COUPLING ERROR EXPLORATION FOR INCOMPLETE MULTI-VIEW MISSING MULTI-LABEL CLASSIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Incomplete Multi-view Missing Multi-label Classification (IM3C) aims to predict potentially relevant labels for each sample by exploiting the consensus and complementary information across available views and labels. Although existing IM3C methods have made notable progress, the missingness of some views makes it difficult for view recovery or feature fusion to fully preserve the view-specific distributional structure, leading to the Representation Bias Problem (RBP). Furthermore, the RBP reduces the reliability of label predictions, while incomplete labels provide insufficient or even erroneous supervision, thereby interfering with representation learning and continuously amplifying prediction errors, *i.e.*, Coupling Error Propagation (CEP). To address these challenges, we propose Coupling Error Exploration (CEE) for incomplete multi-view missing multi-label classification, aiming to achieve robust and accurate label prediction by exploiting sample-level and label-level structural information. Specifically, to alleviate RBP, we first recover the missing views and ensure the view-specific distribution structure through multi-view rectified flow learning. Furthermore, we propose multi-view structure reinforcement that utilizes multi-view contrastive learning to improve semantic consistency across views. Finally, to mitigate CEP, we design collaborative dual-graph structure learning, which enhances sample representations by leveraging sample graph relationships and explores label semantic correlations through label graph relationships. Through dual-graph collaboration, we can uncover comprehensive relational information both between samples and between views. Theoretical analyses and extensive experimental results demonstrate that CEE outperforms state-of-the-art methods.

1 INTRODUCTION

Multi-view data originates from multiple perspectives or different sensors, while the same object often possesses diverse attributes and multiple labels. Data possessing these characteristics is referred to as multi-view multi-label data (Liu et al., 2026a). To handle such data, researchers have proposed the Multi-view Multi-label Learning (MvMlL) method (Hao et al., 2025b; Tu et al., 2026), which aims to explore the consistency and complementarity across views as well as the correlations among labels, thereby learning discriminative sample representations. Applications of this learning paradigm span image annotation (Zhang et al., 2020), multimedia retrieval (Sanghavi & Verma, 2022), and medical diagnosis (Shen et al., 2020; Yang et al., 2023). In recent years, MvMlL has attracted increasing research attention and achieved considerable progress.

Normally, conventional MvMlL methods generally assume that all views and labels are fully collected. However, in practice, equipment or sensor malfunctions often result in missing views (Yuan et al., 2025; 2026). Meanwhile, high annotation costs and annotation errors or omissions may leave sample labels incomplete. This results in collected samples typically exhibiting both missing views and incomplete labels. To address these challenges, Incomplete Multi-view Missing Multi-label Classification (IM3C) (Tan et al., 2018; Xie et al., 2025) was developed. Existing IM3C methods fall into two categories: imputation-based and non-imputation-based methods. The former reconstruct missing views to recover complete multi-view representations. Such as, AIMNet (Liu et al., 2024a) alleviates the feature incompleteness problem through attention-based interpolation, while

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

TACVI-Net (Zhao et al., 2025b) reconstructs missing views based on cross-view shared semantics. Non-imputation-based methods directly utilize available views for representation learning. For example, RANK (Liu et al., 2025) reduces the impact of double missingness by mining and aligning shared semantics in available views, while QUALITY (Liu et al., 2026c) improves representation quality by extracting and fusing shared and private information from views.

Although these methods have achieved promising results, several limitations remain. For example, imputation-based methods generally rely on local sample statistics or cross-view shared semantics to impute missing views but struggle to recover view-specific distributional structures. In contrast, non-imputation-based methods directly fuse available views. When these views contain noise or insufficient semantic information, the resulting representation suffers from distributional shifts and instability. These characteristics make it difficult for multi-view data to guarantee a specific distribution structure and stable quality, *i.e.*, Representation Bias Problem (RBP). Furthermore, RBP also reduces label-prediction accuracy, and incomplete labels will provide insufficient or even incorrect semantic guidance for representation learning, thereby impairing model performance, *i.e.*, Coupling Error Propagation (CEP). Therefore, mitigating representation bias and coupled error propagation under incomplete views and labels is key for robust and accurate label prediction.

To address these issues, we propose Coupling Error Exploration (CEE) for incomplete multi-view missing multi-label classification, which aims to mitigate coupled error propagation under double-missing conditions. Specifically, CEE is trained in two stages: generator pretraining and joint training. During generator pretraining, we design multi-view specific distribution learning, which enables the recovery of reliable representations, thereby preserving view-distribution characteristics and quality stability. During joint training, multi-view structure reinforcement utilizes multi-view contrastive learning to enhance consensus representations among views. To suppress error propagation along instance relationships and label dependencies, we propose collaborative dual-graph structure learning, which enhances sample representation by a sample graph and uses the label graph to supplement label semantics. Through dual-graph collaboration, our CEE can achieve robust and stable relationship learning. In general, our major contributions are summarized as follows:

- We propose a novel Coupling Error Exploration (CEE) for incomplete multi-view missing multi-label classification, which aims to mitigate the accumulation and propagation of errors induced by incomplete views and missing labels during training, thereby enhancing the robustness and stability of the model.
- To address RBP, we propose Multi-view Specific Distribution Learning (MSDL), which recovers view representations by learning a conditional velocity model from source views to target views in the latent space and fusing generation results from multiple sources. Consequently, MSDL enhances the stability of recovered results while preserving the structural characteristics of the view distribution.
- To address CEP, we propose Collaborative Dual-graph Structure Learning (CDSL), in which information-compressed feature graph learning aggregates by compressing neighborhood information to refine instance representations, while missing-aware label graph learning aims to explore and capture the underlying inter-label dependencies under incomplete labels. Consequently, CDSL alleviates unreliable information propagation across sample relationships and label dependencies.
- We verified the robustness and stability of the proposed CEE through both theoretical and experimental methods. Extensive experiments demonstrate that CEE achieves significant results compared to state-of-the-art methods.

2 METHOD

2.1 NOTATION AND OVERVIEW

Given an incomplete multi-view dataset $\{\boldsymbol{X}^{(v)}\}_{v=1}^{m}$ comprising n samples and m views, where $\boldsymbol{X}^{(v)} \in \mathbb{R}^{n \times d_v}$ denotes the feature matrix of the v-th view and d_v is its feature dimensionality. The corresponding label matrix is denoted by $\boldsymbol{Y} \in \{0,1\}^{n \times c}$, where c is the number of labels. To characterize the missing-view pattern, we introduce a view indicator matrix $\boldsymbol{W} \in \{0,1\}^{n \times m}$, where $W_i^v = 1$ indicates that the v-th view of sample i is available, whereas $W_i^v = 0$ indicates that it is

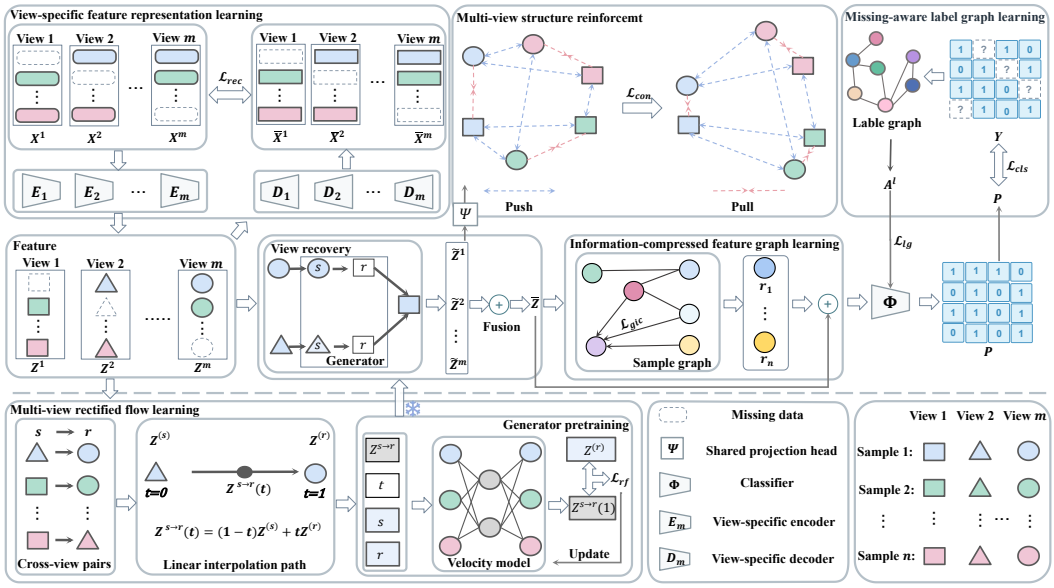


Figure 1: The overall framework of the proposed CEE. The bottom section represents the generator pretraining stage, which consists of multi-view rectified flow learning. The top section represents the joint training stage, which comprises four main modules: view-specific feature representation learning, multi-view structure reinforcement, information-compressed feature graph learning, and missing-aware label graph learning.

missing. Similarly, the label indicator matrix $\boldsymbol{G} \in \{0,1\}^{n\times c}$ describes label availability: $G_{i,j} = 1$ indicates that the j-th label of sample i is available, and a value of zero indicates that it is missing. The set of available views for sample i is denoted by $\mathcal{O}_i = \{v \mid W_i^v = 1\}$.

As illustrated in Fig. 1, our CEE comprises two stages: generator pretraining and joint training. During generator pretraining, we construct ordered training pairs from views jointly observed for the same sample and optimize the conditional velocity model using the rectified-flow loss. Accordingly, the objective for this stage is:

$$\mathcal{L}_1 = \mathcal{L}_{\mathrm{rf}}, \tag{1}$$

where $\mathcal{L}_{\mathrm{rf}}$ represents the rectified-flow loss. After pretraining, we freeze the generator and use it to generate missing-view representations during joint training.

During joint training, view-specific autoencoders produce latent representations, and the frozen generator completes missing views. Multi-view structure reinforcement promotes agreement across the completed views. Subsequently, information-compressed feature graph learning and missing-aware label graph learning explore structural information between the sample and between the label spaces. In summary, the objective loss of CEE in joint training is formulated as follows:

$$\mathcal{L}_2 = \mathcal{L}_{\mathrm{rec}} + \mathcal{L}_{\mathrm{cl}} + \mathcal{L}_{\mathrm{gic}} + \mathcal{L}_{\mathrm{lg}} + \mathcal{L}_{\mathrm{cls}}, \tag{2}$$

where $\mathcal{L}_{\mathrm{rec}}$ represents the reconstruction loss, $\mathcal{L}_{\mathrm{cl}}$ represents the cross-view contrastive loss, $\mathcal{L}_{\mathrm{gic}}$ represents the graph information compression loss, $\mathcal{L}_{\mathrm{lg}}$ represents the label-graph loss, $\mathcal{L}_{\mathrm{cls}}$ represents the binary cross-entropy loss.

