Straight-Path Flow Matching for Incomplete Multi-View Clustering

Yiteng Yuan[1*], Junyan Wang[2*†], Zheyuan Liu[2], Hong Jia[3], Lei Fan[4], Zhulin Tao[5†], and Lianbo Guo[1]

[1] School of Software Engineering, Huazhong University of Science and Technology, Wuhan, China
[2] Australian Institute for Machine Learning, Adelaide University, Adelaide, Australia
[3] University of Auckland, Auckland, New Zealand
[4] University of New South Wales, Sydney, Australia
[5] Communication University of China, Beijing, China
yuanyiteng&lbguo@hust.edu.cn, junyan.wang&zheyuan.liu@adelaide.edu.au
hong.jia@auckland.ac.nz, lei.fan1@unsw.edu.au, taozl@cuc.edu.cn

Abstract. Incomplete Multi-View Clustering addresses the problem of clustering multi-modal data when certain views are missing. Recent end-to-end generative approaches leverage diffusion models to recover missing views via stochastic noise-to-data trajectories. While expressive, such mechanisms are not explicitly designed for clustering, as they initialize from cluster-agnostic noise and rely on stochastic denoising dynamics. In this work, we revisit probability path design in end-to-end generative IMVC. We introduce a flow-matching framework with a linear interpolation path between paired view representations, that replaces diffusion with probability flows between observed and missing views. We provide a formal analysis showing that deterministic ODE flows are inherently better aligned with clustering objectives than diffusion-based stochastic trajectories, especially in terms of transport mechanisms that respect class-conditional data distributions and maintain cluster consistency in finite-step regimes. Building upon this insight, we develop an end-to-end IMVC architecture that integrates straight-path flow-matching view completion with cluster-level and entropy-based alignment to enforce cross-view clustering consistency. Extensive experiments on standard IMVC benchmarks demonstrate that the proposed framework establishes new state-of-the-art performance.

Keywords: Incomplete Multi-View Clustering · Flow Matching · ODE

1 Introduction

The task of Incomplete Multi-View Clustering (IMVC) [8] addresses the challenge of multi-view clustering (MVC) [2], with an additional layer of difficulty, i.e., not all views are consistently available for each sample. Compared to the

⋆ Equal contribution. † Corresponding author.

traditional MVC task setup, IMVC more closely reflects real-world use cases involving multi-view data [40, 52, 53], as such data collection processes may encounter sensor malfunctions, occlusions, disk space limitations, or even storage media corruptions.

Traditional IMVC methods [24, 25, 42, 43, 50, 54] focus on latent alignment and reconstruction under missing-view constraints. Recently, generative modeling [6, 10, 31] has emerged as a powerful alternative paradigm. Diffusion-based approaches [40,53], in particular, formulate missing-view recovery as conditional generation: starting from Gaussian noise, a stochastic denoising process conditioned on the observed view progressively reconstructs the missing representation. Earlier works [40] adopt a two-stage pipeline in which clustering is performed after generative recovery. Recent research has shifted toward end-to-end IMVC frameworks [53] that optimize missing-view recovery and clustering jointly. In the joint framework, the generative module is trained under clustering-oriented supervision, so the intermediate and final recovered features directly determine the clustering embeddings and assignments during training. Consequently, the properties of the generative trajectory directly influence the structural organization of the learned representations, as cross-view discrepancies are progressively reduced and intra-cluster coherence is reinforced under clustering-oriented supervision.

However, diffusion-based generative modeling typically relies on stochastic dynamics, often formulated as stochastic differential equations (SDEs) [32], that transform cluster-agnostic Gaussian noise into data representations through progressive denoising. As illustrated in Fig.1, the reverse process starts from noise that is independent of the observed sample and gradually reconstructs the target representation, causing trajectories to traverse regions of the latent space that are initially unrelated to the underlying cluster structure. Moreover, during the early and intermediate stages of denoising, the representations remain strongly influenced by noise, and cluster-discriminative structure becomes evident only in the low-noise regime (Fig.2). Although such properties are natural for general-purpose generation, IMVC instead requires view-completion mechanisms that are deterministically conditioned on observed representations and preserve the clustering structure of the representations throughout the transport process.

To this end, we adopt flow matching [22] to learn a deterministic ODE transport operator for missing-view recovery. Unlike stochastic diffusion trajectories, flow matching enables direct modeling of continuous transports anchored at the observed representation and consistent with the clustering structure of the representation space. Concretely, we specify a linear interpolation (i.e., straight-path) between paired latent representations of observed and missing views, and train a neural vector field to match the corresponding path velocity. This yields trajectories that evolve consistently with the underlying cluster structure while requiring only a small number of integration steps. A formal analysis comparing diffusion trajectories and flow-based transport is provided to clarify their structural differences. Building on this formulation, we integrate straight-path completion with an end-to-end IMVC framework that combines within-view reconstruction and cross-view alignment, where the cluster-level contrastive loss enforces assignment consistency and the entropy-based alignment stabilizes probability predictions. The resulting design jointly optimizes view recovery and clustering consistency within a unified framework.

Our contributions are summarized as follows:

- We propose a flow-matching framework for incomplete multi-view clustering (IMVC) that performs cross-view completion by learning a vector field to transport an observed latent representation to its missing-view counterpart along a straight-path formulation.
- We present a formal analysis showing straight-path transport and finite-step cluster separability of flow matching under standard assumptions.
- We develop an end-to-end model that integrates within-view reconstruction, cross-view alignment (cluster-level and entropic alignment), and straight-path completion, achieving SOTA performance on IMVC benchmarks.

2 Related Work

Deep Incomplete Multi-View Clustering. Handling missing modalities is a central challenge in incomplete multi-view clustering (IMVC). Autoencoder-based approaches [26, 39, 47, 51] learn latent representations by reconstructing observed views. Graph-based methods [4, 38, 41] propagate information across samples through graph structures. Contrastive learning approaches [7,17,21] improve cross-view representation consistency by maximizing agreement between views. To explicitly reconstruct missing modalities, recent studies introduce generative models [9]. Generative Adversarial Networks (GANs) [19, 35, 37] enable view synthesis but often suffer from training instability and mode collapse. Variational Autoencoders (VAEs) [3, 13, 27] provide a probabilistic framework for view generation. More recently, diffusion models have achieved remarkable progress in generative modeling [10, 16, 29, 36]. They have also been explored for missing-view generation in IMVC [40, 53]. However, diffusion models generate representations through stochastic noise-to-data trajectories that start from cluster-agnostic noise, so the transport process does not explicitly preserve the clustering structure of the representations.

Flow Matching in Generative Modeling. Flow Matching (FM) [14, 22, 23] is a generative framework that learns deterministic vector fields to transport samples between probability distributions through ODE flows. Instead of relying on stochastic noise injection and iterative denoising as in diffusion models, FM directly models probability transport along predefined interpolation paths between data distributions. Several extensions have further expanded its applicability, including latent-space flow matching [5], flow generator matching [11], and hybrid diffusion–flow formulations [30]. The deterministic transport formulation of flow matching provides a natural mechanism for modeling representation transport between paired observations. Motivated by this formulation, we adopt flow matching to model cross-view completion in the latent representation space for incomplete multi-view clustering.

