A Two-Stage Information-Driven Multi-View Multi-Label Learning Method for Incomplete Data with Noisy Labels

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

Abstract—**To address the performance degradation in multi-view multi-label learning caused by missing views, missing labels, and label noise, this paper proposes a two-stage information-driven multi-view multi-label learning framework tailored for incomplete data and noisy labels. Inspired by information theory, the proposed method introduces task-relevance constraints and information compression mechanisms to effectively extract task-relevant information from incomplete and noisy data while suppressing irrelevant information. Specifically, in the first stage, multiple view-specific encoders are trained to extract discriminative representations from each view, focusing on task-relevant features. A label confidence mechanism is incorporated to identify and mitigate the adverse impact of label noise during training. In the second stage, the hidden representations obtained from the first stage are integrated through a fusion alignment mechanism and an autoencoder-based architecture. Additionally, a graph regularization term based on a label co-occurrence graph is introduced to enhance the modeling of structural correlations among labels and improve classifier accuracy for final prediction. Experimental results on fifteen public datasets demonstrate that the proposed method outperforms state-of-the-art approaches across various evaluation metrics, particularly under scenarios involving incomplete data and label noise. The method effectively mitigates performance degradation caused by data incompleteness and noisy labels, resulting in enhanced prediction accuracy and stability in multi-label learning, which is consistent with theoretical expectations.**

Index Terms—**deep learning, multi-view learning, multi-label learning, label noise**

I. Introduction

MULTI-VIEW data arises when an entity is observed from different perspectives, each providing distinct yet potentially complementary information. Such data often captures both shared and view-specific features of the same instance. In addition, a single object or image typically conveys rich semantic content, leading to the association of multiple labels [1]. For instance, a landscape photograph might be represented using descriptors such as SIFT, Gist, and HSV, and annotated with labels like “sea,” “clouds,” and “fish” [2].

To make better use of multi-view multi-label data, the concept of Multi-view Multi-label Classification (MvMLC) has been introduced, leading to the development of several techniques based on assumptions of view and label completion over the past few years [3]. However, it has been increasingly recognized by researchers that, in practical scenarios, multi-view multi-label data is often incomplete due to uncertainties during data collection and manual labeling processes. The absence of certain views or labels poses significant challenges for MvMLC tasks, thereby highlighting the pressing need and importance of studying incomplete Multi-view Multi-label Classification (iMvMLC) [4].

In the context of iMvMLC, traditional approaches such as iMvWL [5] proposed by Tan et al. and NAIML [6] developed by Li et al. have been explored. iMvWL integrates multi-view features and multi-label information into a discriminative shared space simultaneously. On the other hand, NAIML successfully leverages the consistency among multi-view data and the structural relationships between multi-labels. Despite their achievements, these methods still face notable limitations. For instance, both traditional techniques rely on shallow machine learning models, which are unable to capture deep discriminative patterns within the data and exhibit low computational efficiency. Additionally, matrix factorization-based methods like iMvWL encounter difficulties when dealing with new test samples during prediction using trained models [7].

Recent developments in deep neural networks (DNNs), such as MTD [8] and RANK [9] proposed by Liu et al., have shown remarkable advantages in this area. Compared to traditional matrix factorization techniques, DNNs are superior at capturing high-level semantic information and better suited for addressing complex challenges in iMvMLC.

Specifically designed DNN architectures for iMvMLC tasks have seen the widespread adoption of autoencoders, while contrastive learning has emerged as an effective strategy for performance improvement [10]. However, existing iMvMLC methods have failed to fully exploit the core principles of contrastive learning—particularly the extraction of task-relevant information through maximizing cross-view mutual information (MI) [11]. This limitation restricts their ability to fully utilize valuable information.

Manuscript received June 24, 2025; revised August 29, 2025.

This work was supported 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 is a graduate student at the School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan 114051, Liaoning China (e-mail: 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 (e-mail: 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 (e-mail: 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 (e-mail: 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).

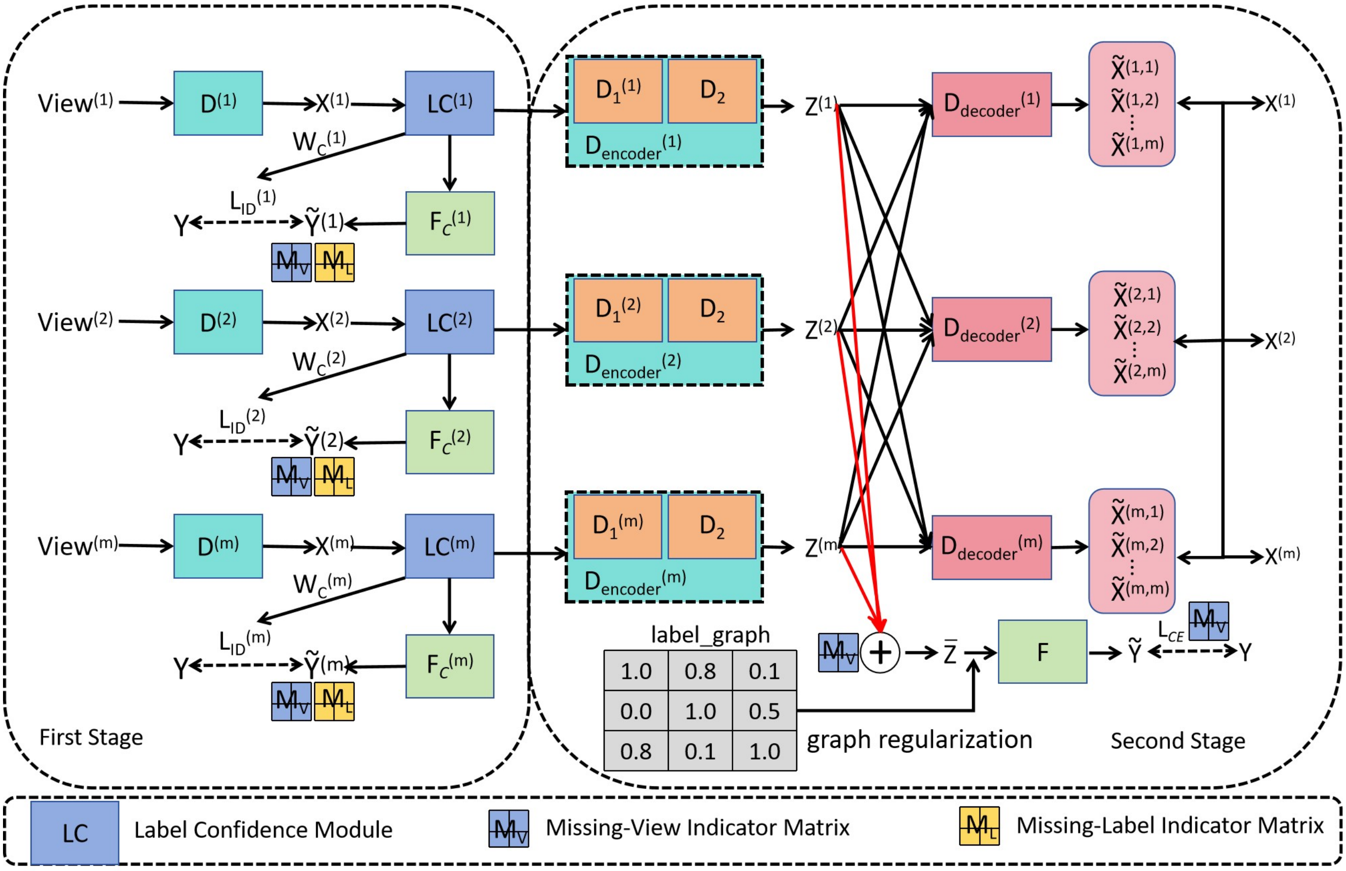


Fig. 1. The Main Framework of TIMMLIN

Moreover, label noise is common in practical applications, especially in optical domains. For example, in remote sensing imagery or medical optical imaging, high image complexity and ambiguous semantic boundaries often lead to annotation biases or inconsistencies during expert labeling, resulting in incorrect tags. Such label noise can corrupt supervisory signals, impair the learning of critical discriminative features, and ultimately degrade classification performance [12]. Despite this, most current methods still assume label accuracy when dealing with incomplete multi-view multi-label data, neglecting the practical challenges posed by label noise.

To tackle these issues, inspired by information-driven theory and contrastive learning analysis, this paper introduces an Information-Driven Network for Incomplete Data Classification with Label Noise (IDLCLN), as shown in Fig. 1. By considering both data incompleteness and the presence of noisy positive labels, our training process focuses not only on classification performance but also on representation quality. Accordingly, we design a two-phase network architecture for IDLCLN: Phase 1: A Noise-aware Enhanced Representation Generation Network. Phase 2: An Incomplete Multi-view Classification Network with Mutual Prediction Capabilities.

Unlike existing approaches, our model aims to integrate task-relevant information compression with mutual information extraction to address IDLCLN challenges. Our main contributions are summarized as follows:

1. A novel DNN framework developed from an information-driven perspective to solve IDLCLN problems. This work highlights the potential of information-driven theory in tackling IDLCLN challenges.

2. An innovative information-driven framework that balances cross-view mutual information and view-specific information. Specifically, our method not only extracts task-relevant features from individual views but also integrates shared and complementary information across views within a unified architecture. The framework effectively addresses pervasive label noise in real-world scenarios through confidence mechanisms and graph-structured constraints, suppressing noise interference while enhancing model robustness and expressiveness under non-ideal data conditions.

