View-Specific Dropout and Weighted View Dual-Channel Decoupling Framework for Unbalanced Multi-View Weak Multi-Label Classification

Minghao Zhou, Yiying Wang, Siying Pan, Yuanyuan Li and Qing Ai*

***Abstract*—Incomplete Multi-View Weak Multi-Label Learning (IMVWML) aims to address Multi-View Multi-Label Learning (MVML) problem with incomplete data. Existing IMVWML algorithms overlook the problem of unbalanced sample quantities across different views. However, in real world, data from varying environments or operational conditions often exhibits significant imbalance, which makes the model rely on views with more samples seriously, while neglecting the fewer. The imbalance leads to a degradation in the performance of the multi-label classification task. To address the above-mentioned issue, we propose the VSDW-DDF model, which is based on a dual-channel learning framework and consists of three components: the View-Specific Weight (VSVW) module, the Sample-Guided Graph Regularization (SGR) module, and the View-Specific Dropout Mechanism (VSDM) module. Firstly, in the VSVW module, we propose the view-specific weights dynamically assign appropriate weights to each view based on their actual contributions, ensuring that important views receive sufficient attention. Secondly, in the SGR module, we design a sample-guided graph regularization loss using sample supervision information, which can effectively preserve the geometric structure among samples, even if some views are severe incomplete. Finally, in the VSDM module, we propose a view-specific dropout mechanism, which is more targeted to individual view compared to traditional dropout. It adaptively adjusts the dropout rate based on the information of each view, preventing overfitting and improving robustness. Experimental results demonstrate that VSDW-DDF significantly outperforms existing methods in handling Unbalanced Multi-View Weak Multi-Label Learning (UMVWML) problems, proving its effectiveness and advancements.**

***Index Terms*—multi-view learning, multi-label learning, incomplete views, weak labels, deep learning**

Manuscript received Mar 15, 2025; revised May 11, 2025.

This work was funded in part by the Basic Research Project of Education Department of Liaoning Province in China (JYTMS20230929) and the Postgraduate Scientific and Technological Innovation Project of University of Science and Technology Liaoning in China (LKDYC202422).

Minghao Zhou and Yiying Wang contributed equally to this work.

Minghao Zhou is a graduate student at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, Liaoning China (ustlzhouminghao@163.com)

Yiying Wang is a graduate student at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, Liaoning China (yiyingwang2024@126.com)

Siying Pan is a graduate student at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, Liaoning China (ustlpansiying@163.com)

Yuanyuan Li is a graduate student at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, Liaoning China (ustlliyuanyuan@163.com)

Qing Ai* is a Professor at University of Science and Technology Liaoning, Anshan 114051, Liaoning China (Corresponding author to provide phone: +86-155-0492-8887; e-mail: lyaiqing@126.com).

I. Introduction

MULTI-LABEL learning has become a research hotspot in recent decades, aiming to solve classification problems where a sample contains multiple labels simultaneously [1]. It has a wide range of applications in text, video, music classification, semantic scene classification, and other fields [2]. The traditional multi-label classification problem only uses single-view data, but with the increase of data sources and feature extraction methods, analyzing data from a single-view dimension can no longer meet the increasing complexity and comprehensiveness requirements in classification tasks [3]. Utilizing multi-view data from different sources and feature extraction methods can provide a more comprehensive and accurate description of the observed object [4]. MVML can utilize information from multiple views to better capture the diversity and complementarity of the data, thereby improving classification performance. Zhao et al. proposed the CDMM [5] model, which extracts both consistent and diverse information from multi-view data, learns latent representations via the Hilbert-Schmidt Independence Criterion (HSIC) [6] to bridge feature space and label semantics, and achieves consistent classification through late-stage fusion.

However, the typical MVML approaches often overlook the challenges posed by incomplete views and weak labels, limiting their effectiveness when applied to such data. In recent years, researchers have proposed several IMVWML models to mitigate the impact of incomplete views and weak labels. For instance, Liu et al. introduced the DMVMLC [7] model, which uses matrix factorization to synchronize multiple views in the label space, while embedding both global high-rank and local low-rank constraints within the predicted multi-label matrix. This method effectively addresses the issue of incomplete views and weak labels, but its reliance on matrix factorization may not fully capture complex non-linear relationships. Therefore, Wen et al. proposed an MTD [8] model that explicitly decouples shared features and view-specific features, and introduced cross-channel contrastive loss to improve the feature decoupling capability.

However, in numerous real-world scenarios, collected data often varies due to environmental or operational shifts, leading to sample size imbalances [9]. For example, when handling patient records, the number of patient records may be unbalanced due to personal factors or differences in laboratory procedures. The imbalance can cause the model to rely on views with more samples and ignore views with fewer samples, resulting in classification results biased towards views with more samples and reducing the generalization ability of the model. Traditional methods do not effectively address this issue.

Therefore, we propose the VSDW-DDF model. VSDW-DDF consists of three modules based on a dual-channel framework: VSVW, VSDM, and SGR. Inspired by Zhang et al. [10], our view-specific representation learning framework constructs advanced feature extraction and reconstruction networks. To address these challenges, we propose a novel cross-channel contrastive loss function that aims to reduce the distance between positive pairs while increasing the distance between negative pairs. In addition, considering the issue of unbalanced views, in the VSVW module, we assign corresponding weights based on the actual contributions of each view and incorporate view-specific weights into the loss function calculation process, which ensures that important but information-deficient views receive sufficient attention, thereby mitigating performance degradation caused by unbalanced views. Meanwhile, we propose a view-specific dropout mechanism in the VSDM module, which dynamically adjusts the feature dropout probability during the training process based on the information content of each view. VSDM not only effectively prevents overfitting, but also promotes the learning ability of available information from other views, improving the overall performance of the model. In summary, the main contributions of this paper are as follows:

1.We propose a novel VSDW-DDF (as illustrated in Fig. 1) model, based on a dual-channel decoupling framework that takes into account the problem of sample size imbalance in different views. In VSDW-DDF, we design three parts: VSVW, SGR, and VSDM to solve the unbalanced multi-view weak multi-label classification (UMVWMLC) task.

2.We conducted a large number of experiments on five benchmark UMVWML datasets, and the experimental results show that VSDW-DDF outperforms the most advanced methods.

II. Related Works

A. Multi-View Multi-Label Learning

MVML is a comprehensive task that enhances classification performance by leveraging richer descriptive capabilities. Nevertheless, developing an efficient model to integrate multiple views and labels remains a significant challenge. The learning procedure inherently leads to diverse correlation challenges, including maintaining consistency and complementarity across multiple views, as well as tackling the alignment discrepancies between views and labels [11]. Wang et al. [12] applied non-negative matrix factorization to explore complementary information from different views and converted the label of each view into a coherent consensus label for MVML. Although Wang et al. examined the complementary relationships present in multiple views, they overlooked the assumption of multi-view consistency. Unlike the method proposed by Wang et al., the M2LD [13] emphasizes the extraction of consensus information from diverse views. By leveraging matrix factorization, this approach constructs a cross-view fused feature space, modeling the consistency relationships between views and labels to enhance the learning of latent subspaces with label information. Zhao et al. [14] developed a partially shared dictionary learning framework that investigates consistency and complementarity across views through flexible shared sparse coefficients. Moreover, Zhao et al. introduced the LVSL [15] model to address asymmetry in MVML learning. This model primarily extracts structural information from unaligned views, which improves the learning of view-specific labels.