3 Methodology

In this section, we first concretely demonstrate that, under three assumptions, flow matching exhibits desirable theoretical properties over conditional diffusion models when applied to the task of IMVC (Sec.3.1). Based on the theoretical analysis, we proceed to propose our method of adopting flow matching for the task in Sec.3.2.

3.1 Motivation of Adopting Flow Matching

Problem Setup. Without loss of generality, we consider a two-view IMVC setting. Let $\mathcal{X} \subset \mathbb{R}^D$ denote the original data space and $\mathcal{Z} \subset \mathbb{R}^d$ the corresponding latent representation space. For each sample, the raw observations from the two views are denoted by $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{X}$, and their encoded latent representations are $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathcal{Z}$, respectively.

Under this formulation, the core objective of IMVC is as follows: given an observed view representation (e.g., $\boldsymbol{z}_1$), recover the missing-view representation (e.g., $\boldsymbol{z}_2$) such that both representations correspond to the same cluster. Equivalently, if $c : \mathcal{Z} \to \{1, \dots, K\}$ denotes the clustering assignment function, the desired consistency condition is $c(\boldsymbol{z}_1) = c(\boldsymbol{z}_2)$.

Assumptions. We establish all arguments under the following three assumptions. Note that they are shared by both the conditional diffusion models and flow matching.

(i) *Cluster Separability.* The cluster sets $\mathcal{S}_k$ $(k = 1, \dots, K)$ in the representation space are compact subsets of $\mathbb{R}^d$. Any two clusters have a strictly positive minimum distance $\delta > 0$, i.e., $\inf_{i \neq j} d(\mathcal{S}_i, \mathcal{S}_j) = \delta > 0$.

(ii) *View Consistency.* Different-view representations of the same sample belong to the same cluster, i.e., the clustering label mapping $c(\cdot)$ satisfies: for any sample representations $\boldsymbol{z}^v$ and $\boldsymbol{z}^{v'}$, $c(\boldsymbol{z}^v) = c(\boldsymbol{z}^{v'})$.

(iii) *Sufficient Model Approximation.* The neural network can sufficiently approximate the target function with bounded approximation error. i.e., the model can capture the true probability flow or the vector field.

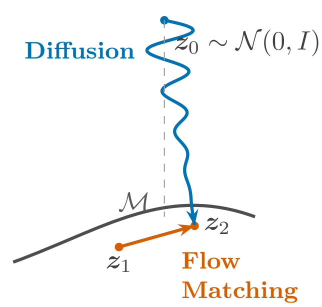


Fig. 1: Comparison between Flow Matching and Diffusion trajectories. Without loss of generality, we illustrate a two-view IMVC setting, where $\boldsymbol{z}_1$ denotes an observed view and $\boldsymbol{z}_2$ denotes a missing view. $\mathcal{M}$ denotes the region of the representation space that contains the cluster data, while $\boldsymbol{z}_0$ represents Gaussian noise. **Diffusion trajectory** deviates from $\mathcal{M}$ due to injected noise, resulting in a long and winding path when recovering $\boldsymbol{z}_2$. In contrast, the **Flow Matching trajectory** remains within the data region $\mathcal{M}$, producing a more direct path that shortens as clustering improves.

Preliminary on Flow Matching. In IMVC, to address the view-completion and the clustering objectives, flow matching directly constructs a deterministic ODE flow from the observed view representation $\boldsymbol{z}_1$ to the completed view representation $\boldsymbol{z}_2$. Here, time domain $t \in [0, 1]$, with $t = 0$ for the observed view, $t = 1$ for the completed view. We note that flow matching does not involve random terms in its formulation, or noise initializations.

Interpolation Process. Flow matching first performs a linear interpolation between the observed view $\boldsymbol{z}_1$ (source distribution $p_0 = p(\boldsymbol{z}_1)$) and the completed view $\boldsymbol{z}_2$ (target distribution $p_1 = p(\boldsymbol{z}_2)$) to construct the interpolated representation $\boldsymbol{x}_t$ at time t. The corresponding target vector field is the velocity of the interpolation path, i.e.,

$$\boldsymbol{x}_t = (1-t)\boldsymbol{z}_1 + t\boldsymbol{z}_2, \quad t \in [0,1], \tag{1}$$

$$\boldsymbol{v}_{\text{target}}(\boldsymbol{x}_t, t) = \boldsymbol{z}_2 - \boldsymbol{z}_1, \tag{2}$$

where Eq.1 represents the shortest path in Euclidean space, i.e., a straight line path, which is the optimal shortcut connecting the observed view and the completed view. Its path length is directly determined by $\|\boldsymbol{z}_2 - \boldsymbol{z}_1\|$.

Vector Field Learning. The learning objective of flow matching is to fit a vector field $\boldsymbol{v}_\theta(\boldsymbol{x}_t, t)$ via a neural network, such that it approximates the target vector field $\boldsymbol{v}_{\text{target}}(\boldsymbol{x}_t, t)$. The training loss uses mean squared error:

$$\mathcal{L}_{\text{FM}} = \mathbb{E}_{t\sim\mathcal{U}(0,1),\boldsymbol{z}_1,\boldsymbol{z}_2} \left\|\boldsymbol{v}_\theta(\boldsymbol{x}_t, t) - \boldsymbol{v}_{\text{target}}(\boldsymbol{x}_t, t)\right\|_2^2, \tag{3}$$

where $\mathcal{U}(0, 1)$ is the uniform distribution on $[0, 1]$. The vector field $\boldsymbol{v}_\theta(\cdot)$ takes the interpolated representation $\boldsymbol{x}_t$, the time step t, and the observed view $\boldsymbol{z}_1$ as input and outputs a velocity vector of the same dimension as the representation.

Reversibility. The ODE framework of flow matching possesses a natural reversibility. If the forward vector field $\boldsymbol{v}_\theta(\cdot)$ from $\boldsymbol{z}_1$ to $\boldsymbol{z}_2$ is learned, the reverse vector field $\boldsymbol{v}_{\theta_R}(\cdot)$ from $\boldsymbol{z}_2$ to $\boldsymbol{z}_1$ is simply $-\boldsymbol{v}_\theta(\cdot)$. The reverse ODE is:

$$\frac{d\boldsymbol{x}_s}{ds} = \boldsymbol{v}_{\theta_R}(\boldsymbol{x}_s, s) = -\boldsymbol{v}_\theta(\boldsymbol{x}_{1-t}, 1-t), \quad x_s = (1-s)\boldsymbol{z}_2 + s\boldsymbol{z}_1. \tag{4}$$

This property grants that flow matching only needs to train the vector field in one direction to achieve bidirectional view completion, where we provide a detailed proof in Appendix D. This is in contrast to methods using diffusion models, which require training two separate models.