II. Related Works

A. Information-Driven Theory

Information-driven theory offers a principled and unified framework for representation learning by emphasizing the suppression of redundant features while preserving discriminative information. This enables the extraction of more distinctive and informative representations. The central principle involves enhancing task-relevant information and suppressing irrelevant signals within the input data—a strategy that has demonstrated effectiveness in single-view scenarios [13]. More recently, efforts have been made to extend this framework to multi-view learning, with the aim of addressing challenges such as view incompleteness. However, current approaches often fail to fully exploit both the shared and view-specific information across multiple perspectives, which constrains their generalization performance in complex, real-world settings.

B. Application of Information Theory in Contrastive Learning

In the field of contrastive learning, numerous researchers have pursued theoretical insights into its underlying mechanisms. Oord et al. [14] demonstrated that the widely adopted Noise-Contrastive Estimation (NCE) [15] loss serves as a lower bound to the mutual information (MI) between different views, thereby leading to its reinterpretation as the InfoNCE loss. This theoretical insight suggests that the performance improvements achieved through contrastive learning are primarily attributed to the maximization of cross-view mutual information.

Building upon this foundation, Tsai et al. [16] provided further theoretical justification from an information-theoretic perspective. Within the multi-view learning framework, non-overlapping information across views is typically regarded as noise or redundancy. Consequently, increasing mutual information between views facilitates the extraction of more discriminative and task-relevant features. However, in practical applications, multi-view data is not always derived through data augmentations or consistent transformations, which may undermine the validity of this assumption. Therefore, it is essential to not only enhance inter-view information sharing but also preserve the semantic integrity of the original inputs. This highlights the necessity

III. Method

This section provides a systematic description of the proposed model, which consists of two sequential phases: Phase 1: A Noise-Aware Enhanced Representation Generation Network, and Phase 2: An Incomplete Multi-View Classification Network with Mutual Prediction Capabilities. Before delving into the technical details, we first formalize the problem definition and introduce the necessary notations to ensure clarity.

A. Problem Definition

We formally define the IDLCLN task as follows: A given dataset with m views and n samples is denoted as: $\{V^{(k)} \in \mathbb{R}^{n\times d^{(k)}}\}_{k=1}^{m}$, where $d^{(k)}$ is the feature dimension of the k-th view. We define $Y \in \{0,1\}^{n\times l}$ as the label matrix, where l is the number of categories. $Y_{i,j} = 1$ indicates that the i-th sample is marked with the j-th category, otherwise $Y_{i,j} = 0$. Considering the compatibility for missing views and missing labels, we define two key prior matrices, the missing-view indicator $\mathbf{M}_V \in \{0,1\}^{n\times m}$ and the missing-label indicator $\mathbf{M}_L \in \{0,1\}^{n\times l}$. Taking the i-th sample as an example, $\mathbf{M}_V^{i,j} = 1$ means its j-th view is available, otherwise $\mathbf{M}_V^{i,j} = 0$. $\mathbf{M}_L^{i,j} = 0$ indicates that we are not sure whether the i-th sample has the j-th label, otherwise $\mathbf{M}_L^{i,j} = 1$. For simplicity, we fill random values for the missing views in the original data and set '0' for unknown labels. Our objective is to train a neural network model on data with missing values and noise, enabling it to perform multi-label classification on samples with incomplete data and label noise. In addition, our representative notations are summarized in Table I.

TABLE I
Notations

Notation	Description
n, m, l	number of samples, views, and categories, respectively
$d_v^{(k)}, d_x^{(k)}, d_z^{(k)}$	different dimensions
$v^{(k)} \in \mathbb{R}^{d_v^{(k)}}$	the original instance from k-th view
$x^{(k)} \in \mathbb{R}^{d_x^{(k)}}$	first-stage representation
$z^{(k)} \in \mathbb{R}^{d_z^{(k)}}$	second-stage representation
$y \in \{0,1\}^l$	label
$V^{(k)}, X^{(k)}, Z^{(k)}$	data matrices
$\widetilde{X}^{(u,k)} \in \mathbb{R}^{n\times d_x^{(u)}}$	reconstruction of $X^{(u)}$ from $Z^{(k)}$
$Y \in \{0,1\}^{n\times l}$	label matrix
$\widetilde{Y}^{(k)}, \widetilde{Y} \in [0,1]^{n\times l}$	prediction of Y
$M_V \in \{0,1\}^{n\times m}$	missing-view indicator matrix
$M_L \in \{0,1\}^{n\times l}$	missing-label indicator matrix
$D_{encoder}, D_{decoder}$	encoder and decoder, respectively
F_C, F	classifier

B. First Stage: A Noise-Aware Enhanced Representation Generation Network

Numerous studies have emphasized that high-quality representations should ideally retain substantial task-relevant information while filtering out task-irrelevant components essentially aligning with the notion of a minimal sufficient statistic. In the context of IDLCLN tasks, raw input data $v^{(k)}$ typically carries a significant amount of irrelevant content, which may hinder the performance of the downstream classification model.

To mitigate this problem, drawing inspiration from information-theoretic principles, we propose to derive a more refined and task-relevant representation $x^{(k)}$ that adheres to the Information-Driven principle, thereby serving as a replacement for the original low-purity instance $v^{(k)}$:

$$\min_{x^{(k)}} -I(x^{(k)}; y) + \beta^{(k)} I(x^{(k)}; v^{(k)}), \tag{1}$$

where $\beta^{(k)} > 0$ serves as a tunable parameter. The term $I(x^{(k)}; y)$ represents the MI between the transformed representation $x^{(k)}$ and the corresponding label y, which can be practically estimated via minimizing the standard cross-entropy loss. On the other hand, $I(x^{(k)}; v^{(k)})$ measures the MI between $x^{(k)}$ and the original input $v^{(k)}$, acting as a regularization term that penalizes the retention of task-irrelevant information in $x^{(k)}$.

Directly optimizing this objective is generally intractable due to the complexity of mutual information estimation. To address this issue, we adopt a variational approximation strategy inspired by [17], which enables us to derive the following tractable upper bound:

$$I(x^{(k)}; v^{(k)}) \leq \mathbb{E}_{p(x^{(k)}, v^{(k)})}\left[\log \frac{p(x^{(k)}|v^{(k)})}{q(x^{(k)})}\right], \tag{2}$$

where $\mathbb{E}$ represents expectation, and $q(x^{(k)})$ represents an approximate distribution to the marginal distribution $p(x^{(k)})$. Focusing on this upper bound, we can arbitrarily set $q(x^{(k)}) = \mathcal{N}(x^{(k)}|\mathbf{0}, \mathbf{I})$ and $p(x^{(k)}|v^{(k)}) = \mathcal{N}(x^{(k)}|\mu^{(k)}, \Sigma^{(k)})$ for computational expedience, where both mean $\mu^{(k)} \in \mathbb{R}^{d_x^{(k)}}$ and diagonal covariance $\Sigma^{(k)} \in \mathbb{R}^{d_x^{(k)}\times d_x^{(k)}}$ are correlated with $v^{(k)}$. Specifically, we utilize a special MLP $\mathbf{D}^{(k)}$ to derive $\mu^{(k)}$ and $\Sigma^{(k)}$. When the input of $\mathbf{D}^{(k)}$ is the data matrix

$\mathbf{V}^{(k)} \in \mathbb{R}^{n\times d_x^{(k)}}$, we let two new symbols represent its output, namely $\mathbf{U}^{(k)} \in \mathbb{R}^{n\times d_x^{(k)}}$ and $\mathbf{S}^{(k)} \in \mathbb{R}^{n\times d_x^{(k)}}$. So $X^{(k)} = \mathbf{U}^{(k)} + \epsilon \odot \mathbf{S}^{(k)}$, where $\epsilon \in \mathbb{R}^{n\times d_x^{(k)}}$ is a standard Gaussian distribution matrix, and $\odot$ denotes the Hadamard product.

By substituting the two Gaussians $p(x^{(k)}|v^{(k)})$ and $q(x^{(k)})$ into formula (2), decomposing them into multiple one-dimensional Gaussians, and approximating the marginal distribution $p(v^{(k)})$ using an empirical distribution, formula (2) can be transformed into the equivalent problem $\min \frac{1}{2nd_{x^{(k)}}} \sum_{i=1}^{n} \sum_{j=1}^{d_{x^{(k)}}} \big(- \log\left(S_{i,j}^{(k)}\right)^2 + \left(U_{i,j}^{(k)}\right)^2 + \left(S_{i,j}^{(k)}\right)^2 - 1\big)$.

To enhance the robustness of the model against label noise, especially for noise present in positive labels, a representation-aware label confidence mechanism is proposed. Specifically, a lightweight MLP module $\Phi_{\text{conf}}^{(k)}$ is designed to take the view-specific representation $\mathbf{x}^{(k)}$ as input and generate a label confidence matrix $\mathbf{C}^{(k)} \in [0,1]^{n\times l}$:

$$\mathbf{C}^{(k)} = \sigma\left(\Phi_{\text{conf}}^{(k)}\left(\mathbf{x}^{(k)}\right)\right), \tag{3}$$

where $\sigma(\cdot)$ represents the sigmoid activation function. The generated confidence weights are embedded into the cross-entropy loss as dynamic weights for each label, which helps reduce the impact of unreliable labels during model training. Furthermore, by integrating the missing-view and missing-label indicator matrices, $\mathbf{M}_V$ and $\mathbf{M}_L$, the loss function excludes the negative effects of unavailable views and labels. The final form of the revised supervised loss is given by:

$$\begin{aligned}\mathcal{L}_{\text{ID}}^{(k)} = \frac{1}{nl}\sum_{i=1}^{n}\sum_{j=1}^{l} M_V^{i,j} M_L^{i,j} \cdot C_{i,j}^{(k)} \cdot \big[(1-Y_{i,j})\\ log(1-\widetilde{Y}_{i,j}^{(k)}) + Y_{i,j}\log(\widetilde{Y}_{i,j}^{(k)})\big],\end{aligned} \tag{4}$$

where $\widetilde{Y}^{(k)}$ is the output of module $F_C^{(k)}$. Accordingly, the overall loss function for the first stage becomes:

$$\begin{aligned}\mathcal{L}_{\text{stage-1}}^{(k)} = \mathcal{L}_{\text{ID}}^{(k)} + \frac{\beta^{(k)}}{2nd_{x^{(k)}}}\sum_{i=1}^{n}\sum_{j=1}^{d_{x^{(k)}}} M_V^{i,k} \cdot \big(\\ -\log\left(\Sigma_{i,j}^{(k)}\right)^2 + \left(\mu_{i,j}^{(k)}\right)^2 + \left(\Sigma_{i,j}^{(k)}\right)^2 - 1\big).\end{aligned} \tag{5}$$

The first component represents the cross-entropy loss, which facilitates the extraction of information relevant to the target task and aligns with $-I(x^{(k)};y)$. The second component serves as a compression constraint, encouraging the representation to retain less redundant information, and corresponds to $I(x^{(k)};v^{(k)})$.

