# Uncertainty-Aware Global-View Reconstruction for Multi-View Multi-Label Feature Selection

**Pingting Hao[1], Kunpeng Liu[2], Wanfu Gao[3,4*]**

[1] College of Computer Science and Technology, Jilin University, China
[2] Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, China
[3] Department of Computer Science, Portland State University, Portland, OR 97201 USA
haopingting@jlu.edu.cn, kunpeng@pdx.edu, gaowf@jlu.edu.cn

## Abstract

In recent years, multi-view multi-label learning (MVML) has gained popularity due to its close resemblance to real-world scenarios. However, the challenge of selecting informative features to ensure both performance and efficiency remains a significant question in MVML. Existing methods often extract information separately from the consistency part and the complementary part, which may result in noise due to unclear segmentation. In this paper, we propose a unified model constructed from the perspective of global-view reconstruction. Additionally, while feature selection methods can discern the importance of features, they typically overlook the uncertainty of samples, which is prevalent in realistic scenarios. To address this, we incorporate the perception of sample uncertainty during the reconstruction process to enhance trustworthiness. Thus, the global-view is reconstructed through the graph structure between samples, sample confidence, and the view relationship. The accurate mapping is established between the reconstructed view and the label matrix. Experimental results demonstrate the superior performance of our method on multi-view datasets.

## Introduction

Feature selection plays a vital role in various domains, including healthcare (Bharati, Mondal, and Podder 2023), anomaly detection (Zhang, Xu, and Zhou 2024), and transportation (Wang et al. 2023). It has been applied in different scenarios, such as unsupervised (Zhang et al. 2024), single-label (Cohen et al. 2023), and multi-label (Klonecki, Teisseyre, and Lee 2023) scenarios, based on specific requirements. The challenge of selecting the optimal subset of features is particularly daunting in the presence of high-dimensional data. However, the multi-view setting offers the advantage of rich semantics combined with more comprehensive data, which in turn introduces additional complexities for feature selection in MVML.

Differentiating from the single-view multi-label scenario, effectively managing the view relationship and integrating information from multiple views is a fundamental aspect in multi-view scenario (Lyu et al. 2024a,b). Previous works in the multi-view scenario primarily focus on identifying commonality across views as an initial step, in accordance with the defined consistency rules between views (Xu, Tao, and Xu 2013; Sun 2013). Typically, the consensus part is derived from the features (Zhang et al. 2018; Liu et al. 2023c; Yin and Zhang 2023; Liu et al. 2015), as depicted in Figure 1a. However, the distinctive part is often overlooked, resulting in a deficiency of information and subsequent performance degradation of the model.

Given the increasing recognition of the significance of the distinctive part within each view, it is imperative to acknowledge the exploration of individual information. Noteworthy works, such as (Liu et al. 2023a; Li and Chen 2021), operate under the assumption of label matrix consistency for each view and establish mappings between each view and the label matrix, as depicted in Figure 1b. Alternatively, Tan et al. (Tan et al. 2019) construct separate mappings for common and individual views, thereby addressing both aspects. In reality, the hypothesis of an inconsistent label matrix based on each view, as illustrated in Figure 1c, can also be considered. Zhao et al. (Zhao et al. 2022) partition the original label matrix into multiple matrices corresponding to each view to preserve the distinctive information within each view. Nonetheless, the accuracy of the process for solving feature weights is compromised due to biased grounds or incomplete constraints. This is primarily because the perspective is not grounded on the ground-truth of the entire view to avoid unclear segmentation.

Although some works such as (Zhu et al. 2020) assume that the label matrix for the whole dataset and each view are observed, and subsequently establish mappings between features and labels for both. However, they often overlook the view relationship and fail to consider the sample-level relationship, as depicted in Figure 1d. Thus, there most likely exists noise and redundancy in direct concatenating matrix. How to utilize effective information to reconstruct the whole dataset becomes challenging.

Indeed, it is important to acknowledge that the label information inherently encompasses valuable comprehensive information. This distinguishes it from the unsupervised multi-view scenario. The effective utilization of this information significantly impacts the performance of feature selection models. Recent works, such as (Zhu, Li, and Zhang 2015) and (Zhang et al. 2020b), leverage the inherent integrity property of the label matrix to establish a mapping relationship between multiple views and a single label matrix.

*correponding author

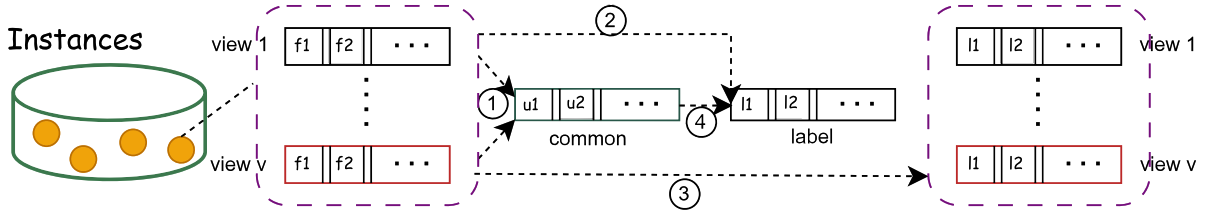



Figure 1: Four typical types for the relationship between features and labels in MVML, namely, a) common-based mapping including ① and ④, b) label consistency mapping including ②, c) view-specific mapping including ③, and d) concatenating view mapping including ④.

Nevertheless, the effective learning of weights for all features remains a challenging task due to the inherent many-to-one mapping relationship.

Moreover, it is worth considering the relaxation of the assumption that the reliability of each view, and even each sample, is uniformly equal to 1. In reality, the quality of histopathological images may vary across different patients (Yagi 2011), and tabular data may contain feature noise (Yelipe, Porika, and Golla 2018). As a result, there can be disparate confidence levels among samples. Therefore, the ability to perceive the informativeness of different samples can enhance the explainability (Han et al. 2022), and facilitate the identification of correct samples for potential data adjustment and retraining. This is particularly crucial in multi-view multi-label feature selection as it helps eliminate interference caused by samples, and mitigates the introduction of noise when establishing the mapping for feature weights.