B. Incomplete Multi-View Weak Multi-Label Learning

In practical scenarios, multi-view data and label information are frequently incomplete. For instance, in social media data, users may share textual descriptions of an event but fail to upload associated images or videos [16]. Similarly, in multi-label data, annotators might skip or incorrectly label certain items due to insufficient prior knowledge or other influencing factors. To tackle the IMVWML problem, Tan et al. [17] developed a robust classifier of IMVWML. This approach employs two matrices derived from incomplete prior knowledge to tackle calculation problems associated with IMVWML. Additionally, it breaks down multi-view features into a hidden joint representation and several view-specific matrices. By leveraging mapping matrices, the approach establishes a connection between the joint representation and label characteristics, which improves the supervision required for learning semantics in scenarios with weak multi-label data. However, as IMVWML operates as a semi-supervised model, it faces difficulties in processing new samples not seen during training. In contrast to IMVWML, the traditional matrix factorization approach NAIML [18], introduced by Li et al., is capable of effectively handling completely new samples. NAIML adopts a composite indicator matrix to address issues of incomplete views and weak labels, combining global high-rank and local low-rank constraints. Nonetheless, such methods depend on traditional feature extraction methodologies. As deep learning continues to evolve rapidly, approaches grounded in deep neural networks have attracted significant interest in numerous representation learning areas. For instance, the method proposed in [19] designs specific encoders and decoders for each view which embed incomplete view and weak label information into the network to mitigate the adverse effects of the dual-incomplete problem. Nevertheless, this approach only carries out a weighted summation of multiple views in the fusion process and does not successfully extract complementary information in high-dimensional feature spaces. Duan et al. proposed the VCMN [20] model, which designed a View Channel Mixer module to model complementary relationships in high-dimensional feature spaces. Additionally, it improves the capacity of network to learn consistent representations across multiple views by employing cross-view and cross-sample contrastive loss, thereby achieving joint learning of complementarity and consistency. Liu et al. created two modules based on the Transformer architecture for cross-view feature extraction and multi-label classification, named LMVCAT [21]. LMVCAT leverages self-attention mechanism to capture complementary information across

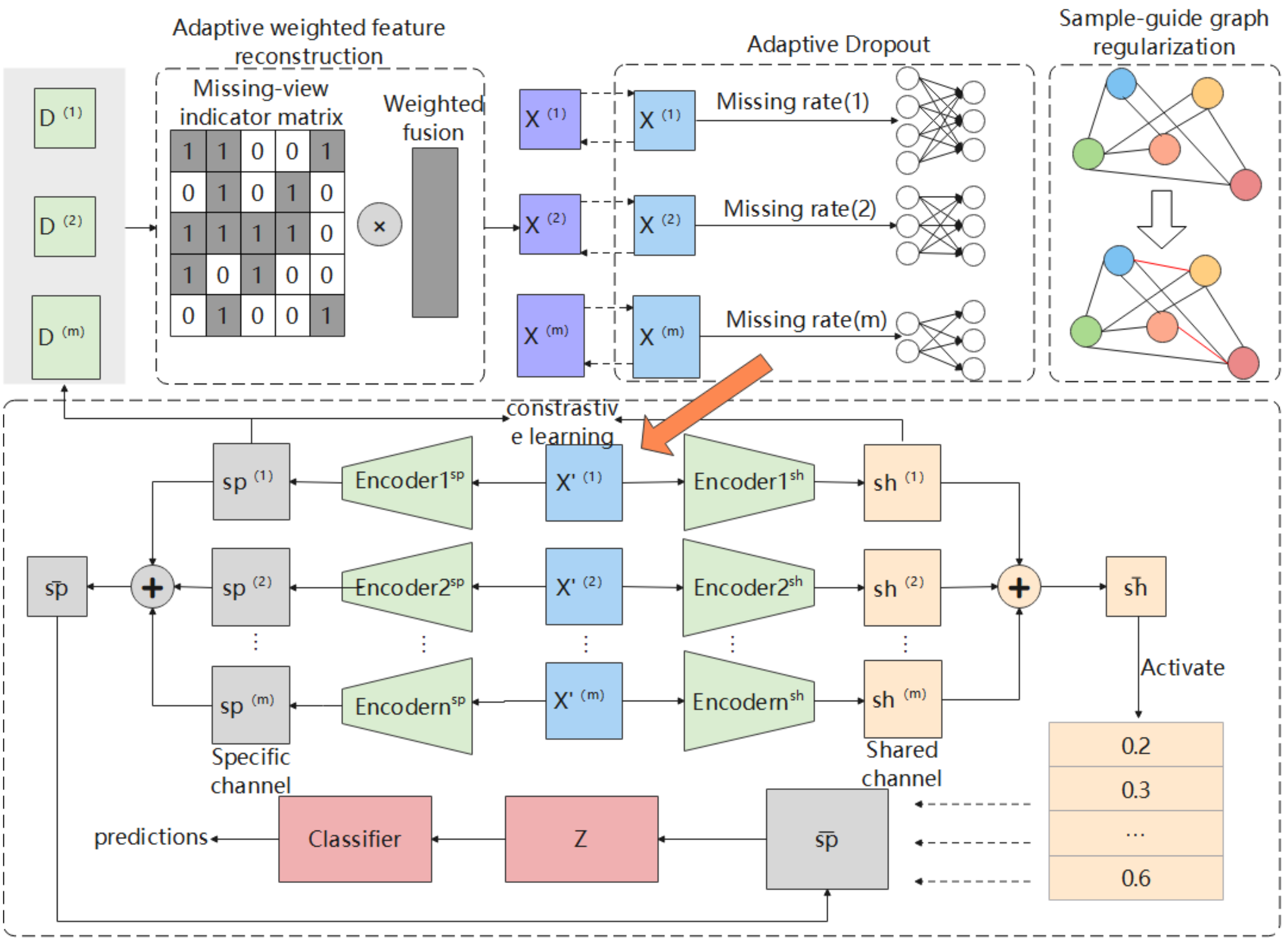


Fig. 1. The Main Framework of VSDW-DDF

different views and applies label-stream shape constraints using incomplete multi-label data. On this basis, Wen et al. proposed the AIM [22] model, which uses a cross-view joint attention mechanism to approximate incomplete instances based on available information and attention scores between instance pairs. They also proposed a multi-view multi-label classification (MVMLC) framework based on label semantic feature learning, using statistical weak label correlation matrices and graph attention networks (GATs) to guide the learning process of label-specific features.

C. Unbalanced Multi-View Clustering Learning

By integrating information from multiple views, multi-view data can more comprehensively and finely depict the characteristics of the sample. However, collected data frequently contains incomplete information. Many traditional methods either ignore incomplete data or use basic imputation techniques, such as replacing incomplete values with the average of available data. For instance, in [23], an index matrix was introduced to remove the representation matrix of unpaired data, ensuring a consistent latent representation matrix. Research [24] introduced a weight matrix for each view and assigned smaller weights to incomplete samples to mitigate their impact. However, these methods find it difficult to accurately recover the information of incomplete data, which may lead to significant bias, especially when handling multi-view data with high rates of incompleteness. In order to extract valuable information from incomplete samples, reference [25] uses non-negative matrix factorization techniques to reconstruct incomplete views and leverages these reconstructions to obtain potential representations. Reference [26] utilizes a shared graph across all views and individual incomplete graphs to reconstruct the complete graph for each view. Another widely used data imputation technique involves inference. In [27], researchers developed a reconstruction term to infer incomplete data in each view and introduced an inverse graph regularization term to maintain local structural consistency between multiple views.

