Content-style Disentanglement Guided Representation Learning for Deep Incomplete Multi-view Clustering

Enze Ji[1] , **Meng Liu**[1] , **Yuzhe Li**[1] and **Zhikui Chen**[1,*]
[1]Dalian University of Technology, Dalian, 116620, China
{jienze, liumeng, 1594873855}@mail.dlut.edu.cn, zkchen@dlut.edu.cn

Abstract

Deep incomplete multi-view clustering methods can mine patterns of incomplete multi-view data without labels, gaining great attention in various domains. However, current methods are obsessed with aligning view-specific representations from available samples with complete views to learn view-invariant information for missing data recovery, ignoring fruitful view-complementary information beneficial to data imputations. Meanwhile, such the view-invariant learning often leads to the problem of dominant view dependency that models tend to over-rely on views with clear clustering structures and neglect weaker ones, causing a performance bottleneck. Therefore, a content-style disentanglement guided representation learning is proposed for the incomplete multi-view clustering (CSMVC). Specifically, CSMVC designs a view-specific dual-representation learning architecture that extracts view-invariant content representations and view-specific style representations from self-supervised reconstructions of each view with the guidance of the Hilbert schmidt independence criterion. Then, it performs a content-centered style-modulated representation imputation mechanism to infer missing data via fully utilizing inter-view complementary and consistent information. Meanwhile, it devises a cross-view consensus structure mining strategy to capture the clustering structure with intra-cluster compactness and inter-cluster separability from attention-based fusion representations that adaptively aggregate content representations and style representations of each view. Finally, comprehensive experiments show that CSMVC outperforms state-of-the-art methods.

1 Introduction

The proliferation of multi-view data, where objects are described by multiple complementary feature sets, has significantly advanced pattern recognition and knowledge discovery [Liu *et al.*, 2026]. This paradigm provides richer representations in fields from computational biology to multimodal AI [Chen *et al.*, 2020]. However, practical applications face a fundamental challenge: the theoretical assumption of complete views rarely holds in real data, where instances often have arbitrary missing views. This undermines conventional clustering methods, making Incomplete Multi-View Clustering (IMVC) a critical research frontier [Chen *et al.*, 2023].

Existing deep learning-based incomplete multi-view clustering methods have achieved significant progress, establishing a dominant paradigm [Yuan *et al.*, 2025]. A key advancement is the integration of deep IMVC approaches, which combine clustering objectives with neural networks' representation learning capabilities—particularly autoencoders and their variants [Yuan *et al.*, 2024]. This end-to-end framework enables simultaneous feature extraction, missing data handling, and cluster structure discovery, overcoming limitations of traditional two-stage methods. The success stems from deep models' strong generalization ability, effectively capturing complex data distributions and cross-view relationships under severe incompleteness. For instance, some works use generative adversarial networks to synthesize missing views by learning underlying data distributions [Wang *et al.*, 2021]. Others integrate contrastive learning to enhance incomplete data reconstruction, leveraging intra-view and inter-view relationships to learn discriminative representations while preserving cluster structures [Yang *et al.*, 2023]. These techniques collectively address data incompleteness and multi-view heterogeneity challenges.

Despite these advancements, significant limitations persist. Current methods often struggle to effectively disentangle and leverage the two inherent types of representations in multi-view data: the view-consistent content representation and the view-specific style representation. The semantic content is shared across views and is crucial for clustering, while stylistic attributes, private to each view, can also provide valuable discriminative signals. A predominant shortcoming of prior approaches is their tendency to over-rely on information from dominant views, forcing features from weaker views to align predominantly with these patterns. This imbalance often results in the loss of valuable stylistic elements, ultimately degrading clustering performance. Furthermore, in the context of incomplete data, conventional imputation methods primarily focus on generating a consensus content representa-

*Corresponding author: Zhikui Chen

tion. The neglect of style-specific elements during imputation leads to reconstructed data that is stylistically distorted or semantically shifted. This distributional mismatch introduces noise and undermines the discriminative power of the learned representations.

To address these challenges, we propose a novel framework for IMVC grounded in representation disentanglement theory. Our approach employs a dual-branch encoder to decompose the features of each view into independent content and style representations, with their independence enforced by the Hilbert-Schmidt Independence Criterion (HSIC). This principled disentanglement facilitates a hierarchical imputation strategy for missing views: for content components, we employ statistical normalization leveraging cross-view distributional consistency, while for style elements, we implement a k-nearest neighbor imputation that preserves local manifold structures. Subsequently, we introduce an attention-based fusion mechanism that dynamically aggregates style information across views, while maintaining semantic consistency through a straightforward averaging of content representations. Finally, we present a multi-granularity contrastive learning framework that enforces both cluster-level and sample-level consistency, providing more stable and discriminative learning signals than probability distribution matching alone.

The main contributions can be summarized as follows:

- We introduce a novel paradigm of representation disentanglement by decoupling content-style representation, which ensures integration of heterogeneous data while preserving semantic integrity.
- We design a novel framework that synergizes dual-branch encoder, Hilbert Schmidt Independence Criterion, hierarchical imputation, attention-based fusion, and multi-granularity contrastive learning to achieve robust incomplete multi-view clustering.
- We conduct comprehensive experiments on four real-world multi-view datasets under different missing ratios. Results demonstrate that our method consistently outperforms thirteen state-of-the-art IMVC baselines.

2 Related Work

The challenge of learning from incomplete multi-view data has catalyzed significant innovation in unsupervised deep learning. Inspired by the powerful feature representation capabilities of deep learning, many Deep IMVC methods have been developed. Existing methods can be broadly categorized into imputation-based and imputation-free approaches.

Imputation-free methods circumvent explicit data completion through strategies like knowledge distillation, adaptive projection, or complementarity mining. For example, DIMVC [Xu *et al.*, 2022] employs nonlinear mapping to transform multi-view embeddings into a high-dimensional feature space, exploiting cluster complementarity to achieve linear separability and clustering consistency. I2MVC [Wu *et al.*, 2025] utilizes pseudo-supervised distillation, separating the task into teacher-network clustering on complete data and student-network classification on incomplete data. APADC [Xu *et al.*, 2023] projects available features into a common space via adaptive feature projection, aligning distributions between complete and incomplete instances without direct imputation.