Flow Matching in IMVC. With the preliminaries of flow matching introduced above, we argue that it is structurally more aligned to the task of IMVC, compared to existing diffusion-based methods. We concretely demonstrate our claim through the two corollaries below, while relating each one to the limitations of the diffusion models. A complete theoretical analysis regarding the properties and limitations of the diffusion model is provided in Appendix A, detailed proofs of the corollaries are in Appendix C.

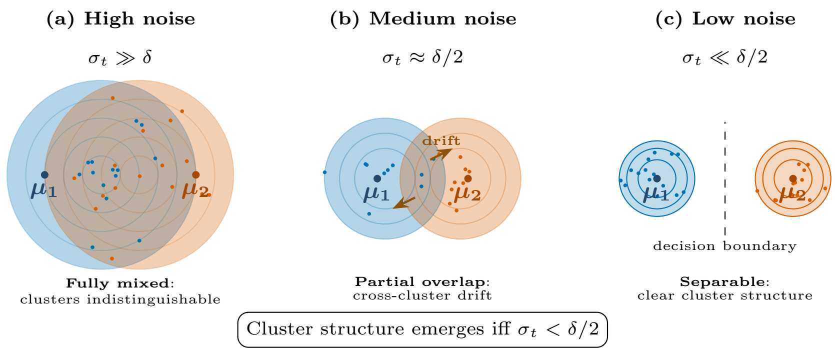


Fig. 2: Cluster separability under diffusion noise. (a) High noise ($\sigma_t \gg \delta$): supports fully overlap, samples mix heavily; **(b)** Medium noise ($\sigma_t \approx \delta/2$): partial overlap may cause cross-cluster drift; **(c)** Low noise ($\sigma_t \ll \delta/2$): clusters are well separated with a clear boundary. In finite-step inference, σ_t is often large due to limited steps, making scenario (b) common and leading to cross-cluster drift. We refer readers to Appendix D, Eq.20 for detailed analysis.

Corollary 1 (Flow Matching's Shortcut Path). *Flow matching constructs an optimal shortcut path between views representation.*

As illustrated in Fig.1, during the training process of IMVC, view representations gradually align as $\|\boldsymbol{z}_1 - \boldsymbol{z}_2\|$ decreases. The path length in flow matching shortens adaptively. When views are perfectly aligned, the path length becomes a small value. On the other hand, the diffusion model's starting point is global Gaussian noise unrelated to the cluster class-conditional data distributions, which encode the clustering structure (Appendix A.2, Eq.18). The reverse process must complete the full evolution of "noise space $\Rightarrow$ representation space $\Rightarrow$ cluster-specific distribution". Even if the observed view representation and completion target view representation in latent space are highly aligned, diffusion cannot utilize the view consistency information. The path length is fixed as the distance from the noise to the cluster set.

Remark 1 (Shortcut Path Advantage). Compared to the diffusion model's Gaussian noise initialization (Eq.18), flow matching offers the theoretical advantages of non-redundant paths and adaptively shortening path length, as shown in Fig.1.

Corollary 2 (Flow Matching's Finite-Step Separability). *Flow matching achieves cluster separability within any finite number of steps.*

In other words, sample trajectories from different cluster sets remain separated throughout the evolution, with no cross-cluster drift, as in Fig.2 **(c)**. In contrast, the cluster separability of the diffusion model relies on a strong asymptotic condition (where $\sigma_t < \delta/2$), which only holds as $t \to 0$, as shown in Fig.2 and is discussed in detail in Appendix A.2, Eq.20. This condition is difficult to satisfy in finite-step practical inference, leading to overlapping supports of noised distributions for different clusters and making samples prone to cross-cluster drift, significantly reducing cluster separability.

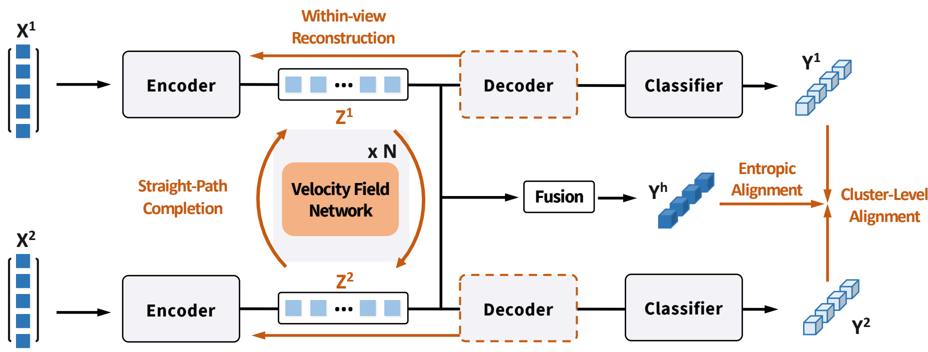


Fig. 3: Overview of the proposed flow-matching framework for incomplete multi-view clustering. Each view is encoded into latent representations and reconstructed via decoders. A velocity field network models straight-path transport between paired latent for cross-view completion. The completed representations are fused for cluster prediction, followed by cluster-level alignment and entropic alignment.

Remark 2 (Finite-Step Separability Advantage). Flow matching offers the theoretical benefit of finite-step cluster separability with no cross-cluster drift compared to the diffusion model's asymptotic separability condition (Fig.2, Eq.20).

3.2 Flow-Matching Formulation of IMVC

Motivated by the theoretical analysis in Sec.3.1, we formulate incomplete multi-view clustering as a cross-view completion problem in the latent representation space. Under the assumptions of cluster separability, view consistency, and sufficient model approximation, our method is designed to preserve clustering structure while recovering missing-view representations. The overall framework consists of three components: within-view reconstruction, cross-view alignment (including cluster-level and entropic alignment), and straight-path completion.

Within-view Reconstruction. Before performing cross-view transport, we first learn compact latent representations for each view via view-specific autoencoders. Given an input feature $\boldsymbol{x}_v \in X$ from view $v \in \{1, 2\}$, an encoder $\mathcal{E}_v(\cdot)$ maps it to a latent representation $\boldsymbol{z}_v = \mathcal{E}_v(\boldsymbol{x}_v) \in Z$, and a decoder $\mathcal{D}_v(\cdot)$ reconstructs it as $\hat{\boldsymbol{x}}_v = \mathcal{D}_v(\boldsymbol{z}_v)$. This step extracts view-specific structural information and learns latent representations for each view. To preserve the intrinsic structure of each view in the latent space, we adopt a standard reconstruction objective in N samples:

$$\mathcal{L}_{\text{rec}} = \frac{1}{N} \sum_{i=1}^{N} \left(\|\mathcal{D}_1(\mathcal{E}_1(\boldsymbol{x}_1^{(i)})) - \boldsymbol{x}_1^{(i)}\|^2 + \|\mathcal{D}_2(\mathcal{E}_2(\boldsymbol{x}_2^{(i)})) - \boldsymbol{x}_2^{(i)}\|^2 \right), \tag{5}$$

Cluster-Level Alignment. Under Assumptions *(i)–(ii)*, different views of the same sample are expected to belong to the same cluster and therefore share a consistent cluster assignment. However, in practice, the view-specific encoders and the learned clustering head can yield view-dependent predictions, leading to inconsistent assignments across views and degrading the clustering structure of the learned representations. We therefore employ a cluster-level contrastive objective to explicitly couple the two views in the clustering space.

