SEMANTICS-GUIDED DISENTANGLED MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Existing deep multi-view clustering methods commonly adopt dual-branch paradigms to extract and fuse view-shared and view-specific features, yet they suffer from inherent structural limitations. Direct aggregation of these features introduces view-irrelevant noise and underutilizes cross-view complementary information, while partitioning features by view origins does not align with clustering objectives, due to redundant view-shared features and underexploited discriminative cues in view-specific features. To address these issues, we propose a Semantics-Guided Disentangled paradigm for Multi-View Clustering (SG-DMVC). Specifically, we explicitly decompose the input of each view into discriminative cluster-relevant features and task-agnostic cluster-irrelevant features. The view-wise cluster-relevant features are aggregated via log-space probability consensus to yield robust cross-view shared semantics, which are further regularized by dual-level topological alignment to preserve global manifold distribution consistency and refine local sample-pair structures. Meanwhile, covariance decorrelation regularization is imposed to enforce second-order decorrelation between cluster-relevant and cluster-irrelevant features, effectively eliminating irrelevant feature interference and latent information leakage. Extensive experiments on both complete and incomplete multi-view datasets show that the proposed SG-DMVC yields superior clustering performance and strong robustness against view missing compared with state-of-the-art methods.

1 INTRODUCTION

Real-world data typically exhibit multi-view properties, where an object is characterized by heterogeneous representations from different sensors, viewpoints or feature-extraction pipelines (Zhang et al., 2019). As a key unsupervised-learning task, multi-view clustering exploits semantic correlations across views to fuse consistent and complementary information for assigning unlabeled samples into clusters, with wide applications in bioinformatics analysis (Li et al., 2025), multimedia mining (Hu et al., 2024), and fault detection (Huang et al., 2024).

Conventional multi-view clustering employs heuristic shallow methods, categorized into latent representation learning, graph learning, nonnegative matrix factorization and tensor learning (Fang et al., 2023). Constrained by linear modeling assumptions, these methods lack sufficient representation capacity for high-dimensional, highly-nonlinear real-world data and struggle with complex view heterogeneity and noise (Zheng et al., 2022; Ma et al., 2024). Driven by the strong nonlinear fitting capability of neural networks, deep multi-view clustering has emerged as a dominant research direction (Wang et al., 2024). Disentangled representation learning offers a promising remedy for multi-view heterogeneity (Zou et al., 2026). Existing works commonly adopt binary decomposition to separate view-shared and view-specific latent components, which mitigates feature entanglement, lowers representation redundancy and boosts clustering performance.

Despite performance gains from the shared-specific disentanglement paradigm, it exhibits fundamental flaws. Existing disentanglement strategies decouple features purely based on view origins instead of optimizing semantic discriminability for downstream clustering (Chen et al., 2025; Xu et al., 2021; Ke et al., 2023; Chen et al., 2026). This coarse-grained partitioning yields two major drawbacks. Direct concatenation of view-shared and view-specific features injects view-irrelevant noise and discards clustering-oriented cross-view complementary information (Xu et al., 2023a).

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.

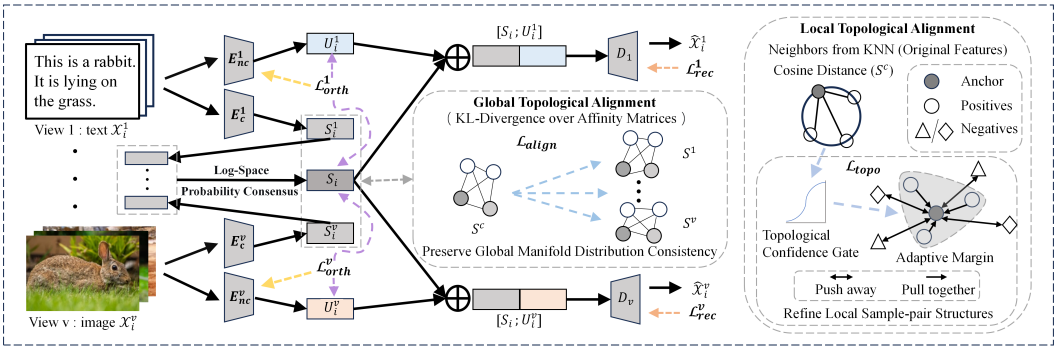


Figure 1: Overview of the proposed SG-DMVC framework. It disentangles multi-view inputs into cluster-relevant and cluster-irrelevant latents, leveraging log-space probability consensus, dual-level topological alignment and covariance decorrelation regularization for clustering assignments.

View-specific components are not pure noise and often contain vital clustering-discriminative semantics (Chen et al., 2025), while view-shared components carry substantial task-irrelevant redundancy (Ke et al., 2023). However, view-origin-driven disentanglement cannot guarantee that view-shared features benefit clustering and view-specific features do not. As a result, decomposed latent representations remain severely entangled, greatly impairing model clustering performance.

To address the aforementioned issues, this paper proposes a Semantics-Guided Disentangled Multi-View Clustering (SG-DMVC). The overall framework is illustrated in Figure 1. Grounded in information-theoretic latent factorization and mutual information upper bound theory, the method adopts a semantics-guided disentanglement paradigm to decompose single-view latent representations into cluster-relevant and cluster-irrelevant latent features. Log-space probability consensus is employed to capture consistent cross-view cluster semantics and suppress view-specific manifold perturbations. A dual-level topological alignment mechanism constrains global manifold distribution consistency and refines local sample-pair structures to stabilize unsupervised clustering optimization. Then, covariance decorrelation regularization enforces second-order decorrelation of dual latent subspaces, eliminating latent information leakage and mitigating geometric misalignment and data noise. Though these building blocks draw upon prior arts, our novelty resides in the semantics-guided disentanglement paradigm, unified theoretical characterization, and end-to-end pipeline adapted for both complete and incomplete multi-view inputs. The main contributions of this paper are summarized as follows:

1) This paper presents an semantics-guided disentanglement paradigm for unsupervised multi-view clustering, which decomposes per-view latents into cluster-relevant and cluster-irrelevant components guided by clustering semantic relevance and uses only cluster-relevant representations at inference, instead of adopting concatenated view-shared and view-specific features.

2) This paper provides a unified theoretical characterization of the proposed method under complete-view assumptions, wherein it suppresses manifold perturbations and cross-view geometric misalignment while enforcing subspace independence to achieve clean cross-view semantic disentanglement.

3) Extensive comparisons against state-of-the-art baselines and ablation studies on mainstream benchmarks demonstrate the superiority and robustness of the proposed method for both complete and incomplete multi-view settings.