Clearly, the training in the first stage actually involves separately utilizing data from each view to predict the label. As shown in Figure 1, the raw data $V^{(k)}$ directly pass through the two MLPs $D^{(k)}$ and $F_C^{(k)}$ to predict the label Y in the first stage. $F_C^{(k)}$ is used solely for training $F_C^{(k)}$ and will be discarded after training. After the first-stage training, we obtain several MLPs $\{F_C^{(k)}\}_{k=1}^{m}$ that can enhance the concentration of task-relevant information.

C. Second Stage: An Incomplete Multi-View Classification Network with Mutual Prediction Capabilities

To more effectively utilize informative signals, contrastive learning is often employed through optimization of the original InfoNCE loss. However, this approach may inadvertently separate positive sample pairs. Although supervised contrastive learning offers improved performance, its direct application to iMvMLC tasks is hindered by incomplete label annotations. Motivated by the ideas in [18], we design an autoencoder-based mutual information extraction framework capable of facilitating cross-view prediction.

Motivated by previous studies, we aim to enhance cross-view MI in multi-view learning while retaining essential original semantics, guided by principles from information theory:

$$\begin{aligned}\max_{\mathbf{z}^{(k)},\,k=1,\ldots,m} \quad & \frac{1}{m(m-1)}\sum_{k=1}^{m}\sum_{u\neq k}^{m} I(\mathbf{z}^{(k)};\mathbf{x}^{(u)})\\ & + \frac{\lambda}{m}\sum_{k=1}^{m} I(\mathbf{z}^{(k)};\mathbf{x}^{(k)}),\end{aligned} \tag{6}$$

where $\lambda > 0$ is a combination factor, and $\mathbf{z}^{(k)}$ is the second-stage representation. According to MI, we have $I(\mathbf{z}^{(k)};\mathbf{x}^{(u)}) = H(\mathbf{x}^{(u)}) - H(\mathbf{x}^{(u)} \mid \mathbf{z}^{(k)})$, where $H(\mathbf{x}^{(u)})$ represents the entropy and $H(\mathbf{x}^{(u)} \mid \mathbf{z}^{(k)})$ represents the conditional entropy. Since the m trained MLPs $\{F_C^{(k)}\}_{k=1}^{m}$ are fixed in the second stage, we can ignore $H(\mathbf{x}^{(u)})$. So our goal is to maximize $-H(\mathbf{x}^{(u)} \mid \mathbf{z}^{(k)})$. By employing variational approximation, we have:

$$-H(\mathbf{x}^{(u)} \mid \mathbf{z}^{(k)}) \geq \mathbb{E}_{p(\mathbf{x}^{(u)},\mathbf{z}^{(k)})}\left[\log q(\mathbf{x}^{(u)} \mid \mathbf{z}^{(k)})\right]. \tag{7}$$

Similarly, we can also set $q(\mathbf{x}^{(u)} \mid \mathbf{z}^{(k)}) = \mathcal{N}(\mathbf{x}^{(u)} \mid \mu^{(u,k)}, \sigma\mathbf{I})$, with a fixed $\sigma \in \mathbb{R}$, and the mean $\mu^{(u,k)} \in \mathbb{R}^{d_{\mathbf{x}}^{(u)}}$ is dependent on $\mathbf{z}^{(k)}$. In this case, maximizing $\mathbb{E}_{p(\mathbf{x}^{(u)},\mathbf{z}^{(k)})}\left[\log q(\mathbf{x}^{(u)} \mid \mathbf{z}^{(k)})\right]$ is equivalent to minimizing $\mathbb{E}_{p(\mathbf{x}^{(u)},\mathbf{z}^{(k)})}\left[\|\mathbf{x}^{(u)} - \mu^{(u,k)}\|_2^2\right]$. In our work, we employ a decoder $D_{decoder}$ to generate $\mu^{(u,k)}$, i.e., $\mu^{(u,k)} = D_{decoder}^{(u)}(\mathbf{z}^{(k)})$, and denote the reconstructed sample as $\widetilde{\mathbf{x}}^{(u,k)} = D_{decoder}^{(u)}(\mathbf{z}^{(k)})$. With the same treatment applied to $I(\mathbf{z}^{(k)};\mathbf{x}^{(k)})$, by introducing $\mathbf{M}_V$, we can obtain the following information-theoretic loss for incomplete multi-view learning:

$$\begin{aligned}L_{CL} = \; & \frac{1}{nm(m-1)}\sum_{i=1}^{n}\sum_{k=1}^{m}\sum_{u\neq k}^{m}\frac{M_V^{i,u}M_V^{i,k}}{d_{\mathbf{x}}^{(u)}}\\ & \left\|\mathbf{x}_i^{(u)} - \widetilde{\mathbf{x}}_i^{(u,k)}\right\|_2^2 + \frac{\lambda}{nm}\sum_{i=1}^{n}\sum_{k=1}^{m}\frac{M_V^{i,k}}{d_{\mathbf{x}}^{(k)}}\\ & \left\|\mathbf{x}_i^{(k)} - \widetilde{\mathbf{x}}_i^{(k,k)}\right\|_2^2 = \; L_{CL1} + \lambda L_{CL2},\end{aligned} \tag{8}$$

where L_{CL1} denotes the first component, aimed at enhancing cross-view MI, and λL_{CL2} indicates the second component, which helps retain the original information. It is worth noting that minimizing the loss in Equation (6) further strengthens cross-view MI in the second stage network, functioning similarly to contrastive learning, yet differing from InfoNCE by not relying on instance-level discrimination.

Now we are discussing the structure of the new autoencoder network in this stage. Each encoder $D_{encoder}^{(k)}$ comprises two parts. The first part is a view-specific module $D_1^{(k)}$ mapping $X^{(k)}$ to a joint semantic space and the second part is a shared module D_2. Therefore, $D_{encoder}^{(k)}(X^{(k)}) = D_2\left(D_1^{(k)}(X^{(k)})\right)$, where $D_1^{(k)}$ and D_2 are both MLPs. Furthermore, we utilize one specific decoder $D_{decoder}^{(u)}$ in conjunction with different representations $Z^{(k)}{}_{k=1}^{m}$ to reconstruct specific $X^{(u)}$, so that $\widetilde{X}^{(u,k)} = D_{decoder}^{(u)}(Z^{(k)})$. In this way, we use the same set of decoders and part of the same encoder D_2 to implicitly align the representations.

It is important to note that our approach reduces the mutual information (MI) between $\mathbf{x}^{(k)}$ and $\mathbf{v}^{(k)}$ during the first stage, while increasing the MI between $\mathbf{z}^{(k)}$ and $\mathbf{x}^{(k)}$ in the second stage. Although this may appear contradictory at first glance, we provide the following rationale to justify its effectiveness. Within the autoencoder framework, $\mathbf{z}^{(k)}$ is intended to retain as much informative content from $\mathbf{x}^{(k)}$ as possible; therefore, the informativeness and purity of $\mathbf{x}^{(k)}$ are of critical importance. When the original instance $\mathbf{v}^{(k)}$ contains substantial noise or redundancy, it becomes necessary to first extract task-relevant information—effectively implementing an initial compression step—before attempting to preserve and reconstruct meaningful features. As a result, the model focuses on reconstructing $\mathbf{x}^{(k)}$ rather than the noisier $\mathbf{v}^{(k)}$, thereby improving the overall quality of the learned representations. This mechanism proves especially advantageous when the integrity of $\mathbf{v}^{(k)}$ is compromised.