Imputation-based methods employ diverse strategies to reconstruct missing views before performing clustering on completed datasets. These deep IMVC approaches encompass predictor-based, neighborhood-based, and prototype-based paradigms. Predictor-based methods minimize conditional entropy between observed and predicted views to reconstruct missing data [Lin *et al.*, 2021]. Some works unify cross-view consistency learning and missing-view recovery through mutual information optimization [Lin *et al.*, 2023]. Neighborhood-based approaches infer missing entries by identifying nearest neighbors within or across views [Tang and Liu, 2022], with frameworks like SURE addressing view-unaligned and sample-missing problems via noise-robust contrastive learning [Yang *et al.*, 2023]. Prototype-based methods learn representations from missing views using sample-prototype relationships [Jin *et al.*, 2023]. Advanced implementations include RPCIC's cross-view contrastive learning with robust discriminative loss to handle prototype misalignment [Yuan *et al.*, 2024], and PMIMC's relational consistency learning combined with prototype contrastive learning to address prototype-unaligned problems while recovering missing data [Yuan *et al.*, 2025].

Existing methods suffer from a key limitation: they treat multi-view representations as indivisible units. This holistic approach creates a bottleneck, lacking mechanisms to separate consistent semantic content from view-specific stylistic elements. As a result, learned embeddings have suboptimal discriminative power and robustness to missing views. Our approach introduces a paradigm shift through explicit content-style decomposition, enabling hierarchical imputation and multi-granular contrastive learning. This addresses core limitations of previous methods for more effective incomplete multi-view clustering.

3 The Proposed Model

In this section, we first introduce the notations and definitions utilized throughout this paper. Subsequently, we focus on the implementation details and loss function of the proposed CSMVC for Deep IMVC. The architecture of CSMVC is shown in Figure.1

3.1 View-specific Dual-representation Learning

View-specific dual-representation learning (VDL) aims to fully capture consistent content information and complementary style information from available data of each view for clustering. To accomplish this, VDL constructs a dual-branch encoding architecture to extract content representations and style representations from data of each view, and then performs self-supervised reconstructions on concatenations of dual representations to preserve manifold structure of data. Meanwhile, VDL introduces the Hilbert schmidt independence criterion constraint to enhance the information disentanglement between content representations and style representations via the statistical correlation minimization.

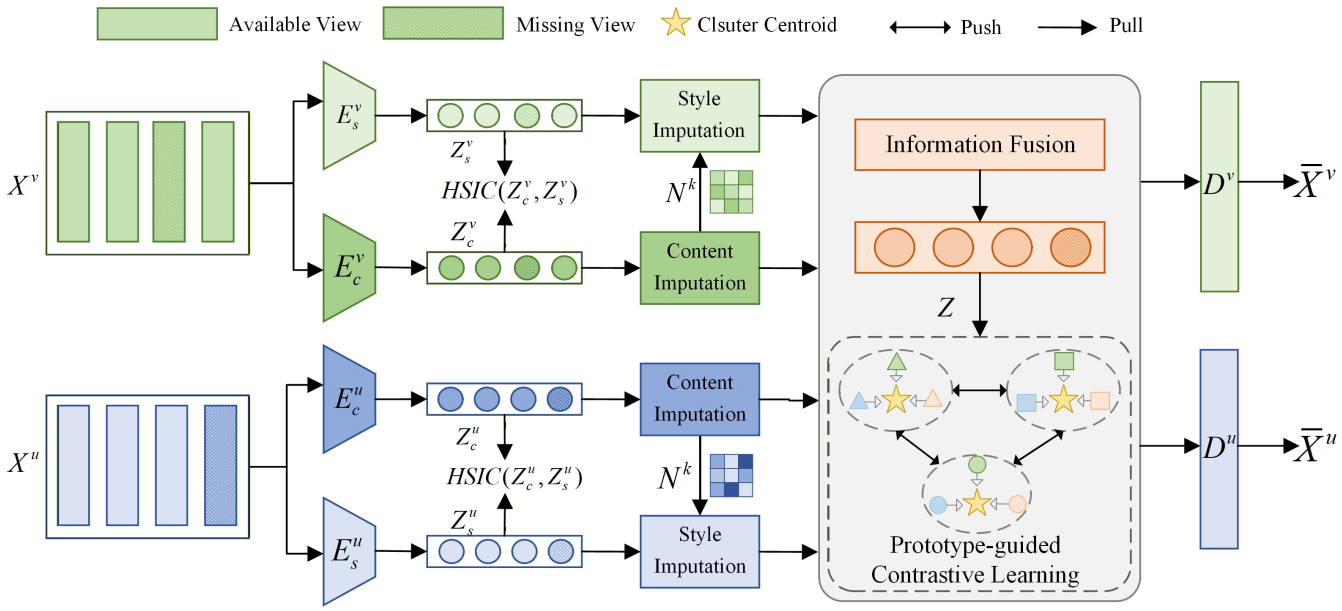


Figure 1: The flowchart of CSMVC.

Specifically, given the available data matrix $\mathbf{X}^v$ of the v-th view, VDL leverages a content encoder $E_c^v(\cdot)$ and a style encoder $E_s^v(\cdot)$ to extract content representations Z_c^v and style representations Z_s^v, respectively:

$$Z_c^v = E_c^v(\mathbf{X}^v), \quad Z_s^v = E_s^v(\mathbf{X}^v) \tag{1}$$

Next, VDL imposes a self-supervised reconstruction loss $\mathcal{L}_{\text{r}}$ between available data and reconstruction data via a decoder $D^v(\cdot)$:

$$\mathcal{L}_{\text{r}} = \sum_{v=1}^{V} \|\mathbf{X}^v - \hat{\mathbf{X}}^v\|_F^2 \tag{2}$$

where $\hat{\mathbf{X}}^v = D^v([Z_c^v \cdot Z_s^v])$ denotes the reconstruction data. $[\cdot]$ denotes concatenation along the feature dimension.