2 Related Work

This section concentrates on deep multi-view clustering from the perspective of latent-space modeling, which can be categorized into three paradigms: consensus representation learning, disentangled representation learning, and probabilistic representation learning.

Consensus representation learning pursues unified cross-view embeddings by boosting inter-view semantic consistency without explicit latent-space decomposition. For example, COMPLETER (Lin et al., 2021) builds a three-branch contrastive prediction framework with view-level recon-

struction constraints for implicit missing-feature completion and cross-view consistency learning under incomplete multi-view inputs. DCP (Lin et al., 2023) utilizes dual contrastive prediction and hierarchical feature propagation to mitigate cross-view semantic discrepancy, enhance feature discriminability and relieve view heterogeneity for incomplete multi-view data. **Disentangled representation learning** separates latent factors into view-shared semantics and view-specific information, implemented implicitly via loss constraints or explicitly through specialized network modules and regularizers. Typically, DFL-Net (Chen et al., 2025) adopts a progressive disentangled feature learning pipeline, narrowing semantic gaps to derive view-shared representations and applying orthogonal subspace constraints to explicitly isolate view-specific features and filter noise. GMAE (Zou et al., 2026) constructs dual-branch autoencoders coupled with cross-view adversarial discriminators and correlation loss, enforcing statistical independence between view-shared and view-specific embeddings and alleviating view-distribution entanglement to produce high-quality clustering representations. **Probabilistic representation learning** leverages variational inference and information-theoretic criteria (e.g., VIB (Alemi et al., 2017), DVIB (Bao, 2021)), treating latent embeddings as stochastic distributions to quantify representation uncertainty and perform cross-view statistical distribution calibration. Notably, Multi-VAE (Xu et al., 2021) develops a VAE-based generative framework that disentangles view-shared and view-specific factors with variational latent variables. SUMVC (Cui et al., 2023) presents an information-theoretic multi-view clustering paradigm, optimizing mutual-information-aware objectives to preserve task-relevant cues while suppressing redundant cross-view information for discriminative embeddings.

Current multi-view disentangled clustering methods rely on heuristic fusion or overly strict independence constraints. Such hard regularization easily causes feature degradation and semantic collapse, failing to balance noise suppression and discriminative information preservation. To address these issues, this work proposes a theoretically grounded two-stage framework integrating log-space probability consensus and dual-level topological alignment. Combined with covariance decorrelation regularization, the framework effectively suppresses manifold perturbations and geometric misalignment, achieving precise latent disentanglement and robust unsupervised multi-view clustering.

3 The Proposed Method

Given a multi-view dataset $\mathcal{X} = \{\{\mathbf{x}_i^v \in \mathbb{R}^{d_v}\}_{v=1}^V\}_{i=1}^N$, where N denotes the number of samples, V is the number of views, $\mathbf{x}_i^v$ denotes the feature vector of the i-th sample under the v-th view, d_v denotes the feature dimension of the v-th view, and Y denotes the cluster-membership matrix.

3.1 Semantic-Aware Disentanglement Formulation

Assumption 1 (Latent Factor Generation) *Each multi-view observation random vector $\boldsymbol{X}^v$ is generated via $\boldsymbol{X}^v = f_v(\boldsymbol{S}, \boldsymbol{U}^v) + \boldsymbol{\mathcal{E}}^v$, where $f_v(\cdot)$ denotes the view-dependent stochastic generation mapping for the v-th view, $\boldsymbol{S}$, $\boldsymbol{U}^v$, and $\boldsymbol{\mathcal{E}}^v$ denote the view-invariant cluster-relevant latent vectors, view-specific cluster-irrelevant latent vectors, and observation-noise random vectors, respectively. This generative process satisfies $I(Y; \boldsymbol{U}^v \mid \boldsymbol{S}) = 0$ and $\boldsymbol{\mathcal{E}}^v \perp\!\!\!\perp (Y, \boldsymbol{S}, \boldsymbol{U}^v)$.*

Theorem 1 *Given Assumption 1, it holds that $I(Y; \boldsymbol{X}^v) \leq I(Y; \boldsymbol{S})$, i.e., for any view v, the mutual information between view observation $\boldsymbol{X}^v$ and cluster label Y is no greater than the mutual information between the view-invariant cluster-relevant latent vectors $\boldsymbol{S}$ and Y.*

We introduce this latent generative assumption for multi-view information partitioning to support our disentangled clustering framework. From Assumption 1, Theorem 1 gives an information-theoretic upper bound: $I(Y; \boldsymbol{X}^v) \leq I(Y; \boldsymbol{S})$ for each view observation $\boldsymbol{X}^v$. Hence, optimizing only the mutual information of $\boldsymbol{S}$ suffices to capture the intrinsic clustering semantics. To implement this objective and achieve latent disentanglement, we use two view-specific encoders to decompose single-view features into dual latent representations: $\mathbf{S}^v = E_c^v(\mathbf{X}^v), \mathbf{U}^v = E_{nc}^v(\mathbf{X}^v)$, where $E_c^v(\cdot)$ and $E_{nc}^v(\cdot)$ denote the cluster-related and cluster-irrelevant encoders for view v, respectively. Based on the information-bottleneck principle, we formulate an ideal unsupervised disentanglement objective to retain cluster-discriminative information within $\mathbf{S}^v$, suppress discriminative leakage from $\mathbf{U}^v$, and enforce statistical independence between the two representations:

$$\max \sum_{v=1}^{V} I(\mathbf{S}^v; Y) - \alpha_v I(\mathbf{U}^v; Y) - \beta_v I(\mathbf{S}^v; \mathbf{U}^v), \tag{1}$$

where $\alpha_v, \beta_v \geq 0$ denote non-negative balancing coefficients. Nevertheless, Y is inaccessible under unsupervised conditions, rendering this objective intractable for direct optimization.

3.2 Cross-View Semantic Aggregation via log-space probability Consensus

Due to the unavailability of Y, the preceding information-allocation objective cannot be directly optimized. For each sample i, we employ log-space probability consensus: we first perform softmax normalization on each view-wise class-related representation $\mathbf{s}_i^v$. Next, we conduct cross-view averaging over log-probabilities, followed by re-normalization via softmax to yield the sample-level consensus representation $\mathbf{s}_i$: $\mathbf{s}_i = \mathrm{softmax}\left(\frac{1}{V} \cdot \sum_{v=1}^{V} \log \mathbf{s}_i^v\right)$. This log-space probability aggregation possesses a consensus-filtering property: it attenuates dimensions with inconsistent cross-view responses to suppress spurious view-specific variations, while preserving consistent cluster-semantic information for unsupervised optimization.