2.2 Multi-view specific distribution learning

2.2.1 View-specific feature representation learning

Multi-view data exhibits heterogeneous feature dimensions and the specificity and diversity of view-specific distributions. Therefore, to remove redundant information while retaining view-specific information, we employ a view-specific autoencoder to obtain latent representations with a common dimensionality. Specifically, each view has its own independent encoder-decoder: the encoder extracts a view-specific representation, and the decoder reconstructs the original input from

that representation. For the v-th view, the encoding and decoding processes are expressed as $\boldsymbol{Z}^{(v)} = E_v(\boldsymbol{X}^{(v)}; \theta_v)$ and $\widehat{\boldsymbol{X}}^{(v)} = D_v(\boldsymbol{Z}^{(v)}; \phi_v)$. Here, $\boldsymbol{Z}^{(v)} \in \mathbb{R}^{n \times d_z}$ is the view-specific latent representation, where d_z denotes the dimensionality of the latent space. E_v and D_v denote the encoder and decoder of the v-th view, while θ_v and ϕ_v are the learnable parameters of E_v and D_v, respectively. We employ the reconstruction loss to minimize reconstruction error while introducing a missing-view matrix to mitigate the impact of missing views:

$$\mathcal{L}_{\text{rec}} = \sum_{v=1}^{m} \frac{1}{n d_v} \sum_{i=1}^{n} W_i^v \left\| \boldsymbol{X}_i^{(v)} - \widehat{\boldsymbol{X}}_i^{(v)} \right\|_2^2, \tag{3}$$

where $\widehat{\boldsymbol{X}}_i^{(v)}$ is the reconstructed feature of sample i in the v-th view.

2.2.2 Multi-view rectified flow learning

In the collected data, missing views result in the model lacking sufficient semantic information, leading to distribution shift and instability, *i.e.*, Representation Bias Problem (RBP). To address this issue, we propose multi-view rectified flow learning, which learns cross-view velocity fields in the latent space and aggregates generation results from multiple source views to recover each missing-view representation. Specifically, let $\boldsymbol{Z}^{(s)}$ and $\boldsymbol{Z}^{(r)}$ represent cross-view pairs that are observed simultaneously in the same sample. We connect them through a linear interpolation path and define the corresponding target velocity as the time derivative of this path:$\mathbf{Z}^{s\to r}(t) = (1-t)\boldsymbol{Z}^{(s)} + t\boldsymbol{Z}^{(r)}$, $\frac{d\mathbf{Z}^{s\to r}(t)}{dt} = \boldsymbol{Z}^{(r)} - \boldsymbol{Z}^{(s)}$. Here, $t \sim \text{Uniform}([0,1])$ is a continuous time variable, and $\mathbf{Z}^{s\to r}(t)$ denotes the state on the linear path at time t when the representation of transmission path from $\boldsymbol{Z}^{(s)}$ to $\boldsymbol{Z}^{(r)}$. After this, we parameterize a conditional velocity model $v_\omega(\mathbf{Z}^{s\to r}(t), t, s, r)$ by concatenating $\mathbf{Z}^{s\to r}(t)$ with the time t and the identity embeddings of the source and target views. Its velocity-matching objective is defined as follows:

$$\mathcal{L}_{\text{rf}} = \frac{1}{|\Omega| d_z} \sum_{(i,s,r)\in\Omega} \left\| v_\omega\left(\mathbf{Z}_i^{s\to r}(t), t, s, r\right) - \left(\boldsymbol{Z}_i^{(r)} - \boldsymbol{Z}_i^{(s)}\right) \right\|_2^2, \tag{4}$$

where $\Omega = \{(i,s,r) \mid s \neq r,\, W_i^s W_i^r = 1\}$ is the training set composed of ordered cross-view pairs which are observed in the same sample; $|\Omega|$ is the number of valid pairs; ω denotes the learnable parameters of the conditional velocity model; $\mathbf{Z}_i^{s\to r}(t)$ is the transmission path of the i-th sample. During generator pretraining, all other network parameters are fixed and only the conditional velocity model is optimized. During joint training, the velocity model is frozen. For a missing target view, we take the representation of any observed source view as the initial state and solve the following ordinary differential equation along the learned velocity field v_ω:

$$\frac{d\widehat{\mathbf{Z}}^{s\to r}(t)}{dt} = v_\omega\left(\widehat{\mathbf{Z}}^{s\to r}(t), t, s, r\right), \qquad \widehat{\mathbf{Z}}^{s\to r}(0) = \boldsymbol{Z}^{(s)}, \tag{5}$$

where $\widehat{\mathbf{Z}}^{s\to r}(t)$ denotes the generation trajectory induced by the learned velocity field. Note that $\mathbf{Z}^{s\to r}(t)$ during generator pretraining is a manually constructed linear supervision path, whereas $\widehat{\mathbf{Z}}^{s\to r}(t)$ is the trajectory obtained by integrating the learned velocity field during joint training. We discretize this process using the explicit Euler method and define its terminal state as $\Phi_\omega^{s\to r}(\boldsymbol{Z}^{(s)}) := \widehat{\mathbf{Z}}^{s\to r}(1) \approx \boldsymbol{Z}^{(r)}$. For each target view, we retain its representation if observed and recover it by averaging the representations generated from all observed source views:

$$\widetilde{\boldsymbol{Z}}_i^{(r)} = \begin{cases} \boldsymbol{Z}_i^{(r)}, & W_{ir} = 1, \\ \dfrac{1}{|\mathcal{O}_i|} \displaystyle\sum_{s\in\mathcal{O}_i} \Phi_\omega^{s\to r}\left(\boldsymbol{Z}_i^{(s)}\right), & W_{ir} = 0. \end{cases} \tag{6}$$

2.3 Multi-view structure reinforcement

Complete latent view representations do not necessarily guarantee semantic consistency across views. We therefore apply multi-view contrastive learning to the completed representations to reinforce their semantic agreement. Specifically, we use a shared MLP to map each completed view

representation $\widetilde{\boldsymbol{Z}}^{(v)}$ into a low-dimensional space, obtaining the projected representation $\boldsymbol{H}^{(v)}$. In the low-dimensional space, we measure the relation between view representations of different samples using cosine similarity:

$$S_{i,j}^{(v,u)} = \frac{\left(\boldsymbol{H}_i^{(v)}\right)^{\mathsf{T}} \boldsymbol{H}_j^{(u)}}{\left\|\boldsymbol{H}_i^{(v)}\right\|_2 \left\|\boldsymbol{H}_j^{(u)}\right\|_2}, \tag{7}$$

where $S_{i,j}^{(v,u)}$ is the cosine similarity between the representation of sample i in view v and that of sample j in view u. $\boldsymbol{H}_i^{(v)}$ represents the projected representation of the v-th view of the i-th sample in the contrastive space. For an anchor $\boldsymbol{H}_i^{(v)}$, the representations of the same sample in the remaining $m-1$ views are positives, whereas those of all other samples across all views are negatives. Based on this, the cross-view contrastive loss is defined as follows:

$$\mathcal{L}_{\mathrm{cl}} = -\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m(m-1)}\sum_{v=1}^{m}\sum_{\substack{u=1\\u\neq v}}^{m}\log\frac{\exp\left(S_{i,i}^{(v,u)}/\tau\right)}{\mathbb{S}_{\mathrm{pos}}^{(i,v)}+\mathbb{S}_{\mathrm{neg}}^{(i,v)}}, \tag{8}$$

where $\mathbb{S}_{\mathrm{pos}}^{(i,v)} = \sum_{\substack{r=1\\r\neq v}}^{m}\exp\left(S_{i,i}^{(v,r)}/\tau\right)$, $\mathbb{S}_{\mathrm{neg}}^{(i,v)} = \sum_{\substack{j=1\\j\neq i}}^{n}\sum_{r=1}^{m}\exp\left(S_{i,j}^{(v,r)}/\tau\right)$, and τ is the temperature parameter, set to 0.2.

2.4 Collaborative dual-graph structure learning

2.4.1 Information-compressed feature graph learning

Missing-view recovery enhances the completeness of intra-sample view representations, but inter-sample representation errors persist, which leads to the propagation of unreliable relationships between samples. To address this, we propose information-compressed feature graph learning to construct sample-level graph relationships and aggregate useful neighbor information, thereby limiting the propagation of coupled errors.

First, we perform equal-weight fusion over all views to obtain the sample-level representation $\bar{\boldsymbol{Z}}_i = \frac{1}{m}\sum_{v=1}^{m}\widetilde{\boldsymbol{Z}}_i^{(v)}$ for the i-th sample. Based on these fused representations, we connect each sample to its nearest neighbors according to the cosine similarity $S_{i,j}$, obtaining the neighborhood set $\mathcal{N}_i$. We then project $\bar{\boldsymbol{Z}}_i$ into query, key, and value to obtain $\boldsymbol{q}_i$, $\boldsymbol{k}_i$, and $\boldsymbol{v}_i$. To enhance the sample representation with information from related samples, we use attention to aggregate neighboring representations:

$$\alpha_{i,j} = \frac{\exp\left(\boldsymbol{q}_i^{\mathsf{T}}\boldsymbol{k}_j/\sqrt{d_z}\right)}{\sum_{l\in\mathcal{N}_i}\exp\left(\boldsymbol{q}_i^{\mathsf{T}}\boldsymbol{k}_l/\sqrt{d_z}\right)}, \qquad \boldsymbol{f}_i = \mathrm{ELU}\left(\sum_{j\in\mathcal{N}_i}\alpha_{i,j}\boldsymbol{v}_j\right), \tag{9}$$

where $\alpha_{i,j}$ denotes the attention weight of sample i over neighbor j, and $\boldsymbol{f}_i$ is the resulting neighborhood representation of sample i.