Specifically, let $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathbb{R}^K$ be the cluster probability vectors predicted from $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$, respectively. We adopt an InfoNCE [28] objective to align cross-view cluster predictions in N samples:

$$\mathcal{L}_{\text{cluster}} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\exp\Big(\cos(\boldsymbol{y}_1^{(i)}, \boldsymbol{y}_2^{(i)})/\tau\Big)}{\sum_{j=1}^{N}\exp\Big(\cos(\boldsymbol{y}_1^{(i)}, \boldsymbol{y}_2^{(j)})/\tau\Big)}, \tag{6}$$

where $\cos(\cdot,\cdot)$ denotes cosine similarity, and τ is a temperature parameter. This objective increases the similarity between corresponding cross-view predictions while discouraging alignment with non-matching samples, thereby promoting view-consistent clustering and preserving inter-cluster separation.

Entropic Alignment. Although the cluster-level contrastive objective promotes cross-view assignment consistency, it operates at the sample level and does not explicitly regularize the structure of the predicted probability distributions. In practice, direct fusion in Euclidean space may distort the relative probability structure and over-smooth cluster confidence. We therefore employ an entropy-weighted alignment objective to regularize probability consistency across views.

Let $\boldsymbol{y}_1$, $\boldsymbol{y}_2$, and $\boldsymbol{y}_h$ denote the cluster probability distributions predicted from $\boldsymbol{z}_1$, $\boldsymbol{z}_2$, and the fused representation $\boldsymbol{h}$, respectively. Since these distributions lie on the probability simplex, we adopt the centered log-ratio (clr) transformation (Appendix E) to perform fusion in a simplex-consistent space. An entropy-based weighting scheme assigns larger weights to more confident predictions, and the target distribution $\boldsymbol{y}^*$ is obtained via a weighted combination in the clr space followed by inverse mapping back to the simplex. We then enforce consistency between the fused prediction and the entropy-weighted target via KL divergence:

$$\mathcal{L}_{\text{entropy}} = \mathrm{KL}(\boldsymbol{y}_h \,\|\, \boldsymbol{y}^*), \tag{7}$$

where $\mathrm{KL}\,(\cdot\|\cdot)$ is the KL divergence. This objective stabilizes cross-view probability structure while preventing uncertain predictions from dominating the fusion, thereby preserving post-alignment cluster separability.

Straight-Path Completion. With cluster-level and entropic alignment regularizing cross-view consistency in the clustering space, we now realize cross-view representation completion via flow matching in the latent space. Given paired latent representations $(\boldsymbol{z}_1, \boldsymbol{z}_2)$ from complete samples, we construct the interpolated state $\boldsymbol{x}_t$ along the straight-line path (Eq.1). A neural vector field $\mathcal{F}(\cdot)$ is trained to approximate the target transport direction, so that integrating the learned ODE maps an observed representation to its missing-view counterpart.

Flow Prediction Loss. Given the straight-line interpolation $\boldsymbol{x}_t$ defined in Eq.1, the target transport direction at any intermediate state is uniquely determined as $\boldsymbol{z}_2 - \boldsymbol{z}_1$ (see Eq.2). We therefore learn a parametric vector field $\mathcal{F}(\boldsymbol{x}_t, t)$ to approximate this target velocity by minimizing

$$\mathcal{L}_{\text{flow-pred}} = \mathbb{E}_{t,\boldsymbol{z}_1,\boldsymbol{z}_2} \left\| \mathcal{F}(\boldsymbol{x}_t, t) - (\boldsymbol{z}_2 - \boldsymbol{z}_1) \right\|_2^2 . \tag{8}$$

This objective enforces the learned vector field to follow the straight-line shortcut path between paired latent representations, as discussed in Corollary 1. By aligning the local velocity with $\boldsymbol{z}_2 - \boldsymbol{z}_1$, the transport process exploits view consistency and adaptively shortens the path length as cross-view representations become aligned. This avoids redundant detours and maintains the cluster structure established by the alignment modules.

Completion Consistency Loss. Although the vector field is trained to approximate the shortcut transport direction, finite-step numerical integration may introduce approximation errors. Based on the reversibility property of flow matching (Eq. 4), forward integration $(t : 0 \to 1)$ maps $\boldsymbol{z}_1$ to the recovered feature $\hat{\boldsymbol{z}}_2$, while reverse integration $(s : 0 \to 1, s = 1 - t)$ maps $\boldsymbol{z}_2$ to $\hat{\boldsymbol{z}}_1$:

$$\hat{\boldsymbol{z}}_2 = \boldsymbol{z}_1 + \int_0^1 \mathcal{F}(\boldsymbol{x}_t, t)\ \mathrm{d}t, \tag{9}$$

$$\hat{\boldsymbol{z}}_1 = \boldsymbol{z}_2 + \int_0^1 -\mathcal{F}(\boldsymbol{x}_{1-t}, 1 - t)\ \mathrm{d}s. \tag{10}$$

We then enforce bidirectional completion consistency by minimizing

$$\mathcal{L}_{\text{flow-comp}} = \mathbb{E}_{\boldsymbol{z}_1,\boldsymbol{z}_2} \left[\|\hat{\boldsymbol{z}}_2 - \boldsymbol{z}_2\|_2^2 + \|\hat{\boldsymbol{z}}_1 - \boldsymbol{z}_1\|_2^2 \right] . \tag{11}$$

This loss constrains the transport trajectory at both endpoints, ensuring that finite-step evolution remains faithful to the true latent representations. Together with the shortcut-path property, it supports stable cross-view completion while maintaining the cluster structure discussed in Corollary 2.

Overall Objective. The model is trained by minimizing a weighted combination of the losses introduced above, which jointly optimize representation reconstruction, cross-view cluster consistency, probability-level alignment, and shortcut-path completion:

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \lambda_1 \mathcal{L}_{\text{flow-pred}} + \lambda_2 \mathcal{L}_{\text{flow-comp}} + \lambda_3 \mathcal{L}_{\text{cluster}} + \lambda_4 \mathcal{L}_{\text{entropy}}, \tag{12}$$

where $\lambda_1, \lambda_2, \lambda_3$, and λ_4 balance the contributions of the transport, alignment, and reconstruction objectives.

4 Experiment

This section presents the experimental evaluation of the proposed framework. We first describe the experimental settings (Sec.4.1), and compare our method with recent incomplete multi-view clustering approaches on multiple benchmarks (Sec.4.2). Then, we conduct ablation studies to analyze the contribution of each component (Sec.4.3). Finally, we provide additional analyses, including hyper-parameter sensitivity, representation visualization, and flow-step behavior, to further understand the properties of the proposed method (Sec.4.4).