To tackle the above issues, we propose a novel method called **U**ncertainty-aware **G**lobal-view **R**econstruction for Multi-view Multi-label **F**eature **S**election (UGRFS). By taking into account the complete information contained in the label matrix, we introduce a regularizer paradigm that captures the relationship between the higher-dimensional label space and the initialization of the fusion feature space using view weights. In addition, our method incorporates constraints that operate on the graph structure of the global-view and its splitting structure with uncertainty. Overall, the main contributions of this paper can be summarized as follows:

- The UGRFS method introduces a sparse model that perceives sample confidence in the multi-view multi-label scenario. Its objective is to tackle noise and redundancy across views. This finer-grained perception not only facilitates evaluation and adjustment based on expert knowledge, but also promotes with the computation of feature weights mutually to improve performance.
- The regularization paradigms for global-view reconstruction aim to establish an accurate mapping that considers the graph structure, sample-relationship, and view-relationship. In our method, the new variable for global-view is automatically learned instead of being fused with predefined weights. This allows for balancing the consistency and complementary within one variable.
- To ensure algorithm convergence, the multiplicative update rules are applied to the optimization process. Subsequently, extensive experiments are conducted on six multi-view multi-label datasets. The experimental results validate the superiority of UGRFS over state-of-the-art methods.

## Related Work

### Multi-Label Feature Selection

In the multi-label scenario, instances are associated with multiple labels to capture their rich semantics (Zhang et al. 2020a). Unlike single-label or unsupervised methods, some existing multi-label feature selection methods focus on the label matrix. Specific to the noises in label matrix, Jian et al. (Jian et al. 2016) utilize low-rank techniques to obtain the latent label matrix. Additionally, the reconstruction of the label matrix utilizes label correlation and the dependence of the original label space to constrain its graph structure, as demonstrated in previous works (Fan et al. 2021a; Huang and Wu 2021; Han et al. 2022).

Considering mainly for feature selection problem, how to deal with the feature weights is of primary importance for multi-label scenario. The design of the regularizer for feature weights considers structure constraints and leverages the benefits of different norms. Some methods incorporate label correlation as a constraint or combine multiple paradigms (Huang et al. 2019; Li et al. 2022; Li, Hu, and Gao 2023). Alternatively, other techniques decompose the feature weight matrix to enhance the relationship between features and labels (Lin et al. 2023), or extract local discriminative features (Fan et al. 2021b). Although the multi-label problem shares similarities with MVML, the distinction lies in the incorporation of multiple views.

### Multi-View Multi-Label Learning

Expansion based on multi-label scenario, the cases for diverse description of data has explosive growth (Sanghavi and Verma 2022). Other expansion cases are also discussed based on MVML such as the non-aligned case (Zhong, Lyu, and Yang 2024) and incomplete views. Combined with label matrix, complex cases with incomplete views include but not limited to missing labels (Liu et al. 2023b), label noises (Liu, Sun, and Feng 2022) and partial labels (Jiang et al. 2024). Existing MVML methods strive to balance consistency and complementarity to preserve the comprehensive information (Lyu et al. 2022; Wu et al. 2019). Zhu et

al. (Zhu et al. 2018) propose a method to incorporate diverse information from the feature space and label space into the latent space of each view, deviating from traditional loss functions in MVML.

## The Proposed Method

### Problem Formulation

Given a multi-view multi-label dataset $(X, Y)$, $X = \left\{X^{(i)}\right\}_{i=1}^{V}$ contains $V$ different views, and $X^{(i)} \in \mathbb{R}^{n\times d(i)}$ with the same number of instances $n$ in different views. Each element $x_{m.}^{(i)} \in X^{(i)}$ is a $d(i)$-dimensional feature vector and denotes the $m$-th instance for $i$-th view. The fusion view with view-relationship is setting as $X^f \in \mathbb{R}^{n\times d}$, and the global-view distribution represents as $D \in \mathbb{R}^{n\times d}$ where $d$ denotes the number of features. The sample confidence matrix is defined as non-negative variable $C = \left\{C^{(i)}\right\}_{i=1}^{V}$ and $C^{(i)} \in \mathbb{R}^{n\times 1}$. In the observed label matrix $Y \in \{0,1\}^{n\times l}$, the $j$-th label is denoted by $y_{.j}$. In this section, the main procedure of UGRFS is introduced, including global-view reconstruction $L_G$ and uncertainty-aware feature selection $L_U$, which can be formalized: $L_{URGFS} = L_G(\cdot) + L_U(\cdot)$.

### Global-View Reconstruction

**View-Relationship Fusion** Due to the uneven distribution of information content among different views, directly concatenating them easily ignores the difference between views. Instead, estimating the contribution of each view to the global-view matrix can be achieved through the computation of similarity in local geometric structures.

To compute the similarity of the local structure, we exploit it in the feature space. The affinity matrix, denoted as $S^{(v)}$, is computed for the features in the $v$-th view as a core component of graph Laplacian (Chung 1997). The element in the matrix is defined as:

$$s_{ij}^{(v)} = \begin{cases} exp(-\frac{||x_{i.}^{(v)} - x_{j.}^{(v)}||_2^2}{\sigma^2}) & if\ x_{i.}^{(v)} \in \mathcal{N}_q(x_{j.}^{(v)})\ or \\ & x_{j.}^{(v)} \in \mathcal{N}_q(x_{i.}^{(v)}) \\ 0 & otherwise \end{cases}, \tag{1}$$

where $\sigma$ represents a predefined parameter, and $\mathcal{N}_q(x_{j.}^{(v)})$ represents the set of $q$ nearest neighbors of $x_{i.}^{(v)}$. Similarly, the meaning of $\mathcal{N}_q(x_{i.}^{(v)})$ can be inferred. The formula based on the smooth assumption (Zhu 2005) can be represented as:

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} s_{ij}^{(v)}\left(y_{i.} - y_{j.}\right)^2. \tag{2}$$

According to the definition of the graph Laplacian, the graph Laplacian matrix $L_x^{(v)} = A^{(v)} - S^{(v)}$, where $A^{(v)}$ is the diagonal matrix, whose elements are $a_{ii}^{(v)} = \sum_{j=1}^{n} s_{ij}^{(v)}$. Thus, Formula (2) can be transformed into Formula (3) as:

$$Tr(y^T\left(A^{(v)} - S^{(v)}\right)y) = Tr(y^T L_x^{(v)} y). \tag{3}$$

We enforce the constraints that view weights are non-negative and the sum of the view weights for all views equals to 1. The view weight $v_i$ for the $i$-th view can be determined as follows:

$$v_i = \frac{1/Tr(y^T L_x^{(i)} y)}{\sum_{i=1}^{V} 1/Tr(y^T L_x^{(i)} y)}. \tag{4}$$