However, neighborhood aggregation also introduces noisy or unreliable information. To reduce this risk, we compress the aggregated information by regularizing the information retained in a stochastic representation $\boldsymbol{r}_i$ conditioned on $\boldsymbol{f}_i$. We model the conditional posterior $q(\boldsymbol{r}_i \mid \boldsymbol{f}_i)$ as a diagonal Gaussian with mean $\boldsymbol{\mu}_i$ and log-variance $\boldsymbol{\ell}_i = \log\boldsymbol{\sigma}_i^2$. The graph-information representation is then obtained through reparameterization:

$$q(\boldsymbol{r}_i \mid \boldsymbol{f}_i) = \mathcal{N}\left(\boldsymbol{r}_i; \boldsymbol{\mu}_i, \mathrm{diag}(\exp(\boldsymbol{\ell}_i))\right), \qquad \boldsymbol{r}_i = \boldsymbol{\mu}_i + \exp(\boldsymbol{\ell}_i/2)\odot\boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{d_z}), \tag{10}$$

where $\boldsymbol{\epsilon}_i$ is random noise. We measure how much neighborhood information the stochastic representation retains using the mutual information $I(F;R)$, where F and R denote the random variables corresponding to $\boldsymbol{f}_i$ and $\boldsymbol{r}_i$, respectively. Directly evaluating $I(F;R)$ requires the log-density of the aggregated posterior $q(\boldsymbol{r}) = \mathbb{E}_{\boldsymbol{f}}[q(\boldsymbol{r} \mid \boldsymbol{f})]$. This posterior is a mixture of conditional Gaussian distributions, and the required log-density expectation generally has no closed-form expression. We therefore introduce a variational upper bound on $I(F;R)$, expressed as an expected KL divergence.

By constraining this upper bound, we limit the neighborhood information retained in R without explicitly evaluating $I(F;R)$:

$$I(F;R) \leq \mathbb{E}_{\boldsymbol{f}}[D_{\mathrm{KL}}(q(\boldsymbol{r} \mid \boldsymbol{f}) \,\|\, p(\boldsymbol{r}))], \tag{11}$$

where $p(\boldsymbol{r}) = \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{d_z})$ is the standard normal prior shared across samples. Since both the conditional posterior $q(\boldsymbol{r} \mid \boldsymbol{f})$ and the prior are Gaussian, we can evaluate each KL term analytically without computing the aggregated posterior $q(\boldsymbol{r})$. The derivations of the upper bound and the KL expression are provided in the Appendix. Averaging the KL divergence over samples and normalizing by d_z yields the graph information compression loss:

$$\mathcal{L}_{\mathrm{gic}} = \frac{1}{nd_z}\sum_{i=1}^{n} D_{\mathrm{KL}}(q(\boldsymbol{r}_i \mid \boldsymbol{f}_i) \,\|\, p(\boldsymbol{r})) = \frac{1}{2nd_z}\sum_{i=1}^{n}\sum_{k=1}^{d_z}\left(\mu_{i,k}^2 + \exp(\ell_{i,k}) - \ell_{i,k} - 1\right), \tag{12}$$

where $\mu_{i,k}$ and $\ell_{i,k}$ denote the k-th elements of the posterior mean and log-variance for sample i.

Finally, we compute a sample-adaptive gate $\boldsymbol{g}$ by passing the concatenation of $\bar{\boldsymbol{Z}}$ and $\boldsymbol{f}$ through a linear layer followed by a sigmoid activation. The gate controls the contribution of the graph-information representation $\boldsymbol{r}$, yielding $\widehat{\boldsymbol{Z}} = \bar{\boldsymbol{Z}} + \boldsymbol{g} \odot \boldsymbol{r}$, where $\widehat{\boldsymbol{Z}}$ denotes the sample representation obtained by incorporating gated graph information into the fused multi-view representation.

2.4.2 Missing-aware label graph learning

Sample graphs optimize representations within the sample space, thereby effectively mitigating error propagation across sample relationships. However, they fail to alleviate error propagation across label relationships caused by insufficient semantic supervision. To address this, we therefore propose missing-aware label graph learning, which estimates directed co-occurrence relations from observed labels and regularizes the classifier weights of related labels to limit the propagation of coupled errors across label dependencies.

First, we constructed a graph with labels as nodes, where edge weights are calculated based on observed co-occurrence probabilities. Specifically, for source label i and target label j, we count $N_{i,j}$, the number of samples for which both labels are observed as positive, and $D_{i,j}$, the number of samples for which label i is observed as positive and label j is observed. We then estimate the directed conditional label co-occurrence probability as $C_{i,j} = N_{i,j}/D_{i,j}$. When $D_{i,j} = 0$, the corresponding $C_{i,j} = 0$. We further normalize $C_{i,j}$ over all target labels associated with source label a, yielding $A^{\mathrm{L}}_{i,j} = C_{i,j}/\sum_{r=1}^{c} C_{i,r}$, where $A^{\mathrm{L}}_{i,j}$ denotes the normalized directed relation weight from label i to label j. To incorporate label relationships into classifier learning, we impose constraints on the differences between classifier weights based on graph edge weights, resulting in the following label graph loss:

$$\mathcal{L}_{\mathrm{lg}} = \sum_{i=1}^{c}\sum_{j=1}^{c} \frac{A^{\mathrm{L}}_{i,j}\,\|\boldsymbol{w}_i - \boldsymbol{w}_j\|_2^2}{\sum_{r=1}^{c}\sum_{s=1}^{c} A^{\mathrm{L}}_{r,s}}, \tag{13}$$

where $\boldsymbol{w}_i$ and $\boldsymbol{w}_j$ denote the classifier weights associated with labels i and j.

Finally, we feed $\widehat{\boldsymbol{Z}}$ into the classifier to obtain the prediction $\boldsymbol{P}$. Because the observed labels are incomplete for each sample, we use the label observation indicator matrix to mask invalid supervision and adopt the following binary cross-entropy loss:

$$\mathcal{L}_{\mathrm{cls}} = -\frac{1}{nc}\sum_{i=1}^{n}\sum_{j=1}^{c} G_{i,j}\left[Y_{i,j}\log P_{i,j} + (1 - Y_{i,j})\log\left(1 - P_{i,j}\right)\right], \tag{14}$$

where $Y_{i,j}$ and $P_{i,j}$ denote the ground-truth value and predicted probability of the j-th label for sample i, respectively.

3 Experiments

3.1 Experimental settings

Datasets. We conduct a series of experiments on five widely used datasets. Concretely, Corel5k (Duygulu et al., 2002), Pascal07 (Everingham et al., 2010), ESPGame (Von Ahn & Dab-

Table 1: Performance comparison (mean/standard deviation) of nine methods with 50% missing views and 50% missing labels on five databases. The best results are marked in **bold**.

