accurately. [Zhang *et al.*, 2020] propose a sparse feature selection MVML method, which exploits both view relations and label correlations to select discriminative features for further multi-label model training. [Wu *et al.*, 2019] propose a view-specific MVML method named SIMM, which simultaneously leverages shared subspace exploitation and view-specific information extraction to enhance the performance of multi-label classifier. Except for the above MVML methods, there are also some weakly supervised MVML methods, including MVML with missing labels [Huang *et al.*, 2019], MVML with missing views [Tan *et al.*, 2018], non-aligned MVML [Zhao *et al.*, 2023; Zhong *et al.*, 2024], etc.

3 The Proposed Method

3.1 Notations

Formally speaking, we denote $\mathcal{X} = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \ldots \times \mathbb{R}^{d_V}$ as the feature space with V views and $\mathcal{Y} = \{c_1, c_2, \ldots, c_q\}$ as the label space with q class labels, where d_t $(1 \leq t \leq T)$ is the feature dimension of t-th view. Given the multi-view multi-label training data $\mathcal{D} = \{(\boldsymbol{X}_i, \boldsymbol{y}_i)|1 \leq i \leq n\}$ with n instances, where each $\boldsymbol{X}_i \in \mathcal{X}$ is represented by V feature vectors $[\boldsymbol{x}_i^{(1)}; \boldsymbol{x}_i^{(2)}; \ldots; \boldsymbol{x}_i^{(V)}]$ and $\boldsymbol{y}_i \in \{0,1\}^{q\times 1}$ is the label vector associated with $\boldsymbol{X}_i$, our proposed CISF aims to integrate these heterogeneous representations from different views to construct a robust multi-label classifier $\boldsymbol{f} : \mathcal{X} \mapsto 2^{\mathcal{Y}}$ and further predicts some proper labels for unseen instances.

3.2 Formulation

Consistences and complementarities are two key ingredients for boosting multi-view multi-label learning. Existing

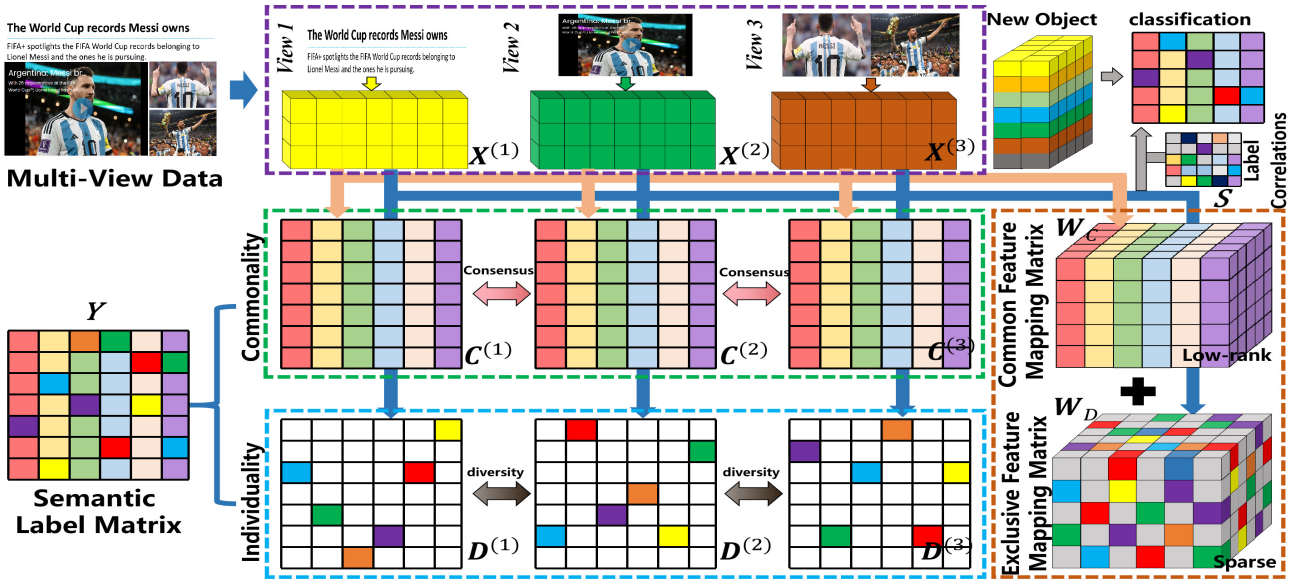

Figure 2: The overview architecture of our proposed CISF method. The semantic label matrix contains two different kinds of semantics information: Common Semantics and Individual Semantics. The common semantics reflects the consensus information across different views and the individual semantics characterizes the complementary information of diverse views.

MVML methods generally advocates different view to predict the same label results to use consistent information across different views and simultaneously considers the different contribution weights of each specific view to learn the complementary information among different views. Eq. (1) illustrates a general MVML framework:

$$\min_{\boldsymbol{W}^{(i)}} \sum_{i=1}^{V} \mu^{(i)} \left\| \boldsymbol{Y} - \boldsymbol{W}^{(i)} \boldsymbol{X}^{(i)} \right\|_F^2 + \Gamma(\boldsymbol{W}^{(i)}), \tag{1}$$

where $\boldsymbol{Y} \in \mathbb{R}^{q \times n}$ is the label matrix, $\boldsymbol{X}^{(i)} \in \mathbb{R}^{d_i \times n}$ is the feature matrix of i-th view, $\boldsymbol{W}^{(i)} \in \mathbb{R}^{q \times d_i}$ is the mapping matrix from features to semantics, $\Gamma(\cdot)$ is the regularization term and $\mu^{(i)}$ is the weight of i-th view.

According to Eq.(1), we can easily see that existing MVML methods mainly formulate multi-view consistent and complementary information in the feature space. Basically, they target to seek one or several good view that can characterize all relevant labels as completely as possible. Obviously, such multi-view fusion strategy cannot respect every view sufficiently and even lead the semantic representation of some rare labels to be overwhelmed by the core labels.