Therefore, we represent the initial global-view as the fusion matrix $X^f$, which takes into account the relationships between views. The concatenation representation can be expressed as:

$$X^f = \left[v_1 X^{(1)}, v_2 X^{(2)} \cdots v_V X^{(V)}\right]. \tag{5}$$

**Global-View Distribution** Given the observed label matrix $Y$, the complete information could be transformed into distribution that facilitates the reconstruction of the global-view (Xu, Liu, and Geng 2019). To achieve this, a nonlinear transformation is applied to the label matrix, mapping the preliminary fusion matrix $X^f$ to a higher-dimensional label space. The higher-dimensional label space $\rho(Y)$ is constructed using Gaussian kernel, which can be expressed as:

$$\rho(Y) = exp(-\frac{J_1 Y^T(\sum_{i=1}^{l}(y_{ij})^2) - YY^T}{(avg(p\_dist))^2}), \tag{6}$$

where $p\_dist$ represents the array of pairwise distances in the label matrix, and $avg(p\_dist)$ means the average value for all pairwise distances. Other definitions such as $J_1 \in \mathbb{R}^{n\times 1}$ are represented as the full one matrix, and $Y(\cdot)$ denotes a series of operations on the element in the label matrix. Then, the global-view distribution can be modeled through the coefficient matrix $\hat{W}_y$ as:

$$D = \rho(Y)\hat{W}_y + b = Y_x W_y, \tag{7}$$

where $b$ is the bias for this model. To facilitate matrix computations, the coefficient matrix and the higher-dimensional label space are transformed into $W_y = \begin{bmatrix}\hat{W}_y \\ b\end{bmatrix} \in \mathbb{R}^{(n+1)\times d}$ and $Y_x = [\rho(Y), J_1] \in \mathbb{R}^{n\times(n+1)}$, respectively.

Specific to the design for the global-view reconstruction $L_G$, two aspects need to be highlighted. First, it is essential to consider the interrelationship between different views. Second, preserving the structure of the global-view is of utmost importance. Since this method is specifically tailored for multi-view scenarios, the roles of the matrices involved in the function $L_G$ cannot be interchanged. Consequently, a more refined regularizer paradigm for $L_G$ can be formulated, taking into account these two aspects:

$$L_G(\cdot) = \left\|D - X^f\right\|_F^2 + Tr(D^T L^Y D), \tag{8}$$

where the way to define $L^Y$ is the same as $L_x^{(v)}$.

### Uncertainty-Aware Feature Selection

In terms of dimensionality, feature selection involves selecting meaningful columns. In order to account for the uncertainty of samples, different proportions of information are

extracted from each row in the view. By considering both dimensions, the loss function for the uncertainty-aware feature selection model is constructed as:

$$loss = \sum_{i=1}^{V} \left\| diag(C^{(i)})X^{(i)}W^{(i)} - Y \right\|_F^2 + \Phi(W), \quad (9)$$

where $W^{(i)} \in \mathbb{R}^{d(i)\times c}$ denotes the feature weight of each view. The $\Phi(W)$ denotes the regularization paradigm of $W$ to control the model complexity, and $W \in \mathbb{R}^{d\times c}$ denotes the global-view feature weight matrix. The inclusion of the learnable confidence vector $C^{(i)}$ effectively mitigates interference, thereby accentuating the significance of pertinent features.

In addition to considering the constraints of view-relationship and structure similarity, the uncertainty perception is also incorporated as a penalty function in this model. Therefore, the penalty function is defined as:

$$penalty = \sum_{i=1}^{V} \left\| Y_x W_y^{(i)} - diag(C^{(i)})X^{(i)} \right\|_F^2, \quad (10)$$

where $W_y = \left\{ W_y^{(i)} \right\}_{i=1}^{V}$ and $W_y^{(i)} \in \mathbb{R}^{(n+1)\times d(i)}$ which is splitting from the coefficient weight $W_y$. Thus, the definition for $Y_x W_y^{(i)}$ can be seen the part of $D$ which can be written as $D^{(i)}$. Simultaneously, each uncertainty-aware view incorporates characteristics of the global distribution, enabling accurate computation of the learnable $C^{(i)}$. Overall, the uncertainty-aware feature selection $L_U$ can be formulated by combining the loss function and the penalty as:

$$\begin{aligned} L_U(\cdot) &= \sum_{i=1}^{V} \left\| diag(C^{(i)})X^{(i)}W^{(i)} - Y \right\|_F^2 \\ &+ \sum_{i=1}^{V} \left\| D^{(i)} - diag(C^{(i)})X^{(i)} \right\|_F^2 + \|W\|_{2,1}. \end{aligned} \quad (11)$$

The term $diag(C^{(i)})X^{(i)}$ can be interpreted as the decomposition of the global-view reconstruction into individual views. This is equivalent to taking into account both the sample-level and view-level relationships in the solution for $W^{(i)}$. Therefore, concatenating the feature weights of each view as a variable for controlling model complexity does not affect the feature order. Following the concatenation procedure, we utilize $l_{2,1}$-norm regularization to induce group sparsity (Yuan and Lin 2006).

## The Objective Function

The model for UGRFS is represented by jointly integrating the procedures for global-view distribution and uncertainty-aware feature selection within a unified framework. The detailed objective function is written as follows:

$$\begin{aligned} &\min_{W^{(i)},C^{(i)},W_y^{(i)}} \sum_{i=1}^{V} \left\| diag(C^{(i)})X^{(i)}W^{(i)} - Y \right\|_F^2 \\ &+ \alpha Tr\left(D^T L^Y D\right) + \beta \sum_{i=1}^{V} \left\| D^{(i)} - diag(C^{(i)})X^{(i)} \right\|_F^2 \\ &+ \gamma \left\| D - X^f \right\|_F^2 + \delta \|W\|_{2,1}, \end{aligned} \quad (12)$$

Algorithm 1: Uncertainty-aware Global-view Reconstruction

**Input**: Data matrices $\left\{ \boldsymbol{X}^{(i)} \right\}_{i=1}^{V}$; Label matrix $\boldsymbol{Y}$.
**Parameter**: Parameters $\alpha$, $\beta$, $\gamma$ and $\delta$.
**Output**: Selected features.