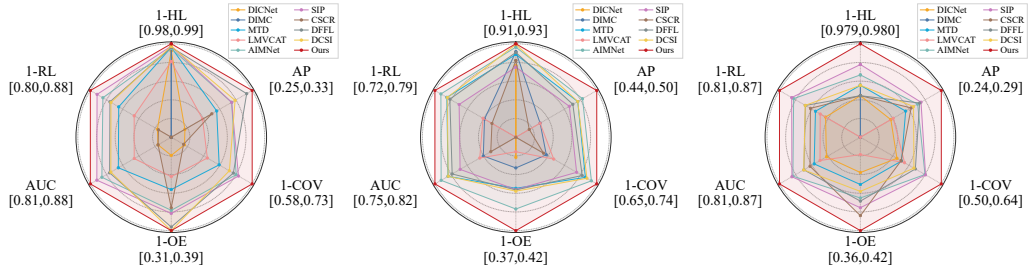
Data	Metric	DICNet	DIMC	MTD	LMVCAT	AIMNet	SIP	CSCR	DFFL	DCSI	Ours
Corel5k	AP ↑	$0.378_{0.009}$	$0.363_{0.009}$	$0.410_{0.008}$	$0.371_{0.008}$	$0.399_{0.010}$	$0.419_{0.008}$	$0.380_{0.024}$	$0.412_{0.008}$	$0.431_{0.011}$	$\mathbf{0.432}_{0.010}$
	1-HL ↑	$0.987_{0.000}$	$0.987_{0.000}$	$0.987_{0.000}$	$0.986_{0.000}$	$0.987_{0.000}$	$0.987_{0.000}$	$0.976_{0.014}$	$0.987_{0.000}$	$\mathbf{0.988}_{0.000}$	$0.987_{0.000}$
	1-RL ↑	$0.881_{0.003}$	$0.869_{0.002}$	$0.890_{0.006}$	$0.876_{0.003}$	$0.904_{0.000}$	$0.911_{0.001}$	$0.871_{0.030}$	$0.896_{0.003}$	$0.899_{0.003}$	$\mathbf{0.917}_{0.002}$
	AUC ↑	$0.884_{0.004}$	$0.873_{0.003}$	$0.893_{0.006}$	$0.878_{0.003}$	$0.907_{0.000}$	$0.914_{0.002}$	$0.874_{0.030}$	$0.900_{0.004}$	$0.901_{0.004}$	$\mathbf{0.919}_{0.002}$
	1-OE ↑	$0.465_{0.010}$	$0.443_{0.003}$	$0.484_{0.014}$	$0.441_{0.017}$	$0.465_{0.002}$	$0.497_{0.009}$	$0.464_{0.023}$	$0.481_{0.013}$	$\mathbf{0.517}_{0.021}$	$0.500_{0.017}$
	1-Cov ↑	$0.723_{0.009}$	$0.696_{0.004}$	$0.746_{0.012}$	$0.722_{0.005}$	$0.777_{0.001}$	$0.791_{0.006}$	$0.702_{0.064}$	$0.759_{0.007}$	$0.763_{0.007}$	$\mathbf{0.808}_{0.005}$
Pascal07	AP ↑	$0.501_{0.011}$	$0.535_{0.012}$	$0.548_{0.008}$	$0.528_{0.004}$	$0.546_{0.007}$	$0.550_{0.014}$	$0.518_{0.005}$	$0.549_{0.005}$	$0.561_{0.004}$	$\mathbf{0.564}_{0.007}$
	1-HL ↑	$0.928_{0.000}$	$0.932_{0.000}$	$0.932_{0.000}$	$0.915_{0.006}$	$0.930_{0.001}$	$0.930_{0.001}$	$0.930_{0.000}$	$0.932_{0.001}$	$0.932_{0.000}$	$\mathbf{0.933}_{0.001}$
	1-RL ↑	$0.782_{0.004}$	$0.812_{0.005}$	$0.828_{0.005}$	$0.812_{0.004}$	$0.829_{0.003}$	$0.827_{0.008}$	$0.798_{0.002}$	$0.826_{0.007}$	$0.835_{0.002}$	$\mathbf{0.839}_{0.003}$
	AUC ↑	$0.810_{0.003}$	$0.833_{0.005}$	$0.848_{0.005}$	$0.833_{0.006}$	$0.850_{0.004}$	$0.848_{0.005}$	$0.823_{0.000}$	$0.845_{0.009}$	$0.853_{0.006}$	$\mathbf{0.860}_{0.003}$
	1-OE ↑	$0.423_{0.013}$	$0.451_{0.018}$	$0.456_{0.013}$	$0.430_{0.005}$	$0.462_{0.011}$	$0.464_{0.012}$	$0.438_{0.007}$	$0.457_{0.012}$	$0.464_{0.010}$	$\mathbf{0.478}_{0.013}$
	1-Cov ↑	$0.730_{0.003}$	$0.763_{0.006}$	$0.781_{0.006}$	$0.764_{0.007}$	$0.781_{0.004}$	$0.780_{0.006}$	$0.745_{0.000}$	$0.779_{0.009}$	$0.789_{0.006}$	$\mathbf{0.793}_{0.003}$
ESPGame	AP ↑	$0.298_{0.003}$	$0.286_{0.005}$	$0.304_{0.004}$	$0.292_{0.002}$	$0.301_{0.004}$	$0.309_{0.006}$	$0.296_{0.005}$	$0.293_{0.005}$	$0.312_{0.005}$	$\mathbf{0.317}_{0.004}$
	1-HL ↑	$\mathbf{0.982}_{0.000}$	$\mathbf{0.982}_{0.000}$	$\mathbf{0.982}_{0.000}$	$0.981_{0.000}$	$\mathbf{0.982}_{0.000}$	$\mathbf{0.982}_{0.000}$	$\mathbf{0.982}_{0.000}$	$\mathbf{0.982}_{0.000}$	$\mathbf{0.982}_{0.000}$	$\mathbf{0.982}_{0.000}$
	1-RL ↑	$0.832_{0.002}$	$0.816_{0.002}$	$0.837_{0.002}$	$0.829_{0.001}$	$0.845_{0.000}$	$0.848_{0.003}$	$0.831_{0.000}$	$0.833_{0.002}$	$0.846_{0.004}$	$\mathbf{0.851}_{0.002}$
	AUC ↑	$0.836_{0.002}$	$0.820_{0.002}$	$0.841_{0.001}$	$0.831_{0.001}$	$0.849_{0.000}$	$0.852_{0.002}$	$0.835_{0.000}$	$0.838_{0.002}$	$0.849_{0.003}$	$\mathbf{0.855}_{0.001}$
	1-OE ↑	$0.440_{0.007}$	$0.427_{0.008}$	$0.444_{0.009}$	$0.425_{0.008}$	$0.432_{0.013}$	$0.447_{0.014}$	$0.445_{0.012}$	$0.434_{0.011}$	$\mathbf{0.459}_{0.010}$	$\mathbf{0.459}_{0.008}$
	1-Cov ↑	$0.596_{0.005}$	$0.553_{0.008}$	$0.599_{0.005}$	$0.591_{0.005}$	$0.619_{0.002}$	$0.626_{0.005}$	$0.581_{0.001}$	$0.594_{0.006}$	$0.622_{0.008}$	$\mathbf{0.635}_{0.004}$
IAPRTC12	AP ↑	$0.328_{0.007}$	$0.303_{0.005}$	$0.330_{0.006}$	$0.316_{0.006}$	$0.325_{0.000}$	$0.334_{0.003}$	$0.297_{0.003}$	$0.317_{0.005}$	$0.331_{0.003}$	$\mathbf{0.346}_{0.005}$
	1-HL ↑	$\mathbf{0.980}_{0.000}$	$\mathbf{0.980}_{0.000}$	$\mathbf{0.980}_{0.000}$	$0.979_{0.000}$	$\mathbf{0.980}_{0.000}$	$\mathbf{0.980}_{0.000}$	$\mathbf{0.980}_{0.000}$	$\mathbf{0.980}_{0.000}$	$\mathbf{0.980}_{0.000}$	$\mathbf{0.980}_{0.000}$
	1-RL ↑	$0.875_{0.003}$	$0.854_{0.000}$	$0.875_{0.000}$	$0.870_{0.000}$	$0.883_{0.001}$	$0.885_{0.001}$	$0.859_{0.001}$	$0.870_{0.001}$	$0.879_{0.001}$	$\mathbf{0.892}_{0.002}$
	AUC ↑	$0.875_{0.002}$	$0.856_{0.001}$	$0.876_{0.000}$	$0.872_{0.000}$	$0.884_{0.001}$	$0.886_{0.001}$	$0.860_{0.001}$	$0.873_{0.001}$	$0.879_{0.001}$	$\mathbf{0.893}_{0.001}$
	1-OE ↑	$0.464_{0.018}$	$0.437_{0.007}$	$0.468_{0.011}$	$0.435_{0.009}$	$0.453_{0.010}$	$0.470_{0.003}$	$0.441_{0.008}$	$0.461_{0.013}$	$0.465_{0.007}$	$\mathbf{0.477}_{0.010}$
	1-Cov ↑	$0.652_{0.005}$	$0.596_{0.002}$	$0.648_{0.001}$	$0.647_{0.003}$	$0.672_{0.001}$	$0.678_{0.004}$	$0.597_{0.006}$	$0.635_{0.001}$	$0.660_{0.003}$	$\mathbf{0.698}_{0.005}$
Mirfickr	AP ↑	$0.587_{0.001}$	$0.592_{0.002}$	$0.608_{0.004}$	$0.598_{0.003}$	$0.601_{0.004}$	$\mathbf{0.616}_{0.005}$	$0.592_{0.001}$	$0.612_{0.002}$	$0.610_{0.003}$	$\mathbf{0.616}_{0.005}$
	1-HL ↑	$0.888_{0.001}$	$0.888_{0.001}$	$0.891_{0.000}$	$0.881_{0.002}$	$0.889_{0.000}$	$\mathbf{0.892}_{0.001}$	$0.888_{0.000}$	$0.890_{0.000}$	$0.891_{0.000}$	$0.891_{0.000}$
	1-RL ↑	$0.861_{0.001}$	$0.867_{0.001}$	$0.875_{0.001}$	$0.865_{0.001}$	$0.873_{0.002}$	$\mathbf{0.879}_{0.002}$	$0.868_{0.002}$	$0.876_{0.001}$	$0.877_{0.002}$	$\mathbf{0.879}_{0.002}$
	AUC ↑	$0.849_{0.000}$	$0.854_{0.001}$	$0.861_{0.001}$	$0.852_{0.001}$	$0.860_{0.001}$	$0.864_{0.002}$	$0.854_{0.003}$	$0.860_{0.002}$	$0.862_{0.001}$	$\mathbf{0.865}_{0.003}$
	1-OE ↑	$0.640_{0.008}$	$0.638_{0.003}$	$0.659_{0.004}$	$0.646_{0.003}$	$0.653_{0.007}$	$\mathbf{0.667}_{0.004}$	$0.642_{0.005}$	$0.658_{0.003}$	$0.657_{0.003}$	$\mathbf{0.667}_{0.009}$
	1-Cov ↑	$0.647_{0.002}$	$0.657_{0.003}$	$0.676_{0.002}$	$0.666_{0.006}$	$0.671_{0.003}$	$0.679_{0.005}$	$0.660_{0.004}$	$0.680_{0.003}$	$0.680_{0.002}$	$\mathbf{0.684}_{0.003}$

bish, 2004), IAPRTC12 (Guillaumin et al., 2010), Mirflickr (Huiskes & Lew, 2008). For detailed information about all datasets, please refer to the Appendix.

Comparison methods. We compared our CEE with nine state-of-the-art IM3C methods. Specifically, they are DICNet (Liu et al., 2023a), DIMC (Wen et al., 2023), MTD (Liu et al., 2024b), LMVCAT (Liu et al., 2023b), AIMNet (Liu et al., 2024a), SIP (Liu et al., 2024c), CSCR (Wang et al., 2025), DFFL (Xiao et al., 2026), and DCSI (Wen et al., 2026). Details of the comparison methods are provided in the Appendix.

Metrics. To evaluate the effectiveness of the method, we use six commonly adopted multi-label classification metrics: Average Precision (AP), Hamming Loss (HL), Ranking Loss (RL), Area Under the ROC Curve (AUC), One-Error (OE), and Coverage (Cov). For consistent interpretation, we report $1-\mathrm{HL}$, $1-\mathrm{RL}$, $1-\mathrm{OE}$, and $1-\mathrm{Cov}$ together with AUC and AP, such that higher values indicate better performance for all metrics.

Implementation details. Each dataset is partitioned into training, validation, and test subsets containing 70%, 15%, and 15% of the samples, respectively. We developed CEE using PyTorch 2.10.0 and ran experiments on Linux on an NVIDIA RTX 4090 GPU with 24 GB of memory. Each view is processed by a fully connected autoencoder, and the dimensions of the encoder (decoder) layers are $d_v - 0.8d_v - 0.64d_v - 1500 - 512(512 - 1500 - 0.64d_v - 0.8d_v - d_v)$. Model optimization uses SGD with a learning rate of 0.1 and a batch size of 64. Generator pretraining runs for 400 epochs, followed by 100 epochs of joint training. We will release the source code after acceptance of the manuscript.



(a) Corel5k (b) Pascal07 (c) IAPRTC12

Figure 2: Experimental results of ten methods on the three databases with 70% missing views and 70% missing labels. In this visualization, the worst results are positioned at the center, while the best results correspond to the vertices.

3.2 Experimental results analysis

Analysis of overall performance: To verify the overall performance of our CEE, we compare CEE with nine state-of-the-art IM3C methods on five datasets, with both the missing-view and missing-label rates set to 50%. The results are reported in Table 1. From this, CEE achieved strong performance, outperforming nine IM3C baselines across all five datasets. This is attributed to our Collaborative Dual-graph Structure Learning (CDSL). Specifically, in CDSL, information-compressed feature graph learning refines sample representations by leveraging relationships within feature graphs and effectively mitigates the propagation of coupled errors across sample relationships. Subsequently, missing-aware label graph learning models label dependencies by constructing a label graph, which effectively alleviates the propagation of coupled errors across label relationships, resulting in more reliable label supervision. Through the collaboration of these dual graphs, we can effectively mitigate CEP, thereby achieving more robust label classification.