Due to the randomness of data loss, incomplete multi-view data is typically categorized as balanced and unbalanced types (as illustrated in Fig. 2). In the balanced state, the incomplete rate of each view is the same; in the unbalanced state, the incomplete rates of each view are different (the blank areas in the graphics represent incomplete data). In fact, imbalances are more common. Unbalanced and incomplete multi-view data display uneven sample distribution between different views, leading to the Unbalanced Multi-View Clustering Learning (UMVL). Reference [28] first proposed a method to solve UMVL, inspired by the principles of biological evolution. In addition, reference [29] proposed a tensor-based approach to address this issue by applying low rank tensor constraints to similar graph matrices to capture potential relationships between different view data.

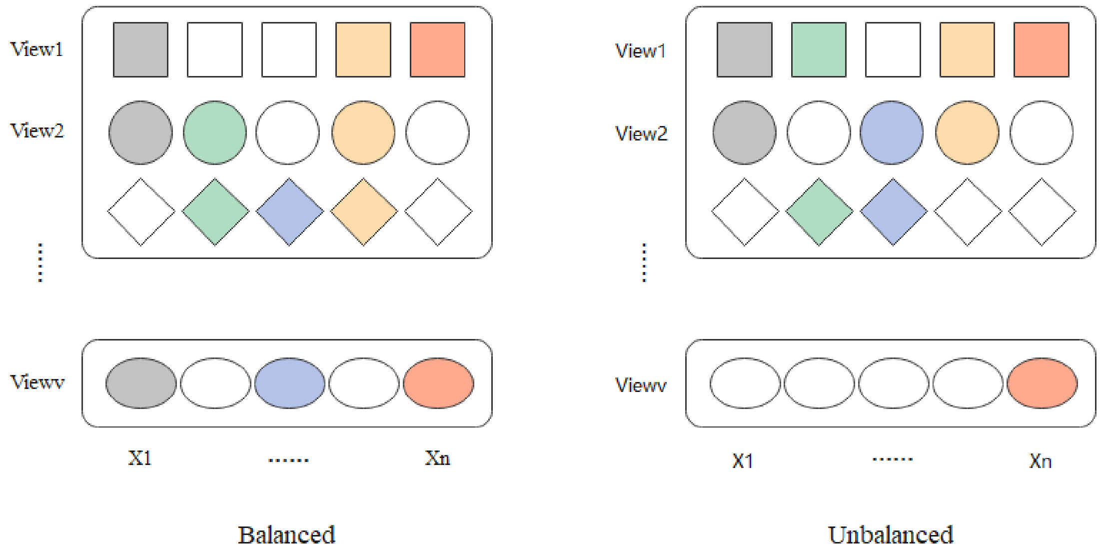


Fig. 2. Two Typical Incomplete Multi-View Data Examples

III. Method

In this section, we provide a detailed introduction to our model, covering four key components: a dual-channel decoupling framework, sample-guided graph regularization, an adaptive dropout mechanism, and multi-label classification. For clarity, we first provide a brief overview of the formal definition of the problem and commonly used notations.

A. Problem Definition

Given a UMVWML dataset $\{X^{(v)}\}_{v=1}^{m}$, $Y_{v=1}^{(m)}$, where $X^{(v)}$ stands for the feature matrix of the v-th view encompassing n samples, and m represents the quantity of views. The matrix $X^{(v)} \in \mathbb{R}^{n\times d_v}$ has its i-th row depicting the instance of the i-th sample in the v-th view, with a dimension of d_v. The matrix $Y \in \{0,1\}^{n\times c}$ is the label matrix with c classes, where $Y_{i,j}$ indicates that the i-th sample is labeled as the j-th class. For unbalanced views, an unbalanced view index matrix $W \in \{0,1\}^{n\times m}$ is introduced, where $W_{i,j} = 1$ indicates that the j-th view of the i-th sample is accessible, otherwise $W_{i,j} = 0$. For weak labels, $G \in \{0,1\}^{n\times c}$ is set as the weak-label index matrix, where $G_{i,j} = 1$ implies that the j-th label of the i-th sample is known, otherwise $G_{i,j} = 0$. Finally, in the data preprocessing stage, the unavailable views and unknown labels are filled with '0' to obtain the final unbalanced multi-view feature matrices $\{X^{(v)}\}_{v=1}^{m}$ and the weak label matrix Y. VSDW-DDF aims to train a classifier capable of conducting multi-label classification inference on unbalanced multi-view data. Subscripts $B_{i,j}$, $B_{i,:}$, and $B_{:,j}$ denote the element, row, and column of any matrix B.

B. VSVW Module Based on Dual-Channel Decoupling Framework

Due to the diversity of multi-view data acquisition methods, the original information in each view may have different feature dimensions, which is not conducive to the parallel execution of deep networks. To overcome this challenge, we project heterogeneous raw data into a unified embedding space with dimension d_e using encoders. Unlike conventional deep multi-view networks, we adopt two groups of multi-layer perceptrons as dual pathways: the shared pathway and the view-specific pathway. These pathways aim to extract shared and view-specific information from the raw data, represented as $\{E_v^{sh} : X^{(v)} \to Sh^{(v)}\}_{v=1}^{m}$ and $\{E_v^{sp} : X^{(v)} \to SP^{(v)}\}_{v=1}^{m}$. Here, E_v^{sh} and E_v^{sp} denote the shared feature encoder and the view-specific feature encoder for the v-th view. $X^{(v)}$ refers to the input after dropout processing. $Sh^{(v)} \in \mathbb{R}^{n\times d_e}$ and $SP^{(v)} \in \mathbb{R}^{n\times d_e}$ are the extracted multi-view shared feature matrix and the view-specific feature matrix. Despite having the same structure, the two pathways fulfill different roles. Specifically, E_v^{sh} aims at exploring cross-view common features, maintaining the fundamental attributes shared by all views, while E_v^{sp} focuses on extracting distinctive features specific to each view. This design divides the discriminative information of each view into two elements, meeting both the consistency and complementarity assumptions across multiple views. Nonetheless, accomplishing these goals with only two encoding channels is difficult because of insufficient guidance. As a result, we introduce a multi-view cross-channel contrastive loss L_{ccc} to facilitate the distinction between these two types of features:

$$L_{ccc} = \sum_{i=1}^{n} \frac{1}{3N^2-N} \frac{2\sum_{u=1}^{m}\sum_{v=1}^{m}[\Upsilon]S(sh_i^{(u)}, sp_i^{(v)})^2}{\sum_{u=1}^{m}\sum_{v=u}^{m}[\Upsilon]\left((S(sh_i^{(u)}, sh_i^{(v)})+1)/2\right)} \tag{1}$$
$$+ \sum_{i=1}^{n} \frac{1}{N^2-N} \frac{\sum_{u=1}^{m}\sum_{v\neq u}^{m}[\Upsilon]S(sp_i^{(u)}, sp_i^{(v)})^2}{\sum_{u=1}^{m}\sum_{v=u}^{m}[\Upsilon]\left((S(sh_i^{(u)}, sh_i^{(v)})+1)/2\right)},$$