Hence, in this paper, we directly take the multi-view fusion in the semantic space and explicitly measure multi-view consistences and complementarities in objective fusion model. Specifically, we assume that each view corresponds to a view-specific label set $\boldsymbol{Y}^{(i)} \in \mathbb{R}^{q \times n}$ and each label set contains two different kinds of semantics - Common Semantics $\boldsymbol{C}^{(i)} \in \mathbb{R}^{q \times n}$ and Individual Semantics $\boldsymbol{D}^{(i)} \in \mathbb{R}^{q \times n}$, i.e.,

$$\min_{\boldsymbol{C},\boldsymbol{D}} \sum_{i=1}^{V} \mu^{(i)} \left\| \boldsymbol{Y}^{(i)} - \left(\boldsymbol{C}^{(i)} + \boldsymbol{D}^{(i)} \right) \right\|_F^2 + \Phi(\boldsymbol{C}) + \Psi(\boldsymbol{D}). \tag{2}$$

The common semantics $\boldsymbol{C}^{(i)} \in \mathbb{R}^{q \times n \times V}$ refers to the shared semantic labels represented by all V views and it reflects the consensus semantic information across different views. In our model, we introduce Hilbert-Schmidt Independence Criterion [Gretton *et al.*, 2005] to constrain the semantic consistence of the V different views, i.e.,

$$\Phi(\boldsymbol{C}) = - \sum_{i,j=1}^{V} \mu^{(i)} \mu^{(j)} \mathcal{HSIC}(\boldsymbol{C}^{(i)}, \boldsymbol{C}^{(j)}). \tag{3}$$

The individual semantics $\boldsymbol{D}^{(i)} \in \mathbb{R}^{q \times n \times V}$ refers to the exclusive semantic labels owned by each specific view, which characterizes the diversities and complementarities among different views. Based on the assumption that the individual semantics is exclusive for each specific view and the diversity is also sparse across different views, we measure such semantic diversities and complementarities by minimizing the sum of the product of each pair of individual semantics, i.e.,

$$\Psi(\boldsymbol{D}) = \sum_{i,j=1}^{V} \mu^{(i)} \mu^{(j)} Tr \left(\boldsymbol{D}^{(i)} \cdot \boldsymbol{D}^{(j)\top} \right). \tag{4}$$

Except for the above semantics consistences and complementarities, another inherent property of learning from multi-view multi-label data is how to utilize label correlations. Different from previous fixed label co-occurrence relationships, we try to leverage a dynamic label correlations $\boldsymbol{S} \in \mathbb{R}^{q \times q}$ and recover all relevant labels by minimizing

$$\min_{\boldsymbol{S}} \sum_{i=1}^{V} \mu^{(i)} \left\| \boldsymbol{Y} - \boldsymbol{S} \boldsymbol{Y}^{(i)} \right\|_F^2. \tag{5}$$

In addition, in order to construct the direct correspondences from features to semantics and obtain a desired multi-label classifier for unseen examples prediction simultaneously, we introduce two feature mapping matrices $\boldsymbol{W}_C^{(i)}, \boldsymbol{W}_D^{(i)} \in \mathbb{R}^{q \times d_i}$ that correspond to the common semantics $\boldsymbol{C}^{(i)}$ and the individual semantics $\boldsymbol{D}^{(i)}$ respectively, i.e., $\boldsymbol{C}^{(i)} =$

$\boldsymbol{W}_C^{(i)}\boldsymbol{X}^{(i)}$ and $\boldsymbol{D}^{(i)} = \boldsymbol{W}_D^{(i)}\boldsymbol{X}^{(i)}$. By integrating the above functions (2)-(5), we can obtain the final framework of our proposed CISF method as follows:

$$\begin{aligned}
&\min_{\boldsymbol{W}_C^{(i)},\boldsymbol{W}_D^{(i)},\boldsymbol{S}} \sum_{i=1}^{V} \mu^{(i)} \left\| \boldsymbol{Y} - \boldsymbol{S}(\boldsymbol{W}_C^{(i)} + \boldsymbol{W}_D^{(i)})\boldsymbol{X}^{(i)} \right\|_F^2 \\
&+ \alpha \sum_{i,j=1}^{V} \mu^{(i)}\mu^{(j)} \left(-\mathcal{HSIC}(\boldsymbol{S}\boldsymbol{W}_C^{(i)}\boldsymbol{X}^{(i)}, \boldsymbol{S}\boldsymbol{W}_C^{(j)}\boldsymbol{X}^{(j)}) \right) \\
&+ \beta \sum_{i,j=1}^{V} \mu^{(i)}\mu^{(j)} Tr\left(\boldsymbol{W}_D^{(i)}\boldsymbol{X}^{(i)} \cdot \left(\boldsymbol{W}_D^{(j)}\boldsymbol{X}^{(j)} \right)^\top \right) \\
&+ \gamma \sum_{i=1}^{V} \|\boldsymbol{W}_C^{(i)}\|_* + \eta \sum_{i=1}^{V} \|\boldsymbol{W}_D^{(i)}\|_F^2.
\end{aligned} \tag{6}$$

Here, the common feature mapping matrix $\boldsymbol{W}_C^{(i)}$ is constrained with nuclear norm to preserve its low-rank property, since its represented shared semantic labels tend to have statistical co-occurrence. The exclusive feature mapping matrix is constrained with F-norm, since the diversities of exclusive semantics are always expressed as sparse. The weights $\mu^{(i)}$ are defined by inverse distance weighting strategy to avoid undesired hyperparameters [Nie *et al.*, 2016].

3.3 Optimization

To optimize (6) conveniently, we introduce an additional variable constraint $\boldsymbol{A}^{(i)} = \boldsymbol{S}\boldsymbol{W}_C^{(i)}\boldsymbol{X}^{(i)}$ and convert (6) to its Augmented Lagrange Multiplier (ALM) form as follows:

$$\begin{aligned}
&\min_{\boldsymbol{W}_C^{(i)},\boldsymbol{W}_D^{(i)},\boldsymbol{A}^{(i)},\boldsymbol{S}} \sum_{i=1}^{V} \mu^{(i)} \|\boldsymbol{Y} - \boldsymbol{S}(\boldsymbol{W}_C^{(i)} + \boldsymbol{W}_D^{(i)})\boldsymbol{X}^{(i)}\|_F^2 \\
&- \alpha \sum_{i,j=1}^{V} \mu^{(i)}\mu^{(i)} HSIC(\boldsymbol{A}^{(i)}, \boldsymbol{A}^{(j)}) + \gamma \sum_{i=1}^{V} \|\boldsymbol{W}_C^{(i)}\|_* \\
&+ \beta \sum_{i,j=1}^{V} \mu^{(i)}\mu^{(j)} Tr\left(\boldsymbol{W}_D^{(i)}\boldsymbol{X}^{(i)} \cdot (\boldsymbol{W}_D^{(j)}\boldsymbol{X}^{(j)})^\top \right) \\
&+ \sum_{i=1}^{V} \frac{\lambda^{(i)}}{2} \left\| \boldsymbol{A}^{(i)} - \boldsymbol{S}\boldsymbol{W}_C^{(i)}\boldsymbol{X}^{(i)} \right\|_F^2 + \eta \sum_{i=1}^{V} \|\boldsymbol{W}_D^{(i)}\|_F^2 \\
&+ \sum_{i=1}^{V} Tr\left({\boldsymbol{M}^{(i)}}^\top \left(\boldsymbol{A}^{(i)} - \boldsymbol{S}\boldsymbol{W}_C^{(i)}\boldsymbol{X}^{(i)} \right) \right).
\end{aligned} \tag{7}$$

Obviously, the above function involves four variables $\boldsymbol{W}_C^{(i)}$, $\boldsymbol{W}_D^{(i)}$, $\boldsymbol{A}^{(i)}$ and $\boldsymbol{S}$, which cannot be optimized simultaneously. Therefore, we adopt the alternating minimization strategy and update these variables iteratively.

Update $\boldsymbol{W}_C^{(i)}$ with other variables fixed. We can calculate $\boldsymbol{W}_C^{(i)}$ by minimizing the following objective function:

$$\min_{\boldsymbol{W}_C^{(i)}} \sum_{i=1}^{V} \zeta^{(i)} \left\| \frac{\Theta}{\zeta^{(i)}} - \boldsymbol{S}\boldsymbol{W}_C^{(i)}\boldsymbol{X}^{(i)} \right\|_F^2 + \gamma \sum_{i=1}^{V} \|\boldsymbol{W}_C^{(i)}\|_* \tag{8}$$

where $\Theta = 2\mu^{(i)}(\boldsymbol{Y} - \boldsymbol{S}\boldsymbol{W}_D^{(i)}\boldsymbol{X}^{(i)}) + (\lambda^{(i)}\boldsymbol{C}^{(i)} + \boldsymbol{M}^{(i)})$ and $\zeta^{(i)} = (2\mu^{(i)} + \lambda^{(i)})/2$. According to [Zhu *et al.*, 2010], (8) has the closed form solution and the variable $\boldsymbol{W}_C^{(i)}$ can be optimized following $\boldsymbol{W}_C^{(i)} = \mathcal{S}_{\frac{\gamma}{2\zeta^{(i)}}}(\frac{\Theta}{2\zeta^{(i)}})$, where $\mathcal{S}$ is the singular value thresholding.

Algorithm 1 The Training Process of **CISF**

Inputs:
$\mathcal{D}$: MVML training data $\{(\boldsymbol{x}_i^{(v)}, \boldsymbol{y}_i)|i \in [n], v \in [V]\}$;
α, β, γ and η: the trade-off parameters;
I_{max} : the number of maximum iterations;
$\boldsymbol{x}^{(i)*}$: the unseen example.
Process:
Initialized $\boldsymbol{W}_C^{(i)}$, $\boldsymbol{W}_D^{(i)}$, $\boldsymbol{A}^{(i)}$, $\boldsymbol{S}$ and $\mu^{(i)}$;
while $t < I_{max}$ **do**
  for $i = 1, 2, \ldots, V$ **do**
    Update $\boldsymbol{W}_C^{(i)}$ by solving (8);
  end for
  for $i = 1, 2, \ldots, V$ **do**
    Update $\boldsymbol{W}_D^{(i)}$ by solving (9);
  end for
  for $i = 1, 2, \ldots, V$ **do**
    Update $\boldsymbol{A}^{(i)}$ by solving (10);
  end for
  Update $\boldsymbol{S}$ by solving (11);
  for $i = 1, 2, \ldots, V$ **do**
    Update $\boldsymbol{M}^{(i)}$ and $\lambda^{(i)}$ by Eq. (12);
  end for
  if converge **then**
    break;
  end if;
end while;
Output:
$\boldsymbol{y}^*$: the predicted label $\sum_{i=1}^{V} \mu^{(i)} \boldsymbol{S}(\boldsymbol{W}_C^{(i)} + \boldsymbol{W}_D^{(i)})\boldsymbol{x}^{(i)*}$.

Update $\boldsymbol{W}_D^{(i)}$ with other variables fixed. The variable $\boldsymbol{W}_D^{(i)}$ can be updated following:

$$\begin{aligned}
&\min_{\boldsymbol{W}_D^{(i)}} \sum_{i=1}^{V} \mu^{(i)} \|\boldsymbol{Y} - \boldsymbol{S}(\boldsymbol{W}_C^{(i)} + \boldsymbol{W}_D^{(i)})\boldsymbol{X}^{(i)}\|_F^2 + \eta \sum_{i=1}^{V} \|\boldsymbol{W}_D^{(i)}\|_F^2 \\
&+ \beta \sum_{i,j=1}^{V} \mu^{(i)}\mu^{(i)} Tr\left(\boldsymbol{W}_D^{(i)}\boldsymbol{X}^{(i)} \cdot (\boldsymbol{W}_D^{(j)}\boldsymbol{X}^{(j)})^\top \right).
\end{aligned} \tag{9}$$

We take the derivative of (9) with respect to $\boldsymbol{W}_D^{(i)}$ to 0. Afterwards, based on KKT conditions, we can easily update $\boldsymbol{W}_D^{(i)}$ in an iterative manner [Tan *et al.*, 2021].