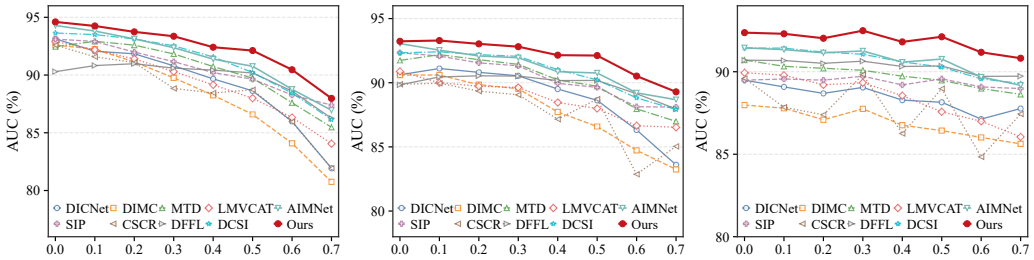
Analysis of extreme missing conditions: To verify the effectiveness of our CEE under conditions of extreme missingness, we conducted experiments on three datasets with both view and label missingness rates set to 70%. As shown in Fig. 2, we can observe that CEE consistently achieves leading performance across all six metrics on the three datasets, demonstrating its ability to maintain a competitive advantage even under severe view and label incompleteness. To elaborate, this is attributed to our multi-view specific distribution learning, which effectively mitigates RPB by recovering reliable representations that match the target view distribution, thereby improving view-distribution characteristics and quality stability. Meanwhile, multi-view structure reinforcement further strengthens the semantic consistency of the recovered representations by reinforcing the correlations between different views of the same sample. Therefore, the model can gain access to more comprehensive feature information, thereby achieving superior classification performance.

3.3 Ablation study

We conducted ablation experiments on three datasets, with both the view missing rate and the label missing rate fixed at 50% to verify the contribution of each component. Specifically, we decomposed CEE into three main components: Multi-view Specific Distribution Learning (MSDL), Multi-view Structure Reinforcement (MSR), and Collaborative Dual-graph Structure Learning (CDSL). As shown in Table 2, the absence of either MSR or CDSL resulted in a significant decline in model performance. The experimental results lead to the following conclusions: (1) MSDL recovers the specific distribution representations of missing views, thereby enhancing the model's classification performance. (2) MSR effectively strengthens the consensus representation across views, thereby improving inter-view consistency. (3) CDSL learns relationships between samples and views, thereby mitigating error propagation between samples and labels.

Table 2: Ablation study of our CEE on three databases under 50% missing views and 50% missing labels.

base	MSDL	MSR	CDSL	Corel5k				Pascal07				IAPRTC12			
				AP↑	1-RL↑	AUC↑	1-OE↑	AP↑	1-RL↑	AUC↑	1-OE↑	AP↑	1-RL↑	AUC↑	1-OE↑
✓				0.3989	0.8851	0.8881	0.4619	0.5457	0.8251	0.8530	0.4599	0.3155	0.8733	0.8747	0.4369
✓	✓			0.4035	0.8904	0.8929	0.4677	0.5435	0.8239	0.8460	0.4570	0.3201	0.8726	0.8743	0.4447
✓	✓	✓		0.4114	0.9081	0.9110	0.4789	0.5578	0.8385	0.8577	0.4727	0.3454	0.8922	0.8925	0.4743
✓	✓	✓	✓	**0.4325**	**0.9170**	**0.9194**	**0.5008**	**0.5641**	**0.8392**	**0.8608**	**0.4782**	**0.3468**	**0.8920**	**0.8934**	**0.4771**



(a) Equal view and label missing rates (b) Label missing rate fixed at 50% (c) View missing rate fixed at 50%

Figure 3: Performance comparison of ten methods on Corel5k in terms of AUC under three settings: (a) equal missing-view and missing-label rates; (b) a fixed missing-label rate of 50% with varying missing-view rates; and (c) a fixed missing-view rate of 50% with varying missing-label rates. Higher values indicate better performance.

3.4 Robust Analysis

We compared CEE with nine state-of-the-art methods on the Corel5k dataset using the AUC metric to verify the effectiveness of CEE. Specifically, we consider three conditions: jointly increasing view and label missing rates, varying missing-label rates with the missing-view rate fixed at 50%, and varying missing-view rates with the missing-label rate fixed at 50%. The experimental results are shown in Fig. 3. As can be seen, increasing either the label-missing or view-missing rate reduces the semantic information available to the models, causing fluctuations in their performance. In contrast, although the AUC performance of our CEE declined as the missing rates increased, it consistently achieved the best performance. This demonstrates the robust nature of CEE and its effectiveness in mitigating CEP. This success stems from our proposed Multi-view Specific Distribution Learning(MSDL) and Collaborative Dual-graph Structure Learning(CDSL). MSDL enables the model to effectively handle variations in the proportion of missing views by restoring the representation of the distribution of missing views. CDSL constructs sample graphs and label graphs to model the relationships of samples and labels, allowing the model to maintain superior performance even under conditions of dual missingness

4 Conclusion

In this paper, we propose a Coupling Error Exploration (CEE) for incomplete multi-view missing multi-label classification, and it addresses RBP and CEP. Specifically, to alleviate RBP, we recover the missing views and ensure the view-specific distribution structure through multi-view rectified flow learning. Subsequently, multi-view structure enhancement improves the semantic consistency of completed views through cross-view contrastive learning. To alleviate RBP, we further introduce collaborative dual-graph structure learning, which enhances sample representations by exploiting feature graph relationships while complementing semantic supervision through label dependencies. Extensive experiments consistently validate the effectiveness and robustness of CEE.

REFERENCES

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 *European Conference on Computer Vision*, pp. 97–112. Springer, 2002.

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 (2):303–338, 2010.

Matthieu Guillaumin, Jakob Verbeek, and Cordelia Schmid. Multimodal semi-supervised learning for image classification. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 902–909. IEEE, 2010.

Pingting Hao, Wanfu Gao, and Liang Hu. Embedded feature fusion for multi-view multi-label feature selection. *Pattern Recognition*, 157:110888, 2025a.

Pingting Hao, Kunpeng Liu, and Wanfu Gao. Uncertainty-aware global-view reconstruction for multi-view multi-label feature selection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 17068–17076, 2025b.

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

Quanjiang Li, Tianxiang Xu, Tingjin Luo, Yan Zhong, Yang Li, Yiyun Zhou, and Chenping Hou. Theory-driven label-specific representation for incomplete multi-view multi-label learning. *Advances in Neural Information Processing Systems*, 38:121996–122037, 2026.

Yiting Li, Jia Zhang, Hanrui Wu, Guodong Du, and Jinyi Long. Consistent and specific multi-view multi-label learning with correlation information. *Information Sciences*, 687:121395, 2025.

Chengliang Liu, Jie Wen, Xiaoling Luo, Chao Huang, Zhihao Wu, and Yong Xu. Dicnet: Deep instance-level contrastive network for double incomplete multi-view multi-label classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 8807–8815, 2023a.

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, pp. 8816–8824, 2023b.

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, pp. 13864–13872, 2024a.

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. In *Advances in Neural Information Processing Systems*, volume 36, 2024b.

Chengliang Liu, Gehui Xu, Jie Wen, Yabo Liu, Chao Huang, and Yong Xu. Partial multi-view multi-label classification via semantic invariance learning and prototype modeling. In *Forty-first International Conference on Machine Learning*, 2024c.

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*, 47(6):4940–4956, 2025.

Chengliang Liu, Bo Li, Bob Zhang, Xiaoling Luo, Yabo Liu, and Jie Wen. Permutation-consistent variational encoding for incomplete multi-view multi-label classification. In *International Conference on Learning Representations*, volume 2026, pp. 156613–156630, 2026a.

Yadong Liu, Chengliang Liu, Jie Wen, Li Shen, Bob Zhang, and Yong Xu. Learning compact semantic information and reliable pseudo-labels for incomplete multi-view multi-label classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(7):7575–7589, 2026b.

Yadong Liu, Waikeung Wong, Yulong Chen, and Jie Wen. Quality-aware and soft consistency driven representation fusion for incomplete multi-view multi-label classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 23882–23890, 2026c.

Rushil Sanghavi and Yashaswi Verma. Multi-view multi-label canonical correlation analysis for cross-modal matching and retrieval. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 4700–4709. IEEE, 2022.

Yi Shan, Liyang Gao, Yuena Lin, Zhen Yang, Gengyu Lyu, and Honggui Han. Hypergraph-based multi-view multi-label classification via adaptive high-order semantic fusion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 25269–25277, 2026.

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

Qiaoyu Tan, Guoxian Yu, Carlotta Domeniconi, Jun Wang, and Zili Zhang. Incomplete multi-view weak-label learning. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, volume 18, pp. 2703–2709, 2018.

Yanqiang Tu, Gengyu Lyu, Dian Zhang, Xiaozhu Jing, Wei Ke, and Wuman Luo. Robust multi-view multi-label learning via uncertainty-gated diffusion and dynamic graph mrmr. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pp. 4683–4694, 2026.

Luis Von Ahn and Laura Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 319–326, 2004.

Yiming Wang, Qun Li, Dongxia Chang, Jie Wen, Fu Xiao, and Yao Zhao. A category-driven contrastive recovery network for double incomplete multi-view multi-label classification. *IEEE Transactions on Multimedia*, 27:4008–4017, 2025.

Hao Wei, Yongjian Deng, Qiuru Hai, Yuena Lin, Zhen Yang, and Gengyu Lyu. Multi-view multi-label classification via view-label matching selection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 21456–21464, 2025.

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.

Jie Wen, Lian Zhao, Xiaohuan Lu, Chengliang Liu, Li Shen, Chao Huang, and Yong Xu. Disentangling consistent and specific information for double incomplete multi-view multi-label classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 7307–7320, 2026.

Xinyu Xiao, Shuhan Qi, Yulin Wu, Bin Chen, Chao Xing, and Xuan Wang. Dual feature fusion for incomplete multi-view multi-label learning. *IEEE Transactions on Multimedia*, 28:1329–1341, 2026.

Wulin Xie, Lian Zhao, Jiang Long, Xiaohuan Lu, and Bingyan Nie. Multi-view factorizing and disentangling: A novel framework for incomplete multi-view multi-label classification. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1914–1923. IEEE, 2025.

Xu Yan, Jun Yin, and Jie Wen. Incomplete multi-view multi-label learning via disentangled representation and label semantic embedding. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 30722–30731. IEEE, 2025.

Shunxiang Yang, Cheng Lian, Zhigang Zeng, Bingrong Xu, Junbin Zang, and Zhidong Zhang. A multi-view multi-scale neural network for multi-label ecg classification. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 7(3):648–660, 2023.

Honglin Yuan, Yuan Sun, Fei Zhou, Jing Wen, Shihua Yuan, Xiaojian You, and Zhenwen Ren. Prototype matching learning for incomplete multi-view clustering. *IEEE Transactions on Image Processing*, 34:828–841, 2025.

Honglin Yuan, Yuan Sun, Xingfeng Li, Xi Wu, Xiaojian You, and Zhenwen Ren. Adaptive hardness-driven dictionary distillation for incomplete streaming view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 13424–13438, 2026.