where S is the similarity measurement function: $S(x,y) = \frac{x^T y}{\|x\|_2\cdot\|y\|_2}$. and $[\Upsilon]$ is the conditional function: if the condition $\{\Upsilon : W_{i,u}W_{i,v} = 1\}$ holds, then $[\Upsilon] = 1$, otherwise 0. $N = \sum_{u,v} W_{i,u}W_{i,v}$ represents the number of valid instance pairs. In brief, we focus solely on instance pairs where neither sample is unbalanced. $sh_i^{(v)}$ and $sp_i^{(v)}$ represent the features of the i-th sample in the shared feature matrix $Sh^{(v)}$ and the view-specific feature matrix $SP^{(v)}$. Based on Equation (1), our cross-channel contrastive loss includes two parts: the numerator reflects the average similarity of negative sample pairs, while the denominator reflects the average similarity of positive sample pairs. More precisely, we generate pairs from the $2N$ channels, with positive sample pairs being composed of shared features across different views, while the remaining pairs are classified as negative samples. Our aim is to increase the similarity among positive sample pairs in the shared feature space while reducing the similarity among negative sample pairs. This approach fulfills the design objective of the dual-channel model by reinforcing the consistency of shared features across multiple views while maintaining clear distinctions between view-specific features and other shared or view-specific features. To enhance the feature extraction ability of the encoders further, we integrate stacked decoders to reconstruct the embedded features extracted by the dual-channel encoders back into the original feature space, i.e., $\{D_v : S^{(v)} \in \mathbb{R}^{n\times d_e} \rightarrow \overline{X}^{(v)} \in \mathbb{R}^{n\times d_o}\}_{v=1}^2$, where D_v indicates the decoder associated with the v-th view, and $\overline{X}^{(v)}$ signifies the reconstructed feature. Finally, we use a weighted mean squared error loss L_{re} to measure the reconstruction quality and incorporate a weight adjustment mechanism based on unbalanced views during the reconstruction process:

$$L_{re} = \frac{1}{n}\sum_{v=1}^{m}\sum_{i=1}^{n}\frac{1}{d_v}\left\|\overline{X}_{i,:}^{(v)} - X_{i,:}^{(v)}\right\|_2^2 W\mathrm{W}_{i,v}, \tag{2}$$

where W represents the view weight determined according to unbalanced views, and W serves to conceal instances that are unavailable. By utilizing $Sh^{(v)}$ and $SP^{(v)}$ obtained from each view, cross-view fusion can be seamlessly conducted to derive the distinct shared representation and view-specific representation for all samples:

$$\overline{Sh}_{i:} = \sum_{v=1}^{m}\frac{Sh_{i:}^{(v)}\mathrm{W}_{i:v}}{\sum_v \mathrm{W}_{i:v}}, \quad \overline{SP}_{i:} = \sum_{v=1}^{m}\frac{SP_{i:}^{(v)}\mathrm{W}_{i:v}}{\sum_v \mathrm{W}_{i:v}}. \tag{3}$$

Next, we aim to integrate shared information with view-specific information to learn a coherent representation of the samples. In this process, rather than using typical addition or concatenation operations, we employ a new feature interaction approach. Specifically, reinforcing the shared information using the view-specific information:

$$Z_{i,j} = \theta(\overline{SP}_{i,j}) \cdot \overline{Sh}_{i,j}, \tag{4}$$

where θ represents the sigmoid activation function, and $Z \in \mathbb{R}^{n\times d_e}$ denotes the final fused representation.

C. SGR Module

Numerous unsupervised multi-view learning methods employ graph regularization to maintain the inherent structure of the data. This approach is grounded in the premise that samples which are similar in the original feature space should also retain their similarity in the latent space, a technique that has demonstrated its efficacy in unsupervised scenarios. When we extend this assumption to multi-label classification tasks with supervised task characteristics, we propose a new hypothesis: the similarity of samples in the original feature space should still be preserved in the embedding space. Based on this hypothesis, we design a sample-guided graph regularization method that guides the feature extraction process by constructing a similarity graph between samples and particularly considers the impact of unbalanced views on model training. Firstly, we compute the similarity matrix based on the sample matrix X:

$$T_{i,j} = \frac{C_{i,j}\|X_i^{(v)} - X_j^{(v)}\|_2}{C_{i,j}\|X_i^{(v)} - X_j^{(v)}\|_2 + \eta}, \tag{5}$$

where $T \in [0,1]^{n\times n}$ denotes the sample similarity graph, which captures the pairwise similarity between samples. $C_{i,j} = G_i \cdot G_j^T$ specifies the views available for samples i and j. The terms $X_i^{(v)}$ and $X_j^{(v)}$ correspond to the feature vectors of the i-th and j-th samples in the v-th view. η is a constant, empirically set to 1000 for simplicity. Given that $C_{i,j}$ is typically much larger than the similarity derived from feature distance $\|X_i^{(v)} - X_j^{(v)}\|_2$ in most datasets. If two samples are jointly present in a few views, they are deemed more similar compared to other pairs of samples, even when the number of views is large. This property ensures that the model can still uncover potential relationships between samples, even when data is unbalanced in some views.

By employing this similarity graph, we can adjust the distance between any two samples in the embedding feature space using the loss function L_{gc}, thereby achieving the goal of retaining structural information.

$$L_{gc} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|Z_{i,:} - Z_{j,:}\|_2^2 T_{i,j}. \tag{6}$$

To boost the computational efficiency of the model on GPUs, we transform formula (6) into a matrix multiplication format:

$$L_{gc} = \frac{1}{n^2}Tr(Z^T L Z), \tag{7}$$

where $Tr(\cdot)$ represents the trace operation. The matrix L indicates the Laplacian matrix, which is determined as $L = D - T$, where D is a diagonal matrix with its diagonal components given by $D_{i,i} = \sum_{j=1}^n T_{i,j}$.

D. VSDM Module

In multi-view learning, the original data contains a large amount of redundant information, and there may be significant differences in the quantity of data across different views, leading to unbalanced view issues that inevitably interfere with the feature extraction process of the model. To address this challenge, we propose a view-specific dropout mechanism based on feature-level dropout. Unlike traditional neuron-level dropout, feature-level dropout adaptively discards input features during training, which can effectively prevent overfitting and promote the model to learn more robust feature representations. For this purpose, for any view v, we dynamically adjust the dropout probability p_v based on its incomplete rate and global parameter σ. This method ensures that the dropout probability for each view reflects its own data quality and incomplete situation, thereby better adapting to the differences between views. Let $X^{(v)}$ be the original input feature matrix for the v-th view, and $\mathrm{Drop}^{(v)}$ be the corresponding binary dropout matrix, where each element is independently set to 0 with probability p_v, and kept as 1 with probability $1 - p_v$. The feature matrix after dropout processing is denoted as:

$$\{X'^{(v)}\}_{v=1}^{m} = \{X^{(v)} \odot \text{Drop}^{(v)}\}_{v=1}^{m}, \tag{8}$$

where $\odot$ denotes element-wise multiplication. By combining the incomplete rate of each view and the global parameter σ, view-specific dropout dynamically adjusts the dropout probability for each view, ensuring that views with less data do not overly influence the learning process of the model. This mechanism effectively balances the contributions of different views to the model, enhances the generalization ability of the model, prevents overfitting on high-incomplete-rate views, and enables the model to extract more consistent and useful feature representations from all views. Experimental results show that using the view-specific dropout mechanism significantly improves model performance, especially when handling multi-view data. This method not only effectively alleviates unbalanced view issues but also enhances the stability and generalization ability of the model.