Update $\boldsymbol{A}^{(i)}$ with other variables fixed. The optimization subproblem with regard to $\boldsymbol{A}^{(i)}$ can be reformulated as:

$$\begin{aligned}
&\min_{\boldsymbol{A}^{(i)}} \sum_{i=1}^{V} \frac{\lambda^{(i)}}{2} \left\| \boldsymbol{A}^{(i)} - \boldsymbol{S}\boldsymbol{W}_C^{(i)}\boldsymbol{X}^{(i)} + \frac{1}{\lambda^{(i)}}\boldsymbol{M}^{(i)} \right\|_F^2 \\
&- \alpha \sum_{i,j=1}^{V} \mu^{(i)}\mu^{(i)} \mathcal{HSIC}(\boldsymbol{A}^{(i)}, \boldsymbol{A}^{(i)}).
\end{aligned} \tag{10}$$

Here, $\mathcal{HSIC}(\boldsymbol{A}^{(i)}, \boldsymbol{A}^{(j)}) = (n-1)^{-2} Tr\left(\boldsymbol{H}\boldsymbol{K}^{(i)}\boldsymbol{H}\boldsymbol{K}^{(j)}\right)$, $\boldsymbol{K}^{(i)} = {\boldsymbol{A}^{(i)}}^\top \boldsymbol{A}^{(i)}$ is the Gram matrix and $\boldsymbol{H}$ centers it to

have zero mean. Theorem 1 (in section 5.3) guarantees the subproblem (10) to be convex and the optimal solution could be obtained by setting its derivative with respect to $\boldsymbol{A}^{(i)}$ to 0.

Update $\boldsymbol{S}$ with other variables fixed. The variable $\boldsymbol{S}$ can be updated by solving the following sub-problem:

$$\begin{aligned}\min_{\boldsymbol{S}} &\sum_{i=1}^{V} \mu^{(i)} \left\| \boldsymbol{Y} - \boldsymbol{S}(\boldsymbol{W}_C^{(i)} + \boldsymbol{W}_D^{(i)})\boldsymbol{X}^{(i)} \right\|_F^2 \\ &+ \sum_{i=1}^{V} \frac{\lambda^{(i)}}{2} \left\| \boldsymbol{A}^{(i)} - \boldsymbol{S}\boldsymbol{W}_C^{(i)}\boldsymbol{X}^{(i)} \right\|_F^2 \\ &+ Tr\left(\boldsymbol{M}^{(i)\top} \left(\boldsymbol{A}^{(i)} - \boldsymbol{S}\boldsymbol{W}_C^{(i)}\boldsymbol{X}^{(i)} \right) \right)\end{aligned} \tag{11}$$

Similar to (9), we also take the derivative of (11) with respect to $\boldsymbol{S}$ to 0, and then we can obtain its closed-form solution.

Update $\boldsymbol{M}^{(i)}$ and $\lambda^{(i)}$ with other variables fixed. Finally, we update the Lagrange multiplier matrices $\boldsymbol{M}^{(i)}$ and penalty scalars $\lambda^{(i)}$ following:

$$\begin{aligned}\boldsymbol{M}^{(i)^{t+1}} &= \boldsymbol{M}^{(i)^t} + \lambda^{(i)^t} \left(\boldsymbol{A}^{(i)^t} - \boldsymbol{S}\boldsymbol{W}_C^{(i)^t}\boldsymbol{X}^{(i)^t} \right) \\ \lambda^{(i)^{t+1}} &= \min\left(\lambda_{max}, \tau\lambda^{(i)^t} \right)\end{aligned} \tag{12}$$

During the process of model training, we first initialize the required variables, and then repeat the above steps until the algorithm converges or reaches the maximum iterations. Finally, we make prediction for unseen instance following $\boldsymbol{y}^* = \sum_{i=1}^{V} \mu^{(i)} \boldsymbol{S}(\boldsymbol{W}_C^{(i)} + \boldsymbol{W}_D^{(i)})\boldsymbol{x}^{(i)^*}$. Algorithm 1 summarizes the whole procedure of our proposed CISF method.

4 Experiments

4.1 Experimental Setting

To evaluate the performance of our proposed CISF method, we implement experiments on seven widely-used MVML data sets, including *Emotions*, *Scene*, *Corel5k*, *Espgame*, *Pascal*, *Iaprtc12* and *Mirflickr* data sets. Table 1 summarizes the detailed characteristics of the above data sets.

Data sets	Instances	Views	$D_{min-max}$	Labels
Emotions	593	2	8 - 64	6
Scene	2407	2	98 - 196	6
Corel5k	4999	4	100 - 3895	260
Pascal	9963	5	512 - 4086	20
Iaprtc12	19627	6	100 - 3985	291
Espgame	20770	4	100 - 4096	268
Mirflickr	25000	5	100 - 4096	457

Table 1: Characteristics of our employed data sets. $D_{min-max}$ is the smallest-largest dimensions of features.

Meanwhile, we compare our proposed CISF with the following five state-of-the-art MVML methods, including **LSPC** [Szymanski *et al.*, 2016], **FIMAN** [Wu *et al.*, 2020], **ICM2L** [Tan *et al.*, 2021], **BEMVL** [Lu *et al.*, 2023] and **NAIM3L** [Li and Chen, 2022]. The configured parameters of the above methods are set according to the suggestions in their corresponding literature.

In addition, five popular multi-label evaluation metrics are employed to measure the performance of each comparing method, including *Hamming Loss (H-L)*, *Ranking Loss (R-L)*, *One Error (O-E)*, *Coverage (COV)* and *Average Precision (A-P)* [Zhang and Zhou, 2013]. For each dataset, we randomly select 70% examples for training, 10% examples for parameter tuning and 20% examples for evaluation, where each algorithm is run 5 times independently. The codes and data sets are provided in https://gengyulyu.github.io/homepage/.

4.2 Experimental Results

Table 2 illustrates the experimental comparisons between our proposed CISF and other five comparing methods on all evaluation metrics, where the average metrics results and standard deviations are recorded respectively. According to Table 2, out of 210 (7 data sets $\times$ 6 methods $\times$ 5 metrics) statistical comparisons can make the following observations:

- Among all five comparing methods, our proposed CISF method is superior to **LSPC**, **FIMAN** and **ICM2L** in almost all cases, and it also outperforms **BEMVL** and **NAIM3L** in 91.42% and 88.57% cases, respectively.
- Among all employed evaluation metrics, our proposed CISF achieves the best performance in 97.14% cases on *Hamming Loss*, *Ranking Loss* and *Average Precision* metrics. And on *One Error* and *Coverage* metrics, it is also superior to other methods over 94% cases.
- Among all employed datasets, CISF outperforms almost all comparing methods on *Emotions*, *Scene*, *Pascal*, *Iaprtc12* and *Mirflickr* datasets. And it also achieve superior performance against other comparing methods over 82% cases on *Corel5k* and *Espgame* data sets.
- Overall, our proposed CISF method can achieve competitive performance against other *feature-fusion* based MVML methods, which demonstrates the effective of our proposed multi-view *semantic-fusion* strategy.