Final Prediction and Overall Objective Loss After acquiring the informative representations $\{Z^{(k)}\}_{k=1}^{m}$ from each view, the next step is to integrate them for final prediction. Due to view incompleteness, direct concatenation is infeasible. Although a naive strategy such as simple averaging can be employed, it fails to consider the differing significance among views. Therefore, we utilize a widely adopted fusion approach designed specifically for scenarios with missing views, formulated as:

$$\bar{\mathbf{Z}}_{i,\cdot} = \sum_{k=1}^{m} w^{(k)} \cdot \frac{\mathbf{Z}_{i,\cdot}^{(k)} \cdot \mathbf{M}_V^{i,k}}{\sum_{r=1}^{m} \mathbf{M}_V^{i,r}}, \tag{9}$$

which is a weighted average of the $\mathbf{Z}_{i,\cdot}^{(k)}$. $\{w^{(k)}\}_{k=1}^{m}$ are learnable weights that have been processed through softmax. For the final classifier F, we employ MLPs. And we utilize the cross-entropy loss function, which is shown as follows:

$$\begin{aligned} L_{CE} = \frac{1}{nl} \sum_{i=1}^{n} \sum_{j=1}^{l} \mathbf{M}_L^{i,j} \Big[&(1 - \mathbf{Y}_{i,j}) \log(1 - \widetilde{\mathbf{Y}}_{i,j}^{*}) \\ &+ \mathbf{Y}_{i,j} \log(\widetilde{\mathbf{Y}}_{i,j}^{*}) \Big] \end{aligned} \tag{10}$$

where the final prediction $\widetilde{\mathbf{Y}} = F \in [0,1]^{n \times l}$ is learnable weights that have been processed through softmax. To further enhance the model's capability in modeling label correlations and suppress the propagation of label noise, we introduce a label co-occurrence graph regularization term. Specifically, we define a label graph $\mathbf{M}_L \in \mathbb{R}^{l \times l}$ based on the co-occurrence statistics of labels in the training data. Given the predicted label distribution $\widetilde{\mathbf{Y}}$, we encourage semantically related labels to have similar predictions via the following graph smoothness regularization:

$$\mathcal{L}_{\text{graph}} = \mathrm{Tr}\left(\widetilde{\mathbf{Y}}^{*} \mathbf{L} \left(\widetilde{\mathbf{Y}}^{*} \right)^{\top} \right), \tag{11}$$

where $\mathbf{L} = \mathbf{D} - \mathbf{G}_L$ is the Laplacian matrix of the label graph. By combining (6) , (8) and (9) through a parameter α, we can obtain the final objective loss for the network in the second stage:

$$\mathcal{L}_{\text{SCL}} = \mathcal{L}_{\text{CE}} + \alpha \mathcal{L}_{\text{CL1}} + \gamma \mathcal{L}_{\text{CL2}} + \lambda_{\text{G}} \mathcal{L}_{\text{graph}}, \tag{12}$$

where λ_{G} is a regularization coefficient that balances the contribution of label structure information.

Algorithm 1 Training Process

```
Input: Incomplete multi-view data $\{\mathbf{V}^{(k)}\}_{k=1}^{m}$, indicator matrices $\mathbf{M}_V, \mathbf{M}_L$, multi-label matrix $\mathbf{Y}$, batch size $B$, parameters $\{\beta^{(k)}\}_{k=1}^{m}$, $\alpha$, $\gamma$
Output: The trained model
Stage 1: Initialize $\{\mathbf{D}^{(k)}, \mathbf{F}_C^{(k)}\}_{k=1}^{m}$
for $k = 1$ to $m$ do
    while not converged do
        Compute $\mathbf{X}^{(k)} = \mathbf{D}^{(k)}(\mathbf{V}^{(k)})$
        $\widetilde{\mathbf{Y}} = \mathbf{F}_C^{(k)}(\mathbf{X}^{(k)})$, $\mathcal{L}_{ID}^{(k)}$
        Update $\mathbf{D}^{(k)}, \mathbf{F}_C^{(k)}$
    end while
end for
Save $\{\mathbf{D}^{(k)}\}_{k=1}^{m}$
Stage 2: Initialize $\{\mathbf{D}_1^{(k)}\}_{k=1}^{m}$, $\mathbf{D}_2$, $\{\mathbf{D}_{decoder}^{(k)}\}_{k=1}^{m}$, classifier $\mathbf{F}$, weights $\{w^{(k)}\}_{k=1}^{m}$
while not converged do
    for $k = 1$ to $m$ do
        Compute $\mathbf{Z}^{(k)} = \mathbf{D}_2(\mathbf{D}_1^{(k)}(\mathbf{D}^{(k)}(\mathbf{V}^{(k)})))$
    end for
    for $u = 1$ to $m$ do
        for $k = 1$ to $m$ do
            Compute $\widetilde{\mathbf{X}}^{(u,k)} = \mathbf{D}_{decoder}^{(u)}(\mathbf{Z}^{(k)})$
        end for
    end for
    Compute $\bar{\mathbf{Z}}, \widetilde{\mathbf{Y}} = \mathbf{F}(\bar{\mathbf{Z}})$, $\mathcal{L}_{SCL}$
    Update parameters
end while
Save final model components
```

IV. Experiments

This section presents a systematic evaluation of the effectiveness and performance of the proposed TIMMLIN model, conducted through extensive experiments on fifteen publicly available datasets.

TABLE II
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 III
AVERAGE PRECISION UNDER DIFFERENT NOISE LEVELS ON FIFTEEN DATASETS

Dataset	Noise Level	AIM	DMVMLC	LMVCAT	MTD	RANK	iMvWL	VCMN	TIMMLIN
Corel5k	30%Sample	0.388 ± 0.013	0.335 ± 0.008	0.359 ± 0.012	0.392 ± 0.005	0.396 ± 0.004	0.252 ± 0.008	0.382 ± 0.008	0.435 ± 0.008
	50%Sample	0.385 ± 0.005	0.332 ± 0.009	0.355 ± 0.011	0.391 ± 0.001	0.394 ± 0.008	0.250 ± 0.011	0.381 ± 0.006	0.430 ± 0.010
	70%Sample	0.381 ± 0.005	0.330 ± 0.001	0.354 ± 0.003	0.388 ± 0.002	0.395 ± 0.007	0.249 ± 0.007	0.381 ± 0.004	0.429 ± 0.002
ESPGame	30%Sample	0.299 ± 0.002	0.292 ± 0.003	0.283 ± 0.009	0.298 ± 0.003	0.310 ± 0.014	0.221 ± 0.003	0.299 ± 0.005	0.323 ± 0.011
	50%Sample	0.297 ± 0.001	0.291 ± 0.002	0.281 ± 0.011	0.297 ± 0.009	0.307 ± 0.008	0.219 ± 0.014	0.299 ± 0.009	0.321 ± 0.002
	70%Sample	0.294 ± 0.012	0.290 ± 0.013	0.278 ± 0.007	0.296 ± 0.013	0.304 ± 0.004	0.217 ± 0.008	0.298 ± 0.008	0.320 ± 0.007
Iaprtc12	30%Sample	0.319 ± 0.009	0.317 ± 0.002	0.303 ± 0.011	0.320 ± 0.010	0.336 ± 0.008	0.214 ± 0.014	0.320 ± 0.013	0.363 ± 0.006
	50%Sample	0.320 ± 0.010	0.315 ± 0.005	0.302 ± 0.012	0.320 ± 0.012	0.337 ± 0.004	0.214 ± 0.007	0.318 ± 0.008	0.362 ± 0.001
	70%Sample	0.320 ± 0.010	0.315 ± 0.006	0.301 ± 0.010	0.321 ± 0.012	0.337 ± 0.008	0.210 ± 0.009	0.319 ± 0.014	0.361 ± 0.001
MirFlickr	30%Sample	0.591 ± 0.004	0.586 ± 0.005	0.581 ± 0.005	0.601 ± 0.007	0.595 ± 0.006	0.473 ± 0.011	0.607 ± 0.015	0.620 ± 0.015
	50%Sample	0.586 ± 0.011	0.586 ± 0.011	0.580 ± 0.005	0.599 ± 0.013	0.594 ± 0.014	0.471 ± 0.013	0.606 ± 0.014	0.616 ± 0.002
	70%Sample	0.588 ± 0.009	0.584 ± 0.006	0.581 ± 0.015	0.598 ± 0.004	0.594 ± 0.005	0.439 ± 0.001	0.606 ± 0.002	0.614 ± 0.007
Pascal07	30%Sample	0.547 ± 0.008	0.529 ± 0.012	0.519 ± 0.003	0.547 ± 0.011	0.551 ± 0.015	0.458 ± 0.013	0.554 ± 0.012	0.579 ± 0.009
	50%Sample	0.545 ± 0.006	0.525 ± 0.013	0.516 ± 0.005	0.547 ± 0.008	0.545 ± 0.007	0.455 ± 0.007	0.548 ± 0.003	0.578 ± 0.013
	70%Sample	0.540 ± 0.012	0.522 ± 0.010	0.508 ± 0.006	0.546 ± 0.010	0.545 ± 0.006	0.452 ± 0.009	0.544 ± 0.010	0.577 ± 0.004