E. Weighted Multi-Label Classification and Overall Loss Function

$$P = \theta(Z\boldsymbol{\omega} + \boldsymbol{\varepsilon}). \tag{9}$$

where $\boldsymbol{\omega} \in \mathbb{R}^{2d\times c}$ and $\boldsymbol{\varepsilon} \in \mathbb{R}^{n\times c}$ represent the learning parameters of the classifier, while $P \in \mathbb{R}^{n\times c}$ denotes our prediction score matrix.

Algorithm 1: VSDW-DDF Training Process

Input: Unbalanced multi-view data$\{X^{(v)}\}_{v=1}^{m}$, unbalanced view index matrix W, weak labels Y, weak-label index matrix G.

Output: Prediction results P.

1) **Initialization:** Initialize model parameters and set hyperparameters (α, β, γ), learning rate and number of training epochs e.
2) Set t=0.
3) While $t < e$:
4) Construct the binary dropout matrix $\text{Drop}^{(v)}$.
5) **repeat:**
   1. Compute the input data after dropout $\{X'^{(v)}\}_{v=1}^{m}$ according to formula (8);
   2. Use dual-channel encoders $\{Sh^{(v)}\}_{v=1}^{m}$ and $\{SP^{(v)}\}_{v=1}^{m}$ to extract shared embedding features $\{E_v^{sh}\}_{v=1}^{m}$ and view-specific embedding features $\{E_v^{sp}\}_{v=1}^{m}$;
   3. Compute the fused shared embedding features $\overline{Sh}$ and fused view-specific features $\overline{SP}$ according to formula (3);
   4. Compute the final fused representation Z according to formula (4);
   5. Compute the similarity graph T and its corresponding Laplacian matrix L according to formula (5);
   6. Compute the cross-channel contrastive loss L_{ccc} according to formula (1), the reconstruction loss L_{re} according to formula (2), and the graph embedding loss L_{gc} according to formula (7);
   7. Obtain the prediction P according to formula (9) and compute the classification loss L_{mc} according to formula (10);
   8. Compute the total loss L_{all} according to formula (11);
   9. Update network parameters;
   10. $t = t + 1$;
6) **End loop**.

In single-label classification tasks, the cross-entropy loss is generally used to guide model training. By contrast, for multi-label classification tasks, the prediction of each category is considered as an independent binary classification issue. Moreover, to handle unknown labels in the label matrix, we employ the following weighted multi-label cross-entropy loss as the primary classification loss:

$$\begin{aligned} L_{mc} = &-\frac{1}{nc}\sum_{i=1}^{n}\sum_{j=1}^{c}(Y_{i,j}\log(P_{i,j}) \\ &+ (1-Y_{i,j})\log(1-P_{i,j}))G_{i,j}, \end{aligned} \tag{10}$$

where G is utilized to conceal the unknown labels.

By combining the cross-channel contrastive loss (equation 1), the reconstruction loss (equation 2), the graph embedding loss (equation 7), and the weighted multi-label classification loss (equation 10), our total loss function can be represented as:

$$L_{all} = L_{mc} + \alpha L_{gc} + \beta L_{ccc} + \gamma L_{re}, \tag{11}$$

where α, β, and γ indicate the corresponding penalty parameters. The specific training process is described in Algorithm 1.

IV. EXPERIMENTS

In this section, we conduct extensive experiments to evalucate the performance of the proposed model, VSDW-DDF, across five public datasets.

A. Experimental Settings

1) Dataset: We utilized five classic MVML datasets to validate our model: Corel5k, ESPGame, IAPRTC12, Pascal07, and MIRFLICKR. From these datasets, six kinds of features were obtained to reflect their six distinct views, such as GIST, HSV, Hue, SIFT, RGB, and LAB features. The sample count in these datasets ranges from 4,999 to 25,000, and the number of classes spans from 20 to 291, as presented in Table I.

2) Evaluation Indicators: Following the practices of numerous existing works, we selected six evaluation metrics to assess all comparative methods, specifically Ranking Loss (RL), Average Precision (AP), Hamming Loss (HL), Area Under the Curve (AUC), One-Error (OE), and Coverage (COV). To facilitate a more intuitive comparison of performance, we report results concerning AP, 1-HL, 1-RL, AUC, 1-OE, and 1-COV; hence, for all evaluation metrics, higher values indicate better performance.

3) Implementation Details: In this experiment, the model was implemented using PyTorch version 1.10.1 and Python version 3.9.18. The hardware configuration included an RTX 3090 GPU and an i7-12900k CPU. The learning rate was set to 0.1, and the SGD optimizer was chosen for model training. For all five datasets, the batch size and momentum were set to 128 and 0.9.

TABLE I
STASTICS OF THE EXPERIMENT DATA SET.

Dataset	#View	#Label	#Sample	#Label/#Sample
Corel5k	6	260	4999	3.40
ESPGame	6	268	20770	4.69
Iaprtc12	6	291	19627	5.72
Pascal07	6	20	9963	1.47
MirFlickr	6	38	25000	4.72

TABLE II
AVERAGE PRECISION OF THE EXPERIMENTAL RESULTS

Data	CDMM	DMVMLC	VCMN	LMVCAT	AIM	MTD	VSDW-DDF
Corel5k	0.312 ± 0.004	0.351 ± 0.001	0.388 ± 0.009	0.369 ± 0.011	0.391 ± 0.008	0.398 ± 0.007	0.414 ± 0.008
ESPGame	0.255 ± 0.001	0.291 ± 0.003	0.290 ± 0.003	0.283 ± 0.003	0.259 ± 0.003	0.296 ± 0.003	0.304 ± 0.002
Iaprtc12	0.259 ± 0.002	0.298 ± 0.004	0.302 ± 0.004	0.295 ± 0.004	0.254 ± 0.004	0.310 ± 0.003	0.318 ± 0.003
Pascal07	0.469 ± 0.003	0.497 ± 0.009	0.515 ± 0.010	0.472 ± 0.011	0.453 ± 0.013	0.520 ± 0.005	0.522 ± 0.005
MirFlickr	0.479 ± 0.001	0.565 ± 0.003	0.585 ± 0.006	0.566 ± 0.007	0.509 ± 0.009	0.583 ± 0.004	0.588 ± 0.003

TABLE III
HAMMING LOSS OF THE EXPERIMENTAL RESULTS

Data	CDMM	DMVMLC	VCMN	LMVCAT	AIM	MTD	VSDW-DDF
Corel5k	0.985 ± 0.000	0.987 ± 0.000	0.988 ± 0.000	0.986 ± 0.000	0.988 ± 0.000	0.988 ± 0.000	0.988 ± 0.000
ESPGame	0.980 ± 0.000	0.983 ± 0.000	0.983 ± 0.000	0.982 ± 0.000	0.983 ± 0.000	0.982 ± 0.000	0.983 ± 0.000
Iaprtc12	0.978 ± 0.000	0.981 ± 0.000	0.981 ± 0.000	0.979 ± 0.000	0.980 ± 0.000	0.980 ± 0.000	0.981 ± 0.000
Pascal07	0.918 ± 0.000	0.927 ± 0.002	0.929 ± 0.002	0.922 ± 0.006	0.925 ± 0.003	0.930 ± 0.001	0.930 ± 0.001
MirFlickr	0.870 ± 0.001	0.885 ± 0.001	0.887 ± 0.002	0.877 ± 0.006	0.878 ± 0.001	0.887 ± 0.001	0.887 ± 0.001