4.1 Experiment Setup

Dataset. We evaluate the proposed method on five multi-view benchmarks: *(i)* Synthetic3D [15], a synthetic Gaussian-mixture dataset with correlated features, where two views are selected following [53]; *(ii)* CUB [34], a fine-grained bird dataset containing 11,788 image–text pairs from 200 categories, where visual features are extracted using GoogLeNet and textual features via doc2vec, and the first 10 categories are used as in [18,53]; *(iii)* HandWritten [1], a digit dataset with 2,000 samples and six feature views, from which two views are adopted following [53]; *(iv)* LandUse-21 [49], a remote-sensing dataset with 2,100 images from 21 categories, where PHOG and LBP descriptors are used as two views; and *(v)* Fashion [44], a product image dataset with 10,000 samples, from which two visual feature views are selected.
Metric. Following the standard incomplete multi-view clustering (IMVC) protocol [21, 53], we simulate missing views by randomly masking one view for a subset of samples. The missing rate is defined as the ratio of incomplete samples to the total dataset size. We follow [21,53] to evaluate method performance using three standard metrics: Accuracy (ACC), Normalized Mutual Information (NMI), and Adjusted Rand Index (ARI).
Implementation details. Our framework consists of view-specific autoencoders, a straight-path flow matching module, and a clustering classifier. The model is trained for 200 epochs using the Adam optimizer with an initial learning rate of 1×10^{-3}. The batch size is set according to the dataset configuration (typically 128 or 256). We simulate incomplete multi-view scenarios by randomly masking one view for a subset of samples based on the target missing rate ($\rho \in \{0.1, 0.3, 0.5\}$). The weighting coefficients $\lambda_1, \lambda_2, \lambda_3$, and λ_4 are configured individually for each dataset. For the flow matching process, we adopt an efficient setting with a single integration step ($Step = 1$) during both training and inference. We refer readers to Appendix. F for detailed implementation details.

4.2 Comparisons with State-of-the-Arts.

Table 1 compares our method with recent incomplete multi-view clustering approaches under three missing rates on five benchmark datasets. Overall, the proposed framework consistently achieves the best or competitive performance across almost all evaluation metrics. On datasets with clear cluster structures,

Table 1: Comparison with state-of-the-art incomplete multi-view clustering (IMVC) methods under different missing rates. ACC, NMI, and ARI denote clustering Accuracy, Normalized Mutual Information, and Adjusted Rand Index, respectively (higher is better). τ indicates the missing rate of views.

Dataset	Method	$\tau = 0.1$			$\tau = 0.3$			$\tau = 0.5$		
		ACC	**NMI**	**ARI**	**ACC**	**NMI**	**ARI**	**ACC**	**NMI**	**ARI**
Synthetic3d	DCP [20]	88.00	65.17	67.87	79.83	55.21	56.69	85.53	58.33	62.56
	DSIMVC [33]	73.11	59.02	58.38	70.78	56.41	55.43	67.44	51.74	49.42
	GCFAGG [48]	72.52	57.87	56.68	70.23	55.51	53.34	69.12	53.21	52.03
	CPSPAN [12]	88.83	65.51	69.12	87.50	61.79	66.68	86.33	59.29	63.65
	APADC [46]	85.73	59.07	62.98	86.47	60.84	64.90	84.38	58.48	61.85
	ProImp [17]	86.55	61.79	65.42	85.50	59.29	62.81	82.78	54.76	57.69
	DVIMVC [45]	50.03	28.43	24.57	46.97	22.07	15.93	50.32	25.64	18.79
	ICMVC [4]	85.03	58.30	61.96	87.17	61.20	66.29	84.20	55.33	59.60
	DCG [53]	91.23	71.35	76.05	88.00	64.11	68.48	87.67	63.23	67.52
	Ours	**94.17**	**78.66**	**83.51**	**91.17**	**69.56**	**75.56**	**88.17**	**63.78**	**68.26**
CUB	DCP [20]	42.77	55.42	33.39	40.60	52.37	31.41	38.87	50.15	31.18
	DSIMVC [33]	63.67	59.85	46.23	49.22	49.93	32.80	48.99	49.61	32.35
	GCFAGG [48]	67.67	64.14	51.14	62.72	60.27	45.03	59.63	55.95	39.91
	CPSPAN [12]	76.67	71.38	58.65	74.33	70.33	56.84	73.33	69.68	57.17
	APADC [46]	53.04	59.05	42.98	70.33	70.33	56.84	48.53	59.93	41.54
	ProImp [17]	69.33	69.65	55.93	71.83	64.45	51.64	69.56	64.19	52.06
	DVIMVC [45]	63.83	63.00	50.50	70.42	68.43	56.35	68.98	68.81	52.61
	ICMVC [4]	32.33	33.36	16.85	45.13	41.32	25.11	41.70	37.78	21.62
	DCG [53]	82.23	77.70	69.21	77.17	71.35	59.85	75.50	**72.21**	59.12
	Ours	**86.00**	**78.19**	**72.28**	**82.67**	**72.80**	**66.31**	**80.50**	71.74	**63.15**
HandWritten	DCP [20]	63.66	70.44	52.82	75.68	79.05	67.45	72.07	76.17	63.81
	DSIMVC [33]	71.37	70.24	60.30	67.47	65.27	54.26	56.90	57.99	44.12
	GCFAGG [48]	75.07	68.19	59.79	69.16	63.25	53.67	61.02	53.57	42.12
	CPSPAN [12]	68.70	69.06	59.17	82.20	75.05	69.58	72.55	68.60	57.51
	APADC [46]	72.58	71.29	59.07	60.66	64.84	48.75	52.70	63.12	39.96
	ProImp [17]	81.05	78.48	70.40	80.60	77.26	69.72	78.98	74.08	66.47
	DVIMVC [45]	29.08	19.36	11.06	26.63	16.57	8.84	25.13	15.69	8.09
	ICMVC [4]	82.70	81.06	74.67	82.27	79.90	72.81	75.04	71.89	63.29
	DCG [53]	82.75	82.63	74.88	82.70	80.54	73.96	80.80	76.21	70.45
	Ours	**85.30**	**82.78**	**76.72**	**86.50**	**82.09**	**77.04**	**83.40**	**76.26**	**70.76**
LandUse-21	DCP [20]	26.19	31.20	13.02	25.48	30.18	11.08	21.52	26.32	11.11
	DSIMVC [33]	16.56	16.40	4.32	16.46	16.57	4.35	16.70	16.84	4.50
	GCFAGG [48]	19.05	19.99	6.32	18.53	19.89	6.24	18.43	19.47	5.98
	CPSPAN [12]	20.05	27.20	8.14	18.05	25.75	6.90	20.38	26.99	8.05
	APADC [46]	20.92	26.74	7.97	19.99	24.40	7.44	19.15	23.91	7.08
	ProImp [17]	24.43	29.22	11.44	24.63	28.45	11.78	24.45	27.43	10.94
	DVIMVC [45]	13.90	22.48	2.66	13.24	23.21	2.74	12.86	19.67	2.54
	ICMVC [4]	26.81	30.71	13.60	26.12	29.83	12.12	26.12	29.83	12.12
	DCG [53]	27.52	31.36	14.57	27.33	**32.09**	14.47	25.76	29.07	13.17
	Ours	**29.10**	**31.95**	**15.15**	**28.19**	31.71	**14.55**	**27.24**	**30.84**	**13.66**
Fashion	DCP [20]	83.70	84.30	76.50	71.80	70.90	52.50	60.80	59.50	33.10
	DSIMVC [33]	88.00	86.40	81.10	87.30	85.00	78.90	83.50	80.30	73.70
	GCFAGG [48]	78.21	74.50	66.28	76.34	72.53	63.98	74.47	69.83	60.37
	CPSPAN [12]	66.16	68.45	55.73	64.80	68.22	55.55	54.81	64.13	48.84
	APADC [46]	81.40	86.50	73.30	80.90	85.01	73.10	75.40	81.50	67.60
	ProImp [17]	92.88	88.34	86.09	74.11	76.97	66.42	89.76	81.94	79.93
	DVIMVC [45]	79.38	80.22	71.50	82.26	79.82	72.46	80.17	77.09	69.51
	ICMVC [4]	92.41	87.05	85.11	89.31	83.06	79.93	79.37	74.44	68.46
	DCG [53]	95.83	91.29	91.19	93.13	86.99	86.00	90.04	82.25	79.99
	Ours	**96.07**	**91.99**	**91.78**	**93.66**	**87.93**	**87.07**	**90.09**	**82.83**	**80.58**