TABLE IV
HAMMING LOSS UNDER DIFFERENT NOISE LEVELS ON FIFTEEN DATASETS

Dataset	Noise Level	AIM	DMVMLC	LMVCAT	MTD	RANK	iMvWL	VCMN	TIMMLIN
Corel5k	30%Sample	0.987 ± 0.005	0.987 ± 0.002	0.986 ± 0.002	0.987 ± 0.004	0.987 ± 0.003	0.978 ± 0.008	0.987 ± 0.008	0.988 ± 0.005
	50%Sample	0.987 ± 0.004	0.987 ± 0.006	0.986 ± 0.014	0.987 ± 0.014	0.987 ± 0.011	0.978 ± 0.012	0.987 ± 0.003	0.988 ± 0.001
	70%Sample	0.987 ± 0.004	0.987 ± 0.003	0.986 ± 0.009	0.987 ± 0.008	0.987 ± 0.009	0.978 ± 0.008	0.987 ± 0.008	0.988 ± 0.001
ESPGame	30%Sample	0.982 ± 0.005	0.982 ± 0.012	0.981 ± 0.001	0.983 ± 0.015	0.982 ± 0.014	0.970 ± 0.010	0.982 ± 0.002	0.983 ± 0.002
	50%Sample	0.982 ± 0.012	0.982 ± 0.011	0.981 ± 0.006	0.983 ± 0.012	0.982 ± 0.001	0.970 ± 0.011	0.981 ± 0.008	0.983 ± 0.014
	70%Sample	0.982 ± 0.009	0.982 ± 0.006	0.981 ± 0.013	0.983 ± 0.015	0.982 ± 0.013	0.970 ± 0.009	0.981 ± 0.014	0.983 ± 0.001
Iaprtc12	30%Sample	0.980 ± 0.008	0.980 ± 0.007	0.979 ± 0.003	0.981 ± 0.009	0.980 ± 0.014	0.966 ± 0.004	0.980 ± 0.007	0.981 ± 0.005
	50%Sample	0.980 ± 0.002	0.980 ± 0.004	0.979 ± 0.010	0.981 ± 0.004	0.980 ± 0.009	0.966 ± 0.005	0.980 ± 0.011	0.981 ± 0.003
	70%Sample	0.980 ± 0.001	0.980 ± 0.007	0.979 ± 0.004	0.981 ± 0.006	0.980 ± 0.007	0.966 ± 0.002	0.980 ± 0.007	0.981 ± 0.001
MirFlickr	30%Sample	0.888 ± 0.009	0.888 ± 0.008	0.880 ± 0.015	0.890 ± 0.002	0.890 ± 0.005	0.830 ± 0.011	0.890 ± 0.014	0.894 ± 0.011
	50%Sample	0.889 ± 0.009	0.888 ± 0.013	0.880 ± 0.008	0.890 ± 0.011	0.890 ± 0.013	0.830 ± 0.005	0.890 ± 0.008	0.894 ± 0.011
	70%Sample	0.888 ± 0.007	0.888 ± 0.005	0.881 ± 0.004	0.890 ± 0.002	0.890 ± 0.015	0.830 ± 0.013	0.880 ± 0.009	0.894 ± 0.013
Pascal07	30%Sample	0.930 ± 0.004	0.928 ± 0.002	0.919 ± 0.009	0.930 ± 0.013	0.932 ± 0.008	0.880 ± 0.003	0.930 ± 0.010	0.933 ± 0.005
	50%Sample	0.930 ± 0.003	0.927 ± 0.015	0.917 ± 0.006	0.930 ± 0.009	0.932 ± 0.009	0.880 ± 0.012	0.930 ± 0.009	0.933 ± 0.009
	70%Sample	0.930 ± 0.002	0.927 ± 0.006	0.920 ± 0.005	0.930 ± 0.002	0.932 ± 0.002	0.880 ± 0.006	0.930 ± 0.007	0.933 ± 0.005

TABLE V
RANKING LOSS UNDER DIFFERENT NOISE LEVELS ON FIFTEEN DATASETS

Dataset	Noise Level	AIM	DMVMLC	LMVCAT	MTD	RANK	iMvWL	VCMN	TIMMLIN
Corel5k	30%Sample	0.888 ± 0.003	0.871 ± 0.008	0.863 ± 0.003	0.877 ± 0.009	0.887 ± 0.011	0.861 ± 0.007	0.887 ± 0.009	0.913 ± 0.013
	50%Sample	0.887 ± 0.010	0.869 ± 0.013	0.863 ± 0.014	0.877 ± 0.010	0.886 ± 0.012	0.861 ± 0.006	0.887 ± 0.008	0.914 ± 0.002
	70%Sample	0.888 ± 0.013	0.872 ± 0.009	0.863 ± 0.015	0.878 ± 0.010	0.886 ± 0.007	0.861 ± 0.004	0.887 ± 0.012	0.915 ± 0.003
ESPGame	30%Sample	0.839 ± 0.008	0.841 ± 0.006	0.820 ± 0.007	0.831 ± 0.003	0.845 ± 0.009	0.803 ± 0.005	0.837 ± 0.010	0.857 ± 0.012
	50%Sample	0.839 ± 0.015	0.840 ± 0.011	0.818 ± 0.002	0.831 ± 0.006	0.842 ± 0.005	0.803 ± 0.002	0.836 ± 0.006	0.857 ± 0.015
	70%Sample	0.839 ± 0.008	0.840 ± 0.004	0.818 ± 0.013	0.830 ± 0.013	0.842 ± 0.010	0.803 ± 0.012	0.836 ± 0.004	0.857 ± 0.001
Iaprtc12	30%Sample	0.877 ± 0.009	0.875 ± 0.015	0.858 ± 0.011	0.868 ± 0.010	0.882 ± 0.006	0.829 ± 0.007	0.876 ± 0.005	0.897 ± 0.009
	50%Sample	0.878 ± 0.009	0.876 ± 0.007	0.860 ± 0.002	0.868 ± 0.004	0.883 ± 0.005	0.828 ± 0.004	0.875 ± 0.001	0.897 ± 0.013
	70%Sample	0.877 ± 0.005	0.876 ± 0.014	0.860 ± 0.014	0.868 ± 0.011	0.883 ± 0.015	0.828 ± 0.014	0.874 ± 0.002	0.897 ± 0.014
MirFlickr	30%Sample	0.868 ± 0.005	0.867 ± 0.002	0.858 ± 0.010	0.872 ± 0.014	0.871 ± 0.010	0.799 ± 0.002	0.874 ± 0.008	0.884 ± 0.003
	50%Sample	0.870 ± 0.014	0.867 ± 0.014	0.858 ± 0.012	0.872 ± 0.009	0.871 ± 0.010	0.799 ± 0.008	0.874 ± 0.004	0.885 ± 0.012
	70%Sample	0.868 ± 0.003	0.866 ± 0.009	0.860 ± 0.008	0.872 ± 0.014	0.871 ± 0.005	0.799 ± 0.008	0.875 ± 0.011	0.885 ± 0.007
Pascal07	30%Sample	0.822 ± 0.015	0.801 ± 0.007	0.800 ± 0.012	0.827 ± 0.007	0.824 ± 0.009	0.732 ± 0.009	0.830 ± 0.014	0.845 ± 0.008
	50%Sample	0.830 ± 0.014	0.800 ± 0.012	0.804 ± 0.013	0.828 ± 0.014	0.827 ± 0.003	0.731 ± 0.002	0.830 ± 0.009	0.845 ± 0.014
	70%Sample	0.827 ± 0.006	0.804 ± 0.010	0.807 ± 0.014	0.828 ± 0.008	0.830 ± 0.005	0.731 ± 0.010	0.830 ± 0.010	0.844 ± 0.005

TABLE VI

AUC UNDER DIFFERENT NOISE LEVELS ON FIFTEEN DATASETS

Dataset	Noise Level	AIM	DMVMLC	LMVCAT	MTD	RANK	iMvWL	VCMN	TIMMLIN
Corel5k	30%Sample	0.890 ± 0.003	0.874 ± 0.013	0.866 ± 0.011	0.881 ± 0.003	0.890 ± 0.010	0.863 ± 0.005	0.890 ± 0.014	0.918 ± 0.008
	50%Sample	0.890 ± 0.008	0.873 ± 0.006	0.866 ± 0.001	0.881 ± 0.014	0.889 ± 0.001	0.863 ± 0.004	0.890 ± 0.014	0.915 ± 0.006
	70%Sample	0.890 ± 0.002	0.872 ± 0.013	0.866 ± 0.001	0.881 ± 0.006	0.890 ± 0.015	0.863 ± 0.013	0.888 ± 0.009	0.916 ± 0.005
ESPGame	30%Sample	0.844 ± 0.008	0.846 ± 0.002	0.825 ± 0.003	0.836 ± 0.002	0.849 ± 0.004	0.810 ± 0.008	0.842 ± 0.007	0.862 ± 0.006
	50%Sample	0.844 ± 0.014	0.844 ± 0.010	0.823 ± 0.007	0.836 ± 0.012	0.848 ± 0.013	0.810 ± 0.010	0.842 ± 0.012	0.862 ± 0.015
	70%Sample	0.844 ± 0.005	0.844 ± 0.007	0.823 ± 0.004	0.836 ± 0.004	0.847 ± 0.013	0.810 ± 0.011	0.842 ± 0.009	0.861 ± 0.014
Iaprtc12	30%Sample	0.879 ± 0.014	0.874 ± 0.002	0.860 ± 0.009	0.870 ± 0.007	0.882 ± 0.004	0.831 ± 0.002	0.876 ± 0.002	0.897 ± 0.006
	50%Sample	0.879 ± 0.013	0.874 ± 0.007	0.860 ± 0.007	0.870 ± 0.012	0.883 ± 0.012	0.831 ± 0.012	0.875 ± 0.013	0.896 ± 0.004
	70%Sample	0.878 ± 0.012	0.876 ± 0.002	0.860 ± 0.014	0.870 ± 0.011	0.883 ± 0.005	0.831 ± 0.008	0.874 ± 0.003	0.868 ± 0.012
MirFlickr	30%Sample	0.856 ± 0.008	0.833 ± 0.007	0.845 ± 0.008	0.858 ± 0.010	0.857 ± 0.014	0.784 ± 0.004	0.862 ± 0.011	0.871 ± 0.008
	50%Sample	0.857 ± 0.011	0.837 ± 0.013	0.843 ± 0.011	0.858 ± 0.012	0.858 ± 0.012	0.784 ± 0.005	0.860 ± 0.013	0.869 ± 0.002
	70%Sample	0.858 ± 0.013	0.836 ± 0.011	0.849 ± 0.010	0.860 ± 0.010	0.858 ± 0.001	0.784 ± 0.001	0.860 ± 0.005	0.894 ± 0.003
Pascal07	30%Sample	0.844 ± 0.009	0.839 ± 0.009	0.825 ± 0.008	0.847 ± 0.004	0.846 ± 0.008	0.764 ± 0.005	0.851 ± 0.004	0.867 ± 0.007
	50%Sample	0.852 ± 0.008	0.840 ± 0.007	0.839 ± 0.006	0.848 ± 0.011	0.848 ± 0.008	0.763 ± 0.003	0.850 ± 0.005	0.866 ± 0.007
	70%Sample	0.850 ± 0.011	0.840 ± 0.013	0.840 ± 0.010	0.848 ± 0.008	0.848 ± 0.002	0.763 ± 0.003	0.850 ± 0.007	0.866 ± 0.008