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

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.

Dawei Zhao, Hong Li, Yixiang Lu, Dong Sun, and Qingwei Gao. An effective bipartite graph fusion and contrastive label correlation for multi-view multi-label classification. *Pattern Recognition*, 162:111430, 2025a.

Lian Zhao, Jie Wen, Xiaohuan Lu, Wai Keung Wong, Jiang Long, and Wulin Xie. Task-augmented cross-view imputation network for partial multi-view incomplete multi-label classification. *Neural Networks*, 187:107349, 2025b.

A Appendix

This section provides additional methodological details and experimental results to support the main paper. We first describe our use of large language models in Section A.1.In Section A.2, we briefly explain our motivation. Section A.3 reviews related work on multi-view multi-label learning and incomplete multi-view missing multi-label classification. In Section A.4, we derive the variational upper bound and the KL divergence used in information-compressed feature graph learning. Next, Section A.5 presents the pseudocode for CEE. Section A.6 provides detailed information about the datasets and the missingness settings used in our experiments. In Section A.7, we introduce the baseline methods used for comparison. Finally, Section A.8 presents additional experiments evaluating the robustness of CEE across various missingness scenarios.

A.1 Large language model use statement

We used a large language model solely to improve the clarity, grammar, and style of the manuscript. It was not involved in research design, implementation, analysis, or interpretation. All technical content was developed and verified by the authors, who take full responsibility for the content of this paper.

A.2 Motivation

The incomplete data in IM3C tasks primarily consists of incomplete views and incomplete labels. Based on how missing views are handled, existing IM3C methods are categorized into imputation-based methods and non-imputation-based methods. The former utilizes local sample information or cross-view shared semantics to impute views; this often results in generated representations that fail to capture the specific characteristics of the target view. The latter directly fuses available views, which can result in fused features containing noise or insufficient semantic information. These challenges make it difficult for models to obtain view features with specific distribution structures and stable quality, leading to the Representation Bias Problem (RBP). Meanwhile, unreliable representations reduce classification performance, and incomplete labels provide incomplete and erroneous semantic guidance for representation learning, leading to Coupled Error Propagation (CEP). Therefore, as shown in the Fig 4, enabling the model to achieve more accurate classification under the influence of RBP and CEP is an urgent issue that needs to be addressed.

A.3 Related work

Multi-view multi-label learning. Multi-view Multi-label Learning(MvMlL) integrates heterogeneous features and label relationships to predict multiple semantic categories for each sample (Hao

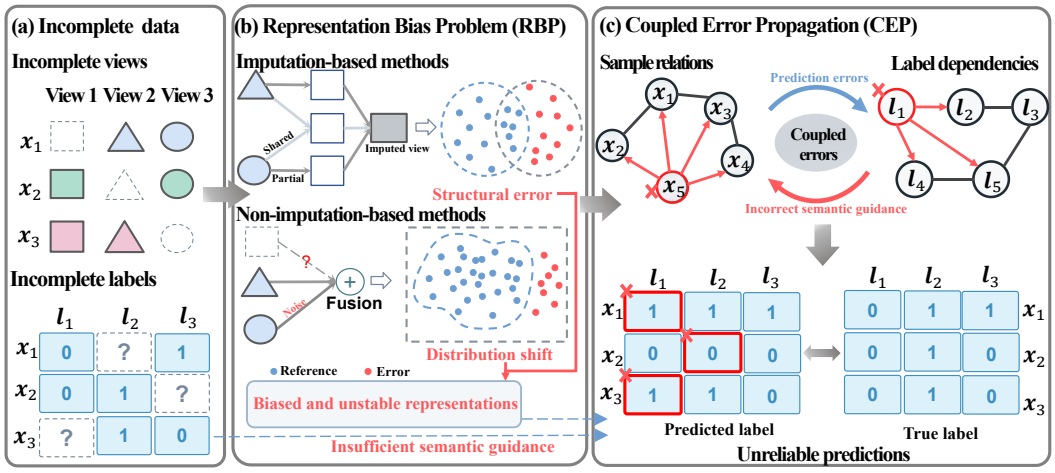


Figure 4: Examples of the Representation Bias Problem(RBP) and Coupling Error Propagation (CEP). (a) Incomplete data in IM3C consists of incomplete views and incomplete labels. (b) Imputation-based methods fail to recover the specific distributional structure of the views, leading to structural errors. Non-imputation-based methods directly fuse views containing noise and incomplete representations, resulting in distribution shift. Both methods yield representations that are biased and unstable, *i.e.*, Representation Bias Problem(RBP). (c) Unreliable representations increase label prediction errors, while incomplete labels provide incorrect semantic supervision for representation learning. Together, these factors cause coupled errors to propagate across sample relations and label dependencies, resulting in unreliable predictions, *i.e.*, Coupled Error Propagation (CEP).

et al., 2025a; Zhao et al., 2022). Recent studies primarily investigate complementary representation learning and structure-aware fusion. For example, CSVL (Li et al., 2025) jointly learns shared and view-specific subspaces while preserving feature manifolds and label correlations. From a graph-learning perspective, BCML (Zhao et al., 2025a) combines anchor-based bipartite graph fusion with contrastive label correlation learning to integrate complementary information and enrich label semantics. Rather than treating feature fusion and label prediction separately, VAMS (Wei et al., 2025) formulates classification as a view-label matching problem and jointly models sample relationships and label dependencies through a unified graph. Moving beyond pairwise relationships, Hyper-AHSF (Shan et al., 2026) constructs view-specific and cross-view hypergraphs to capture group-level semantic interactions, while label-driven contrastive learning strengthens the discriminability of consensus representations. These studies highlight the importance of preserving view-specific information and incorporating semantic relationships into multi-view fusion. However, equipment or sensor failures can lead to missing views. Meanwhile, high annotation costs and erroneous annotations result in incomplete sample labels. Missing views limit the complementary information available for representation learning, while missing labels weaken semantic supervision for classification. These two types of missingness often coexist, significantly impairing the model's label prediction capability. Therefore, there is an urgent need for methods specifically designed for this situation.

Incomplete multi-view missing multi-label classification. Incomplete Multi-view Multi-label Classification (IM3C) aims to predict potentially relevant labels for each sample by leveraging consensus and complementary information across collected views, alongside limited available label supervision (Liu et al., 2026b). Depending on whether missing views are explicitly recovered for downstream prediction, existing methods can be broadly categorized into imputation-based and non-imputation-based approaches. Imputation-based methods recover missing features or latent representations using information from observed views: TDLSR (Li et al., 2026) establishes a connection between sample topology and prototypes to achieve proximity-based data completion, whereas CSCR (Wang et al., 2025) constructs a label coincidence matrix from observed positive labels and uses the resulting sample similarities to initialize missing-view recovery. Non-imputation-based methods learn and fuse representations from available views without explicitly completing the miss-

ing views: MVFD (Xie et al., 2025) learns shared and view-specific factors through masked cross-view prediction and graph disentanglement, then fuses the observed representations for classification. DRLS (Yan et al., 2025) separates shared and specific information within a variational framework and learns label semantic embeddings from a label graph. Despite these advances, similarity-based or prototype-based completion does not necessarily preserve target-view distributional structures, while fusion restricted to observed views may produce unstable representations as view availability changes. These limitations give rise to the Representation Bias Problem (RBP). Furthermore, RBP leads to inaccurate label prediction, while incomplete labels fail to provide sufficient semantic guidance for representation learning, *i.e.*, Coupling Error Propagation (CEP), thereby undermining model performance. Therefore, simultaneously overcoming the RBP and CEP problems in the case of missing views and labels is a challenge that urgently requires exploration.

A.4 Information-compression feature graph learning

To constrain the amount of aggregated neighborhood information retained in sample graph representations, we derive a variational upper bound on the mutual information between neighborhood and graph representations. This bound uses the expected conditional KL divergence to provide a tractable basis for regulating information retention through posterior regularization. Such regularization encourages compact graph representations to mitigate coupled error propagation across samples.

Let F and R denote the random variables corresponding to the neighborhood and stochastic graph-information representations, with realizations $\boldsymbol{f}$ and $\boldsymbol{r}$, respectively. Let $q(\boldsymbol{f})$ denote the marginal distribution of F. Their joint distribution factorizes as $q(\boldsymbol{f}, \boldsymbol{r}) = q(\boldsymbol{f})q(\boldsymbol{r} \mid \boldsymbol{f})$, and the aggregated posterior is $q(\boldsymbol{r}) = \int q(\boldsymbol{f})q(\boldsymbol{r} \mid \boldsymbol{f})\, d\boldsymbol{f}$. Mutual information is the KL divergence between the joint distribution and the product of its marginals:

$$\begin{aligned} I(F;R) &= D_{\mathrm{KL}}(q(\boldsymbol{f},\boldsymbol{r}) \,\|\, q(\boldsymbol{f})q(\boldsymbol{r})) \\ &= \iint q(\boldsymbol{f},\boldsymbol{r}) \log \frac{q(\boldsymbol{f},\boldsymbol{r})}{q(\boldsymbol{f})q(\boldsymbol{r})}\, d\boldsymbol{r}\, d\boldsymbol{f} \\ &= \int q(\boldsymbol{f}) \left[\int q(\boldsymbol{r} \mid \boldsymbol{f}) \log \frac{q(\boldsymbol{r} \mid \boldsymbol{f})}{q(\boldsymbol{r})}\, d\boldsymbol{r} \right] d\boldsymbol{f} \\ &= \mathbb{E}_{\boldsymbol{f}} \left[\int q(\boldsymbol{r} \mid \boldsymbol{f}) \log \frac{q(\boldsymbol{r} \mid \boldsymbol{f})}{q(\boldsymbol{r})}\, d\boldsymbol{r} \right]. \end{aligned} \tag{15}$$

For a prior $p(\boldsymbol{r})$ independent of F, assume that the expected KL divergence is finite. Introducing this prior into the logarithmic ratio gives

$$\log \frac{q(\boldsymbol{r} \mid \boldsymbol{f})}{q(\boldsymbol{r})} = \log \frac{q(\boldsymbol{r} \mid \boldsymbol{f})}{p(\boldsymbol{r})} - \log \frac{q(\boldsymbol{r})}{p(\boldsymbol{r})}. \tag{16}$$

Taking expectations and using $\mathbb{E}_{\boldsymbol{f}}[q(\boldsymbol{r} \mid \boldsymbol{f})] = q(\boldsymbol{r})$, we obtain

$$\begin{aligned} I(F;R) &= \mathbb{E}_{\boldsymbol{f}} \left[\int q(\boldsymbol{r} \mid \boldsymbol{f}) \log \frac{q(\boldsymbol{r} \mid \boldsymbol{f})}{p(\boldsymbol{r})}\, d\boldsymbol{r} \right] - \int q(\boldsymbol{r}) \log \frac{q(\boldsymbol{r})}{p(\boldsymbol{r})}\, d\boldsymbol{r} \\ &= \mathbb{E}_{\boldsymbol{f}}[D_{\mathrm{KL}}(q(\boldsymbol{r} \mid \boldsymbol{f}) \,\|\, p(\boldsymbol{r}))] - D_{\mathrm{KL}}(q(\boldsymbol{r}) \,\|\, p(\boldsymbol{r})). \end{aligned} \tag{17}$$

Since KL divergence is nonnegative,

$$I(F;R) \le \mathbb{E}_{\boldsymbol{f}}[D_{\mathrm{KL}}(q(\boldsymbol{r} \mid \boldsymbol{f}) \,\|\, p(\boldsymbol{r}))]. \tag{18}$$

Thus, imposing an upper limit on the expected conditional KL divergence also limits the neighborhood information retained in R. This provides a theoretical basis for controlling information retention through posterior regularization.