To analyze its estimation error under manifold smoothness, we treat each view encoder as a locally bi-Lipschitz transformation acting on the sample-level ground-truth cluster-invariant latent factor $\mathbf{s}_i^*$. Each view representation decomposes into $\mathbf{s}_i^v = T_v(\mathbf{s}_i^*) + \boldsymbol{\xi}_i^v$, where $T_v(\cdot)$ is the view-dependent bi-Lipschitz embedding modeling cross-view geometric misalignment, and $\boldsymbol{\xi}_i^v$ stands for the linearized stochastic manifold perturbation derived from first-order Taylor linearization, capturing manifold fluctuations and Jacobian-induced residuals. This decomposition holds under small-perturbation conditions for the nonlinear aggregation operator; complete proofs and auxiliary conditions are given in the supplementary material.

Assumption 2 (Manifold Perturbation Model) *For each view v and sample i, the perturbation $\boldsymbol{\xi}_i^v$ satisfies $\mathbb{E}[\boldsymbol{\xi}_i^v] = \mathbf{0}$, $\mathbb{E}\left[(\boldsymbol{\xi}_i^v)^\top \boldsymbol{\xi}_i^u\right] = 0$ for $u \neq v$, and $\mathbb{E}\|\boldsymbol{\xi}_i^v\|_2^2 \leq \sigma^2$. Perturbations are independent of $\mathbf{s}_i^*$. There exists $\delta > 0$ such that all entries of $T_v(\mathbf{s}_i^*) \geq \delta$, and sufficiently small $\boldsymbol{\xi}_i^v$ keep $T_v(\mathbf{s}_i^*) + \boldsymbol{\xi}_i^v$ entrywise positive.*

Theorem 2 *Under Assumption 2, suppose each view-wise encoding follows $\mathbf{s}_i^v = T_v(\mathbf{s}_i^*) + \boldsymbol{\xi}_i^v$, where each $T_v(\cdot)$ is a deterministic mapping. Let $\mathbf{s}_i$ be the output of log-space probability consensus for sample i. The expected squared ℓ_2 estimation error satisfies $\mathbb{E}\left[\|\mathbf{s}_i - \mathbf{s}_i^*\|_2^2\right] \leq (C_1\sigma^2)/V + C_2/V \cdot \sum_{v=1}^{V} \|T_v(\mathbf{s}_i^*) - \mathbf{s}_i^*\|_2^2$, where $C_1, C_2 > 0$ are constants depending on the compact feasible domain, the Lipschitz constants of $\log(\cdot)$ and $\mathrm{softmax}(\cdot)$, and the uniform Jacobian bound of $\log(\cdot)$.*

As shown in Theorem 2, the estimation error decomposes into two terms. The first, scaling as σ^2/V, arises from stochastic perturbations and decays with more views, as log-space probability consensus averages out random noise. The second term averages view-specific geometric misalignment from $T_v(\cdot)$ and converges to a non-zero constant as $V \to \infty$; probability consensus alone cannot remove this bias. To mitigate this systematic distortion, we adopt dual-level topological alignment for cross-view manifold consistency.

3.3 Manifold Purification via Dual-Level Topological Alignment

To compensate for the persistent geometric misalignment remaining after log-space probability consensus aggregation, we take the stop-gradient consensus representation $\mathbf{s}_i$ as a label-free manifold baseline template and conduct topological alignment at both the global distribution and local sample-pair levels. We construct two sets of affinity matrices: $\mathbf{P}_{ij}^c$ (superscript c denotes consensus) built from $\mathbf{s}_i$, and $\mathbf{P}_{ij}^v$ (superscript v denotes view) built from $\mathbf{s}_i^v$. They are computed upon ℓ_2-normalized representations $\tilde{\mathbf{s}}_i = \mathbf{s}_i/\|\mathbf{s}_i\|_2$, $\tilde{\mathbf{s}}_i^v = \mathbf{s}_i^v/\|\mathbf{s}_i^v\|_2$, where self-similarity terms are masked to obtain valid row-wise probability distributions.

At the global level, global topological alignment regularizes each view's neighborhood probability toward the consensus topology via KL-divergence:

$$\mathcal{L}_{\mathrm{align}} = -\frac{1}{V}\sum_{v=1}^{V}\frac{1}{N}\sum_{i=1}^{N}\sum_{j\neq i}\mathbf{P}_{ij}^c \log \mathbf{P}_{ij}^v. \tag{2}$$

The pairwise affinity scores and corresponding probability distributions are defined as $a_{ij}^c = \tilde{\mathbf{s}}_i^\top \tilde{\mathbf{s}}_j/\tau,\ a_{ii}^c = -\infty, \mathbf{P}_{ij}^c \propto \exp\left(a_{ij}^c\right), a_{ij}^v = \tilde{\mathbf{s}}_i^{v\top} \tilde{\mathbf{s}}_j^v/\tau,\ a_{ii}^v = -\infty, \mathbf{P}_{ij}^v \propto \exp\left(a_{ij}^v\right)$, where τ

is a temperature hyper-parameter used to scale the pairwise affinity values. **At the local level**, local topological alignment refines local sample-pair structures using positive and negative neighbors derived from raw-feature neighborhood structures. A topological confidence gate reduces interference from false-positive samples, and an adaptive manifold margin suppresses intra-cluster similarity while enlarging inter-cluster margins. The local contrastive loss is given by:

$$\mathcal{L}_{\text{topo}} = \frac{1}{B} \sum_{i \in \mathcal{B},\ p \in \mathcal{P}_i} -g_{i,p} \cdot \log \frac{\exp\left((a^c_{i,p} - \delta_{i,p})/\tau\right)}{\sum_{n \in \mathcal{N}_i} \exp(a^c_{i,n}/\tau) + \epsilon}, \tag{3}$$

where $\mathcal{B}$ denotes the set of samples within the current training batch, $\delta_{i,p} = \lambda_m \cdot \max(a^c_{i,p}, 0)$ denotes an adaptive manifold margin that varies with each sample pair (i, p), with λ_m being a fixed base-margin hyper-parameter; $g_{i,p}$ denotes the topological confidence gate ranging from 0 to 1; $\mathcal{P}_i$ and $\mathcal{N}_i$ are respectively the sets of positive and negative neighbors for sample i, obtained from raw-feature neighborhoods; and $\epsilon > 0$ prevents numerical overflow.

Instead of direct embedding-value matching, topological alignment enforces cross-view consistency for neighborhood manifold structures. Global topological alignment stabilizes training by regularizing per-view affinity graphs, while local topological alignment mitigates detrimental effects from false-positive samples and refines local sample-pair structures.