In order to comprehensively evaluate the superiority of CISF, *Friedman test* [Demšar, 2006] is utilized as the statistical test to analyze the relative performance among the comparing algorithms. According to Table 3, the null hypothesis of distinguishable performance among the comparing algorithms is rejected at 0.05 significance level. Thus, we further employ the post-hoc Bonferroni-Dunn test [Demšar, 2006] to show the relative performance among the comparing algorithms. Figure 3 illustrates the CD diagrams on each evaluation metric, where the average rank of each algorithm is marked along the axis. According to Figure 3, it is observed that CISF always ranks 1st on all evaluation metrics.

5 Further Analysis

5.1 Ablation Study

In order to evaluate the effect of the each components of our proposed CISF, we conduct the Ablation Study between CISF and its three degenerated algorithms CISFnC, CISFnI and CISFnL, where each degenerated algorithm ignores the common semantics, individual semantics and label correlations,

H-L	Emotions	Scene	Corel5k	Pascal	Iaprtc12	Espgame	Mirflickr
CISF	**0.203±0.020**	**0.169±0.003**	0.029±0.000	**0.073±0.000**	**0.019±0.000**	**0.017±0.000**	**0.005±0.000**
LSPC	0.263±0.017	0.253±0.011	0.028±0.001	0.239±0.005	0.036±0.000	0.029±0.000	0.018±0.002
FIMAN	0.240±0.015	0.313±0.009	0.019±0.000	0.123±0.000	0.035±0.000	0.036±0.000	-
ICM2L	0.306±0.026	0.279±0.016	0.055±0.005	0.153±0.004	0.054±0.000	0.020±0.000	0.010±0.000
BEMVL	0.386±0.019	0.219±0.004	0.025±0.000	0.130±0.008	0.031±0.000	0.028±0.000	0.013±0.000
NAIM3L	0.242±0.044	0.178±0.021	**0.013±0.000**	0.087±0.007	**0.019±0.000**	**0.017±0.000**	0.006±0.000
R-L	Emotions	Scene	Corel5k	Pascal	Iaprtc12	Espgame	Mirflickr
CISF	**0.164±0.010**	**0.115±0.008**	**0.081±0.000**	**0.123±0.008**	**0.104±0.001**	0.184±0.006	**0.218±0.014**
LSPC	0.199±0.026	0.251±0.022	0.889±0.015	0.881±0.004	0.995±0.001	0.992±0.000	0.721±0.008
FIMAN	0.185±0.007	0.241±0.013	0.141±0.004	0.149±0.002	0.138±0.003	0.186±0.004	-
ICM2L	0.330±0.048	0.349±0.067	0.115±0.064	0.216±0.033	0.179±0.004	0.204±0.005	0.284±0.002
BEMVL	0.408±0.036	0.126±0.008	0.394±0.012	0.355±0.062	0.272±0.006	0.281±0.004	0.469±0.036
NAIM3L	0.244±0.090	0.173±0.052	0.113±0.008	0.180±0.037	0.144±0.011	**0.174±0.010**	0.266±0.006
O-E	Emotions	Scene	Corel5k	Pascal	Iaprtc12	Espgame	Mirflickr
CISF	**0.253±0.018**	0.326±0.018	**0.591±0.252**	**0.450±0.020**	0.495±0.005	**0.549±0.023**	**0.837±0.044**
LSPC	0.308±0.037	0.429±0.033	0.912±0.018	0.936±0.009	0.992±0.001	0.989±0.003	0.944±0.004
FIMAN	0.279±0.013	0.547±0.028	0.602±0.020	0.468±0.005	0.559±0.005	0.659±0.008	-
ICM2L	0.439±0.064	0.665±0.061	0.746±0.143	0.585±0.011	0.652±0.022	0.709±0.029	0.873±0.007
BEMVL	0.334±0.072	**0.151±0.011**	0.602±0.018	0.599±0.061	**0.272±0.007**	0.867±0.032	0.873±0.007
NAIM3L	0.500±0.471	0.333±0.235	0.842±0.036	0.525±0.244	0.829±0.039	0.873±0.054	0.978±0.010
COV	Emotions	Scene	Corel5k	Pascal	Iaprtc12	Espgame	Mirflickr
CISF	**1.791±0.118**	**0.724±0.053**	101.8±4.736	**3.100±0.155**	**93.44±0.995**	120.2±1.426	**131.5±9.822**
LSPC	2.182±0.129	1.398±0.117	279.2±9.160	19.57±0.089	286.1±0.036	281.3±0.109	331.3±2.831
FIMAN	1.948±0.123	1.298±0.076	83.26±1.928	4.069±0.059	116.2±1.673	113.5±1.892	-
ICM2L	2.700±0.177	1.852±0.328	113.96±6.985	5.625±0.790	139.7±2.300	128.6±2.223	226.1±1.326
BEMVL	2.944±0.197	0.777±0.046	185.9±3.660	8.385±1.268	193.6±2.313	167.6±1.329	320.5±15.61
NAIM3L	2.173±0.389	0.950±0.270	**65.59±4.159**	4.650±0.828	112.7±6.610	**110.9±6.533**	154.9±3.200
A-P	Emotions	Scene	Corel5k	Pascal	Iaprtc12	Espgame	Mirflickr
CISF	**0.805±0.005**	**0.789±0.010**	**0.355±0.205**	**0.620±0.016**	**0.314±0.003**	**0.366±0.001**	0.132±0.020
LSPC	0.741±0.022	0.618±0.023	0.059±0.005	0.109±0.006	0.020±0.000	0.021±0.005	**0.168±0.008**
FIMAN	0.783±0.004	0.649±0.017	0.332±0.009	0.605±0.004	0.309±0.003	0.267±0.003	-
ICM2L	0.658±0.049	0.546±0.051	0.221±0.114	0.486±0.031	0.230±0.007	0.221±0.014	0.102±0.001
BEMVL	0.590±0.026	0.786±0.008	0.146±0.006	0.376±0.047	0.232±0.004	0.227±0.004	0.055±0.010
NAIM3L	0.726±0.083	0.729±0.066	0.328±0.017	0.547±0.075	0.267±0.020	0.253±0.009	0.101±0.005

Table 2: Experimental comparisons of our proposed CISF with other comparing methods on six evaluation metrics, where the best performances on each metric are shown in bold face. "-" indicates that FIMAN needs over 128G of RAM on *Mirflickr* data set.