For sample i, let $q_i(\boldsymbol{r}) = q(\boldsymbol{r} \mid \boldsymbol{f}_i)$ denote the diagonal Gaussian posterior in Eq. (10). We adopt the standard normal prior $p(\boldsymbol{r}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{d_z})$. From the reparameterization,

$$r_{i,k} = \mu_{i,k} + \exp(\ell_{i,k}/2)\epsilon_{i,k}, \qquad \epsilon_{i,k} \sim \mathcal{N}(0,1). \tag{19}$$

Since $\epsilon_{i,k}$ follows a standard normal distribution, we have $\mathbb{E}[\epsilon_{i,k}] = 0$ and $\mathbb{E}[\epsilon_{i,k}^2] = 1$. Substituting these identities into the reparameterization gives

$$\begin{aligned} \mathbb{E}[r_{i,k}^2 \mid \boldsymbol{f}_i] &= \mu_{i,k}^2 + \exp(\ell_{i,k}), \\ \mathbb{E}[(r_{i,k} - \mu_{i,k})^2 \mid \boldsymbol{f}_i] &= \exp(\ell_{i,k}). \end{aligned} \tag{20}$$

Algorithm 1: The Algorithm of CEE

Input: Training dataset $\{\boldsymbol{X}^{(v)}\}_{v=1}^{m}$ of size N; pretraining epochs E_p; joint-training epochs E_j; batch size B.

Generator pretraining

while $e < E_p$ **do**
  Construct valid latent pairs from jointly observed views of the same sample
  Optimize only the generator using Eq. 1, with the rectified-flow loss $\mathcal{L}_{\mathrm{rf}}$ defined in Eq. 4
end while

Joint training

while $e < E_j$ **do**
  Encode and reconstruct observed views; compute the reconstruction loss $\mathcal{L}_{\mathrm{rec}}$ by Eq. 3
  Obtain completed representations $\widetilde{\boldsymbol{Z}}_i^{(r)}$ via Eq. 6
  Compute the contrastive loss $\mathcal{L}_{\mathrm{cl}}$ by Eq. 8
  Compute the graph information compression loss $\mathcal{L}_{\mathrm{gic}}$ and the label-graph loss $\mathcal{L}_{\mathrm{lg}}$ by Eq. 12 and Eq. 13
  Obtain predictions $\boldsymbol{P}$ and compute the binary cross-entropy loss $\mathcal{L}_{\mathrm{cls}}$ by Eq. 14
  Optimize the objective function by Eq. 2
end while

To compute the KL divergence, we use the above results to evaluate the following expectations:

$$\begin{aligned}\mathbb{E}[\log q(\boldsymbol{r}_i \mid \boldsymbol{f}_i) \mid \boldsymbol{f}_i] &= -\frac{d_z}{2}\log(2\pi) - \frac{1}{2}\sum_{k=1}^{d_z}(\ell_{i,k}+1),\\ \mathbb{E}[\log p(\boldsymbol{r}_i) \mid \boldsymbol{f}_i] &= -\frac{d_z}{2}\log(2\pi) - \frac{1}{2}\sum_{k=1}^{d_z}\left(\mu_{i,k}^2 + \exp(\ell_{i,k})\right).\end{aligned} \tag{21}$$

Accordingly, the KL divergence between the posterior and the prior is given by

$$\begin{aligned}&D_{\mathrm{KL}}(q(\boldsymbol{r}_i \mid \boldsymbol{f}_i) \,\|\, p(\boldsymbol{r}_i))\\ &\quad= \mathbb{E}[\log q(\boldsymbol{r}_i \mid \boldsymbol{f}_i) - \log p(\boldsymbol{r}_i) \mid \boldsymbol{f}_i]\\ &\quad= \frac{1}{2}\sum_{k=1}^{d_z}\left(\mu_{i,k}^2 + \exp(\ell_{i,k}) - \ell_{i,k} - 1\right).\end{aligned} \tag{22}$$

In summary, the variational bound provides a tractable objective for neighborhood information compression. Neighborhood aggregation supplies complementary information from similar samples, while posterior regularization controls its retention in graph representations. Together, these mechanisms support more reliable learning of sample relationships and help mitigate coupled error propagation across samples.

A.5 Algorithm

The complete training procedure for CEE is presented in Algorithm 1.

A.6 Statistics for datasets

The details of all datasets are shown in TABLE 3. To construct the doubly incomplete setting, we randomly introduce missing views and labels. Specifically, view masks are sampled independently according to the specified missing-view rate, with the additional constraint that each sample retains at least one observed view. For labels, the same masking probability is applied to both positive and negative entries.

A.7 comparison methods

To evaluate the performance of our CEE, we compare CEE with nine state-of-the-art IM3C methods on the five incomplete multi-view missing multi-label datasets. The details of these nine methods are as follows:

Table 3: Detailed information on the five multi-view multi-label datasets.

Dataset	# Sample	# Label	# View	# Label/# Sample
Corel5k	4999	260	6	3.40
Pascal07	9963	20	6	1.47
ESPGame	20770	268	6	4.69
IAPRTC12	19627	291	6	5.72
Mirflickr	25000	38	6	4.72

- **DICNet** (Liu et al., 2023a) is an autoencoder-based contrastive network that combines view-specific encoders with instance-level contrastive learning to reduce cross-view semantic inconsistency under missing views and labels.
- **DIMC** (Wen et al., 2023) addresses the simultaneous view and label missingness problem by designing a deep framework that incorporates observation masks into feature fusion and classification.
- **LMVCAT** (Liu et al., 2023b) is a transformer-based framework with a category-aware transformer to capture cross-view complementarity and label correlations under incomplete observations.
- **MTD** (Liu et al., 2024b) addresses the shared-specific feature entanglement by introducing two-channel decoupling and addressing the category collisions problem by cross-channel contrastive learning.
- **AIMNet** (Liu et al., 2024a) addresses the missing-view problem by attention-induced embedding imputation guided by the relationships among available instances.
- **SIP** (Liu et al., 2024c) introduces information bottleneck learning to address the excessive non-shared information, while using label prototype modeling to solve the weak representation label coupling problem.
- **CSCR** (Wang et al., 2025) proposes a category-driven contrastive recovery network that constructs a label coincidence matrix to guide semi-supervised contrastive learning.
- **DFFL** (Xiao et al., 2026) employs unique and consistent feature learning and dual prediction to alleviate feature suppression during multi-view fusion.
- **DCSI** (Wen et al., 2026) addresses the shared-specific information entanglement problem by representation disentanglement and eliminates the unequal view reliability problem by confidence-aware fusion.

A.8 Robust analysis

To verify the robustness of CEE, we evaluated it using five performance metrics across three datasets. We considered three experimental settings: simultaneously increasing both view and label missing rates; varying the view missing rate while keeping the label missing rate fixed at 50% and varying the label missing rate while keeping the view missing rate fixed at 50%. First, as shown in Fig. 5, CEE maintains superior overall performance across all three datasets, even under conditions of simultaneous view and label incompleteness and high missing rates. Second, as illustrated in Fig. 6, with the label missing rate fixed at 50%, CEE continues to outperform other methods across all three datasets as the view missing rate increases. Finally, as shown in Fig. 7, CEE maintains relatively stable performance across all three datasets as the label missing rate rises. It is evident that model performance declines as the missingness rate increases; however, CEE maintains its advantage even under high missingness rates. These results stem from the interplay between MSDL and CDSL. First, MSDL recovers representations containing view-specific distributions, thereby providing more complete view information for subsequent tasks. Meanwhile, upon obtaining this more complete information, CDSL further enhances sample representations by constructing a sample-level feature graph and models label dependencies to mitigate the impact of insufficient supervisory information caused by missing labels. Through the synergistic effect of MSDL and CDSL, our CEE maintains stable performance across various missing-data scenarios.