TABLE VII

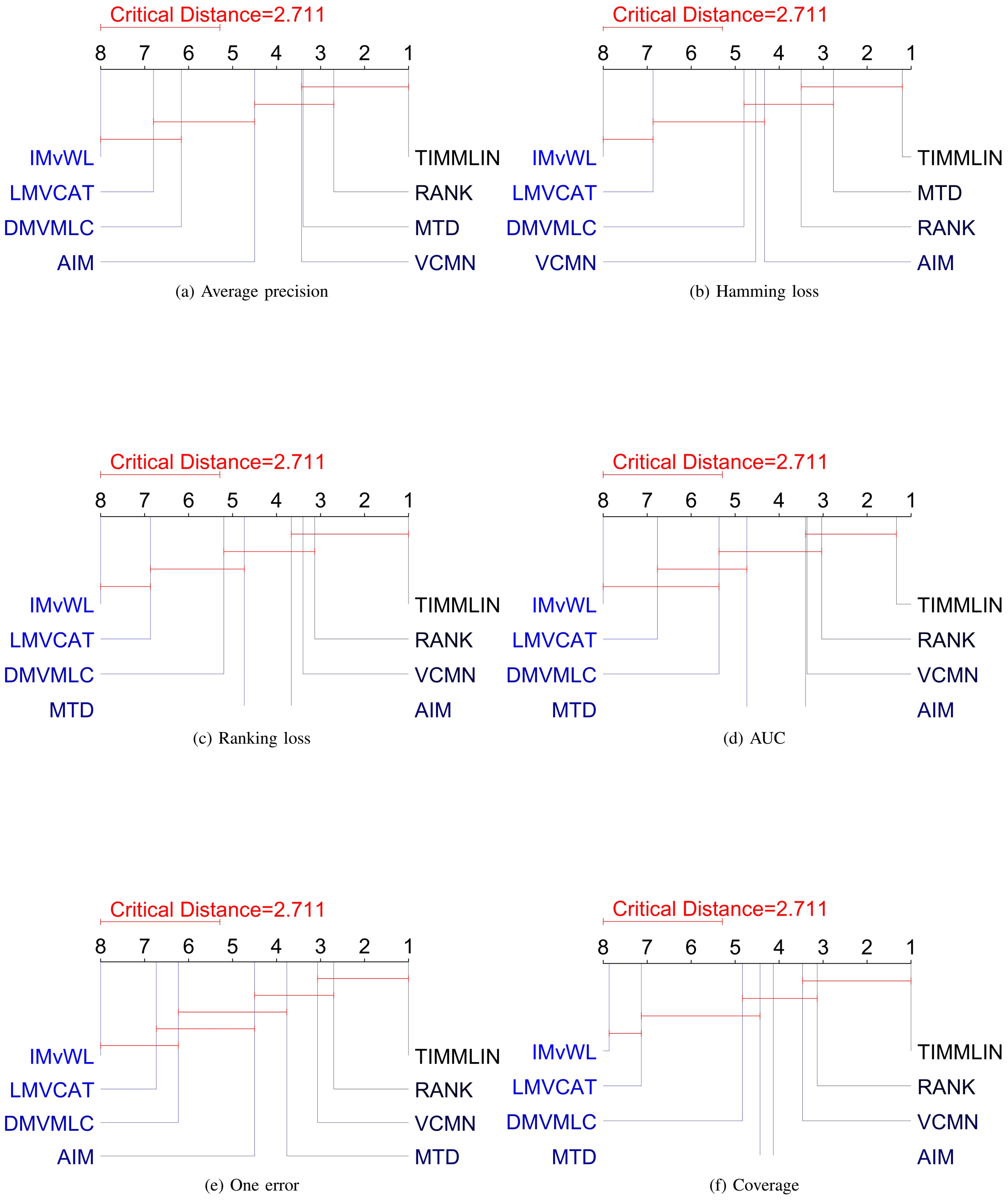
ONE ERROR UNDER DIFFERENT NOISE LEVELS ON FIFTEEN DATASETS

Dataset	Noise Level	AIM	DMVMLC	LMVCAT	MTD	RANK	iMvWL	VCMN	TIMMLIN
Corel5k	30%Sample	0.462 ± 0.007	0.398 ± 0.006	0.428 ± 0.006	0.473 ± 0.009	0.478 ± 0.013	0.308 ± 0.001	0.460 ± 0.015	0.501 ± 0.005
	50%Sample	0.460 ± 0.010	0.396 ± 0.001	0.428 ± 0.008	0.461 ± 0.012	0.472 ± 0.007	0.305 ± 0.003	0.463 ± 0.014	0.497 ± 0.004
	70%Sample	0.462 ± 0.014	0.398 ± 0.004	0.422 ± 0.005	0.467 ± 0.003	0.474 ± 0.005	0.302 ± 0.013	0.463 ± 0.004	0.495 ± 0.007
ESPGame	30%Sample	0.433 ± 0.010	0.420 ± 0.004	0.421 ± 0.012	0.436 ± 0.003	0.455 ± 0.011	0.340 ± 0.003	0.442 ± 0.014	0.468 ± 0.010
	50%Sample	0.428 ± 0.010	0.421 ± 0.008	0.414 ± 0.011	0.438 ± 0.006	0.451 ± 0.014	0.343 ± 0.012	0.449 ± 0.009	0.467 ± 0.012
	70%Sample	0.423 ± 0.010	0.416 ± 0.012	0.409 ± 0.013	0.433 ± 0.005	0.452 ± 0.005	0.345 ± 0.003	0.451 ± 0.002	0.467 ± 0.002
Iaprtc12	30%Sample	0.451 ± 0.010	0.449 ± 0.007	0.428 ± 0.004	0.455 ± 0.011	0.474 ± 0.003	0.355 ± 0.007	0.454 ± 0.001	0.497 ± 0.007
	50%Sample	0.449 ± 0.010	0.443 ± 0.001	0.419 ± 0.002	0.453 ± 0.004	0.475 ± 0.012	0.358 ± 0.002	0.455 ± 0.002	0.495 ± 0.002
	70%Sample	0.449 ± 0.010	0.449 ± 0.005	0.436 ± 0.015	0.456 ± 0.012	0.474 ± 0.001	0.353 ± 0.014	0.458 ± 0.007	0.497 ± 0.010
MirFlickr	30%Sample	0.633 ± 0.012	0.629 ± 0.013	0.628 ± 0.005	0.644 ± 0.013	0.645 ± 0.012	0.507 ± 0.007	0.656 ± 0.001	0.668 ± 0.007
	50%Sample	0.641 ± 0.009	0.627 ± 0.005	0.626 ± 0.006	0.647 ± 0.003	0.648 ± 0.006	0.505 ± 0.009	0.653 ± 0.002	0.667 ± 0.006
	70%Sample	0.636 ± 0.001	0.630 ± 0.003	0.625 ± 0.013	0.649 ± 0.011	0.649 ± 0.011	0.505 ± 0.012	0.651 ± 0.013	0.668 ± 0.015
Pascal07	30%Sample	0.452 ± 0.001	0.450 ± 0.012	0.411 ± 0.001	0.454 ± 0.002	0.451 ± 0.004	0.370 ± 0.006	0.457 ± 0.013	0.499 ± 0.002
	50%Sample	0.463 ± 0.005	0.445 ± 0.006	0.430 ± 0.002	0.454 ± 0.014	0.453 ± 0.011	0.373 ± 0.005	0.455 ± 0.012	0.496 ± 0.004
	70%Sample	0.462 ± 0.004	0.450 ± 0.010	0.421 ± 0.005	0.454 ± 0.009	0.460 ± 0.012	0.373 ± 0.015	0.455 ± 0.011	0.498 ± 0.014