Afterwards, VDL utilizes the Hilbert schmidt independence criterion to minimize the independence between content representations Z_c^v and style representations Z_s^v by evaluating the norm of the cross-covariance operator in a reproducing kernel Hilbert space. In detail, VDL firstly constructs radial basis function kernel matrices $\mathbf{K}_c^v$ and $\mathbf{K}_s^v$ for content representations Z_c^v and style representations Z_s^v in the v-th view, respectively:

$$\mathbf{K}_c^v(i,j) = \exp\left(-\frac{\mathbf{D}(z_{c,i}^v, z_{c,j}^v)}{2\delta^{v2} + \epsilon}\right), \tag{3}$$

$$\mathbf{K}_s^v(i,j) = \exp\left(-\frac{\mathbf{D}(z_{s,i}^v, z_{s,j}^v))}{2\delta_s^{v2} + \epsilon}\right) \tag{4}$$

where $\mathbf{K}_c^v(i,j)$, $\mathbf{K}_s^v(i,j)$ denotes the content, style similarity kernel between the i-th and j-th samples in view v, respectively. $\mathbf{D}(\cdot)$ denotes the Euclidean distance function. $z_{c,i}^v$ and $z_{s,i}^v$ denote the content and style representations of the i-th sample, respectively. The parameters δ_c^{v2} and δ_s^{v2} are set adaptively to the median of the pairwise squared Euclidean distances, ensuring robustness to variations in feature scale. $\epsilon > 0$ is a small constant for numerical stability. $\exp(\cdot)$ denotes the exponential function. Then, VDL computes the Hilbert schmidt independence criterion score between content representations Z_c^v and style representations Z_s^v via:

$$\text{HSIC}(Z_c^v, Z_s^v) = \frac{tr(\tilde{\mathbf{K}}_c^v \tilde{\mathbf{K}}_s^v)}{\sqrt{tr((\tilde{\mathbf{K}}_c^v)^2) \cdot tr((\tilde{\mathbf{K}}_s^v)^2) + \epsilon}} \tag{5}$$

where $tr(\cdot)$ denotes the trace of the matrix. $\tilde{\mathbf{K}}_c^v = \mathbf{H}\mathbf{K}_c^v\mathbf{H}$ and $\tilde{\mathbf{K}}_s^v = \mathbf{H}\mathbf{K}_s^v\mathbf{H}$ denote the centered kernel matrices. $\mathbf{H} = \mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^\top$. $\mathbf{I}$ and $\mathbf{1}$ denote the identity matrix and the all-ones vector, respectively. To implement the information disentanglement between content representations and style representations, VDL defines a HSIC-based regularization loss $\mathcal{L}_{\text{h}}$, as follows:

$$\mathcal{L}_{\text{h}} = \sum_{v=1}^{V} \text{HSIC}(Z_c^v, Z_s^v) \tag{6}$$

3.2 Content-centered Style-modulated Representation Imputation

Content-centered style-modulated representation imputation (CSRI) introduces a unified framework for handling incomplete multi-view data through two objectives: (1) enforcing semantic invariance of content features across views, and (2) recovering missing representations through a imputation strategy that respects the distinct properties of content and style.

The first objective of CSRI is to ensure that semantic content representations exhibit consistent probabilistic cluster assignments across all available data views. To operationalize this principle, CSRI introduces a cross-view content-invariant

learning mechanism applied to the content features from each view, which enforces a coherent, sample-wise alignment of probabilistic assignments between individual views and a unified representation space. Specifically, by formulating a column-wise contrastive objective that treats each cluster as an independent classification task, the framework directly minimizes the distributional divergence between views for shared semantic content. This approach ensures that the learned representations are both discriminative and invariant to view-specific variations, thereby robustly capturing the essential, content-based identity of each sample.

On the top of each view's content features, CSRI constructs a clustering layer k-means with learnable parameters. c_k^v represents the k-th cluster center of the v-th view which can be expressed as:

$$c_k^v = \frac{\sum_{i=1}^n y_{ik}^v z_{c,i}^v}{\sum_{i=1}^n y_{ik}^v} \tag{7}$$

Additionally, DEC [Xie *et al.*, 2016] and IDEC [Guo *et al.*, 2017] are popular single-view deep clustering methods which apply Student's t-distribution to generate soft assignments. A similar way is applied to calculate the output Q^v of the v-th view's clustering layer, which can be described as:

$$q_{ij}^v = \frac{\left(1 + \|z_{c,i}^v - c_j^v\|^2\right)^{-1}}{\sum_{k=1}^K \left(1 + \|z_{c,i}^v - c_k^v\|^2\right)^{-1}} \tag{8}$$

CSRI constructs distribution enhancement on content unified representations $z_{c,i}$

$$p_{ij} = \frac{(q_{ij})^2 / \sum_{i=1}^n q_{ij}}{\sum_{k=1}^K ((q_{ik})^2 / \sum_{i=1}^n q_{ik})} \tag{9}$$

Content invariance principle The key insight is that **content representations should exhibit identical clustering structures across all views**, as they capture the view-invariant semantic information. This principle can be formalized as: $\forall v \in \{1, ..., V\}, q_{ij}^v = p_{ij}$, where q_{ij}^v and p_{ij} represent the probability of sample i belonging to cluster k in view v and the unified representation, respectively.

To accomplish this, CSRI formulates the column contrastive loss as a cross-entropy objective that treats each cluster as an independent classification task. For each cluster j, CSRI utilizes a column-wise contrastive loss that treats the unified distribution P_j as a positive target and the view-specific distribution Q_j^v as its counterpart:

$$\mathcal{L}_{\text{inv}}^v(j) = -\log \frac{\exp(\text{sim}(P_j, Q_j^v))}{\sum_{k=1}^K \exp(\text{sim}(P_j, Q_k^v))} \tag{10}$$

where $\text{sim}(a, b) = a^\top b$ represents the Cosine similarity with temperature $\tau = 1$. The complete column-contrastive loss for view v is obtained by averaging over all clusters:

$$\mathcal{L}_{\text{inv}} = \frac{1}{K} \sum_{v=1}^V \sum_{k=1}^K \mathcal{L}_{\text{inv}}^v(k) \tag{11}$$

The second objective of CSRI is to recover plausible imputations for missing views. To address this joint challenge of imputation, CSRI implements through a strategy that imputation is performed distinctly for content and style representations, guided by their underlying statistical and geometric properties. Specifically, missing content features are recovered via a global, distribution-aware normalization scheme, while absent style attributes are reconstructed using a local, manifold-preserving k-nearest neighbors method.