Assumption 3 (Dual-Level Topological Alignment Idealized Conditions) *We state high-level regularity assumptions for dual-level topological alignment. All gradient-descent iterates lie within a compact optimization domain, and row-wise logit centering eliminates the shift-invariance ambiguity of softmax normalization, rendering the temperature-scaled softmax mapping locally invertible on this domain. We adopt an idealized topological surrogate $\mathcal{L}^{\text{ideal}}_{\text{topo}}$ as a mathematically tractable proxy for the practical contrastive topological loss, where $\mathcal{L}^{\text{ideal}}_{\text{topo}} = \mathbb{E}_{\boldsymbol{s}^* \sim \mathcal{D}}\left[1/N \cdot \sum_{i=1}^{N} \sum_{p \in \mathcal{P}_i,\, y_p = y_i} -\log\left(\exp(a^c_{i,p}/\tau) \Big/ \sum_{n \neq i} \exp(a^c_{i,n}/\tau)\right)\right]$. The mismatch between practical loss and this surrogate is uniformly bounded in expectation. Together with the global cross-view alignment loss, this surrogate jointly regularizes perturbations of view-specific and consensus affinity matrices during optimization.*

Theorem 3 (Dual-Loss Joint Perturbation Bound) *Under Assumption 3, let $\mathcal{L}_{\text{align}}$ denote the population-level cross-view alignment loss, and $\mathcal{L}^{\text{ideal}}_{\text{topo}}$ the population-level expected idealized topological surrogate loss (proxy for practical training loss). Let $\mathbf{P}^v$ denote the view-specific affinity matrix, $\bar{\mathbf{P}}^c \in \mathbb{R}^{N \times N}$ the block-diagonal ground-truth consensus affinity matrix. Define $\mathcal{L}^*_{\text{topo}} \triangleq \frac{1}{N} \sum_{i=1}^{N} k_i \log k_i$, where k_i is the number of ground-truth same-cluster neighbors for sample i, and let $k_{\min} \triangleq \min_{1 \le i \le N} k_i$ denote the minimal value over $\{k_i\}_i$; $\mathcal{L}^*_{\text{topo}}$ is the dataset-driven theoretical minimum of $\mathcal{L}^{\text{ideal}}_{\text{topo}}$, a constant independent of model parameters. We obtain the joint expected perturbation bound*

$$\mathbb{E}_{\boldsymbol{s}^* \sim \mathcal{D}}\left[\|\mathbf{P}^v(\boldsymbol{s}^*) - \bar{\mathbf{P}}^c\|_F\right] \le C_{\text{align}} \sqrt{\mathcal{L}_{\text{align}}} + C_{\text{local}} \sqrt{\mathcal{L}^{\text{ideal}}_{\text{topo}} - \mathcal{L}^*_{\text{topo}}},$$

where $C_{\text{align}} = \sqrt{2N} > 0$, $C_{\text{local}} = \sqrt{2N/k_{\min}} > 0$ are data-dependent constants.

The above results build upon idealized manifold-theoretic assumptions, including compact optimization domains and regularity conditions for the temperature-scaled softmax mapping. While enabling tractable information-theoretic analysis, these assumptions are rarely strictly satisfied for real-world multi-view data. Instead of pursuing sharp quantitative bounds valid over arbitrary inputs, our theorems deliver qualitative mechanistic insights by disentangling the two core optimization objectives: cross-view alignment and topological structure matching. Together with the full formal result in supplementary material, this theorem demonstrates that dual-level topological alignment jointly constrains the Frobenius perturbation of view-specific affinity matrices and enforces cross-view consistency of cluster-relevant eigenspaces. Proofs, complete technical assumptions, discussion on assumption violations, and empirical validations are deferred to the supplementary material.

3.4 Subspace Isolation via Covariance Decorrelation

Following log-space probability consensus and dual-level topological alignment, subspace disentanglement separates cluster-relevant and cluster-irrelevant representations. Each view-specific decoder

reconstructs inputs by concatenating $\mathbf{s}_i$ and $\mathbf{u}_i^v$: $\hat{\mathbf{x}}_i^v = D_v\big(\big[\mathbf{s}_i; \mathbf{u}_i^v\big]\big)$. The reconstruction loss is formulated as

$$\mathcal{L}_{\text{rec}} = \frac{1}{V}\sum_{v=1}^{V}\frac{1}{N}\sum_{i=1}^{N}\left\|\mathbf{x}_i^v - \hat{\mathbf{x}}_i^v\right\|_2^2. \tag{4}$$

To remove statistical coupling between the two latent subspaces, feature-wise z-score standardization yields $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{U}}^v$ with zero mean and unit variance per dimension. A covariance decorrelation regularizer is defined as

$$\mathcal{L}_{\text{orth}} = \frac{1}{V}\sum_{v=1}^{V}\frac{1}{d_1 d_2}\left\|\frac{(\tilde{\mathbf{S}})^\top\tilde{\mathbf{U}}^v}{N-1}\right\|_F^2, \tag{5}$$

where d_1 is the feature dimension of $\tilde{\mathbf{S}}$, and d_2 that of $\tilde{\mathbf{U}}^v$. Minimizing $\mathcal{L}_{\text{orth}}$ induces second-order decorrelation $\text{Cov}\left(\tilde{\mathbf{S}}, \tilde{\mathbf{U}}^v\right) \approx \mathbf{0}$: $\mathcal{L}_{\text{rec}}$ retains raw-feature information, while $\mathcal{L}_{\text{orth}}$ suppresses cross-subspace information leakage. Combining these loss terms yields the overall objective

$$\mathcal{L}_{\text{total}} = \sum_{v=1}^{V}\mathcal{L}_{\text{rec}}^v + \alpha\left(\mathcal{L}_{\text{align}} + \lambda\,\mathcal{L}_{\text{orth}}\right) + \beta\,\mathcal{L}_{\text{topo}}, \tag{6}$$

where $\alpha, \beta, \lambda \geq 0$ are trade-off hyper-parameters weighting different loss components.

Table 1: Clustering performance comparison on different MVC datasets (ACC, NMI, and PUR).