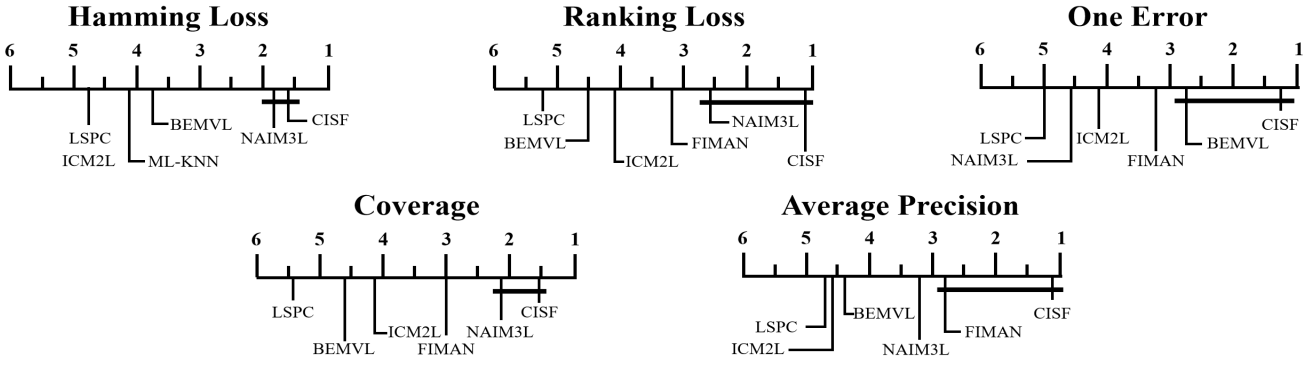

Figure 3: Experimental Comparisons of all comparing algorithms with the Bonferroni-Dunn test (CD = 2.576 at 0.05 significance level).

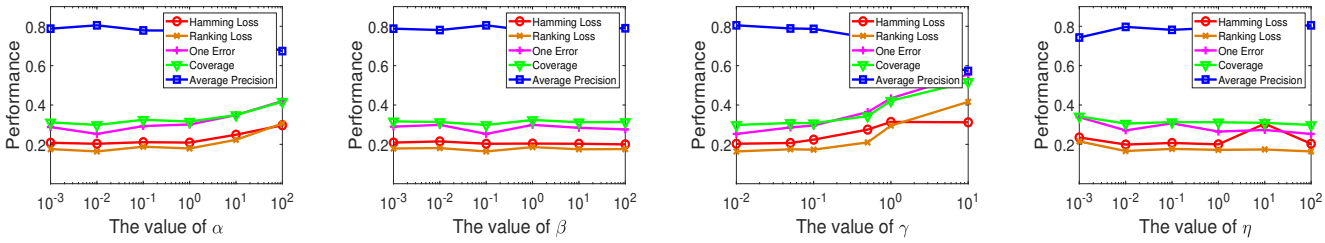

Figure 4: The parameter analysis of CISF on *Emotions* data set, where the *Coverage* results are normalized by the number of class labels (q) so as to make all metric results be characterized in a unified figure.

Evaluation Metric	τ_F	critical value
Hamming Loss	5.536	2.533
Ranking Loss	10.705	
One Error	6.927	
Coverage	10.924	Methods: 6, Data sets: 7
Average Precision	7.125	

Table 3: Friedman statics τ_F in terms of each evaluation metric.

Methods	H-L	R-L	O-E	COV	A-P
CISFnC	0.296	0.174	0.293	1.905	0.778
CISFnI	0.264	0.172	0.286	1.895	0.789
CISFnL	0.208	0.175	0.290	1.852	0.785
CISF	**0.203**	**0.164**	**0.253**	**1.791**	**0.805**

Table 4: The ablation study of CISF on *Emotions* data set.

respectively. Table 4 reports the experimental comparison between these methods on *Emotions* data set. According to Table 4, we can find that CISFnI achieves better performance than CISFnC method, which shows that common semantics has greater contribution than individual semantics to the effectiveness of learning model. Besides, our proposed CISF is superior to both CISFnC and CISFnI methods, which demonstrates that the common semantics and individual semantics can jointly improve the performance of MVML model.

5.2 Parameter Sensitivity

We study the sensitivity analysis of our proposed CISF with respect to its four employed parameters α, β, γ and η. Figure 4 shows the performance of CISF under different parameter configurations on *Emotions* data set. According to Figure 4, we can find that α and γ usually have great influence on the performance of the proposed model, and we select the optimal values of them from $\{10^{-3}, 10^{-2}, \ldots, 10^{2}\}$ and $\{0.01, 0.05, \ldots, 10\}$, respectively. Meanwhile, other parameters often follow the optimal configurations $\beta = 0.1$ and $\eta = 100$ but vary with minor adjustments on different data sets. In addition, in our experiments, the value of λ_{max} is set to $1e^6$ and the maximum iterations I_{max} is set to 50.

5.3 Complexity Analysis

At each iteration, the computational cost mainly comes from the derivative calculation and singular value decomposition (SVD) operations, and the complexity of our optimization procedure mainly comes from the optimization of three sub-problems with regard to $\boldsymbol{W}_C^{(i)}$, $\boldsymbol{W}_D^{(i)}$ and $\boldsymbol{A}^{(i)}$. For simplicity, we suppose the dimensionality of each view is d. And the complexity of these sub-problems are $\mathcal{O}(dnr)$, $\mathcal{O}(Tnqd)$ and $\mathcal{O}(q^2d + qdn + q^3)$ respectively, where r is the rank of $\boldsymbol{W}_C^{(i)}$ and T is the iteration number for updating $\boldsymbol{W}_D^{(i)}$.