Content imputation via statistical normalization: For each view v with missing samples, CSRI computes statistics from available samples. For a view v with missing samples, CSRI computes robust estimates of its first- and second-order statistics using only the available samples. The mean μ^v and standard deviation σ^v are calculated as per Eq.(12),

$$\begin{aligned} \mu^v &= \frac{\sum_{i=1}^N M_i^v z_{c,i}^v}{\sum_{i=1}^N M_i^v}, \\ \sigma^v &= \sqrt{\frac{\sum_{i=1}^N M_i^v (z_{c,i}^v - \mu^v)^2}{\sum_{i=1}^N M_i^v} + \epsilon} \end{aligned} \tag{12}$$

where $M_i^v \in \{0, 1\}$ is an availability indicator. A missing content representation $\tilde{z}_{c,i}^v$ is then imputed by conditionally transferring and scaling the normalized content features from other available views, followed by re-centering and re-scaling using the target view's statistics (Eq.(13)). This process leverages the shared semantic content across views under a consistent statistical structure.

$$\tilde{z}_{c,i}^v = \begin{cases} z_{c,i}^v & \text{if } M_i^v = 1 \\ \frac{\sigma^v}{V-1} \sum_{u \neq v} M_i^u \frac{z_{c,i}^u - \mu^u}{\sigma^u} + \mu^v & \text{otherwise} \end{cases} \tag{13}$$

Style imputation via KNN with inverse distance weighting: For missing style representations, CSRI uses content-based KNN imputation. The recovery of missing style representations prioritizes the local geometric structure within the same view. CSRI identifies the top-knearest neighbors for the target sample i within view v based on the Euclidean distance between their (imputed) content features.

$$\mathcal{N}_k(i) = \{j | M_j^v = 1, j \in \text{top-}k \text{ by } \|\tilde{z}_{c,i}^v - \tilde{z}_{c,j}^v\|_2^2\} \tag{14}$$

The missing style representation is then computed as the inverse-distance-weighted average of the style features from these neighbors. This local imputation ensures that the recovered style features are coherent with the local data manifold.

$$\tilde{z}_{s,i}^v = \frac{\sum_{j \in \mathcal{N}_k(i)} w_{ij} z_{s,j}^v}{\sum_{j \in \mathcal{N}_k(i)} w_{ij} + \epsilon}, \quad w_{ij} = \frac{1}{\|\tilde{z}_{c,i}^v - \tilde{z}_{c,j}^v\|_2^2 + \epsilon} \tag{15}$$

3.3 Cross-view Consensus Structure Mining

Cross-view consensus structure mining (CCSM) presents a unified framework for discovering consensus structures across incomplete views through two complementary mechanisms: (1) an attention-based fusion strategy that dynamically

integrates style representations while preserving content semantics, and (2) a prototype-guided contrastive learning approach that enhances cluster structure discriminability.

To integrate information across views, CCSM first proposes a parameter-efficient attention-based fusion mechanism that learns to dynamically weigh and combine style representations while preserving the semantic consistency of content features. The process consists of three principled stages: dual projection, gated attention weighting, and normalized fusion.

Dual projection: To enable structured cross-view interaction without excessive parameter growth, CCSM projects the style representation $\tilde{z}_s^v$ for each view vinto two distinct latent spaces using globally shared projection matrices:

$$Z_p^v = W_p \tilde{z}_s^v, \quad Z_o^v = W_o \tilde{z}_s^v \tag{16}$$

These projections separate the roles of similarity computation Z_p^v and value aggregation Z_o^v, promoting efficiency.

Gated attention weights: Inter-view affinities are captured via the similarity matrix $\mathbf{Y}_{uv} = Z_p^u {Z_p^v}^T$. Rather than applying a standard soft attention, CCSM utilizes a gating mechanism to modulate attention scores nonlinearly. The raw score for view v is computed as:

$$s_v = \sum_{u \neq v} \mathbf{Y}_{uv} \cdot \sigma(\gamma \mathbf{Y}_{uv} + \beta) \tag{17}$$

where $\sigma(\cdot)$ denotes the sigmoid activation, and γ, β are globally gating parameters, which allows the model to suppress or amplify cross-view interactions based on the magnitude of similarity, providing a form of adaptive conditioning.

Normalized fusion: The scores s_v are normalized across views via softmax to obtain attention weights α_v, which are used to fuse the style projections. Meanwhile, CCSM applies a simple average to maintain stability and avoid distorting semantically consistent information for content representations:

$$Z_s^{\mathrm{f}} = \sum_{v=1}^{V} \alpha_v Z_o^v, \quad Z_c^{\mathrm{f}} = \frac{1}{V} \sum_{v=1}^{V} \tilde{z}_c^v \tag{18}$$

The final unified multimodal representation is the concatenation of the fused content and style components: $Z = [Z_c^{\mathrm{f}}, Z_s^{\mathrm{f}}] \in \mathbb{R}^{N \times (D_c + D_s)}$. This design ensures that view-invariant content is preserved, while view-specific stylistic information is integrated through a conditioned attention mechanism, yielding a compact yet expressive joint representation.

Then, CCSM leverages prototype-guided contrastive learning to explicitly optimize the alignment between samples and their corresponding cluster centroids. Specifically, the assigned centroid of each sample acts as a positive anchor, while all other centroids serve as negative counterparts. This formulation drives the latent representations to form a well-structured and separable manifold, thereby simultaneously enhancing intra-cluster compactness and inter-cluster separation. As a result, the proposed approach not only sharpens cluster boundaries but also systematically promotes representation discriminability in an unsupervised manner.