Initialize $\boldsymbol{W^{(i)}}$, $\boldsymbol{C^{(i)}}$, $\boldsymbol{W_y^{(i)}}$;
**repeat**
  Update the matrix $\boldsymbol{W^{(i)}}$ according to Formula (14);
  Update the matrix $\boldsymbol{C^{(i)}}$ according to Formula (15);
  Update the matrix $\boldsymbol{W_y^{(i)}}$ according to Formula (16);
  Update the objective function (13);
**until** Convergence
Obtain the ordered feature sequence by calculating $||\boldsymbol{W}_{(j)}||_2$ where $j = 1, 2, 3, ..., d$;
**return** Top ranked features as ***s-UGRFS-f***.

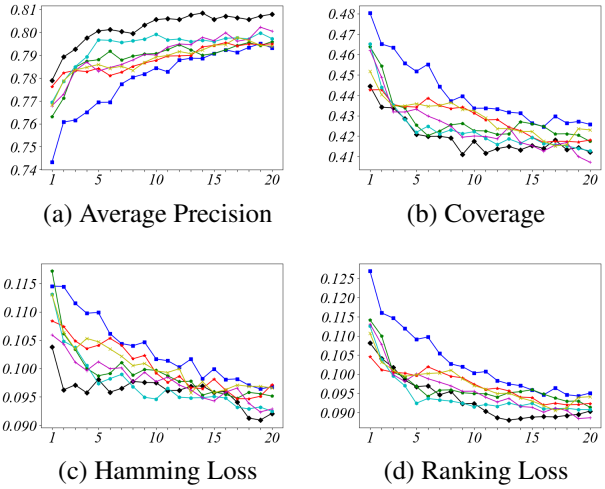
where $\alpha$, $\beta$, $\gamma$ and $\delta$ are trade-off parameters to keep the balance of the model. It can be seen that the $D$ is associated with three regularizer paradigms, which fuses the information for global structure, sample-confidence and view-relationship. The constraints on $C^{(i)}$ primarily aim to minimize the disparity between the splitting structure of the global-view and the uncertainty-aware view. Due to the comprehensive considerations from $D$ to $C^{(i)}$, accurate mapping between features and labels is achieved, leading to precise feature weights acquisition.

## Optimization

Considering the smoothness of the objective function, unless otherwise stated, the related definition used in this paper are listed as follows. The Frobenius norm is denoted as $\|Y\|_F = \sqrt{\sum_{p=1}^{n}\sum_{q=1}^{l} y_{(pq)}^2}$, and it can be transformed to $\|Y\|_F^2 = Tr(Y^T Y)$. The $l_{2,1}$-norm is denoted as $\|W\|_{2,1} = \sum_{p=1}^{d}\sqrt{\sum_{q=1}^{l} w_{(pq)}^2}$, and we can relax the corresponding terms as $\|W\|_{2,1} = 2Tr(W^T E W)$ where the diagonal matrix $e_{ii} = 1/(2\|W_i\|_2)$. The transformation based on Formula (12) can be written as:

$$\begin{aligned} \Theta &= \sum_{i=1}^{V} Tr(F^T F) + \alpha Tr\left(D^T L^Y D\right) \\ &+ \beta \sum_{i=1}^{V} Tr(G^T G) + \gamma Tr(D - X^f)^T (D - X^f) \\ &+ 2\delta Tr\left(W^T E W\right) - Tr\left(\varphi W^{(i)^T}\right) \\ &- Tr(\psi W_y^{(i)^T}) - Tr(\tau diag(C^{(i)})^T), \end{aligned} \quad (13)$$

where $\varphi \in \mathbb{R}^{d(i)\times c}$, $\psi \in \mathbb{R}^{(n+1)\times d(i)}$, $\tau \in \mathbb{R}^{n\times n}$ denote Lagrangian multipliers, and $F = diag(C^{(i)})X^{(i)}W^{(i)} - Y$, $G = D^{(i)} - diag(C^{(i)})X^{(i)}$. To solve the minimization



(a) Average Precision (b) Coverage

(c) Hamming Loss (d) Ranking Loss

Figure 2: Seven methods on SCENE in terms of Average Precision, Coverage, Hamming Loss and Ranking loss.

for Formula (13), we take advantage of a non-negative alternating iterative optimization algorithm where $A^{(i)} = diag(C^{(i)})X^{(i)}$. The description of the procedure is presented in Algorithm 1.

$$W^{(i)} \leftarrow W^{(i)} \circ \frac{{A^{(i)}}^T Y}{{A^{(i)}}^T A^{(i)} W^{(i)} + 2\delta E W^{(i)}}. \tag{14}$$

$$C^{(i)} \leftarrow C^{(i)} \circ \frac{Y {W^{(i)}}^T {X^{(i)}}^T + \beta D^{(i)} {X^{(i)}}^T}{A^{(i)} W^{(i)} {W^{(i)}}^T {X^{(i)}}^T + \beta A^{(i)} {X^{(i)}}^T}. \tag{15}$$

$$W_y^{(i)} \leftarrow W_y^{(i)} \circ \frac{\alpha {Y_x}^T S^Y D^{(i)} + \beta {Y_x}^T A^{(i)} + \gamma {Y_x}^T X^{f(i)}}{\alpha {Y_x}^T A^Y D^{(i)} + (\beta + \gamma) {Y_x}^T D^{(i)}}. \tag{16}$$

## Complexity Analysis

The optimization procedure consists of three primary components, assuming a feature matrix dimension of $d$ for each view. When updating $W^{(i)}$, the computation complexity is $O(d^2 n)$. The computation complexity for updating $C^{(i)}$ is also $O(n^2 d)$. Updating $W_y^{(i)}$ has a computation complexity of $O(n^3 + n^2 d)$. Considering these complexities, the entire training procedure can be conservatively approximated as $O(n^3 + n^2 d + d^2 n)$ per iteration.

# Experiments

## Experimental Setup

**Datasets** As shown in Table 1, we evaluate our method on six widely used multi-view multi-label datasets, namely yeast (Elisseeff and Weston 2001), SCENE (Chua et al. 2009), VOC07 (Everingham and Winn 2010), MIRFlickr

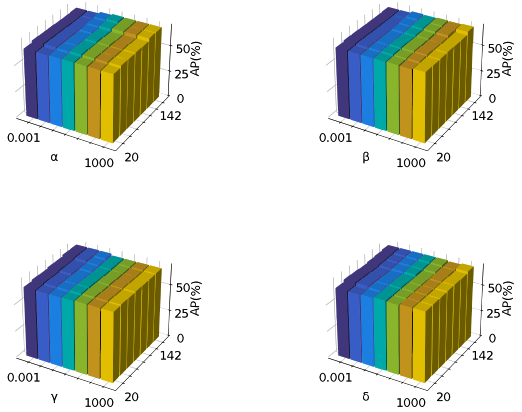


Figure 3: Parameter sensitivity studies on the MIRFlickr dataset.