Table 2: Ablation study of the proposed loss components on the HandWritten dataset under missing rate $\tau = 0.3$. $\mathcal{L}_{\text{flow-pred}}$ and $\mathcal{L}_{\text{flow-comp}}$ denote the flow-based shortcut completion losses, $\mathcal{L}_{\text{cluster}}$ represents the cluster-level alignment loss, and $\mathcal{L}_{\text{entropy}}$ refers to the entropy-based alignment loss. ✓ and ✗ indicate whether a component is enabled or removed. All variants retain the reconstruction loss.

Variants	**Loss Components**				**Metrics (%)**		
	$\mathcal{L}_{\text{flow-pred}}$	$\mathcal{L}_{\text{flow-comp}}$	$\mathcal{L}_{\text{cluster}}$	$\mathcal{L}_{\text{entropy}}$	**ACC**	**NMI**	**ARI**
(w/o) $\mathcal{L}_{\text{flow-pred}}$	✗	✓	✓	✓	73.95	70.92	61.42
(w/o) $\mathcal{L}_{\text{flow-comp}}$	✓	✗	✓	✓	76.15	71.86	63.38
(w/o) $\mathcal{L}_{\text{flow-pred}}$ & $\mathcal{L}_{\text{flow-comp}}$	✗	✗	✓	✓	77.35	71.89	64.16
(w/o) $\mathcal{L}_{\text{cluster}}$	✓	✓	✗	✓	27.75	37.13	20.05
(w/o) $\mathcal{L}_{\text{entropy}}$	✓	✓	✓	✗	77.20	71.79	66.69
(w/o) $\mathcal{L}_{\text{cluster}}$ & $\mathcal{L}_{\text{entropy}}$	✓	✓	✗	✗	28.70	17.67	7.75
Full model	✓	✓	✓	✓	**86.50**	**82.09**	**77.04**

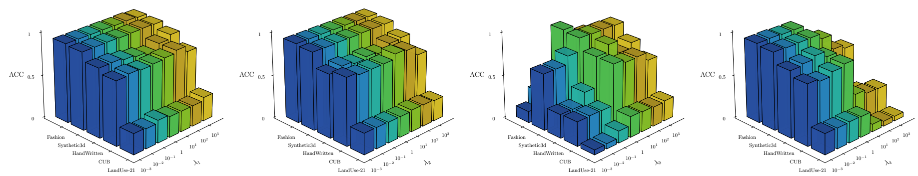
such as Synthetic3D and HandWritten, our method achieves notable improvements across all metrics. These results suggest that the straight-path completion learned by flow matching can recover missing representations while maintaining the underlying cluster geometry.

The advantage of our method becomes more pronounced under higher missing rates. As τ increases from 0.1 to 0.5, many existing methods experience notable performance degradation due to reduced cross-view correspondence, whereas our approach remains relatively stable across datasets. This robustness arises from the combination of alignment and flow-based completion: the cluster-level and entropic alignment modules enforce view-consistent assignments, while the flow-matching transport recovers missing representations by following the straight-path transport between paired latent representations. This behavior is consistent with our theoretical analysis that flow matching maintains cluster separability under finite-step evolution, resulting in stronger robustness under severe incompleteness.

4.3 Ablation Study

We conduct ablation experiments on HandWritten with a missing rate $\tau = 0.3$ to analyze the contribution of each component in the proposed flow-matching formulation, as summarized in Table 2.

Effect of $\mathcal{L}_{\text{cluster}}$. Removing the cluster-level contrastive loss causes the most severe performance drop (ACC decreases to 27.75%). This loss explicitly aligns cross-view cluster assignments and establishes view-consistent clustering structure in the latent space. Without this constraint, representations from different views may produce inconsistent cluster predictions, violating the cluster separability assumption required by the flow-matching formulation. Consequently, reliable cross-view completion cannot be achieved.



(a) ACC vs. λ_1 **(b)** ACC vs. λ_2 **(c)** ACC vs. λ_3 **(d)** ACC vs. λ_4

Fig. 4: Hyperparameter analysis of the loss coefficients on five datasets with missing rate $\tau = 0.3$. (a) λ_1 for the flow prediction loss $\mathcal{L}_{\text{flow-pred}}$; (b) λ_2 for the completion consistency loss $\mathcal{L}_{\text{flow-comp}}$; (c) λ_3 for the cluster-level contrastive loss $\mathcal{L}_{\text{cluster}}$; (d) λ_4 for the entropy-based alignment loss $\mathcal{L}_{\text{entropy}}$.

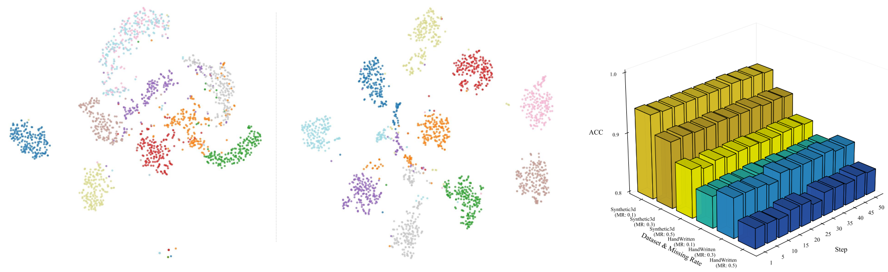
Effect of $\mathcal{L}_{\text{entropy}}$. The entropy-based alignment complements the cluster-level contrastive objective by further stabilizing cross-view probability predictions. While $\mathcal{L}_{\text{cluster}}$ enforces assignment consistency at the sample level, $\mathcal{L}_{\text{entropy}}$ regularizes the probability distributions across views. Removing this term leads to performance degradation, indicating that probability-level alignment is also important for maintaining stable cluster assignments and cross-view completion.