Given the unified representation Z and the assigned cluster centroids $C \in \mathbb{R}^{K \times D}$, the row contrastive loss is formulated as an InfoNCE (Noise-Contrastive Estimation) loss that treats each sample's assigned centroid as a positive pair and all other centroids as negative pairs:

$$\mathcal{L}_{\mathrm{c}} = -\frac{1}{N} \sum_{i=1}^{N} \log \frac{\exp(\mathrm{sim}(z_i, c_j))}{\sum_{j'=1}^{N} \exp(\mathrm{sim}(z_i, c_{j'}))} \tag{19}$$

where c_j, $c_{j'}$ is the centroid assigned to cluster j, j', respectively. This formulation promotes intra-cluster compactness and inter-cluster separation, effectively reinforcing semantic coherence across views.

The total training objective integrates multiple complementary learning signals:

$$\mathcal{L} = \mathcal{L}_{\mathrm{r}} + \lambda_1 \mathcal{L}_{\mathrm{h}} + \lambda_2 \mathcal{L}_{\mathrm{inv}} + \lambda_3 \mathcal{L}_{\mathrm{c}} \tag{20}$$

where λ_1, λ_2 and λ_3 are trade-off parameters

4 Experiment

4.1 Experimental Setup

Datasets. The experimental evaluation in this study utilizes four benchmark multi-view clustering datasets widely adopted in the literature: Caltech101-7 [Fei-Fei *et al.*, 2007], HandWritten [LeCun *et al.*, 1989], AWA-7 [Romera-Paredes and Torr, 2015], and ALOI-100 [Geusebroek *et al.*, 2005]. These datasets are selected for their diversity in modality composition, sample sizes, and the number of available views, thereby comprehensively evaluating the robustness and generalization ability of the proposed method. Detailed statistical descriptions of these datasets are provided in Table.1. To show the performance of CSMVC on missing multi-view data, we set the MR(missing rate) to [0.1, 0.3, 0.5, 0.7].

Dataset	Samples	Clusters	Dimensionality
Caltech101-7	1,400	7	1984/512/928/254/40
HandWritten	2,000	10	240/76/216/47/64/6
AWA-7	10,158	50	2688/2000/2000/2000/2000/4096/4096
ALOI-100	10,800	100	77/13/64/125

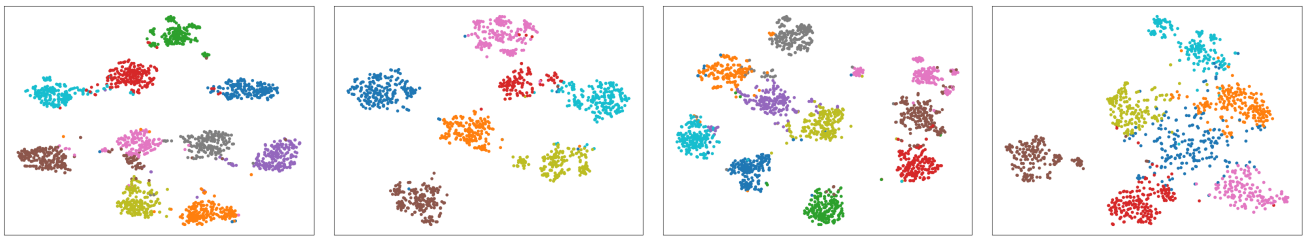
Table 1: Statistics of four benchmark datasets.

Comparison methods. To validate the superiority of the proposed method, we conduct comprehensive comparisons with 11 state-of-the-art approaches, including COMPLETER [Lin *et al.*, 2021], SURE [Yang *et al.*, 2023], DSIMVC [Tang and Liu, 2022], DIMVC [Xu *et al.*, 2022], ProImp [Li *et al.*, 2023], DCP [Lin *et al.*, 2023], APADC [Xu *et al.*, 2023], CPSPAN [Jin *et al.*, 2023], IMVC-IE [Huang *et al.*, 2024], ICMVC [Chao *et al.*, 2024], GHICMC [Chao *et al.*, 2025], PMIMC [Yuan *et al.*, 2025], CAPIMAC [Ma *et al.*, 2025]. In all experiments, each baseline is implemented according to its original configuration, and a grid search is applied to its suggested hyperparameter space to ensure optimal performance.

Evaluation metrics. To assess the performance of our proposed framework, we employ three widely adopted evaluation metrics in clustering tasks: ACC, NMI, and F-mea. These metrics provide complementary perspectives on clustering quality, enabling a comprehensive evaluation from different aspects. The larger value of the metrics indicates the better clustering performance.