TABLE VIII

COVERAGE UNDER DIFFERENT NOISE LEVELS ON FIFTEEN DATASETS

Dataset	Noise Level	AIM	DMVMLC	LMVCAT	MTD	RANK	iMvWL	VCMN	TIMMLIN
Corel5k	30%Sample	0.736 ± 0.007	0.703 ± 0.012	0.691 ± 0.008	0.719 ± 0.007	0.737 ± 0.014	0.693 ± 0.012	0.739 ± 0.006	0.803 ± 0.002
	50%Sample	0.737 ± 0.005	0.700 ± 0.015	0.692 ± 0.004	0.720 ± 0.014	0.735 ± 0.009	0.693 ± 0.004	0.737 ± 0.004	0.802 ± 0.014
	70%Sample	0.736 ± 0.003	0.699 ± 0.006	0.692 ± 0.014	0.719 ± 0.002	0.734 ± 0.012	0.688 ± 0.006	0.735 ± 0.010	0.802 ± 0.008
ESPGame	30%Sample	0.609 ± 0.009	0.619 ± 0.012	0.573 ± 0.010	0.586 ± 0.003	0.620 ± 0.002	0.541 ± 0.012	0.603 ± 0.008	0.648 ± 0.014
	50%Sample	0.610 ± 0.011	0.616 ± 0.010	0.573 ± 0.011	0.586 ± 0.014	0.622 ± 0.014	0.540 ± 0.003	0.601 ± 0.007	0.646 ± 0.009
	70%Sample	0.610 ± 0.004	0.619 ± 0.010	0.573 ± 0.005	0.587 ± 0.010	0.622 ± 0.012	0.543 ± 0.015	0.601 ± 0.014	0.645 ± 0.004
Iaprtc12	30%Sample	0.660 ± 0.014	0.661 ± 0.002	0.622 ± 0.012	0.630 ± 0.010	0.669 ± 0.011	0.555 ± 0.008	0.659 ± 0.006	0.705 ± 0.002
	50%Sample	0.659 ± 0.003	0.660 ± 0.010	0.621 ± 0.005	0.629 ± 0.008	0.668 ± 0.009	0.553 ± 0.014	0.658 ± 0.008	0.704 ± 0.006
	70%Sample	0.659 ± 0.014	0.659 ± 0.015	0.621 ± 0.009	0.629 ± 0.007	0.670 ± 0.012	0.550 ± 0.013	0.657 ± 0.012	0.702 ± 0.009
MirFlickr	30%Sample	0.666 ± 0.009	0.663 ± 0.005	0.628 ± 0.013	0.673 ± 0.003	0.667 ± 0.005	0.578 ± 0.001	0.677 ± 0.002	0.696 ± 0.007
	50%Sample	0.665 ± 0.012	0.661 ± 0.001	0.628 ± 0.003	0.678 ± 0.015	0.668 ± 0.001	0.576 ± 0.010	0.675 ± 0.005	0.696 ± 0.001
	70%Sample	0.663 ± 0.013	0.661 ± 0.009	0.627 ± 0.002	0.675 ± 0.013	0.666 ± 0.004	0.573 ± 0.011	0.676 ± 0.013	0.695 ± 0.006
Pascal07	30%Sample	0.775 ± 0.005	0.769 ± 0.013	0.753 ± 0.002	0.779 ± 0.004	0.776 ± 0.014	0.654 ± 0.012	0.782 ± 0.009	0.843 ± 0.004
	50%Sample	0.773 ± 0.010	0.768 ± 0.008	0.751 ± 0.005	0.780 ± 0.004	0.776 ± 0.014	0.655 ± 0.010	0.780 ± 0.006	0.841 ± 0.013
	70%Sample	0.771 ± 0.010	0.768 ± 0.013	0.751 ± 0.006	0.780 ± 0.006	0.775 ± 0.006	0.652 ± 0.015	0.779 ± 0.002	0.840 ± 0.006



(a) Average precision

(b) Hamming loss

(c) Ranking loss

(d) AUC

(e) One error

(f) Coverage

Fig. 2. CD Values on Fifteen Benchmark Datasets under Different Metrics.

TABLE IX
WE CONDUCTED ABLATION EXPERIMENTS ON THE COREL5K DATASET USING THE PROPOSED TIMMLIN MODEL, UNDER THE SETTING WHERE 50% OF THE DATA IS INCOMPLETE AND THREE POSITIVE LABELS ARE ADDED TO 50% OF THE SAMPLES.

Method	Average precision	Hamming loss	Ranking loss	AUC	One error	Coverage
$Backbone$	0.369	0.985	0.871	0.874	0.398	0.706
$Backbone + L_{CL}$	0.409	0.987	0.898	0.897	0.478	0.780
$Backbone + L_{ID}$	0.388	0.987	0.888	0.890	0.462	0.736
$Backbone + L_{CL} + L_{ID}$	0.429	0.987	0.904	0.906	0.53	0.781
$Backbone + L_{CL} + L_{ID} + L_{graph}$	0.435	0.988	0.913	0.918	0.501	0.803

TABLE X
AN OVERVIEW OF THE FRIEDMAN STATISTIC $F_F(K = 8, N = 15)$ ALONG WITH ITS CRITICAL VALUE BASED ON SIX ASSESSMENT METRICS.

Evaluation Metric	F_F	Critical Value (0.05)
Average precision	31.7962963	2.104448
Hamming Loss	27.387037	
Ranking Loss	28.688887	
AUC	27.388889	
One Error	31.937037	
Coverage	28.087037	

A. Experimental Settings

1) Dataset: We employed five classical multi-view multi-label (MVML) datasets to validate our model: Corel5k [19], ESPGame [20], Iaprtc12 [21], Pascal07 [22], and MirFlickr [23]. These datasets encompass six distinct views or feature representations: GIST, HSV, Hue, SIFT, RGB, and LAB, with additional details summarized in Table II**.** To simulate real-world scenarios involving incomplete data and label noise, we randomly removed 50% of the data in each view while ensuring that each sample retained at least one available view. Subsequently, 70% of the samples were randomly selected as the training set. Within the training data, 50% of positive labels and 50% of negative labels in each class were randomly masked or deleted to simulate label incompleteness. Finally, 30%, 50%, and 70% of the training samples were randomly selected and injected with three incorrect positive labels to simulate label noise.

2) Evaluation Indicators: In our experiments, we evaluated model performance using six widely adopted multi-label classification metrics: Average Precision (AP), Hamming Loss (HL), Ranking Loss (RL), adapted Area Under the Curve (AUC), OneError (OE), and Coverage (COV). To facilitate interpretation, we transformed the original OE and COV by using 1-OE and 1-COV, respectively, so that higher values consistently reflect superior model performance.

3) Implementation Details: All experiments were conducted using PyTorch 1.10.1 and Python 3.9.18. The computations were performed on a system equipped with an NVIDIA RTX 3090 GPU and an Intel i7-12900K processor. The model was trained using the SGD optimizer with an initial learning rate of 0.1. Across all fifteen datasets, a consistent batch size of 128 and a momentum coefficient of 0.9 were employed throughout the experiments.

B. Comparison Methods

To evaluate the effectiveness of our proposed model, we selected seven representative comparison methods for the experiments: AIM [24], DMVMLC [25], LMVCAT [26], MTD, iMvWL, RANK, and VCMN [27]. The proposed TIMMLIN model was compared with these baseline methods across fifteen IDLCLN datasets.

At present, multi-view multi-label classification approaches capable of addressing both incomplete data and label noise simultaneously remain limited. Consequently, we incorporated several relevant multi-label classification methods as baseline models in our experimental evaluation. Among these, AIM, DMVMLC, LMVCAT, MTD, and VCMN are all applicable to IDLCLN tasks. Although iMvWL is a multi-view multi-label classification method, it is not designed to handle incomplete data in any form. To ensure a fair comparison, we applied a simple imputation strategy during evaluation: missing views were filled using the mean of the available views, and unknown labels were replaced with ‘0’.

C. Experimental Results and Analysis

To validate the effectiveness of our proposed method under conditions of incomplete data and label noise, we conducted comparative experiments against eight state-of-the-art algorithms across fifteen distinct datasets. Table III-VIII present the mean and standard deviation of various evaluation metrics for each method.

In addition to the six evaluation metrics previously introduced, we further employed the Nemenyi test to perform pairwise comparisons among the algorithms. The average rank differences are evaluated based on the Critical Difference (CD), calculated as:

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6n}}$$

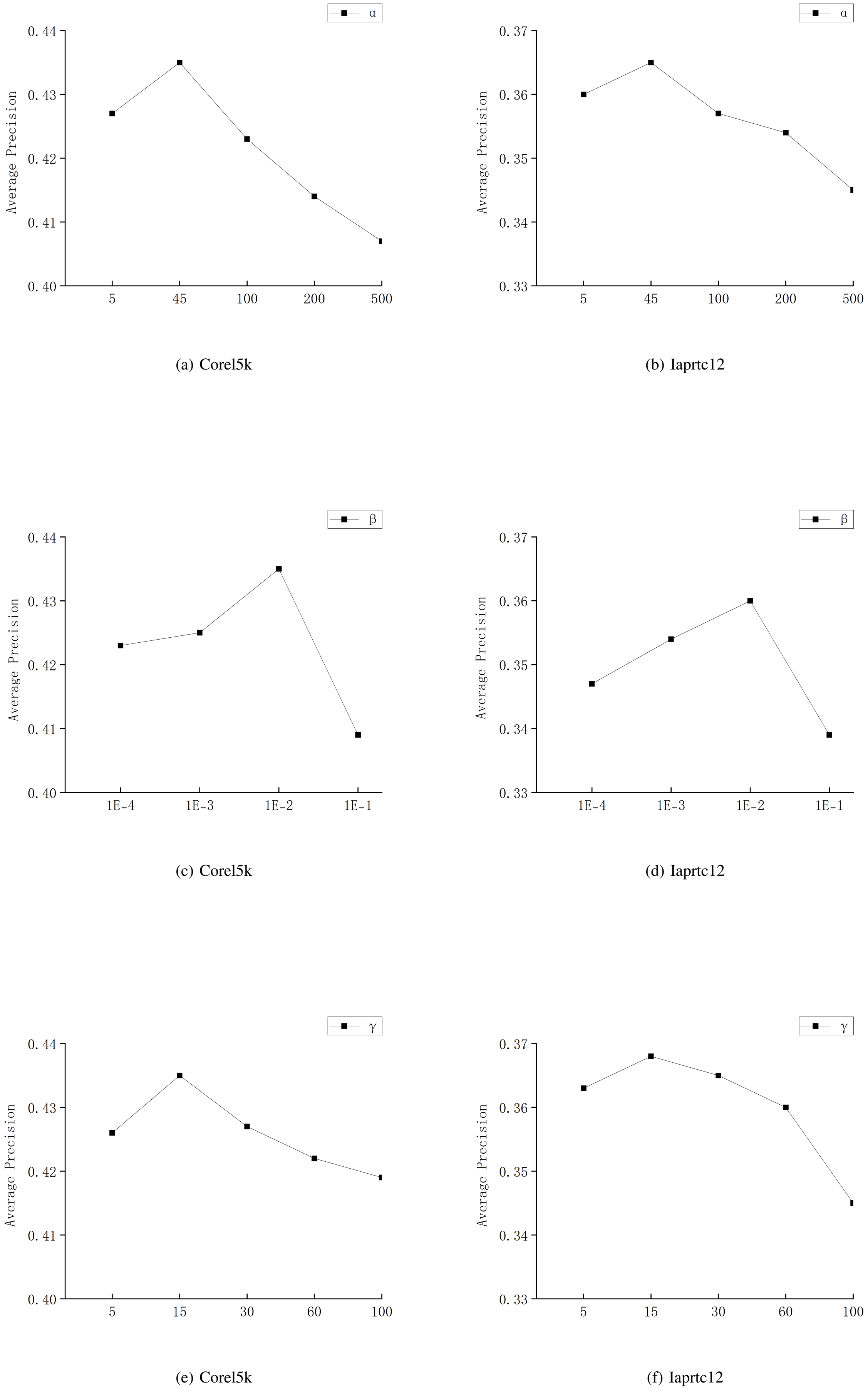
where k denotes the number of algorithms, n represents the number of datasets, and q_α is the critical value. For the Nemenyi test conducted at a significance level of $\alpha = 0.05$, the critical value is $q_\alpha = 2.949$, resulting in a critical difference of $CD = 2.711$ ($k = 8$, $n = 15$).