Methods	BRCA (398, 4, 4)			LGG (267, 3, 4)			Dermatology (358, 6, 2)			MSRCV1 (210, 7, 6)		
	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
COMPLETER	55.53	34.65	65.33	80.15	49.25	80.15	77.65	80.11	82.12	90.00	87.90	90.00
DCP	57.29	39.51	60.55	59.55	44.82	73.03	72.91	77.22	80.73	25.71	23.25	27.14
MFLVC	55.53	27.74	60.05	79.03	49.73	79.03	58.10	56.20	62.85	63.33	66.11	64.29
DSMVC	54.52	33.53	68.84	82.77	54.13	82.77	92.74	87.82	92.74	64.29	54.29	64.29
SURE	39.70	12.85	48.99	62.92	38.01	65.17	88.27	77.03	88.55	91.43	85.84	91.43
DealMVC	59.55	32.79	61.56	72.28	40.55	72.28	45.53	31.13	45.53	82.00	75.54	82.00
GCFAgg	51.51	32.41	61.31	55.06	22.95	61.80	88.27	79.25	88.27	39.52	31.91	42.86
CPSPAN	66.83	34.48	74.12	63.30	30.53	63.30	76.26	84.63	85.20	67.62	69.83	89.52
SDMVC	57.79	33.80	64.57	63.67	43.86	67.79	70.67	83.30	84.92	59.52	52.51	45.24
CVCL	61.98	34.68	68.49	58.20	23.73	58.20	56.25	56.01	67.97	90.62	84.57	90.62
SCMVC	50.25	30.70	60.80	73.41	39.76	73.41	93.85	88.44	93.85	90.95	83.92	90.95
MVCAN	57.79	35.70	32.24	59.55	42.57	27.18	58.38	66.73	51.58	71.54	60.19	71.54
DCG	54.52	37.49	68.34	84.64	59.89	84.64	72.35	74.97	80.45	52.38	41.66	53.81
SparseMVC	70.10	44.90	70.85	83.15	54.62	83.15	95.25	89.86	95.25	**97.14**	**94.22**	**97.14**
BRIDGE	58.31	35.40	65.67	56.67	23.21	69.17	82.92	85.72	93.48	55.03	51.51	57.14
GMAE	67.59	41.20	**84.67**	**92.16**	**75.48**	**98.91**	91.62	83.80	91.62	94.29	88.16	94.29
SG-DMVC	**73.37**	**45.09**	73.37	90.26	67.15	90.26	**95.81**	**90.79**	**95.81**	94.29	87.79	94.29

Methods	Wikipedia (693, 10, 2)			RGB-D (1499, 13, 2)			STL-10 (13000, 10, 3)			NUSWIDEOBJ (30000, 31, 5)		
	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
COMPLETER	57.14	53.10	59.31	34.78	18.06	40.65	11.11	0.20	11.22	12.73	8.51	19.53
DCP	45.31	43.16	46.32	42.24	31.78	49.69	11.08	0.18	11.20	10.38	8.91	20.06
MFLVC	40.12	27.52	41.70	34.23	15.88	38.65	20.18	11.46	21.72	16.59	10.06	23.15
DSMVC	60.32	54.74	62.19	39.68	31.36	49.28	96.85	92.40	96.85	12.79	10.01	22.66
SURE	50.65	39.97	54.11	28.99	26.75	48.31	11.13	0.16	11.23	11.26	10.11	20.51
DealMVC	38.96	37.09	38.96	30.50	11.95	30.57	60.40	48.96	60.40	13.64	2.31	13.64
GCFAgg	51.80	45.87	56.57	25.67	21.00	44.17	17.58	4.46	18.24	18.30	19.41	31.12
CPSPAN	22.08	8.35	24.39	38.78	33.77	57.49	20.76	9.54	21.72	10.32	9.21	22.65
SDMVC	55.99	53.98	62.05	44.51	38.83	51.07	63.32	55.36	63.32	13.20	14.99	24.97
CVCL	42.81	32.69	48.44	25.14	16.31	39.28	15.56	1.13	15.63	14.17	14.83	27.44
SCMVC	53.54	35.59	55.84	35.75	32.83	52.73	79.68	73.49	79.68	17.74	19.01	30.35
MVCAN	59.02	55.81	45.92	43.41	**41.02**	26.28	66.62	72.25	58.49	15.49	13.09	4.06
DCG	54.40	50.09	58.44	40.03	29.89	51.48	77.44	84.87	77.45	15.19	12.25	24.52
SparseMVC	61.04	54.79	62.91	35.40	26.35	49.07	43.68	44.69	44.64	**19.05**	**20.29**	31.52
BRIDGE	45.35	39.23	52.88	29.68	27.60	31.67	28.29	20.61	30.23	15.53	16.38	18.67
GMAE	62.18	56.67	**70.35**	45.07	40.06	**58.94**	96.25	90.26	96.37	17.28	17.05	**31.59**
SG-DMVC	**63.06**	**57.11**	63.78	**45.34**	34.64	54.87	**97.98**	**94.72**	**97.98**	17.36	13.06	24.29

4 EXPERIMENTS

To validate the effectiveness of the proposed method, we perform comprehensive experiments on both complete and incomplete multi-view clustering tasks, following the experimental protocol of a representative multi-view clustering work (Zou et al., 2026).

Table 2: Clustering comparisons on Digits with increased views. "-#V" denotes its count.

Methods	Digits-2V			Digits-3V			Digits-4V			Digits-6V		
	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
COMPLETER	84.10	85.03	84.45	88.85	81.94	88.85	73.55	75.17	77.60	73.60	73.58	75.30
DCP	79.45	79.61	80.65	80.15	83.02	82.65	75.35	78.69	79.20	76.80	79.57	81.00
MFLVC	70.40	69.55	72.45	91.85	85.92	91.85	82.50	79.88	82.50	85.00	84.06	85.00
DSMVC	79.70	75.23	79.70	95.40	90.54	95.40	88.15	85.10	88.15	95.45	92.52	95.45
SURE	81.20	69.51	81.20	74.10	65.58	74.10	61.80	55.34	61.80	66.75	62.72	69.65
DealMVC	50.30	50.13	50.30	45.50	49.07	45.50	62.60	65.57	62.60	57.10	68.04	57.10
GCFAgg	61.55	66.43	62.90	90.80	85.75	90.80	79.55	77.84	79.55	83.95	82.76	83.95
CPSPAN	89.65	80.67	89.65	87.25	79.39	87.25	88.95	81.27	88.95	89.55	82.30	89.95
SDMVC	80.60	74.52	80.60	77.15	73.83	77.85	82.40	77.79	82.40	84.85	75.05	84.85
CVCL	85.42	77.17	85.42	83.96	76.48	83.96	79.22	75.45	79.22	89.15	80.60	89.15
SCMVC	87.50	78.46	87.50	89.35	80.87	89.35	86.15	78.00	86.15	94.15	87.91	94.15
MVCAN	90.00	82.82	79.82	91.90	85.26	83.20	91.90	85.03	83.14	94.90	89.63	89.09
DCG	74.15	68.36	74.25	78.75	80.87	82.40	79.05	80.91	82.65	81.70	82.59	82.70
SparseMVC	84.20	81.45	84.20	93.40	87.23	93.40	84.25	81.35	84.25	91.00	86.28	91.00
BRIDGE	86.28	77.88	86.28	88.94	81.51	88.94	92.06	84.92	92.06	96.01	91.52	96.01
GMAE	95.55	90.88	95.90	95.90	91.51	96.00	96.50	92.07	96.65	97.45	94.16	97.45
SG-DMVC	**96.15**	**92.17**	**96.15**	**97.65**	**94.28**	**97.65**	**97.40**	**93.91**	**97.40**	**98.15**	**95.60**	**98.15**