(Huiskes and Lew 2008), IAPRTC12 (Escalante et al. 2010), and 3Sources (Greene and Cunningham 2009). These datasets cover various domains including genes, vision, and news, and have different numbers of views ranging from 2 to 5. The number of features varies from a few hundred to several thousand. In cases where a view is absent, it is indicated by '-'.

**Comparing Methods** Comparison methods include two multi-view multi-label methods, i.e., M2LD('blue') (Liu et al. 2023c), and MSFS('green') (Zhang et al. 2020b), and four multi-label methods, i.e., MoRE('red') (Liu et al. 2022), MRDM('cyan') (Huang and Wu 2021), MIFS('pink') (Jian et al. 2016) and CLML('yellow') (Li et al. 2022). These methods share a common emphasis on embedded learning, which primarily focuses on the feature space. Thus, the feature weights could be acquired for sorting.

**Evaluation Metrics** We evaluate the effectiveness of our method using four widely adopted metrics (Zhang and Zhou 2013; Gibaja and Ventura 2015), i.e., Average Precision (AP), Coverage, Hamming Loss (HL) and Ranking Loss (RL). A higher value of AP indicates better performance, while the remaining three metrics operate in the opposite direction. The evaluation results are presented as mean accuracy along with standard deviation obtained from five-fold cross-validation. The parameters for each method are tuned within the range of $\{10^{-3}, 10^{-2}, ..., 10^{3}\}$. These methods adhere to the above standards to ensure the validity of the comparative results presented in this paper.

## Experimental Results

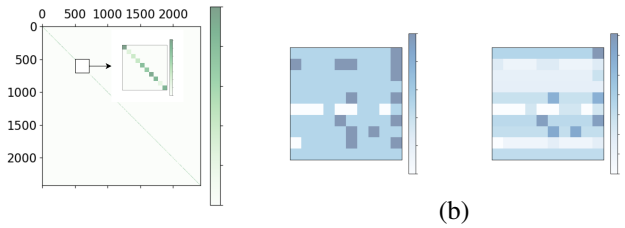
**Feature Selection Performance** Table 2-3 shows the performance of UGRFS, compared with other methods. Each value is calculated based on results obtained using one to twenty percentage features. As can be seen, our method demonstrates superior performance compared to all baseline methods on six datasets. Although the performance of

| Views | yeast | SCENE | VOC07 | MIRFlickr | IAPRTC12 | 3Sources |
|---|---|---|---|---|---|---|
| View-1($d_1$) | GE(79) | CH(64) | DH(100) | DH(100) | DH(100) | BBC(1000) |
| View-2($d_2$) | PP(24) | CM(225) | GIST(512) | GIST(512) | DHV3H1(300) | Reuters(1000) |
| View-3($d_3$) | - | CORR(144) | HH(100) | HH(100) | GIST(512) | Guardian(1000) |
| View-4($d_4$) | - | EDH(73) | - | - | HHV3H1(300) | - |
| View-5($d_5$) | - | WT(128) | - | - | HH(100) | - |
| Instances($n$) | 2417 | 4400 | 3817 | 4053 | 4999 | 169 |
| Features($d$) | 103 | 634 | 712 | 712 | 1312 | 3000 |
| Labels($l$) | 14 | 33 | 20 | 38 | 260 | 6 |

Table 1: The detailed information for the datasets in our experiments.

| Dataset | UGRFS | M2LD | MSFS | MoRE | MRDM | MIFS | CLML |
|---|---|---|---|---|---|---|---|
| | | | | *AP* ↑ | | | |
| yeast | **0.6725±0.014** | 0.6587±0.015 | 0.6495±0.012 | 0.6552±0.016 | 0.6595±0.016 | 0.6445±0.014 | 0.6496±0.012 |
| SCENE | **0.8010±0.009** | 0.7792±0.014 | 0.7887±0.008 | 0.7877±0.007 | 0.7929±0.010 | 0.7896±0.010 | 0.7870±0.008 |
| VOC07 | **0.5871±0.010** | 0.5763±0.009 | 0.5751±0.006 | 0.5664±0.016 | 0.5836±0.008 | 0.5690±0.010 | 0.5790±0.011 |
| MIRFlickr | **0.6770±0.007** | 0.6436±0.013 | 0.6417±0.010 | 0.6322±0.011 | 0.6407±0.008 | 0.6398±0.009 | 0.6442±0.007 |
| IAPRTC12 | **0.1474±0.003** | 0.1370±0.006 | 0.1385±0.005 | 0.1263±0.006 | 0.1400±0.005 | 0.1416±0.003 | 0.1352±0.002 |
| 3Sources | **0.4728±0.055** | 0.4368±0.038 | 0.4252±0.035 | - | 0.4196±0.041 | 0.4623±0.035 | 0.4232±0.046 |
| | | | | *Coverage* ↓ | | | |
| yeast | **0.6239±0.022** | 0.6425±0.017 | 0.6375±0.021 | 0.6357±0.022 | 0.6426±0.023 | 0.6339±0.029 | 0.6407±0.019 |
| SCENE | **0.4214±0.014** | 0.4416±0.016 | 0.4271±0.012 | 0.4295±0.010 | 0.4252±0.017 | 0.4252±0.014 | 0.4308±0.012 |
| VOC07 | **0.4425±0.016** | 0.4547±0.015 | 0.4484±0.016 | 0.4763±0.027 | 0.4531±0.014 | 0.4503±0.022 | 0.4569±0.015 |
| MIRFlickr | **0.5885±0.007** | 0.6069±0.009 | 0.6173±0.013 | 0.6397±0.010 | 0.6128±0.017 | 0.6325±0.013 | 0.6191±0.015 |
| IAPRTC12 | **0.4996±0.008** | 0.5131±0.011 | 0.5098±0.009 | 0.5877±0.020 | 0.5097±0.017 | 0.5004±0.007 | 0.5149±0.005 |
| 3Sources | **0.5301±0.039** | 0.6397±0.030 | 0.6602±0.037 | - | 0.6723±0.042 | 0.6310±0.031 | 0.6727±0.046 |

Table 2: Experimental results (mean ± std) in terms of AP and Coverage, where the 1st/2nd best results are shown in boldface/underline.