	Method	MR=0.1			MR=0.3			MR=0.5			MR=0.7		
		ACC	NMI	F-mea	ACC	NMI	F-mea	ACC	NMI	F-mea	ACC	NMI	F-mea
Caltech101-7	COMPLETER(CVPR'21)	54.24	59.28	53.72	68.94	60.37	66.80	58.13	63.53	56.12	55.53	56.79	53.48
	SURE(TPAMI'22)	82.00	76.89	81.86	73.79	67.83	73.16	75.86	69.66	74.45	75.64	65.67	75.11
	DSIMVC(ICLR'22)	67.44	57.97	67.55	68.23	58.14	68.19	52.36	46.18	52.15	43.91	38.63	42.53
	DIMVC(AAAI'22)	82.21	67.22	66.31	77.32	71.50	62.26	75.21	56.09	55.62	60.43	41.47	46.34
	ProImp(IICAI'23)	75.17	65.40	75.02	81.69	69.34	81.58	74.94	63.85	74.90	71.97	58.61	72.10
	DCP(PAMI'23)	56.53	58.70	55.93	54.66	54.92	53.47	55.37	59.64	52.51	33.66	38.16	28.95
	APADC(TIP'23)	73.70	68.54	73.92	75.57	71.66	74.35	83.46	75.07	82.18	59.14	56.00	53.44
	CPSPAN(CVPR'23)	89.36	82.00	89.05	86.00	78.04	85.68	86.00	75.40	85.44	85.57	75.74	85.22
	IMVC-IE(ICASSP'24)	69.78	51.98	68.57	62.42	51.65	61.71	59.21	45.14	47.65	51.42	34.16	51.92
	ICMVC(AAAI'24)	84.93	77.70	85.14	83.21	74.87	83.50	79.29	69.21	79.22	72.07	65.64	71.95
	GHICMC(AAAI'25)	88.00	81.49	87.64	86.00	81.67	75.93	85.14	69.29	75.22	76.43	67.31	75.83
	PMIMC(TIP'25)	90.14	81.10	89.82	89.36	81.97	89.20	89.29	80.37	88.90	88.71	79.92	**88.36**
	CAPIMAC(IJCAI'25)	84.32	78.35	84.74	83.11	77.32	80.14	82.14	77.12	78.63	79.31	76.62	76.69
	Ours	**90.57**	**83.30**	**90.44**	**89.93**	**82.55**	**89.93**	**89.79**	**81.20**	**89.41**	**88.71**	**79.97**	87.80
HandWritten	COMPLETER(CVPR'21)	76.11	79.53	71.66	73.76	76.16	73.32	72.50	71.57	73.16	81.39	78.50	80.83
	SURE(TPAMI'22)	66.85	58.02	66.28	73.70	63.31	73.26	73.35	63.22	72.98	69.85	59.95	68.36
	DSIMVC(ICLR'22)	79.00	74.81	79.08	78.27	74.15	78.33	71.17	68.53	71.09	69.06	64.57	68.95
	DIMVC(AAAI'22)	63.15	58.72	61.85	60.15	53.35	55.88	59.75	48.78	57.01	41.45	29.83	38.55
	ProImp(IICAI'23)	84.64	82.07	84.37	83.12	78.78	82.83	80.25	74.62	68.53	80.78	70.96	80.44
	DCP(PAMI'23)	72.95	75.33	54.78	72.00	72.11	51.07	71.46	74.16	54.69	58.43	63.44	32.10
	APADC(TIP'23)	81.67	83.80	80.02	81.65	83.75	79.56	78.37	79.86	75.26	57.03	65.26	51.90
	CPSPAN(CVPR'23)	88.75	81.94	88.68	91.05	83.29	91.08	78.85	78.85	77.05	87.55	78.15	87.51
	IMVC-IE(ICASSP'24)	76.95	68.54	75.45	73.15	69.34	73.80	68.95	56.11	69.40	69.45	53.00	69.34
	ICMVC(AAAI'24)	86.05	82.57	85.86	85.20	83.77	85.20	80.85	77.00	81.47	74.05	70.37	75.15
	GHICMC(AAAI'25)	90.75	88.08	90.47	87.25	86.62	87.13	87.85	85.48	88.04	85.65	83.83	85.66
	PMIMC(TIP'25)	90.70	83.82	90.68	92.75	86.34	92.74	90.85	83.74	90.89	89.95	82.05	89.92
	CAPIMAC(IJCAI'25)	89.27	87.98	90.62	88.34	86.69	87.77	87.24	84.93	87.52	86.37	83.30	86.24
	Ours	**93.10**	**88.84**	**93.14**	**94.15**	**87.94**	**94.16**	**92.30**	**85.94**	**92.33**	**91.70**	**84.02**	**91.74**
ALOI-100	COMPLETER(CVPR'21)	56.64	79.03	54.28	45.59	73.31	42.92	33.72	65.82	31.49	38.74	68.12	35.87
	SURE(TPAMI'22)	33.71	71.34	30.09	29.82	67.75	25.07	28.94	67.29	23.25	36.19	64.92	22.06
	DSIMVC(ICLR'22)	64.72	72.90	45.64	63.75	70.47	44.57	60.26	71.93	42.48	57.02	66.60	39.68
	DIMVC(AAAI'22)	67.50	81.29	63.56	59.70	73.81	57.87	55.17	68.91	52.59	54.32	67.78	51.71
	ProImp(IICAI'23)	67.17	81.71	66.31	43.22	70.21	42.78	32.44	64.33	33.06	28.64	61.13	29.83
	DCP(PAMI'23)	50.77	78.35	43.11	46.67	73.43	44.48	41.80	70.91	39.82	39.24	68.32	36.84
	APADC(TIP'23)	35.81	59.96	33.22	34.54	58.11	30.58	29.02	54.74	27.11	23.39	50.79	23.62
	CPSPAN(CVPR'23)	71.69	85.30	68.55	67.96	82.19	66.34	66.62	82.39	63.74	65.44	82.13	62.50
	IMVC-IE(ICASSP'24)	37.75	64.65	39.37	35.13	60.40	36.90	29.74	55.76	31.59	27.12	53.83	28.95
	ICMVC(AAAI'24)	73.06	84.61	69.81	71.99	82.99	69.13	66.99	80.95	64.93	58.46	74.26	57.90
	GHICMC(AAAI'25)	75.56	84.89	74.63	71.92	83.26	71.32	63.34	79.17	62.22	66.44	79.55	65.70
	PMIMC(TIP'25)	74.71	87.41	71.92	75.30	87.92	72.31	71.18	86.35	68.26	72.11	86.79	68.90
	CAPIMAC(IJCAI'25)	73.93	81.46	74.72	72.15	82.26	69.99	70.58	82.32	72.78	70.44	82.46	70.65
	Ours	**83.79**	**91.78**	**82.61**	**80.20**	**89.32**	**79.04**	**79.18**	**88.11**	**77.86**	**76.86**	**87.19**	**75.94**
AWA-7	COMPLETER(CVPR'21)	35.30	51.76	22.70	31.10	47.37	19.90	31.63	47.62	20.10	30.33	45.39	19.81
	SURE(TPAMI'22)	34.22	52.61	25.85	27.28	44.28	18.13	28.91	42.11	20.00	30.55	46.10	20.18
	DSIMVC(ICLR'22)	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M
	DIMVC(AAAI'22)	24.28	37.90	24.45	24.60	37.89	24.63	23.41	37.14	23.60	24.20	37.20	24.25
	ProImp(IICAI'23)	42.75	37.23	41.69	40.75	36.89	38.90	37.80	29.78	34.14	39.70	28.90	34.76
	DCP(PAMI'23)	32.81	45.74	29.06	30.69	43.91	27.27	30.70	43.07	27.13	30.40	42.36	26.86
	APADC(TIP'23)	27.32	47.02	18.38	23.71	42.76	15.67	23.45	43.06	15.15	22.39	40.11	13.34
	CPSPAN(CVPR'23)	44.91	55.12	47.11	44.74	57.23	46.30	43.94	53.82	45.28	38.48	49.03	39.88
	IMVC-IE(ICASSP'24)	24.96	39.65	36.95	24.85	39.88	32.09	21.58	36.33	28.41	19.88	29.88	18.31
	ICMVC(AAAI'24)	47.77	61.81	45.13	49.08	63.05	45.17	49.45	61.72	46.17	46.35	57.82	43.12
	GHICMC(AAAI'25)	60.43	67.45	64.36	57.79	65.59	63.07	60.86	66.82	63.50	55.00	62.53	59.86
	PMIMC(TIP'25)	57.37	67.13	59.94	56.18	66.57	57.96	55.41	65.99	57.56	52.51	64.56	54.14
	CAPIMAC(IJCAI'25)	58.22	66.78	63.47	58.23	66.87	63.52	60.75	67.54	63.97	57.11	64.24	60.02
	Ours	**63.03**	**70.84**	**66.09**	**60.83**	**68.73**	**63.99**	**61.51**	**68.87**	**64.98**	**57.31**	**65.69**	**61.49**