TABLE IV
RANKING LOSS OF THE EXPERIMENTAL RESULTS

Data	CDMM	DMVMLC	VCMN	LMVCAT	AIM	MTD	VSDW-DDF
Corel5k	0.881 ± 0.000	0.875 ± 0.003	0.886 ± 0.003	0.866 ± 0.005	0.886 ± 0.004	0.884 ± 0.002	0.891 ± 0.002
ESPGame	0.817 ± 0.001	0.837 ± 0.001	0.825 ± 0.001	0.816 ± 0.002	0.811 ± 0.002	0.826 ± 0.002	0.834 ± 0.001
Iaprtc12	0.845 ± 0.000	0.870 ± 0.002	0.869 ± 0.001	0.856 ± 0.002	0.842 ± 0.002	0.861 ± 0.002	0.871 ± 0.001
Pascal07	0.773 ± 0.003	0.782 ± 0.007	0.807 ± 0.008	0.771 ± 0.009	0.746 ± 0.010	0.806 ± 0.005	0.807 ± 0.004
MirFlickr	0.827 ± 0.001	0.855 ± 0.002	0.865 ± 0.003	0.852 ± 0.003	0.835 ± 0.003	0.864 ± 0.002	0.866 ± 0.001

B. Unbalanced Multi-View Weak Multi-Label Dataset Setup

We manually developed the UMVWML datasets based on the previously mentioned five complete datasets. This was done with the objective of evaluating the performance of all methods in situations characterized by unbalanced views and weak labels. Specifically, we randomly selected 20%, 30%, 40%, 60%, 70%, and 80% of instances from each of the six views as unavailable instances, which were then replaced with ‘0’ values. We ensured that no invalid samples existed in the dataset, meaning every sample retained at least one available view.

For weak labels, half of the positive and negative labels for each class were set as unknown. A random selection of 70% of the samples with unbalanced views and weak labels was used as the training set. To reduce the randomness of the experiments, the construction process of the UMVWML datasets was repeated multiple times.

C. Comparison Methods

To assess the superiority of our model, we selected six comparison methods in our experiments: DMVMLC [7], VCMN [20], CDMM [5], LMVCAT [21], AIM [22], and MTD [8]. These methods were compared with our proposed VSDW-DDF on the five UMVWML datasets. Currently, there are very few MVMLC methods that can simultaneously address both unbalanced views and weak label issues; therefore, we incorporated several related multi-label classification methods into our experiments. More specifically, among these six methods, DMVMLC, VCMN, LMVCAT, AIM, and MTD are fully applicable to the UMVWMLC tasks. CDMM is an MVMLC method but cannot handle any incomplete data problems. Therefore, we simply used the mean of available instances to fill in the incomplete views and replaced unknown labels with ‘0’.

D. Experimental Results and Analysis

To evaluate the effectiveness of our approach in situations featuring unbalanced views and weak labels, we performed comparisons against six advanced algorithms using five different datasets. The mean and standard deviation of the results are presented in Tables II–VIII (the standard deviation is displayed in decimal form in the bottom-right corner of each cell.) In addition to the six performance metrics mentioned earlier, we also computed the average ranking of each algorithm based on these metrics.

In addition to the intuitive comparison metric "Ave.R", based on the results presented in Table IX and Fig. 3, we draw the following observations:

1. The VSDW-DDF, MTD, AIM, LMVCAT, VCMN, and DMVMLC models generally outperform the CDMM model

TABLE V
AUC OF THE EXPERIMENTAL RESULTS

Data	CDMM	DMVMLC	VCMN	LMVCAT	AIM	MTD	VSDW-DDF
Corel5k	0.882 ± 0.000	0.878 ± 0.002	0.889 ± 0.005	0.869 ± 0.004	0.889 ± 0.003	0.887 ± 0.003	0.894 ± 0.002
ESPGame	0.818 ± 0.001	0.842 ± 0.000	0.830 ± 0.002	0.821 ± 0.002	0.816 ± 0.001	0.831 ± 0.001	0.840 ± 0.001
Iaprtc12	0.847 ± 0.000	0.872 ± 0.001	0.872 ± 0.005	0.858 ± 0.002	0.845 ± 0.002	0.863 ± 0.001	0.873 ± 0.001
Pascal07	0.774 ± 0.001	0.809 ± 0.006	0.832 ± 0.006	0.799 ± 0.008	0.776 ± 0.009	0.829 ± 0.005	0.832 ± 0.003
MirFlickr	0.792 ± 0.002	0.841 ± 0.002	0.854 ± 0.002	0.839 ± 0.004	0.826 ± 0.002	0.851 ± 0.002	0.854 ± 0.001

TABLE VI
ONE ERROR OF THE EXPERIMENTAL RESULTS

Data	CDMM	DMVMLC	VCMN	LMVCAT	AIM	MTD	VSDW-DDF
Corel5k	0.334 ± 0.001	0.426 ± 0.007	0.445 ± 0.011	0.441 ± 0.014	0.466 ± 0.015	0.478 ± 0.018	0.495 ± 0.019
ESPGame	0.293 ± 0.001	0.427 ± 0.001	0.426 ± 0.005	0.426 ± 0.005	0.378 ± 0.006	0.439 ± 0.007	0.454 ± 0.004
Iaprtc12	0.312 ± 0.006	0.422 ± 0.008	0.429 ± 0.011	0.412 ± 0.010	0.358 ± 0.009	0.445 ± 0.006	0.447 ± 0.008
Pascal07	0.407 ± 0.006	0.427 ± 0.010	0.427 ± 0.013	0.386 ± 0.019	0.391 ± 0.015	0.434 ± 0.007	0.434 ± 0.008
MirFlickr	0.487 ± 0.001	0.598 ± 0.008	0.630 ± 0.010	0.614 ± 0.007	0.521 ± 0.015	0.628 ± 0.005	0.635 ± 0.004

TABLE VII
COVERAGE OF THE EXPERIMENTAL RESULTS

Data	CDMM	DMVMLC	VCMN	LMVCAT	AIM	MTD	VSDW-DDF
Corel5k	0.720 ± 0.001	0.709 ± 0.005	0.712 ± 0.009	0.700 ± 0.009	0.731 ± 0.011	0.726 ± 0.007	0.747 ± 0.005
ESPGame	0.538 ± 0.001	0.609 ± 0.001	0.588 ± 0.003	0.426 ± 0.005	0.378 ± 0.006	0.439 ± 0.007	0.454 ± 0.004
Iaprtc12	0.568 ± 0.002	0.646 ± 0.003	0.615 ± 0.005	0.616 ± 0.004	0.584 ± 0.004	0.616 ± 0.004	0.645 ± 0.003
Pascal07	0.711 ± 0.003	0.727 ± 0.006	0.757 ± 0.009	0.716 ± 0.013	0.687 ± 0.010	0.755 ± 0.006	0.758 ± 0.004
MirFlickr	0.603 ± 0.003	0.647 ± 0.004	0.662 ± 0.004	0.648 ± 0.005	0.614 ± 0.003	0.659 ± 0.002	0.665 ± 0.002

TABLE VIII
AVERAGE RANK OF THE EXPERIMENTAL RESULTS

Data	CDMM	DMVMLC	VCMN	LMVCAT	AIM	MTD	VSDW-DDF
Corel5k	5.83	5.83	3.00	6.17	2.17	2.67	1.00
ESPGame	6.33	1.67	3.33	5.00	5.67	3.17	1.50
Iaprtc12	6.50	2.50	2.83	4.83	6.17	3.17	1.17
Pascal07	6.00	3.83	2.17	5.67	6.33	2.17	1.00
MirFlickr	7.00	4.33	1.67	4.67	5.83	2.67	1.00

across five datasets. This superiority can be attributed to the remarkable feature extraction capabilities of neural networks.