(a) (b)

Figure 4: Heatmap of the (a) sample confidence and (b) original features and uncertainty-aware features on yeast.

our method is not significant different from some second-ranking methods, there remains a substantial gap of 4.84% on AP compared to the second-ranking method. The gap can increase to 14.3% when comparing to the last-ranking method. Upon closer analysis, we observe that MIFS and MRDM achieve better performance, ranking second in 37.5% and 29.17% of cases, respectively. The effective combination of feature and label information proves crucial, especially in the multi-view scenario when comparing MIFS and MSFS. We also provide a clear illustration of our performance for all metrics on one dataset in Figure 2.

**Parameter Analysis** As four parameters $\alpha$, $\beta$, $\gamma$ and $\delta$ of UGRFS are involved in two parts, including global-view reconstruction and sparsity in row. We individually adjust each parameter while keeping the others fixed. Figure 3 illustrates the tuning range on the X-axis, ranging from 0.001 to 1000, while the number of features is represented on the Y-axis. The plot demonstrates the impact of these four parameters on our method. Overall, our method exhibits relatively low sensitivity to these parameters, with weak fluctuations primarily observed as the number of features increases.

**Ablation Study** To verify the effectiveness of our design for the objective function, we propose three types of transformations for the original function. The first transformation, UGRFS$_{v1}$, focuses on evaluating the impact of sample confidence, which is beneficial for uncertainty-aware feature selection. The remaining two transformations, UGRFS$_{v2}$ and UGRFS$_{v3}$, prioritize the impact of global-view reconstruction. The main difference between them is that UGRFS$_{v2}$ disregards the variation between views in the initial global-view, while UGRFS$_{v3}$ excludes the related regularizer paradigm for global-view reconstruction, setting parameters $\alpha$, $\beta$ and $\gamma$ as 0. Table 4 presents the performance of the three transformations compared to the original method in terms of AP, Coverage, and RL on yeast,

| Dataset | UGRFS | M2LD | MSFS | MoRE | MRDM | MIFS | CLML |
|---|---|---|---|---|---|---|---|
| | | | | *HL* ↓ | | | |
| yeast | **0.2257±0.005** | 0.2341±0.002 | 0.2299±0.004 | 0.2259±0.006 | 0.2358±0.007 | 0.2279±0.003 | 0.2310±0.004 |
| SCENE | **0.0978±0.004** | 0.1040±0.006 | 0.0999±0.005 | 0.1007±0.005 | 0.0979±0.005 | 0.0987±0.004 | 0.1010±0.005 |
| VOC07 | **0.0847±0.001** | 0.0862±0.002 | 0.0893±0.001 | 0.0890±0.001 | 0.0880±0.001 | 0.0886±0.001 | 0.0884±0.001 |
| MIRFlickr | **0.1775±0.005** | 0.2029±0.007 | 0.1944±0.002 | 0.1931±0.003 | 0.1946±0.003 | 0.1925±0.004 | 0.1917±0.003 |
| IAPRTC12 | **0.01496±0.00** | 0.01653±0.00 | 0.01561±0.00 | 0.01550±0.00 | 0.01560±0.00 | 0.01560±0.00 | 0.01554±0.00 |
| 3Sources | **0.2097±0.020** | 0.2491±0.025 | 0.2357±0.017 | - | 0.2358±0.017 | 0.2332±0.020 | 0.2471±0.018 |
| | | | | *RL* ↓ | | | |
| yeast | **0.2480±0.010** | 0.2605±0.012 | 0.2614±0.010 | 0.2563±0.017 | 0.2574±0.015 | 0.2661±0.015 | 0.2628±0.010 |
| SCENE | **0.0942±0.007** | 0.1037±0.009 | 0.0970±0.006 | 0.0974±0.005 | 0.0949±0.007 | 0.0962±0.007 | 0.0980±0.006 |
| VOC07 | **0.2111±0.006** | 0.2142±0.010 | 0.2133±0.010 | 0.2298±0.017 | 0.2124±0.009 | 0.2162±0.014 | 0.2157±0.011 |
| MIRFlickr | **0.1629±0.005** | 0.1672±0.008 | 0.1855±0.006 | 0.1925±0.006 | 0.1858±0.007 | 0.1880±0.006 | 0.1844±0.006 |
| IAPRTC12 | **0.2020±0.005** | 0.2070±0.009 | 0.2093±0.006 | 0.2452±0.010 | 0.2069±0.013 | 0.2028±0.005 | 0.2134±0.003 |
| 3Sources | **0.4137±0.049** | 0.5045±0.039 | 0.5253±0.043 | - | 0.5360±0.051 | 0.4830±0.037 | 0.5345±0.055 |

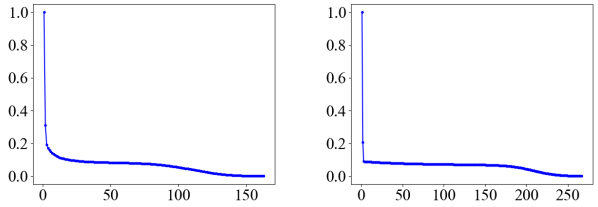
Table 3: Experimental results (mean ± std) in terms of HL and RL, where the 1st/2nd best results are shown in boldface/underline.

| | Metric | UGRFS | UGRFS$_{v1}$ | UGRFS$_{v2}$ | UGRFS$_{v3}$ |
|---|---|---|---|---|---|
| yeast | AP | 0.6725 | 0.6616 | 0.6603 | 0.6610 |
| | Coverage | 0.6239 | 0.6354 | 0.6245 | 0.6237 |
| | RL | 0.2480 | 0.2565 | 0.2496 | 0.2516 |
| MIRFLickr | AP | 0.6770 | 0.6646 | 0.6729 | 0.6470 |
| | Coverage | 0.5885 | 0.6058 | 0.5924 | 0.6249 |
| | RL | 0.1629 | 0.1713 | 0.1659 | 0.1849 |
| IAPRTC12 | AP | 0.1474 | 0.1471 | 0.1472 | 0.1406 |
| | Coverage | 0.4996 | 0.5332 | 0.5358 | 0.5022 |
| | RL | 0.2020 | 0.2034 | 0.2076 | 0.2073 |

Table 4: Ablation studies on the yeast, MIRFlickr and IAPRTC12 datasets for Average Precision, Coverage and Ranking loss.