Table 2: Performance comparison across four datasets with four distinct missing rates. The optimal results are highlighted in bold. 'O/M' represents out-of-memory



(a) HandWritten at MR=0.5 (b) Caltech101-7 at MR=0.5 (c) HandWritten at MR=W0.7 (d) Caltech101-7 at MR=0.7

Figure 2: T-SNE visualization of the fusion representation of samples obtained by different methods on the four datasets with 70% missing views.

4.2 Clustering performance

Based on the comprehensive experimental results presented in Table.2, the proposed framework demonstrates statistically significant improvements over all compared methods, achieving top rankings in ACC, NMI, and F-measure metrics. Three key conclusions emerge: (1) Our method consistently outperforms state-of-the-art approaches across all missing rates on four benchmark datasets, showing balanced improvements in ACC, NMI, and F-measure. This indicates our approach better captures intrinsic data structure rather than optimizing for a single metric. (2) Performance degradation with increasing missing rates is effectively mitigated, with the method maintaining stable, high scores. The advantage becomes particularly pronounced at higher missing rates (MR=0.5 and 0.7), demonstrating enhanced robustness to severe data incompleteness. (3) The method shows remarkable stability and generalizability, maintaining leading performance across datasets of varying scales, modalities, and cluster complexities. Superior results on the challenging AWA-7 dataset highlight the framework's scalability and discriminative power.

4.3 Ablation Study

To validate the effectiveness of the key components in our proposed method, we designed several variant models: CSMVC w/o hsic, CSMVC w/o inv, and CSMVC w/o con represent removal of hsic, content-invariant learning, and consensus structure mining, respectively. From Table.3, we find: (1) All three ablated variants consistently inferior performance compared to the complete MASTER framework across both datasets. This systematic performance degradation upon removal of any individual component confirms that each module contributes essentially to the overall clustering efficacy. (2) The full model achieves the highest accuracy (ACC) and normalized mutual information (NMI) scores on both benchmark datasets, substantiating not only the effectiveness of each constituent module but also the synergistic integration of the complete architectural design.

		ACC			NMI		
Dataset	**Method**	**10%**	**30%**	**50%**	**10%**	**30%**	**50%**
Caltech101-7	CSMVC	90.14	89.93	89.79	83.30	82.55	81.20
	w/o hsic	89.21	87.86	84.29	81.94	78.65	72.04
	w/o hsic + inv	85.78	80.86	84.50	77.66	70.88	72.51
	w/o inv + con	87.92	79.93	82.35	79.14	70.05	68.43
HandWritten	CSMVC	93.10	94.15	92.30	88.84	87.94	85.94
	w/o hsic	88.08	87.32	86.18	87.63	85.16	83.14
	w/o hsic + inv	87.66	84.92	82.53	85.89	81.48	81.29
	w/o inv + con	84.77	82.29	80.57	81.21	80.42	77.04

Table 3: Ablation study results.

4.4 Model Analyses

Hyper-parameters. Figure.3 presents a comprehensive hyperparameter sensitivity analysis on the Caltech101-7 dataset at MR = 0.7 and ALOI-100 at MR = 0.3, evaluating the influence of the trade-off parameters λ_1, λ_2, and λ_3, on clustering accuracy (ACC). In the analysis, each trade-off parameter is varied over the set $\{0.1, 0.01, 0.001, 0.0001, 0.00001\}$ while the other two are adjusted accordingly. These experimental results show that our method is insensitive to changes in parameters and can achieve great performance in a suitable parameter range: it maintains stable, near-optimal performance when $\lambda_2 \in \{0.1, 0.01, 0.001\}$, and $\lambda_3 \in \{0.1, 0.01, 0.001\}$.

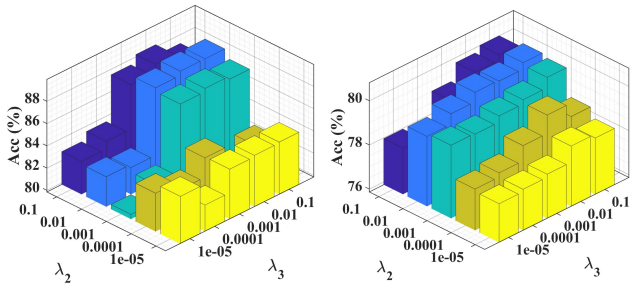


Figure 3: Hyper-parameters analysis on the Caltech101-7 dataset at MR = 0.7 and ALOI-100 at MR = 0.3.

Visualization. Figure.2 presents a qualitative visualization of the clustering results produced by the proposed CSMVC framework on HandWritten and Caltech101-7 datasets at MR = 0.5 and 0.7. For each dataset, the approach performs content-style fusion after completing the missing data imputation across the entire dataset. These representations are subsequently projected into a two-dimensional space via t-SNE for visual assessment. The resulting visualizations demonstrate that samples belonging to the same cluster are tightly grouped with clear spatial separation between different clusters. This distinct, well-separated cluster structure provides visual evidence that CSMVC effectively captures discriminative patterns, even in high missing rates, ensuring both intra-cluster compactness and inter-cluster separation in the learned representation space.