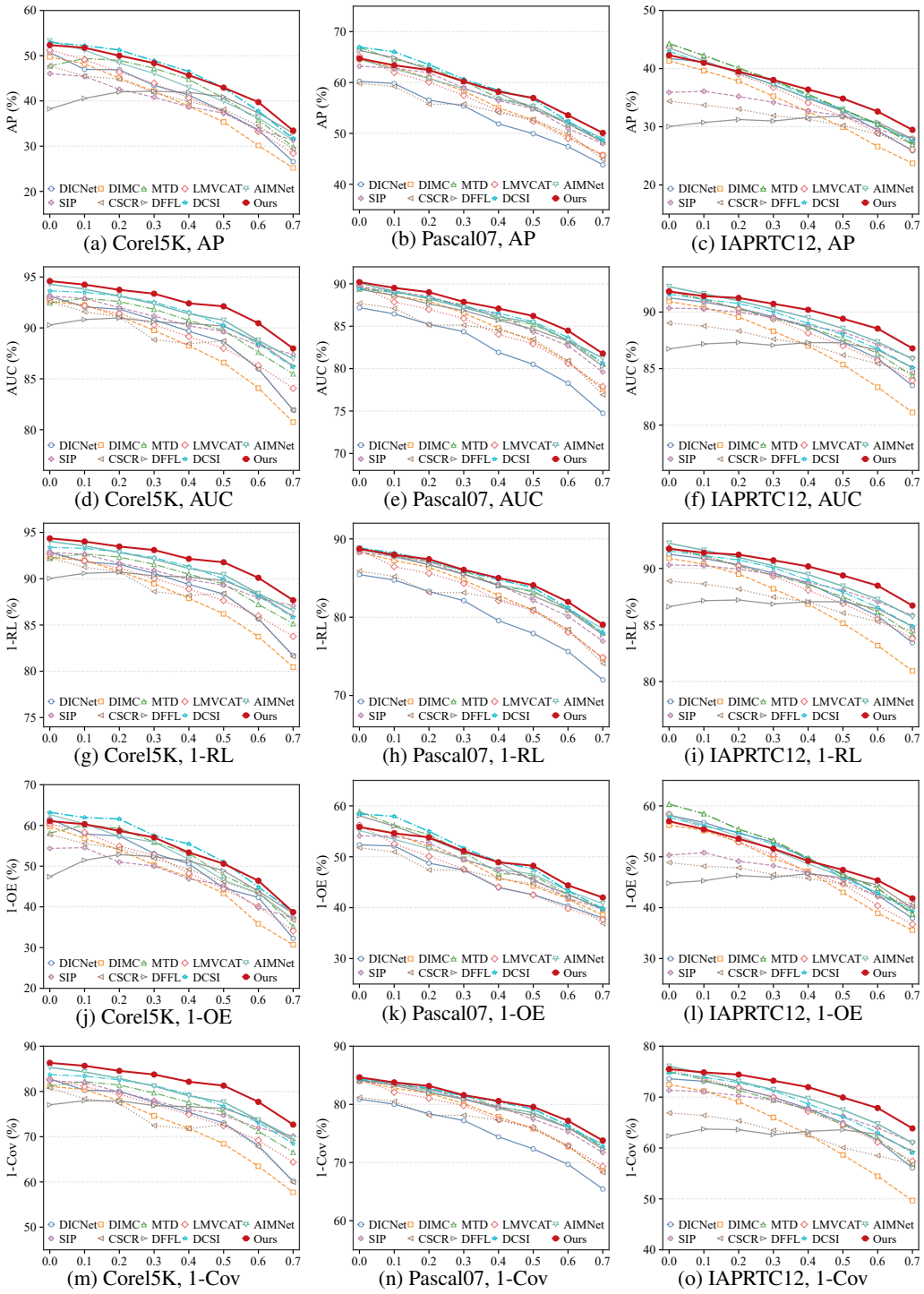


Figure 5: Experimental results of ten methods on three datasets when the label and view missing rates vary simultaneously at the same level. Each column corresponds to a dataset, reporting AP, AUC, 1-RL, 1-OE, and 1-Cov, respectively.

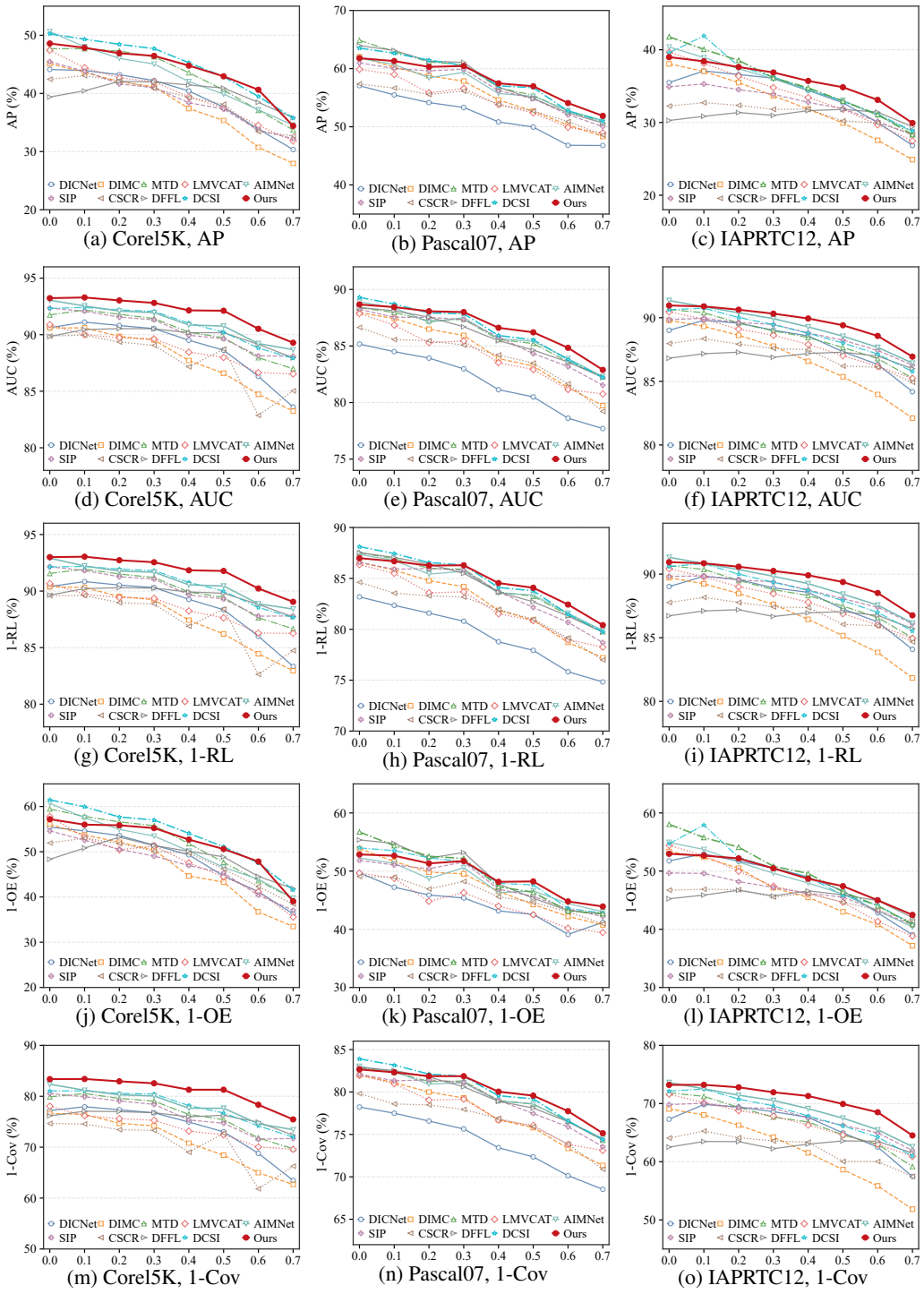


Figure 6: Experimental results of ten methods on three datasets under varying view-missing rates, with the label missing rate fixed at 50%. Each column corresponds to a dataset, reporting AP, AUC, 1-RL, 1-OE, and 1-Cov, respectively.

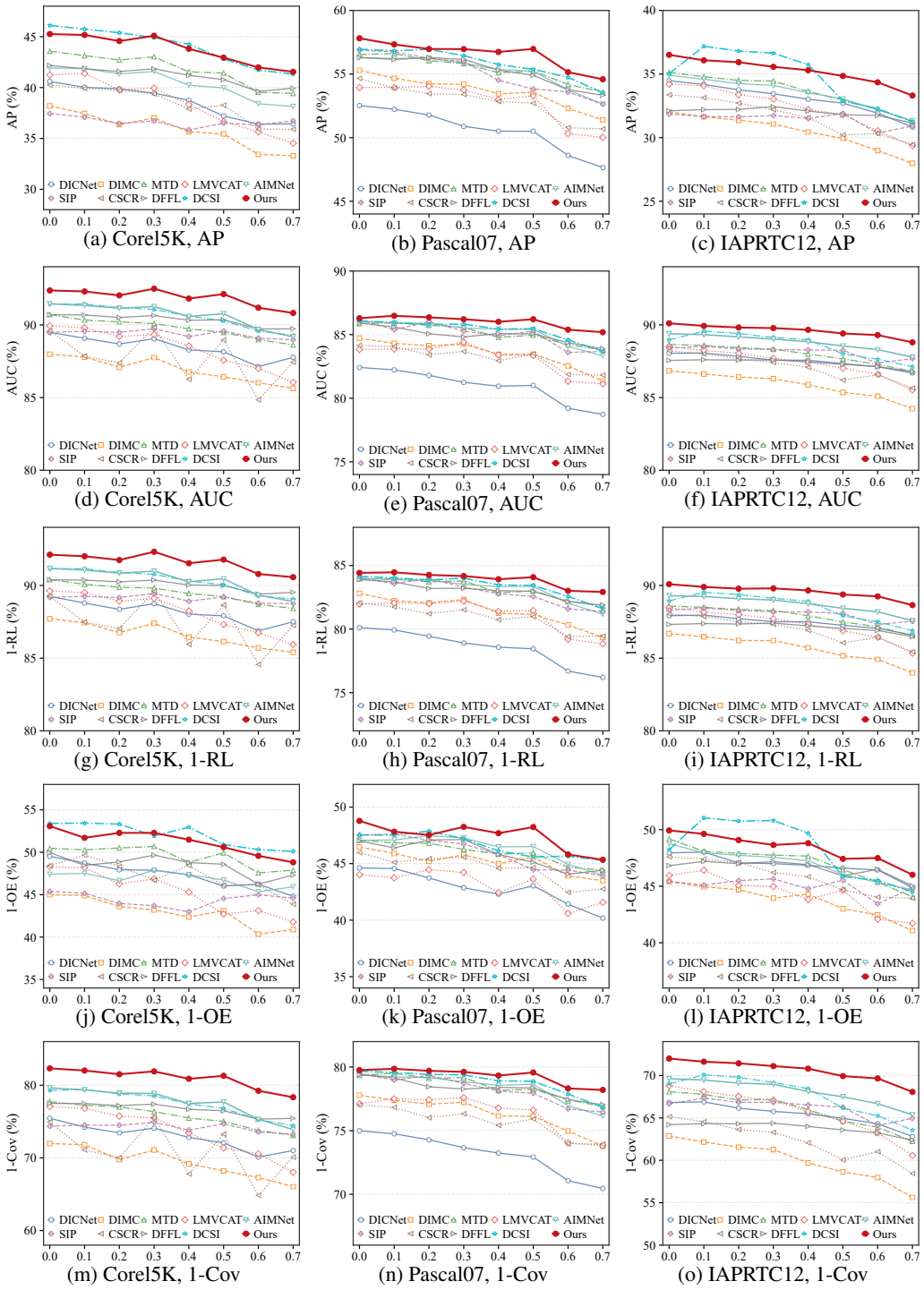


Figure 7: Experimental results of ten methods on three datasets under varying label missing rates, with the view missing rate fixed at 50%. Each column corresponds to a dataset, reporting AP, AUC, 1-RL, 1-OE, and 1-Cov, respectively.