MIRFlickr, and IAPRTC12. UGRFS achieves the best overall performance, except for a minor difference in terms of Coverage on the yeast dataset. UGRFS$_{v3}$ exhibits the lowest performance in terms of AP, as it ignores all reconstructed components. However, UGRFS$_{v2}$ occasionally performs worse than UGRFS$_{v3}$. This finding emphasizes the importance of correctly computation of view weights. In conclusion, each component within the objective function of our method holds irreplaceable significance.

**Visualization of Sample Confidence** To further visualize the variation in sample confidence, we utilize heatmaps to illustrate the differences between samples, as shown in Figure 4a. It is evident that some samples exhibit similar confidence. But individual sample confidence is not exactly equal due to factors such as noise. In Figure 4b, the distinction lies in the inclusion or exclusion of sample confidence. This results in certain features becoming uninformative while a few features remain significant. Consequently, these changes alter the graph structure of the features.




(a) MIRFlickr (b) VOC07

Figure 5: Convergence curves analysis of UGRFS on (a) MIRFlickr and (b) VOC07.

**Convergence** In Figure 5, we represent the Y-axis as the value of $(z^{t-1} - z^t)/z^{t-1}$, where $z^t$ denotes the objective values at time t. The X-axis denotes the number of iterations, and the stopping criterion depends on whether the value is small, such as 0.001. Additionally, the setting for Y-axis could also reflect the speed of the convergence. To facilitate observation of the oscillation, we set the initial point to approximately one. It can be seen that the curve rapidly declines in the initial iterations without any protrusions.

## Conclusion

This paper explores the consistency and complementary in MVML from a novel perspective. By leveraging the label matrix, it reconstructs the global-view, taking into account the graph structure similarity, sample confidence, and view-relationship to preserve informative features and samples. Extensive experiments demonstrate its stability and the superior performance in feature selection. Moving forward, we aim to further investigate valuable methods for view reconstruction.

## Acknowledgments

This work is funded by: by Science Foundation of Jilin Province of China under Grant No. 20230508179RC, and China Postdoctoral Science Foundation funded project under Grant No. 2023M731281, and Changchun Science and Technology Bureau Project 23YQ05.

## References

Bharati, S.; Mondal, M. R. H.; and Podder, P. 2023. A Review on Explainable Artificial Intelligence for Healthcare: Why, How, and When? *IEEE Transactions on Artificial Intelligence*.

Chua, T.-S.; Tang, J.; Hong, R.; Li, H.; Luo, Z.; and Zheng, Y. 2009. Nus-wide: a real-world web image database from national university of singapore. In *Proceedings of the ACM international conference on image and video retrieval*, 1–9.

Chung, F. R. 1997. *Spectral graph theory*, volume 92. American Mathematical Soc.

Cohen, D.; Shnitzer, T.; Kluger, Y.; and Talmon, R. 2023. Few-sample feature selection via feature manifold learning. In *International Conference on Machine Learning*, 6296–6319. PMLR.

Elisseeff, A.; and Weston, J. 2001. A kernel method for multi-labelled classification. *Advances in neural information processing systems*, 14.

Escalante, H. J.; Hernández, C. A.; Gonzalez, J. A.; López-López, A.; Montes, M.; Morales, E. F.; Sucar, L. E.; Villasenor, L.; and Grubinger, M. 2010. The segmented and annotated IAPR TC-12 benchmark. *Computer vision and image understanding*, 114(4): 419–428.

Everingham, M.; and Winn, J. 2010. The PASCAL visual object classes challenge 2007 (VOC2007) development kit. *Int. J. Comput. Vis*, 88(2): 303–338.

Fan, Y.; Liu, J.; Liu, P.; Du, Y.; Lan, W.; and Wu, S. 2021a. Manifold learning with structured subspace for multi-label feature selection. *Pattern Recognition*, 120: 108169.

Fan, Y.; Liu, J.; Weng, W.; Chen, B.; Chen, Y.; and Wu, S. 2021b. Multi-label feature selection with local discriminant model and label correlations. *Neurocomputing*, 442: 98–115.

Gibaja, E.; and Ventura, S. 2015. A tutorial on multilabel learning. *ACM Computing Surveys (CSUR)*, 47(3): 1–38.

Greene, D.; and Cunningham, P. 2009. A matrix factorization approach for integrating multiple data views. In *Joint European conference on machine learning and knowledge discovery in databases*, 423–438. Springer.

Han, Z.; Yang, F.; Huang, J.; Zhang, C.; and Yao, J. 2022. Multimodal dynamics: Dynamical fusion for trustworthy multimodal classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20707–20717.

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

Huang, R.; and Wu, Z. 2021. Multi-label feature selection via manifold regularization and dependence maximization. *Pattern Recognition*, 120: 108149.

Huiskes, M. J.; and Lew, M. S. 2008. The mir flickr retrieval evaluation. In *Proceedings of the 1st ACM international conference on Multimedia information retrieval*, 39–43.

Jian, L.; Li, J.; Shu, K.; and Liu, H. 2016. Multi-label informed feature selection. In *IJCAI*, volume 16, 1627–33.

Jiang, B.; Wu, X.; Zhou, X.; Cohn, A. G.; Liu, Y.; Sheng, W.; and Chen, H. 2024. Semi-Supervised Multi-View Feature Selection with Adaptive Graph Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 35(3): 3615–3629.

Klonecki, T.; Teisseyre, P.; and Lee, J. 2023. Cost-constrained feature selection in multilabel classification using an information-theoretic approach. *Pattern Recognition*, 141: 109605.

Li, J.; Li, P.; Hu, X.; and Yu, K. 2022. Learning common and label-specific features for multi-Label classification with correlation information. *Pattern Recognition*, 121: 108259.

Li, X.; and Chen, S. 2021. 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.

Li, Y.; Hu, L.; and Gao, W. 2023. Multi-label feature selection via robust flexible sparse regularization. *Pattern Recognition*, 134: 109074.

Lin, Y.; He, Z.; Guo, L.; and Ding, W. 2023. Multi-Label Feature Selection via Positive or Negative Correlation. *IEEE Transactions on Emerging Topics in Computational Intelligence*.