5 Conclusion

In this paper, we address the critical challenge of incomplete multi-view clustering by proposing CSMVC, a novel framework based on content-style disentanglement guided representation learning. Unlike existing methods that struggle with dominant view dependency and neglect view-complementary information, CSMVC introduces a principled approach to extract and leverage both view-invariant content representations and view-specific style representations. Our key contributions are: (1) A dual-representation learning architecture using Hilbert-Schmidt Independence Criterion to disentangle content and style via statistical correlation minimization. (2) A content-centered style-modulated imputation mechanism that separately recovers content and style, preserving semantic consistency and local structures. (3) A cross-view consensus mining strategy with attention fusion and prototype-guided contrastive learning, adaptively aggregating information while enhancing cluster separability.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China(No.62476038).

References

[Chao *et al.*, 2024] Guoqing Chao, Yi Jiang, and Dianhui Chu. Incomplete contrastive multi-view clustering with high-confidence guiding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 11221–11229, 2024.

[Chao *et al.*, 2025] Guoqing Chao, Kaixin Xu, Xijiong Xie, and Yongyong Chen. Global graph propagation with hierarchical information transfer for incomplete contrastive multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 15713–15721, 2025.

[Chen *et al.*, 2020] Man-Sheng Chen, Ling Huang, Chang-Dong Wang, and Dong Huang. Multi-view clustering in latent embedding space. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3513–3520, 2020.

[Chen *et al.*, 2023] Jin Chen, Huafu Xu, Jingjing Xue, Quanxue Gao, Cheng Deng, and Ziyu Lv. Incomplete multi-view clustering based on hypergraph. *Information Fusion*, page 102155, 2023.

[Fei-Fei *et al.*, 2007] Li Fei-Fei, Rob Fergus, and Pietro Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, 106(1):59–70, 2007.

[Geusebroek *et al.*, 2005] JM Geusebroek, GJ Burghouts, and AWM Smeulders. The amsterdam library of object images. *International Journal of Computer Vision*, 61(1):103–112, 2005.

[Guo *et al.*, 2017] Xifeng Guo, Long Gao, Xinwang Liu, and Jianping Yin. Improved deep embedded clustering with local structure preservation. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1753–1759, 2017.

[Huang *et al.*, 2024] Binqiang Huang, Zhijie Huang, Shoujie Lan, Qinghai Zheng, and Yuanlong Yu. Incomplete multi-view clustering via inference and evaluation. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8180–8184, 2024.

[Jin *et al.*, 2023] Jiaqi Jin, Siwei Wang, Zhibin Dong, Xinwang Liu, and En Zhu. Deep incomplete multi-view clustering with cross-view partial sample and prototype alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11600–11609, 2023.

[LeCun *et al.*, 1989] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989.

[Li *et al.*, 2023] Haobin Li, Yunfan Li, Mouxing Yang, Peng Hu, Dezhong Peng, and Xi Peng. Incomplete multi-view clustering via prototype-based imputation. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 3911–3919, 2023.

[Lin *et al.*, 2021] Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. Completer: Incomplete multi-view clustering via contrastive prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11169–11178, 2021.

[Lin *et al.*, 2023] Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4447–4461, 2023.

[Liu *et al.*, 2026] Meng Liu, Yuzhe Li, Zhikui Chen, and Yilong Lin. Information-driven complementarity and consistency mining for multi-view clustering. *IEEE Signal Processing Letters*, 33:216–220, 2026.

[Ma *et al.*, 2025] Shubin Ma, Liang Zhao, Mingdong Lu, Yifan Guo, and Bo Xu. Consistency-aware padding for incomplete multi-modal alignment clustering based on self-repellent greedy anchor search. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 5923–5931, 2025.

[Romera-Paredes and Torr, 2015] Bernardino Romera-Paredes and Philip H. S. Torr. An embarrassingly simple approach to zero-shot learning. In *Proceedings of the International Conference on Machine Learning*, volume 37, pages 2152–2161, 2015.

[Tang and Liu, 2022] Huayi Tang and Yong Liu. Deep safe incomplete multi-view clustering: Theorem and algorithm. In *Proceedings of International Conference on Machine Learning*, 2022.

[Wang *et al.*, 2021] Qianqian Wang, Zhengming Ding, Zhiqiang Tao, Quanxue Gao, and Yun Fu. Generative partial multi-view clustering with adaptive fusion and cycle consistency. *IEEE Transactions on Image Processing*, 30:1771–1783, 2021.

[Wu *et al.*, 2025] Benyu Wu, Wei Du, Jun Wang, and Guoxian Yu. Imputation-free incomplete multi-view clustering via knowledge distillation. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 6570–6578, 2025.

[Xie *et al.*, 2016] Junyuan Xie, Ross Brook Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *Proceedings of International Conference on Machine Learning*, pages 478–487, 2016.

[Xu *et al.*, 2022] Jie Xu, Chao Li, Yazhou Ren, Liang Peng, Yujie Mo, Xiaoshuang Shi, and Xiaofeng Zhu. Deep incomplete multi-view clustering via mining cluster complementarity. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8761–8769, 2022.

[Xu *et al.*, 2023] Jie Xu, Chao Li, Liang Peng, Yazhou Ren, Xiaoshuang Shi, Heng Tao Shen, and Xiaofeng Zhu. Adaptive feature projection with distribution alignment for deep incomplete multi-view clustering. *IEEE Transactions on Image Processing*, 32:1354–1366, 2023.

[Yang *et al.*, 2023] Mouxing Yang, Yunfan Li, Peng Hu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Robust multi-

view clustering with incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):1055–1069, 2023.

[Yuan *et al.*, 2024] Honglin Yuan, Shiyun Lai, Xingfeng Li, Jian Dai, Yuan Sun, and Zhenwen Ren. Robust prototype completion for incomplete multi-view clustering. In *Proceedings of the ACM International Conference on Multimedia*, pages 10402–10411, 2024.

[Yuan *et al.*, 2025] 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.