Datasets and Evaluation Metrics. Experiments are conducted on eight multi-view benchmarks spanning biomedical multi-omics, natural-image, and image-text domains. Dataset statistics are summarized in Table 1, where each tuple denotes the number of samples, classes, and views; per-view feature dimensions are deferred to the Appendix. Quantitative results are reported using three standard clustering metrics: clustering accuracy (ACC), normalized mutual information (NMI), and purity (PUR), where higher values denote better clustering quality.

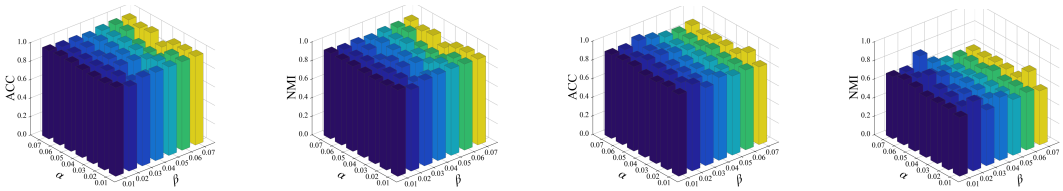
Compared Baselines. We compare our method against a suite of state-of-the-art deep multi-view clustering methods (DMVC) for complete and incomplete scenarios, covering multiple technical paradigms. These include contrastive-learning-based DMVC methods: COMPLETER (Lin et al., 2021), DCP (Lin et al., 2023), MFLVC (Xu et al., 2022), CVCL (Chen et al., 2023), DCG (Zhang et al., 2025); incomplete-focused DMVC methods: SURE (Yang et al., 2022), DealMVC (Yang et al., 2023), CPSPAN (Jin et al., 2023); self-training and aggregation-driven methods: DSMVC (Tang & Liu, 2022), GCFAgg (Yan et al., 2023), SDMVC (Xu et al., 2023b), SCMVC (Wu et al., 2024), MVCAN (Xu et al., 2024), SparseMVC (Liu et al., 2025), BRIDGE (Jiang et al., 2025). We evaluate all methods on eight heterogeneous multi-view benchmarks via ACC, NMI and PUR. For fair comparison, we adopt official implementations and stick to original hyper-parameter configurations without dataset-specific fine-tuning. Full quantitative results are summarized in Table 1.

Implementation Details. All experiments are implemented in PyTorch with fully-connected encoder-decoder backbones (**?**). We use the Adam optimizer with an initial learning rate of 10^{-3} and train all models for 500 epochs. Global random seed is set to 42, and deterministic CUDNN is enabled for reproducibility. Following the full-sample batch scheme, the entire dataset is fed as one batch per iteration. Positive-negative neighbor indices for the contrastive module are pre-computed via nearest-neighbor search. At inference, we run K-means with $n_init = 20$ on normalized shared latent embeddings to generate cluster assignments. We set the shared embedding dimension to 128, the disentanglement loss weight $\alpha = 0.05$, the contrastive loss weight $\beta = 0.01$, the covariance decorrelation loss weight $\lambda = 6 \times 10^3$, and adopt 20 positive neighbors for contrastive learning. To ensure fair comparison, we avoid dataset-specific fine-tuning.

4.1 Robustness and Ablation Experiments

Overall performance on complete multi-view datasets. Table 1 reports the clustering results over eight complete multi-view benchmarks. Owing to disentangled representation learning and multi-objective optimization, our method obtains competitive or state-of-the-art performance across most datasets, showing strong generalization for diverse complete multi-view clustering scenarios.

Multi-view safeness under increasing view numbers. As shown in Table 2, we evaluate model performance on Digits with progressively increased view numbers. Competitors including SURE and DSMVC suffer from negative-transfer effects: their clustering metrics fluctuate or degrade with extra views, as they fail to suppress view-specific noise and cross-view conflicts in feature



(a) Dermatology (b) Dermatology (c) ACC (LGG) (d) NMI (LGG)

Figure 2: Sensitivity analysis of hyper-parameters α and β on two datasets.

fusion. By disentangling view-common semantics from view-specific variations, our method obtains steady performance gains from 2 to 6 views, effectively leveraging multi-view complementarity and mitigating per-view interference. It achieves state-of-the-art ACC, NMI, and PUR under all Digits-2V/3V/4V/6V setups, verifying the strong multi-view safeness of our disentangled framework for dynamically-augmented multi-view inputs.

Performance on incomplete multi-view datasets. To further evaluate the generalization and robustness under real-world challenging conditions, we conduct experiments on four incomplete multi-view datasets: BRCA, LGG, Dermatology, and MSRCV1. Table 3 summarizes the averaged quantitative comparisons over ten independent runs against SOTA baselines at a missing ratio of 0.5. Leveraging disentangled representation learning and cross-view consistent alignment, our method adapts stably to incomplete inputs and mitigates performance drops induced by view missing. It outperforms competitors on most metrics. Comparisons over different missing ratios are given in the Appendix, demonstrating the strong fault-tolerance of our model for incomplete multi-view data.

Ablation analysis of loss functions. To quantitatively validate each loss component within our multi-objective optimization framework, we perform ablation experiments on three representative datasets: Digits-6V, Dermatology, and Wikipedia. Table 4 reports clustering results under different loss combinations. The baseline using only reconstruction loss $\mathcal{L}_{\text{rec}}$ produces limited performance, since reconstruction alone cannot yield discriminative clustering features. Adding either $\mathcal{L}_{\text{align}} + \mathcal{L}_{\text{orth}}$ or $\mathcal{L}_{\text{topo}}$ brings substantial gains: $\mathcal{L}_{\text{align}} + \mathcal{L}_{\text{orth}}$ enables cross-view distribution alignment and subspace decorrelation, whereas $\mathcal{L}_{\text{topo}}$ improves latent topological properties and boosts intra-cluster compactness as well as inter-cluster separability. The full model equipped with all three losses achieves the best results across all datasets, confirming the complementarity and necessity of each component and justifying our overall loss-function design.