Effect of $\mathcal{L}_{\text{flow-pred}}$. The flow prediction loss supervises the learned vector field to follow the straight-path transport direction between paired latent representations. Removing this loss weakens the ability of the model to learn a consistent transport direction, resulting in degraded completion quality. This confirms that accurate velocity-field estimation is necessary for modeling cross-view transport.

Effect of $\mathcal{L}_{\text{flow-comp}}$. The completion consistency loss enforces bidirectional completion consistency between paired latent representations. Without this constraint, the recovered features may deviate from the true latent representations during finite-step integration, leading to degraded performance. This loss therefore improves the stability of cross-view completion. When both flow losses are removed, the model reduces to an alignment-only variant and achieves moderate performance. Introducing the flow-matching formulation substantially improves all metrics, demonstrating that straight-path transport effectively exploits the clustering structure established by the alignment modules.

4.4 Analysis

Hyperparameters Analysis. We analyze the sensitivity of the loss coefficients λ_1, λ_2, λ_3, and λ_4 under the missing rate $\tau = 0.3$, as shown in Fig. 4. The performance remains stable across a wide range of λ_1 and λ_2, corresponding to the flow prediction loss $\mathcal{L}_{\text{flow-pred}}$ and the completion consistency loss $\mathcal{L}_{\text{flow-comp}}$. In contrast, the method is more sensitive to λ_3 and λ_4, which control the cluster-level contrastive loss $\mathcal{L}_{\text{cluster}}$ and the entropy-based alignment loss $\mathcal{L}_{\text{entropy}}$. This observation is consistent with our formulation: alignment losses enforce cross-view assignment consistency and stabilize probability predictions, which are required by the assumptions underlying the flow-matching formulation.



(a) t-SNE Visualization **(b)** Flow-Step Analysis

Fig. 5: Representation visualization and flow-step analysis. (a) t-SNE visualization on the HandWritten dataset with missing rate $\tau = 0.3$, showing the raw features (left) and the representations learned by our method (right). (b) Clustering accuracy under different flow step numbers on the HandWritten and Synthetic3D datasets.

Visualization Analysis. We visualize the learned representations on the HandWritten dataset using t-SNE under the missing rate $\tau = 0.3$, as shown in Fig. 5. Compared with the scattered and mixed raw features, the representations learned by our framework form clearly separated clusters. This improvement arises because the cluster-level contrastive loss and entropy-based alignment enforce cross-view assignment consistency and stabilize probability predictions, while the flow-matching formulation performs cross-view completion by following the straight-path transport between paired latent representations.

Flow-Step Analysis. We analyze the influence of the number of flow steps used for cross-view completion. As shown in Fig. 5, the clustering performance remains stable across different step numbers. This observation is consistent with our theoretical Corollary 2 that flow matching achieves cluster separability within any finite number of steps. Once the vector field is learned, the straight-path transport learned by the vector field produces stable completion under different finite-step integration settings.

5 Conclusion

We presented a flow-matching framework for incomplete multi-view clustering that formulates missing-view recovery as a straight-path transport problem in the latent representation space. Through theoretical analysis, we showed that flow matching provides two desirable properties for IMVC: shortcut-path transport and finite-step cluster separability. Guided by this analysis, we developed an end-to-end model that integrates within-view reconstruction, cross-view alignment, and flow-based completion. Extensive experiments on multiple benchmarks demonstrate that the proposed approach consistently achieves strong clustering performance and remains robust under high missing rates.

References

1. Asuncion, A., Newman, D., et al.: Uci machine learning repository (2007)
2. Bickel, S., Scheffer, T.: Multi-view clustering. In: Icdm. vol. 4, pp. 19–26 (2004)
3. Cai, H., Huang, W., Yang, S., Ding, S., Zhang, Y., Hu, B., Zhang, F., Cheung, Y.M.: Realize generative yet complete latent representation for incomplete multi-view learning. IEEE Trans. Pattern Anal. Mach. Intell. **46**(5), 3637–3652 (2023)
4. Chao, G., Jiang, Y., Chu, D.: Incomplete contrastive multi-view clustering with high-confidence guiding. In: AAAI. vol. 38, pp. 11221–11229 (2024)
5. Dao, Q., Phung, H., Nguyen, B., Tran, A.: Flow matching in latent space. arXiv preprint arXiv:2307.08698 (2023)
6. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Adv. Neural Inform. Process. Syst. **34**, 8780–8794 (2021)
7. Feng, W., Sheng, G., Wang, Q., Gao, Q., Tao, Z., Dong, B.: Partial multi-view clustering via self-supervised network. In: AAAI. vol. 38, pp. 11988–11995 (2024)
8. Gao, H., Peng, Y., Jian, S.: Incomplete multi-view clustering. In: ACM Int. Conf. Multimedia. pp. 245–255. Springer (2016)
9. Gui, J., Sun, Z., Wen, Y., Tao, D., Ye, J.: A review on generative adversarial networks: Algorithms, theory, and applications. IEEE transactions on knowledge and data engineering **35**(4), 3313–3332 (2021)
10. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Adv. Neural Inform. Process. Syst. **33**, 6840–6851 (2020)
11. Huang, Z., Geng, Z., Luo, W., Qi, G.j.: Flow generator matching. arXiv preprint arXiv:2410.19310 (2024)
12. Jin, J., Wang, S., Dong, Z., Liu, X., Zhu, E.: Deep incomplete multi-view clustering with cross-view partial sample and prototype alignment. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 11600–11609 (2023)
13. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
14. Klein, L., Krämer, A., Noé, F.: Equivariant flow matching. Adv. Neural Inform. Process. Syst. **36**, 59886–59910 (2023)
15. Kumar, A., Rai, P., Daume, H.: Co-regularized multi-view spectral clustering. Adv. Neural Inform. Process. Syst. **24** (2011)
16. Labs, B.F., Batifol, S., Blattmann, A., Boesel, F., Consul, S., Diagne, C., Dockhorn, T., English, J., English, Z., Esser, P., et al.: Flux. 1 kontext: Flow matching for in-context image generation and editing in latent space. arXiv preprint arXiv:2506.15742 (2025)
17. Li, H., Li, Y., Yang, M., Hu, P., Peng, D., Peng, X.: Incomplete multi-view clustering via prototype-based imputation. arXiv preprint arXiv:2301.11045 (2023)
18. Li, H., Lin, Y., Hu, P., Yang, M., Peng, X.: Community-aware multi-view representation learning with incomplete information. IEEE Trans. Pattern Anal. Mach. Intell. (2025)
19. Lin, R., Du, S., Wang, S., Guo, W.: Consistent graph embedding network with optimal transport for incomplete multi-view clustering. Information Sciences **647**, 119418 (2023)
20. Lin, Y., Gou, Y., Liu, X., Bai, J., Lv, J., Peng, X.: Dual contrastive prediction for incomplete multi-view representation learning. IEEE Trans. Pattern Anal. Mach. Intell. **45**(4), 4447–4461 (2022)
21. Lin, Y., Gou, Y., Liu, Z., Li, B., Lv, J., Peng, X.: Completer: Incomplete multi-view clustering via contrastive prediction. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 11174–11183 (2021)

22. Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. arXiv preprint arXiv:2210.02747 (2022)
23. Liu, X., Gong, C., Liu, Q.: Flow straight and fast: Learning to generate and transfer data with rectified flow. arXiv preprint arXiv:2209.03003 (2022)
24. Liu, X., Li, M., Tang, C., Xia, J., Xiong, J., Liu, L., Kloft, M., Zhu, E.: Efficient and effective regularized incomplete multi-view clustering. IEEE Trans. Pattern Anal. Mach. Intell. **43**(8), 2634–2646 (2020)
25. Liu, X., Zhu, X., Li, M., Tang, C., Zhu, E., Yin, J., Gao, W.: Efficient and effective incomplete multi-view clustering. In: AAAI. vol. 33, pp. 4392–4399 (2019)
26. Luo, J., Zhao, W., Tang, Y., Zhou, Z., Xiong, H., Tao, Z.: Lightgbm using enhanced and de-biased item representation for better session-based fashion recommender systems. In: Proceedings of the Recommender Systems Challenge 2022, pp. 24–28 (2022)
27. Ma, Y., Liu, X., Wei, Y., Tao, Z., Wang, X., Chua, T.S.: Leveraging multimodal features and item-level user feedback for bundle construction. In: Proceedings of the 17th ACM International Conference on Web Search and Data Mining. pp. 510–519 (2024)
28. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
29. Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., Rombach, R.: Sdxl: Improving latent diffusion models for high-resolution image synthesis. arXiv preprint arXiv:2307.01952 (2023)
30. Schusterbauer, J., Gui, M., Fundel, F., Ommer, B.: Diff2flow: Training flow matching models via diffusion model alignment. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 28347–28357 (2025)
31. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models (2021)
32. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456 (2020)
33. Tang, H., Liu, Y.: Deep safe incomplete multi-view clustering: Theorem and algorithm. In: Int. Conf. Mach. Learn. pp. 21090–21110. PMLR (2022)
34. Wah, C., Branson, S., Welinder, P., Perona, P., Belongie, S., et al.: The caltech-ucsd birds-200-2011 dataset. Tech. rep., Technical Report CNS-TR-2011-001, California Institute of Technology (2011)
35. Wang, J., Xu, Z., Yang, X., Guo, D., Liu, L.: Self-supervised image clustering from multiple incomplete views via constrastive complementary generation. IET Computer Vision **17**(2), 189–202 (2023)
36. Wang, J., Sun, Z., Tan, Z., Chen, X., Chen, W., Li, H., Zhang, C., Song, Y.: Towards effective usage of human-centric priors in diffusion models for text-based human image generation. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 8446–8455 (2024)
37. Wang, Q., Ding, Z., Tao, Z., Gao, Q., Fu, Y.: Generative partial multi-view clustering. arXiv preprint arXiv:2003.13088 (2020)
38. Wang, Y., Chang, D., Fu, Z., Wen, J., Zhao, Y.: Incomplete multiview clustering via cross-view relation transfer. IEEE Trans. Circuit Syst. Video Technol. **33**(1), 367–378 (2022)
39. Wei, S., Wang, J., Yu, G., Domeniconi, C., Zhang, X.: Deep incomplete multi-view multiple clusterings. In: 2020 IEEE International Conference on Data Mining (ICDM). pp. 651–660. IEEE (2020)

40. Wen, J., Deng, S., Wong, W., Chao, G., Huang, C., Fei, L., Xu, Y.: Diffusion-based missing-view generation with the application on incomplete multi-view clustering. In: Int. Conf. Mach. Learn. (2024)
41. Wen, J., Wu, Z., Zhang, Z., Fei, L., Zhang, B., Xu, Y.: Structural deep incomplete multi-view clustering network. In: Proceedings of the 30th ACM international conference on information & knowledge management. pp. 3538–3542 (2021)
42. Wen, J., Xu, G., Tang, Z., Wang, W., Fei, L., Xu, Y.: Graph regularized and feature aware matrix factorization for robust incomplete multi-view clustering. IEEE Trans. Circuit Syst. Video Technol. **34**(5), 3728–3741 (2023)
43. Wen, J., Xu, Y., Liu, H.: Incomplete multiview spectral clustering with adaptive graph learning. IEEE transactions on cybernetics **50**(4), 1418–1429 (2018)
44. Xiao, H., Rasul, K., Vollgraf, R.: Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv:1708.07747 (2017)
45. Xu, G., Wen, J., Liu, C., Hu, B., Liu, Y., Fei, L., Wang, W.: Deep variational incomplete multi-view clustering: Exploring shared clustering structures. In: AAAI. vol. 38, pp. 16147–16155 (2024)
46. Xu, J., Li, C., Peng, L., Ren, Y., Shi, X., Shen, H.T., Zhu, X.: Adaptive feature projection with distribution alignment for deep incomplete multi-view clustering. IEEE Trans. Image Process. **32**, 1354–1366 (2023)
47. Xu, J., Li, C., Ren, Y., Peng, L., Mo, Y., Shi, X., Zhu, X.: Deep incomplete multi-view clustering via mining cluster complementarity. In: AAAI. vol. 36, pp. 8761–8769 (2022)
48. Yan, W., Zhang, Y., Lv, C., Tang, C., Yue, G., Liao, L., Lin, W.: Gcfagg: Global and cross-view feature aggregation for multi-view clustering. In: IEEE Conf. Comput. Vis. Pattern Recog. pp. 19863–19872 (2023)
49. Yang, Y., Newsam, S.: Bag-of-visual-words and spatial extensions for land-use classification. In: Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems. pp. 270–279 (2010)
50. Yin, J., Sun, S.: Incomplete multi-view clustering with reconstructed views. IEEE Transactions on Knowledge and Data Engineering **35**(3), 2671–2682 (2021)
51. Zhang, C., Han, Z., Fu, H., Zhou, J.T., Hu, Q., et al.: Cpm-nets: Cross partial multi-view networks. Adv. Neural Inform. Process. Syst. **32** (2019)
52. Zhang, C., Wei, J., Wang, B., Li, Z., Chen, C., Li, H.: Robust spectral embedding completion based incomplete multi-view clustering. In: ACM Int. Conf. Multimedia. pp. 300–308 (2023)
53. Zhang, Y., Lin, Y., Yan, W., Yao, L., Wan, X., Li, G., Zhang, C., Ke, G., Xu, J.: Incomplete multi-view clustering via diffusion contrastive generation. In: AAAI. vol. 39, pp. 22650–22658 (2025)
54. Zhao, H., Ding, Z., Fu, Y.: Multi-view clustering via deep matrix factorization. In: AAAI. vol. 31 (2017)