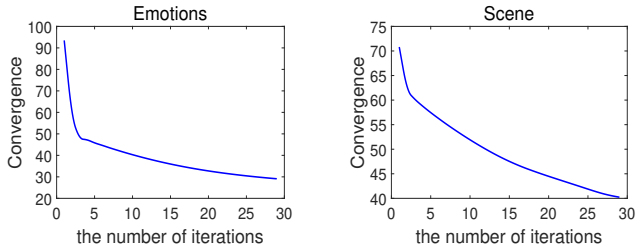

Figure 5: The convergence curves on *Emotions* and *Scene* data sets.

5.4 Convergence Analysis

The convergence of the whole optimization problem (6) depends on how to guarantee the subproblem (10) is convex, especially its HSIC term is negative. Theorem 1 provides the theoretical guarantee for the convexity of (10) under proper parameter setting. Besides, Figure 5 illustrates the convergence curves on *Emotions* and *Scene* data sets, which also empirically demonstrates the convergence of our model.

Theorem 1: The problem (10) is convex given the parameter $\lambda^{(i)} \geq 8q(V-1)\alpha\mu_i\mu_j$, where V is the number of views.

Proof: Due to the page limitation, we provided the proof of Theorem 1 in https://gengyulyu.github.io/homepage/.

6 Conclusion

In this paper, we proposed a Common-Individual Semantic Fusion Multi-View Multi-Label Learning Method. Different from previous feature-fusion based MVML methods, it is the first attempt to conduct multi-view fusion under the guidance of the semantic fusion, where both common semantics and individual semantics are simultaneously incorporated into the multi-view fusion process to learn a desired multi-label classification model. Extensive experimental results on various MVML datasets has demonstrated the effectiveness of our proposed multi-view semantic-fusion strategy.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (No. 2023YFB3107100), the National Natural Science Foundation of China (No. 62306020), the China Postdoctoral Science Foundation (No. 2022M720320), the Beijing Postdoctoral Science Foundation (No. 2023-zz-78), the Fundamental Research Funds for the Central universities (No. 2022JBZY019), the Beijing Natural Science Foundation (No. 4242046), the Major Research Plan of National Natural Science Foundation of China (No. 92167102).

References

[Andrew *et al.*, 2013] G. Andrew, R. Arora, J. Bilmes, and K. Livescu. Deep canonical correlation analysis. In *International Conference on Machine Learning*, pages 1247–1255, 2013.

[Bickel and Scheffer, 2004] S. Bickel and T. Scheffer. Multi-view clustering. In *International Conference on Data Mining*, volume 4, pages 19–26, 2004.

[Burkhardt and Kramer, 2018] S. Burkhardt and S. Kramer. Online multi-label dependency topic models for text classification. *Machine Learning*, 107(5):859–886, 2018.

[Cao *et al.*, 2015] X. Cao, C. Zhang, H. Fu, S. Liu, and H. Zhang. Diversity-induced multi-view subspace clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–594, 2015.

[Demšar, 2006] J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7(Jan):1–30, 2006.

[Flanagan *et al.*, 2021] A. Flanagan, W. Oyomno, A. Grigorievskiy, K. Tan, S. Khan, and M. Ammad-Ud-Din. Federated multi-view matrix factorization for personalized recommendations. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 324–347, 2021.

[Gao *et al.*, 2015] H. Gao, F. Nie, X. Li, and H. Huang. Multi-view subspace clustering. In *IEEE International Conference on Computer Vision*, pages 4238–4246, 2015.

[Gretton *et al.*, 2005] A. Gretton, O. Bousquet, A. Smola, and B. Schölkopf. Measuring statistical dependence with hilbert-schmidt norms. In *International Conference on Algorithmic Learning Theory*, pages 63–77, 2005.

[Gu *et al.*, 2023] Z. Gu, S. Feng, R. Hu, and G. Lyu. Onion: Joint unsupervised feature selection and robust subspace extraction for graph-based multi-view clustering. *ACM Transactions on Knowledge Discovery from Data*, 17(5):1–23, 2023.

[Han *et al.*, 2022] Z. Han, C. Zhang, H. Fu, and J. Zhou. Trusted multi-view classification. In *International Conference on Learning Representations*, pages 2551–2566, 2022.

[Huang *et al.*, 2019] J. Huang, F. Qin, X. Zheng, Z. Cheng, Z. Yuan, W. Zhang, and Q. Huang. Improving multi-label classification with missing labels by learning label-specific features. *Information Sciences*, 492:124–146, 2019.

[Kang *et al.*, 2020] Z. Kang, W. Zhou, Z. Zhao, J. Shao, M. Han, and Z. Xu. Large-scale multi-view subspace clustering in linear time. In *AAAI Conference on Artificial Intelligence*, pages 4412–4419, 2020.

[Li and Chen, 2022] 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*, 44(10):5918–5932, 2022.

[Li *et al.*, 2017] Y. Li, Y. Song, and J. Luo. Improving pairwise ranking for multi-label image classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3617–3625, 2017.

[Li *et al.*, 2019] Z. Li, Q. Wang, Z. Tao, Q. Gao, and Z. Yang. Deep adversarial multi-view clustering network. In *International Joint Conference on Artificial Intelligence*, pages 2952–2958, 2019.

[Li *et al.*, 2021] Z. Li, G. Lyu, and S. Feng. Partial multi-label learning via multi-subspace representation. In *International Conference on International Joint Conferences on Artificial Intelligence*, pages 2612–2618, 2021.

[Liu *et al.*, 2013] J. Liu, C. Wang, J. Gao, and J. Han. Multi-view clustering via joint nonnegative matrix factorization. In *SIAM International Conference on Data Mining*, pages 252–260, 2013.

[Lu *et al.*, 2023] X. Lu, S. Feng, G. Lyu, Y. Jin, and C. Lang. Distance-preserving embedding adaptive bipartite graph multi-view learning with application to multi-label classification. *ACM Transactions on Knowledge Discovery from Data*, 17(2):1–21, 2023.

[Lyu *et al.*, 2020] G. Lyu, S. Feng, and Y. Li. Partial multi-label learning via probabilistic graph matching mechanism. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 105–113, 2020.

[Lyu *et al.*, 2022a] G. Lyu, X. Deng, Y. Wu, and S. Feng. Beyond shared subspace: A view-specific fusion for multi-view multi-label learning. In *AAAI Conference on Artificial Intelligence*, pages 7647–7654, 2022.