Table 3: Comparison of clustering results of different methods on incomplete multi-view datasets.

Methods	BRCA			LGG			Dermatology			MSRCV1		
	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
COMPLETER	37.69	7.52	48.49	51.31	18.46	56.18	46.37	39.53	55.31	41.43	38.57	44.76
DCP	40.20	11.85	47.74	56.93	17.68	56.93	43.58	37.23	51.40	36.67	29.55	38.57
DSMVC	40.70	4.95	42.96	52.81	18.69	57.68	47.49	33.67	53.91	23.33	6.43	23.33
SURE	38.44	10.09	49.50	53.56	16.31	54.31	38.27	18.68	42.18	42.38	26.61	45.24
DealMVC	40.70	4.59	42.96	67.42	34.34	67.42	71.51	55.46	72.35	48.10	41.04	48.10
GCFAgg	46.23	21.19	59.05	12.20	**49.44**	56.55	48.50	59.22	62.85	15.62	28.10	29.05
SCMVC	48.49	24.71	59.80	54.31	23.98	63.30	48.60	35.85	54.75	39.52	36.72	44.29
GMAE	51.01	**32.42**	64.82	74.16	34.91	74.16	55.59	62.15	73.46	54.62	49.00	57.62
SG-DMVC	**59.75**	31.40	**65.28**	**76.33**	43.07	**76.33**	**74.92**	**68.91**	**76.98**	**74.57**	**64.54**	**74.86**

4.2 Further Model Analysis

Hyper-parameter sensitivity analysis. We investigate the influence of the trade-off weights α and β via sensitivity experiments. As illustrated in Fig. 2, within a reasonable interval, clustering performance remains stable against moderate variations of α and β. Excessively large weights will dominate the optimization objective and degrade final clustering quality. Our default setting $\alpha = 0.05$, $\beta = 0.01$ lies in this stable region.

Table 4: Ablation study in terms of loss components on different datasets

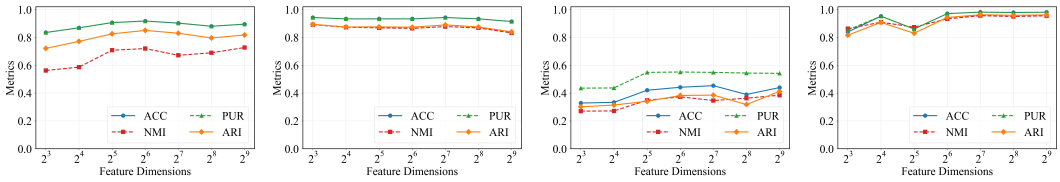
No.	Loss components			Digits-6V			Dermatology			Wikipedia		
	$\mathcal{L}_{\text{rec}}$	$\mathcal{L}_{\text{align}} + \mathcal{L}_{\text{orth}}$	$\mathcal{L}_{\text{topo}}$	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
1	√	–	–	78.80	75.21	78.80	82.68	70.55	82.68	44.16	42.50	49.93
2	√	√	–	96.05	91.70	96.05	86.31	80.52	86.31	48.48	43.77	52.24
3	√	–	√	96.85	92.79	96.85	94.13	89.54	94.13	62.63	56.00	**64.07**
4	√	√	√	**98.15**	**95.60**	**98.15**	**95.81**	**90.79**	**95.81**	**63.06**	**57.11**	63.78

"√" denotes model with the component.



(a) Rel. (Digits-6V) (b) Irrel. (Digits-6V) (c) Rel. (STL-10) (d) Irrel. (STL-10)

Figure 3: t-SNE visualization of disentangled cluster-relevant (Rel.) and cluster-irrelevant (Irrel.) representations on Digits-6V and STL-10.



(a) LGG (b) MSRCV1 (c) RGB-D (d) Digits-6V

Figure 4: Effect of embedding dimension d on different MVC datasets.

t-SNE visualization of disentangled representations. To intuitively observe the disentangling behavior of our framework, we provide t-SNE projections for cluster-relevant and cluster-irrelevant latent features. As shown in Fig. 3, cluster-relevant features form well-separated category clusters. By contrast, cluster-irrelevant features intermix different categories and preserve view-specific local structures instead of collapsing into meaningless trivial noise, which validates the effectiveness of our disentanglement mechanism.

Impact of embedding dimension. Fig. 4 studies how embedding dimension d affects clustering performance. Too small d limits representation capacity and hurts performance; performance rises with d until a moderate value, while overly large d introduces redundant parameters and overfitting risks. Empirically, $d = 128$ achieves a good trade-off for our tasks.

5 Conclusion

This paper presents SG-DMVC, a semantics-guided disentangled multi-view clustering framework. Instead of view-origin-driven decomposition, it separates latent representations into cluster-relevant and cluster-irrelevant components guided by clustering semantics. Log-space probability consensus, dual-level topological alignment, and covariance decorrelation regularization are jointly adopted to suppress manifold perturbations, mitigate cross-view geometric misalignment and reduce inter-subspace statistical coupling. Experiments on diverse complete and incomplete multi-view benchmarks validate the competitive performance and missing-view robustness of SG-DMVC, and theoretical analysis verifies the efficacy of the proposed method. Future work will explore the inherent geometric structures of cluster-irrelevant features to explicitly model view-specific variations and further boost semantics-guided multi-view disentangled clustering.

Ethics Statement

This study does not involve human participants, human subjects, or sensitive personal data. All experimental datasets utilized in this work are publicly available benchmark datasets and are used in full compliance with their original release licenses and usage regulations. This research contains no discriminatory, biased, or privacy-threatening content, involves no potential conflicts of interest, and adheres to the ICLR Code of Ethics.

Reproducibility Statement

All model architectures, hyperparameter settings, training strategies, and experimental protocols are elaborated in the main paper and appendix. Complete theoretical derivations and proof details are provided in the supplementary materials. All public datasets used in this work are clearly cited and publicly accessible. Source code and data preprocessing pipelines will be released upon acceptance to facilitate full reproducibility of all experimental results.

References

Alexander A. Alemi, Ian Fischer, Joshua V. Dillon, and Kevin Murphy. Deep variational information bottleneck. In International Conference on Learning Representations, 2017.

Feng Bao. Disentangled variational information bottleneck for multiview representation learning. In Artificial Intelligence: First CAAI International Conference, CICAI 2021, pp. 91–102, 2021.

Jie Chen, Hua Mao, Wai Lok Woo, and Xi Peng. Deep multiview clustering by contrasting cluster assignments. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 16706–16715, October 2023.