Liu, B.; Li, W.; Xiao, Y.; Chen, X.; Liu, L.; Liu, C.; Wang, K.; and Sun, P. 2023a. Multi-view multi-label learning with high-order label correlation. *Information Sciences*, 624: 165–184.

Liu, C.; Wen, J.; Luo, X.; Huang, C.; Wu, Z.; and Xu, Y. 2023b. DICNet: Deep Instance-Level Contrastive Network for Double Incomplete Multi-View Multi-Label Classification. *arXiv preprint arXiv:2303.08358*.

Liu, M.; Luo, Y.; Tao, D.; Xu, C.; and Wen, Y. 2015. Low-rank multi-view learning in matrix completion for multi-label image classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 29.

Liu, S.; Song, X.; Ma, Z.; Ganaa, E. D.; and Shen, X. 2022. MoRE: multi-output residual embedding for multi-label classification. *Pattern Recognition*, 126: 108584.

Liu, W.; Yuan, J.; Lyu, G.; and Feng, S. 2023c. Label driven latent subspace learning for multi-view multi-label classification. *Applied Intelligence*, 53(4): 3850–3863.

Liu, X.; Sun, L.; and Feng, S. 2022. Incomplete multi-view partial multi-label learning. *Applied Intelligence*, 52(3): 3289–3302.

Lyu, G.; Deng, X.; Wu, Y.; and Feng, S. 2022. Beyond shared subspace: A view-specific fusion for multi-view

multi-label learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 7647–7654.

Lyu, G.; Kang, W.; Wang, H.; Li, Z.; Yang, Z.; and Feng, S. 2024a. Common-Individual Semantic Fusion for Multi-View Multi-Label Learning. In *IJCAI*.

Lyu, G.; Yang, Z.; Deng, X.; and Feng, S. 2024b. L-VSM: Label-Driven View-Specific Fusion for Multiview Multilabel Classification. *IEEE Transactions on Neural Networks and Learning Systems*.

Sanghavi, R.; and Verma, Y. 2022. Multi-view multi-label canonical correlation analysis for cross-modal matching and retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4701–4710.

Sun, S. 2013. A survey of multi-view machine learning. *Neural computing and applications*, 23: 2031–2038.

Tan, Q.; Yu, G.; Wang, J.; Domeniconi, C.; and Zhang, X. 2019. Individuality-and commonality-based multiview multilabel learning. *IEEE transactions on cybernetics*, 51(3): 1716–1727.

Wang, Z.; Wang, N.; Zhang, H.; Jia, L.; Qin, Y.; Zuo, Y.; Zhang, Y.; and Dong, H. 2023. Segmentalized mRMR features and cost-sensitive ELM with fixed inputs for fault diagnosis of high-speed railway turnouts. *IEEE Transactions on Intelligent Transportation Systems*.

Wu, X.; Chen, Q.-G.; Hu, Y.; Wang, D.; Chang, X.; Wang, X.; and Zhang, M.-L. 2019. Multi-View Multi-Label Learning with View-Specific Information Extraction. In *IJCAI*, 3884–3890.

Xu, C.; Tao, D.; and Xu, C. 2013. A survey on multi-view learning. *arXiv preprint arXiv:1304.5634*.

Xu, N.; Liu, Y.-P.; and Geng, X. 2019. Label enhancement for label distribution learning. *IEEE Transactions on Knowledge and Data Engineering*, 33(4): 1632–1643.

Yagi, Y. 2011. Color standardization and optimization in whole slide imaging. In *Diagnostic pathology*, volume 6, 1–12. Springer.

Yelipe, U.; Porika, S.; and Golla, M. 2018. An efficient approach for imputation and classification of medical data values using class-based clustering of medical records. *Computers & Electrical Engineering*, 66: 487–504.

Yin, J.; and Zhang, W. 2023. Multi-view multi-label learning with double orders manifold preserving. *Applied Intelligence*, 53(12): 14703–14716.

Yuan, M.; and Lin, Y. 2006. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(1): 49–67.

Zhang, C.; Fang, Y.; Liang, X.; Zhang, H.; Zhou, P.; Wu, J., Xingyu andYang; Jiang, B.; and Sheng, W. 2024. Efficient Multi-view Unsupervised Feature Selection with Adaptive Structure Learning and Inference. In *Proceedings of the 33rd International Joint Conference on Artificial Intelligence*, 5443–5452.

Zhang, C.; Yu, Z.; Hu, Q.; Zhu, P.; Liu, X.; and Wang, X. 2018. Latent semantic aware multi-view multi-label classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Zhang, J.; Lin, Y.; Jiang, M.; Li, S.; Tang, Y.; and Tan, K. C. 2020a. Multi-label Feature Selection via Global Relevance and Redundancy Optimization. In *IJCAI*, 2512–2518.

Zhang, M.-L.; and Zhou, Z.-H. 2013. A review on multi-label learning algorithms. *IEEE transactions on knowledge and data engineering*, 26(8): 1819–1837.

Zhang, X.; Xu, M.; and Zhou, X. 2024. RealNet: A feature selection network with realistic synthetic anomaly for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16699–16708.

Zhang, Y.; Wu, J.; Cai, Z.; and Philip, S. Y. 2020b. Multi-view multi-label learning with sparse feature selection for image annotation. *IEEE Transactions on Multimedia*, 22(11): 2844–2857.

Zhao, D.; Gao, Q.; Lu, Y.; and Sun, D. 2022. Learning view-specific labels and label-feature dependence maximization for multi-view multi-label classification. *Applied Soft Computing*, 124: 109071.

Zhong, Q.; Lyu, G.; and Yang, Z. 2024. Align While Fusion: A Generalized Nonaligned Multiview Multilabel Classification Method. *IEEE Transactions on Neural Networks and Learning Systems*.

Zhu, C.; Miao, D.; Wang, Z.; Zhou, R.; Wei, L.; and Zhang, X. 2020. Global and local multi-view multi-label learning. *Neurocomputing*, 371: 67–77.

Zhu, P.; Hu, Q.; Hu, Q.; Zhang, C.; and Feng, Z. 2018. Multi-view label embedding. *Pattern recognition*, 84: 126–135.

Zhu, X. 2005. *Semi-supervised learning with graphs*. Carnegie Mellon University.

Zhu, X.; Li, X.; and Zhang, S. 2015. Block-row sparse multiview multilabel learning for image classification. *IEEE transactions on cybernetics*, 46(2): 450–461.