2. Our proposed VSDW-DDF model stands out convincingly, showing better performance than comparative methods in almost all metrics across the five datasets. Whether compared with traditional approaches like CDMM or deep learning-based methods such as LMVCAT or MTD, our VSDW-DDF demonstrates excellent compatibility with unbalanced views and weak labels.

3. Among comparisons within deep learning methods, our method specifically designed to address unbalanced views exhibit notable advantages over other methods. This indicates that it is essential to consider the combined impact of potential unbalanced views and weak labels during the model design phase.

E. Ablation Study

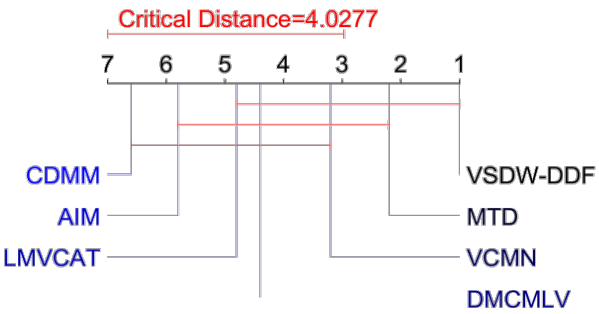
Our model integrates several strategies. To obtain a deeper understanding of the impact of each component, we carried out ablation analyses on the Corel5k dataset (which contains 50% weak labels and 50% unbalanced views) and report the findings in Table X. More precisely, we used a conventional single-channel framework as the baseline and further extended it to a dual-channel baseline architecture. We iteratively added our sample graph regularization module, view weighting module, and dropout module to this baseline and performed module removal operations. From Table X, we observe that all designed components contributed to performance improvements to varying degrees, with the dropout mechanism showing the most significant performance enhancement on the Corel5k dataset. This confirms the effectiveness of our improved dropout mechanism. However, the underlying mechanism behind the effectiveness of this strategy requires further investigation.

TABLE IX
AN OVERVIEW OF THE FRIEDMAN STATISTIC $F_F(K=6, N=5)$ ALONG WITH ITS CRITICAL VALUE BASED ON SIX ASSESSMENT METRICS.

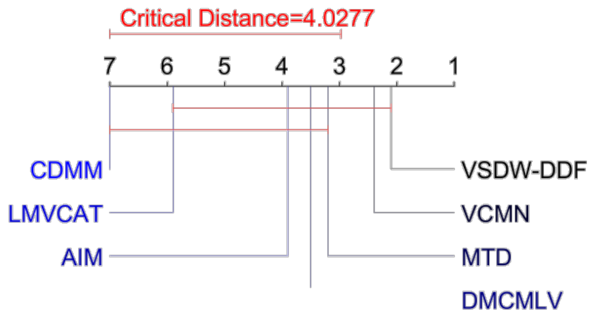
Evaluation Metric	F_F	Critical Value (0.05)
Average precision	21.925926	2.508189
Hamming Loss	9.461538	
Ranking Loss	8.993039	
AUC	9.365155	
One Error	16.000000	
Coverage	5.017773	

TABLE X
ABLATION STUDY RESULTS OF OUR PROPOSED VSDW-DDF MODEL ON THE COREL5K DATASET WITH 50% UNBALANCED VIEWS AND 50% WEAK LABELS. THE ABBREVIATIONS USED ARE AS FOLLOWS: 'D_C' DENOTES DUAL-CHANNEL DECOUPLING FRAMEWORK, 'S_G' REPRESENTS SAMPLE GRAPH REGULARIZATION, 'V_W' INDICATES VIEW WEIGHTING, 'D_O' DENOTES DROPOUT MECHANISM.

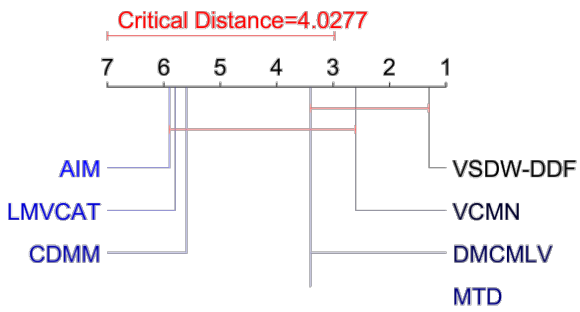
Method	Average precision	Hamming loss	Ranking loss	AUC	One error	Coverage
d_c	0.398	0.988	0.884	0.887	0.478	0.726
d_c+s_g	0.401	0.988	0.886	0.890	0.478	0.737
d_c+s_g+v_w	0.411	0.988	0.889	0.892	0.488	0.743
d_c+s_g+v_w+d_o	0.414	0.988	0.891	0.894	0.495	0.747



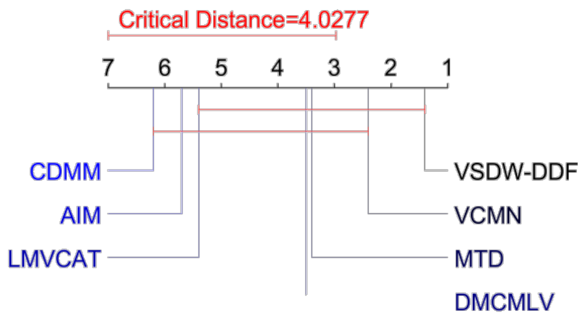
(a) Average precision



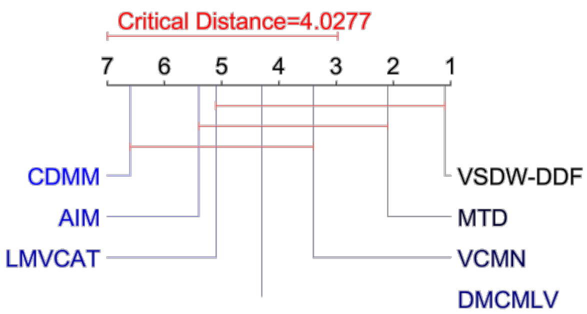
(b) Hamming loss



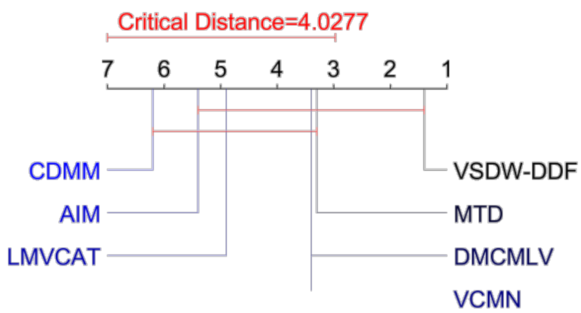
(c) Ranking loss



(d) AUC



(e) One error



(f) Coverage

Fig. 3. CD Values on Six Benchmark Datasets under Different Metrics.