[Lyu *et al.*, 2022b] G. Lyu, S. Feng, W. Liu, S. Liu, and C. Lang. Redundant label learning via subspace representation and global disambiguation. *ACM Transactions on Intelligent Systems and Technology*, 14(1):1–19, 2022.

[Lyu *et al.*, 2024] G. Lyu, Z. Yang, X. Deng, and S. Feng. L-vsm: Label driven view-specific fusion for multi-view multi-label classification. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–15, 2024.

[Madjarov *et al.*, 2010] G. Madjarov, D. Gjorgjevikj, and T. Delev. Efficient two stage voting architecture for pairwise multi-label classification. In *Australasian Joint Conference on Artificial Intelligence*, pages 164–173, 2010.

[Nie *et al.*, 2016] F. Nie, J. Li, and X. Li. Parameter-free auto-weighted multiple graph learning: a framework for

multiview clustering and semi-supervised classification. In *International Joint Conference on Artificial Intelligence*, pages 1881–1887, 2016.

[Szymanski *et al.*, 2016] P. Szymanski, T. Kajdanowicz, and K. Kersting. How is a data-driven approach better than random choice in label space division for multi-label classification? *Entropy*, 18(8):282, 2016.

[Tan *et al.*, 2018] Q. Tan, G. Yu, C. Domeniconi, J. Wang, and Z. Zhang. Incomplete multi-view weak-label learning. In *International Joint Conference on Artificial Intelligence*, pages 2703–2709, 2018.

[Tan *et al.*, 2021] Q. Tan, G. Yu, and J. Wang. Individuality- and commonality-based multiview multilabel learning. *IEEE Transactions on Cybernetics*, 51(3):1716–1727, 2021.

[Wang *et al.*, 2015] W. Wang, R. Arora, K. Livescu, and J. Bilmes. On deep multi-view representation learning. In *International Conference on Machine Learning*, pages 1083–1092, 2015.

[Wang *et al.*, 2016] W. Wang, X. Yan, H. Lee, and K. Livescu. Deep variational canonical correlation analysis. *arXiv preprint arXiv:1610.03454*, 2016.

[Wang *et al.*, 2020] L. Wang, Y. Liu, C. Qin, G. Sun, and Y. Fu. Dual relation semi-supervised multi-label learning. In *AAAI Conference on Artificial Intelligence*, pages 6227–6234, 2020.

[Wang *et al.*, 2021] S. Wang, X. Liu, X. Zhu, P. Zhang, Y. Zhang, F. Gao, and E. Zhu. Fast parameter-free multi-view subspace clustering with consensus anchor guidance. *IEEE Transactions on Image Processing*, 31:556–568, 2021.

[Wang *et al.*, 2023] H. Wang, S. Yang, G. Lyu, W. Liu, T. Hu, K. Chen, S. Feng, and G. Chen. Deep partial multi-label learning with graph disambiguation. In *International Joint Conference on Artificial Intelligence*, pages 4308–4316, 2023.

[Wen *et al.*, 2022] J. Wen, Z. Zhang, L. Fei, B. Zhang, Y. Xu, Z. Zhang, and J. Li. A survey on incomplete multiview clustering. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(2):1136–1149, 2022.

[Wu *et al.*, 2019] X. Wu, Q. Chen, Y. Hu, D. Wang, X. Chang, X. Wang, and M. Zhang. Multi-view multi-label learning with view-specific information extraction. In *International Joint Conference on Artificial Intelligence*, pages 3884–3890, 2019.

[Wu *et al.*, 2020] J. Wu, X. Wu, Q. Chen, and M. Zhang. Feature-induced manifold disambiguation for multi-view partial multi-label learning. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 557–565, 2020.

[Xu *et al.*, 2023] J. Xu, Y. Ren, H. Tang, Z. Yang, L. Pan, Y. Yang, X. Pu, S. Philip, and L. He. Self-supervised discriminative feature learning for deep multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 35(7):7470–7482, 2023.

[Yan *et al.*, 2020] C. Yan, B. Gong, Y. Wei, and Y. Gao. Deep multi-view enhancement hashing for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(4):1445–1451, 2020.

[Zhang and Zhou, 2013] M. Zhang and Z. Zhou. A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 26(8):1819–1837, 2013.

[Zhang *et al.*, 2017] H. Zhang, V. Patel, and R. Chellappa. Hierarchical multimodal metric learning for multimodal classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3057–3065, 2017.

[Zhang *et al.*, 2018a] C. Zhang, Z. Yu, Q. Hu, P. Zhu, and X. Wang. Latent semantic aware multi-view multi-label classification. In *AAAI Conference on Artificial Intelligence*, pages 4414–4421, 2018.

[Zhang *et al.*, 2018b] M. Zhang, Y. Li, X. Liu, and X. Geng. Binary relevance for multi-label learning: an overview. *Frontiers of Computer Science*, 12(2):191–202, 2018.

[Zhang *et al.*, 2020] Y. Zhang, J. Wu, Z. Cai, and P. Yu. Multi-view multi-label learning with sparse feature selection for image annotation. *IEEE Transactions on Multimedia*, 22(11):2844–2857, 2020.

[Zhao *et al.*, 2017] H. Zhao, Z. Ding, and Y. Fu. Multi-view clustering via deep matrix factorization. In *AAAI Conference on Artificial Intelligence*, pages 2921–2927, 2017.

[Zhao *et al.*, 2023] Dawei Zhao, Qingwei Gao, Yixiang Lu, and Dong Sun. Non-aligned multi-view multi-label classification via learning view-specific labels. *IEEE Transactions on Multimedia*, 25:7235–7247, 2023.

[Zhong *et al.*, 2024] Q. Zhong, G. Lyu, and Z. Yang. Align while fusion: A generalized non-aligned multi-view multi-label classification method. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–10, 2024.

[Zhu *et al.*, 2010] G. Zhu, S. Yan, and Y. Ma. Image tag refinement towards low-rank, content-tag prior and error sparsity. In *ACM International Conference on Multimedia*, pages 461–470, 2010.

[Zhu *et al.*, 2018] P. Zhu, Q. Xu, Q. Hu, C. Zhang, and H. Zhao. Multi-label feature selection with missing labels. *Pattern Recognition*, 74:488–502, 2018.