If the average rank difference between two algorithms falls below the CD threshold, there is no statistically significant difference in their classification performance; otherwise, a statistically significant difference exists. The CD diagrams for each evaluation metric are illustrated in Fig. 2. In each subfigure, a line connecting two algorithms indicates that their average rank difference lies within the CD threshold.

As shown in Fig. 2, TIMMLIN demonstrates superior performance compared to all other algorithms in terms of AP, 1-HL, 1-RL, AUC, 1-OE, and 1-COV.

Based on the results presented in Table III-VIII and Fig. 2, the following conclusions can be drawn:

1. Compared with other methods, our proposed information-theoretic approach achieves superior



(a) Corel5k

(b) Iaprtc12

(c) Corel5k

(d) Iaprtc12

(e) Corel5k

(f) Iaprtc12

Fig. 3. The AP values of α, β, and γ on the Corel5K and Iaprtc12 datasets.

performance across all datasets and ranks first in all evaluation metrics. This highlights the significant advantage of applying information theory to address the IDLCLN task.

2. Across the fifteen datasets, the MTD, AIM, LMVCAT, VCMN, RANK, and DMVMLC models consistently outperform the iMvWL method. This superiority can be primarily attributed to the strong feature extraction capabilities offered by neural networks.

3. Among deep learning-based approaches, our model—specifically designed to address both incomplete data and noisy labels—demonstrates clear advantages in the majority of cases. This further underscores the importance of explicitly considering both data incompleteness and label noise during the model design phase, as it significantly contributes to improved overall performance.

D. Ablation Study

To validate the effectiveness of our proposed method, we conducted ablation experiments on the Corel5k dataset under the following settings: only 50% of instances were retained for each view, 50% of the labels were missing, three positive labels were assigned to 50% of the samples, and 70% of the data was used for training. We evaluated the contribution of each loss component by selectively removing parts of the loss function, namely $\mathcal{L}_{ID}^{(k)}$, $\mathcal{L}_{graph}$, and $\mathcal{L}_{CL}$. The results, along with the baseline using only the cross-entropy loss $Backbone$, are summarized in Table IX.

Our main findings are as follows: (i) The inclusion of each additional loss component consistently improved the performance metrics; (ii) Among these components, $\mathcal{L}_{CL}$ yielded the most substantial improvement, underscoring its critical role in effective feature extraction.

It is worth noting that our two-stage framework could, in theory, be simplified into a one-stage model by applying a shared parameter to each $\mathcal{L}_{ID}^{(k)}$ in the first stage and incorporating them into the second-stage loss $\mathcal{L}_{SCL}$. However, as shown in Table IX, this simplification results in a noticeable decline in performance.

We provide the following explanation: the use of a shared parameter causes the model to disproportionately focus on optimizing the view with the largest $\mathcal{L}_{ID}^{(k)}$, thereby leading to insufficient information extraction from the remaining views. For instance, if $\mathcal{L}_{ID}^{(1)}$ is significantly larger than the others, the model tends to prioritize optimizing this term, at the expense of contributions from other views. Although assigning separate parameters to each view could mitigate this issue, it would considerably complicate the parameter tuning process.

E. Parameter Analysis

In our proposed two-stage model, three parameters α, β, and γ must be specified prior to training. To investigate the model's sensitivity to these parameters, we conducted experiments on the Corel5k and Iaprtc12 datasets. In these experiments, 50% of the data in each view were retained, 50% of the labels were randomly removed, three positive labels were assigned to 50% of the samples, and 70% of the samples were used for training. Fig. 3 presents the AP performance of the model under different settings of α, β, and γ.

To ensure the reliability of the results, all other parameters were held constant while one parameter was varied. The experimental results demonstrate that the model maintains consistently strong performance when α is within the range [1, 100], with peak performance observed at $\alpha = 15$. For γ, the model exhibits stable behavior across the range [5, 500], achieving the best result at $\gamma = 45$. Regarding β, the model also shows robustness within the range [5, 100], with optimal performance achieved when β is set to 1×10^{-2}.

V. Conclusion

In this paper, we propose a novel theory-driven two-stage information extraction framework to address the IDLCLN problem. Unlike existing approaches, our method is grounded in information-driven theory, with the objective of effectively modeling and extracting informative signals from incomplete and noisy data. Extensive experimental evaluations on benchmark datasets demonstrate that our method consistently outperforms state-of-the-art techniques, thereby validating the feasibility and potential of leveraging information-theoretic principles for tackling the IDLCLN task.

Despite the notable progress achieved, several important research directions remain unexplored. Future work may focus on developing more efficient strategies for incomplete data recovery, designing more adaptive weighting mechanisms to better handle complex scenarios involving incomplete views and label noise, or incorporating high-order label correlation modeling to further improve classification performance. Additionally, exploring more effective utilization of unlabeled or mislabeled data, as well as evaluating model generalization on large-scale datasets, are also promising areas for further investigation.

In conclusion, this work presents a novel perspective and technical solution to the IDLCLN problem. It not only advances the theoretical understanding at the intersection of multi-view learning and multi-label classification but also provides practical support for real-world applications.

References

[1] Qing Ai, Fei Li, Qingyun Gao, and Fei Zhao. An improved lst-ksvc based on energy model. *Engineering Letters*, 30(4):1436–1443, 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] Pingting Hao, Weiping Ding, Wanfu Gao, and Jialong He. Exploring view-specific label relationships for multi-view multi-label feature selection. *Information Sciences*, 681:121215, 2024.

[4] Changming Zhu and Liju Han. A multi-view multi-label learning with incomplete data and self-adaptive correlations. *International Journal of Machine Learning and Cybernetics*, pages 1–19, 2025.

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

[6] Xiang Li and Songcan Chen. A concise yet effective model for non-aligned incomplete multi-view and missing multi-label learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):5918–5932, 2021.

[7] Jiarui Chen, Wulin Xie, Mengqing Wang, Yinghao Ye, and Xiaohuan Lu. Confidence-enhanced dual-space semantic alignment for partial multi-view incomplete multi-label classification. *Knowledge-Based Systems*, page 113507, 2025.

[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] Chengliang Liu, Jie Wen, Yong Xu, Bob Zhang, Liqiang Nie, and Min Zhang. Reliable representation learning for incomplete multi-view missing multi-label classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 4940–4956, 2025.

[10] Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Dual contrastive pretion for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4447–4461, 2022.

[11] Zhibin Dong, Siwei Wang, Jiaqi Jin, Xinwang Liu, and En Zhu. Cross-view topology based consistent and complementary information for deep multi-view clustering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19440–19451, 2023.

[12] 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.

[13] Qi Wang, Claire Boudreau, Qixing Luo, Pang-Ning Tan, and Jiayu Zhou. Deep multi-view information bottleneck. In *Proceedings of the 2019 SIAM International Conference on Data Mining*, pages 37–45. SIAM, 2019.

[14] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[15] Shikun Li, Xiaobo Xia, Shiming Ge, and Tongliang Liu. Selective-supervised contrastive learning with noisy labels. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 316–325, 2022.

[16] Hwanjun Song, Minseok Kim, Dongmin Park, Yooju Shin, and Jae-Gil Lee. Learning from noisy labels with deep neural networks: A survey. *IEEE transactions on neural networks and learning systems*, 34(11):8135–8153, 2022.

[17] Kenji Kawaguchi, Zhun Deng, Xu Ji, and Jiaoyang Huang. How does information bottleneck help deep learning? In *International Conference on Machine Learning*, pages 16049–16096. PMLR, 2023.

[18] Artemy Kolchinsky, Brendan D Tracey, and David H Wolpert. Nonlinear information bottleneck. *Entropy*, 21(12):1181, 2019.

[19] Pinar Duygulu, Kobus Barnard, Joao FG de Freitas, and David A Forsyth. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In *Computer vision—ECCV 2002: 7th European Conference on Computer Vision Copenhagen, Denmark, May 28–31, 2002 Proceedings, Part IV 7*, pages 97–112. Springer, 2002.

[20] Luis Von Ahn and Laura Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 319–326, 2004.

[21] Michael Grubinger, Paul Clough, Henning Müller, and Thomas Deselaers. The iapr tc-12 benchmark: A new evaluation resource for visual information systems. In *International Workshop ontoImage*, volume 2, 2006.

[22] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88:303–338, 2010.

[23] Mark J Huiskes and Michael S Lew. The mir flickr retrieval evaluation. In *Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval*, pages 39–43, 2008.

[24] 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.

[25] 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.

[26] 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.

[27] 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.