V. Conclusion

In this paper, we address the complex and highly realistic task of UMVWML. We propose VSDW-DDF that integrates view-specific weighting and dropout strategies. Our model effectively mitigates the unbalanced information caused by unbalanced views by decoupling single-channel view-level representations into shared and view-specific representations. Two key strategies are introduced: view-specific weighting and view-specific dropout mechanism, to enhance model performance under unbalanced view conditions. Extensive experimental validation demonstrates that our approach significantly outperforms existing state-of-the-art methods, showing excellent performance in handling arbitrary MVML data with unbalanced views and weak labels.

Ablation studies also confirm the effectiveness of each component, including dual-channel decoupling framework, view-specific weights, sample-guided graph regularization losses, and view-specific dropout mechanism. These results not only showcase the superior performance of our method in tackling UMVWMLC tasks but also provide a solid foundation for future research. Despite the significant achievements in addressing UMVWML problems, there remain several directions worthy of further exploration. For instance, future work could investigate more effective methods for incomplete data recovery, optimize weighting strategies to adapt to more complex unbalanced scenarios, or model higher-order correlations among multiple labels to further improve model performance. Additionally, exploring how to better utilize unlabeled data or assessing the generalization capability of models on larger datasets are also important questions to consider.

In conclusion, we offer a new perspective and technical solution for addressing complex UMVWML problems. It enriches the theoretical intersection of multi-view learning and multi-label classification while providing robust support for practical applications.

References

[1] Qing Ai, Fei Li, Qingyun Gao, and Fei Zhao. An improved lst-ksvc based on energy model. *Engineering Letters*, vol.40, no.4, pp1436-1443, 2022.

[2] Fei Zhao, Qing Ai, Xiangna Li, Wenhui Wang, Qingyun Gao, and Yichun Liu. Tlc-xml: transformer with label correlation for extreme multi-label text classification. *Neural Processing Letters*, 56(1):25, 2024.

[3] Hao Li, Bart Adams, Leonidas J Guibas, and Pauly. Robust single-view geometry and motion reconstruction. *ACM Transactions on Graphics (ToG)*, 28(5):1–10, 2009.

[4] Kaibiao Lin, Runze Chen, Jinpo Chen, Ping Lu, and Fan Yang. Multi-view block matrix-based graph convolutional network. *Engineering Letters*, vol.32, no.6, pp1073-1082, 2024.

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

[6] Arthur Gretton, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf. Measuring statistical dependence with hilbert-schmidt norms. In *International Conference on Algorithmic Learning Theory*, pages 63–77. Springer, 2005.

[7] Yanchen Liu and Changming Zhu. Dmvmlc-vt: Deep incomplete multi-view multi-label image classification with view translation and pseudo-label enhancement. *The Visual Computer*, pages 1–17, 2024.

[8] Chengliang Liu, Jie Wen, Yabo Liu, Chao Huang, Zhihao Wu, Xiaoling Luo, and Yong Xu. Masked two-channel decoupling framework for incomplete multi-view weak multi-label learning. *Advances in Neural Information Processing Systems*, 36:32387–32400, 2023.

[9] Alireza Entezami, Hassan Sarmadi, and Bahareh Behkamal. A novel double-hybrid learning method for modal frequency-based damage assessment of bridge structures under different environmental variation patterns. *Mechanical Systems and Signal Processing*, 201:110676, 2023.

[10] Changqing Zhang, Yeqing Liu, and Huazhu Fu. Ae2-nets: Autoencoder in autoencoder networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2577–2585, 2019.

[11] Fei Li, Qing Ai, Xiangna Li, Wenhui Wang, Qingyun Gao, and Fei Zhao. Intuitionistic fuzzy least squares mltsvm for noisy label data using label-specific features and local label correlation. *Expert Systems with Applications*, 260:125351, 2025.

[12] Guangxia Wang, Changqing Zhang, Pengfei Zhu, and Qinghua Hu. Semi-supervised multi-view multi-label classification based on nonnegative matrix factorization. In *Artificial Neural Networks and Machine Learning–ICANN 2017: 26th International Conference on Artificial Neural Networks, Alghero, Italy, September 11-14, 2017, Proceedings, Part II 26*, pages 340–348. Springer, 2017.

[13] Wei Liu, Jiazheng Yuan, Gengyu Lyu, and Songhe Feng. Label driven latent subspace learning for multi-view multi-label classification. *Applied Intelligence*, 53(4):3850–3863, 2023.

[14] Haoli Zhao, Zhenni Li, Wuhui Chen, Zibin Zheng, and Shengli Xie. Accelerated partially shared dictionary learning with differentiable scale-invariant sparsity for multi-view clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 34(11):8825–8839, 2022.

[15] 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, 2022.

[16] Qing Ai, Fei Li, Xiangna Li, Ji Zhao, Wenhui Wang, Qingyun Gao, and Fei Zhao. An improved mltsvm using label-specific features with missing labels. *Applied Intelligence*, 53(7):8039–8060, 2023.

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

[18] Chaoran Li, Xiyin Wu, Pai Peng, Zhuhong Zhang, and Xiaohuan Lu. Incomplete multi-view partial multi-label classification via deep semantic structure preservation. *Complex & Intelligent Systems*, 10(6):7661–7679, 2024.

[19] Jie Wen, Chengliang Liu, Shijie Deng, Yicheng Liu, Lunke Fei, Ke Yan, and Yong Xu. Deep double incomplete multi-view multi-label learning with incomplete labels and missing views. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[20] Kaibo Duan, Shaodong Cui, Hiroyuki Shinnou, and Shi Bao. View-channel mixer network for double incomplete multi-view multi-label learning. *Neurocomputing*, 617:129013, 2025.

[21] Chengliang Liu, Jie Wen, Xiaoling Luo, and Yong Xu. Incomplete multi-view multi-label learning via label-guided masked view-and category-aware transformers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 8816–8824, 2023.

[22] Chengliang Liu, Jinlong Jia, Jie Wen, Yabo Liu, Xiaoling Luo, Chao Huang, and Yong Xu. Attention-induced embedding imputation for incomplete multi-view partial multi-label classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 13864–13872, 2024.

[23] Jie Wen, Zheng Zhang, Yong Xu, and Zuofeng Zhong. Incomplete multi-view clustering via graph regularized matrix factorization. In *European Conference on Computer Vision*, 2019.

[24] Weixiang Shao, Lifang He, Chun-ta Lu, and S Yu Philip. Online multi-view clustering with incomplete views. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 1012–1017. IEEE, 2016.

[25] Jun Yin and Shiliang Sun. Incomplete multi-view clustering with reconstructed views. *IEEE Transactions on Knowledge and Data Engineering*, 35(3):2671–2682, 2021.

[26] Shiliang Sun and Nan Zhang. Incomplete multiview nonnegative representation learning with graph completion and adaptive neighbors. *IEEE Transactions on Neural Networks and Learning Systems*, 35(3):4017–4031, 2022.

[27] Jie Wen, Zheng Zhang, Yong Xu, Bob Zhang, Lunke Fei, and Hong Liu. Unified embedding alignment with missing views inferring for incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 5393–5400, 2019.

[28] Xiang Fang, Yuchong Hu, Pan Zhou, and Dapeng Oliver Wu. Unbalanced incomplete multi-view clustering via the scheme of view evolution: Weak views are meat; strong views do eat. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 6(4):913–927, 2021.

[29] Guangyan Ji, Gui-Fu Lu, Bing Cai, and Yangfan Du. Unbalanced incomplete multi-view clustering based on low-rank tensor graph learning. *Expert Systems with Applications*, 225:120055, 2023.