Jingyi Chen, Zhibin Dong, Tiejun Li, and Yibo Han. Dc-span: A dual contrastive attention network for multi-view clustering. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 40, pp. 22134–22142, 2026.

Zhe Chen, Xiao-Jun Wu, Tianyang Xu, and Josef Kittler. Dfl-net: Disentangled feature learning network for multi-view clustering. IEEE Transactions on Knowledge and Data Engineering, 37 (8):4537–4547, 2025.

Chen Cui, Yazhou Ren, Jun Pu, Jiaming Li, Xiaorong Pu, Tong Wu, Yuxin Shi, and Liang He. A novel approach for effective multi-view clustering with information-theoretic perspective. In Advances in Neural Information Processing Systems, NeurIPS, 2023.

Uno Fang, Man Li, Jianxin Li, Longxiang Gao, Tao Jia, and Yanchun Zhang. A comprehensive survey on multi-view clustering. IEEE Transactions on Knowledge and Data Engineering, 35 (12):12350–12368, 2023.

Sheng Hu, Gang Zou, and Yu Pan. Adaptive confidence multi-view hashing for multimedia retrieval. In IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 7900–7904, 2024.

Wenjun Huang, Junpeng Mi, Zhi Zhang, Yang Li, and Tao Liu. A coarse and fine-grained deep multi-view subspace clustering method for unsupervised fault diagnosis of rolling bearings. Measurement Science and Technology, 35(10):105104, 2024.

Xiaorui Jiang, Zeyu Ma, Yuhong Fu, Xihong Liu, Yi Wen, Suyuan Liu, Sihang Zhou, Xinwang Liu, and En Zhu. A unified framework to bridge complete and incomplete deep multi-view clustering under non-iid missing patterns. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 594–603, October 2025.

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 (CVPR), pp. 11600–11609, June 2023.

Guanzhou Ke, Guoqing Chao, Xiaoli Wang, Chenyang Xu, Yongqi Zhu, and Shengfeng He. Disentangling multi-view representations beyond inductive bias. In Proceedings of the 31st ACM International Conference on Multimedia, pp. 2582–2590, 2023.

Meng Li, Hao Zhang, and Lin Zhou. Multi-view clustering for single-cell rna-seq data based on graph fusion. Briefings in Bioinformatics, 26(3), 2025.

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 (CVPR), pp. 11306–11315, 2021.

Yijie Lin, Yuanbiao Gou, Xinhang 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 (TPAMI), 45(4):4447–4461, 2023.

Ruimeng Liu, Xin Zou, Chang Tang, Xiao Zheng, Xingchen Hu, Kun Sun, and Xinwang Liu. Sparsemvc: Probing cross-view sparsity variations for multi-view clustering. In Advances in Neural Information Processing Systems (NeurIPS), volume 38, pp. 154594–154622, 2025.

Xi Ma, Ming Zhang, Man-On Pun, and Meng Liu. Subspace-contrastive multi-view clustering. In Proceedings of the 32nd ACM International Conference on Multimedia, pp. 3428–3436, 2024.

Huayi Tang and Yong Liu. Deep safe multi-view clustering: Reducing the risk of clustering performance degradation caused by view increase. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 202–211, 2022.

Jiatai Wang, Zhiwei Xu, Xuewen Yang, Hailong Li, Bo Li, and Xuying Meng. Self-supervised multi-view clustering in computer vision: A survey. IET Computer Vision, 18(6):709–734, 2024.

Song Wu, Yan Zheng, Yazhou Ren, Lifang He, Liang Peng, and Xiaofeng Zhu. Self-weighted contrastive fusion for deep multi-view clustering. IEEE Transactions on Multimedia, 26:3334–3347, 2024.

Chao Xu, Pengfei Li, Li Yao, Xinhang Wan, Chao Zhang, Guanzhou Ke, and Jie Xu. Investigating and mitigating the side effects of noisy views for self-supervised clustering algorithms in practical multi-view scenarios. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 22957–22966, June 2024.

Jie Xu, Yazhou Ren, Huayi Tang, Xiaorong Pu, Xiaofeng Zhu, Meng Zeng, and Zhihai He. Multi-vae: Learning disentangled view-common and view-peculiar visual representations for multi-view clustering. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 9234–9243, 2021.

Jie Xu, Huayi Tang, Yazhou Ren, Liang Peng, Xiaofeng Zhu, and Lifang He. Multi-level feature learning for contrastive multi-view clustering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 16051–16060, June 2022.

Jie Xu, Yazhou Ren, Xiaoshuang Shi, Heng Tao Shen, and Xiaofeng Zhu. Untie: Clustering analysis with disentanglement in multi-view information fusion. Information Fusion, 100:101937, 2023a.

Jie Xu, Huayi Tang, Yazhou Ren, Liang Peng, Xiaofeng Zhu, and Lifang He. Self-supervised discriminative feature learning for deep multi-view clustering. IEEE Transactions on Knowledge and Data Engineering, 35(7):7470–7482, 2023b.

Weiqing Yan, Yuanyang Zhang, Chenlei Lv, Chang Tang, Guanghui Yue, Liang Liao, and Weisi Lin. Gcfagg: Global and cross-view feature aggregation for multi-view clustering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 19863–19872, June 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, 44(12):9005–9020, 2022.

Xihong Yang, Jiaqi Jin, Siwei Wang, Ke Liang, Yue Liu, Yi Wen, Suyuan Liu, Sihang Zhou, Xinwang Liu, and En Zhu. Dealmvc: Dual contrastive calibration for multi-view clustering. In Proceedings of the 31st ACM International Conference on Multimedia (MM), pp. 337–346, 2023.

Yuanyang Zhang, Yijie Lin, Weiqing Yan, Li Yao, Xinhang Wan, Guangyuan Li, Chao Zhang, Guanzhou Ke, and Jie Xu. Incomplete multi-view clustering via diffusion contrastive generation. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 39, pp. 22650–22658, 2025.

Zhaoyang Zhang, Li Liu, Fumin Shen, Heng Tao Shen, and Ling Shao. Binary multi-view clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(7):1774–1782, Jul 2019.

Qi Zheng, Jihong Zhu, Zhibin Li, and Hao Tang. Graph-guided unsupervised multiview representation learning. IEEE Transactions on Circuits and Systems for Video Technology, 33(1):146–159, 2022.

Xin Zou, Ruimeng Liu, Chang Tang, Zhenglai Li, Xinwang Liu, Kunlun He, and Wanqing Li. Learning disentangled representations for generalized multi-view clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 48(9):10489–10503, 2026.