SEMANTIC-GUIDED CROSS-VIEW LATENT TRANSPORT VIA FLOW MATCHING FOR INCOMPLETE MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper presents a seman**T**ic-guided c**R**oss-view l**A**tent transport framework via **C**onditional flow matching for incompl**E**te multi-view clustering (**TRACE**). Specifically, TRACE first learns a dual-level representation that separates compact, cross-view semantic codes from view-specific latent representations. The semantic codes provide clustering-oriented conditions, while the complete latent representations preserve view-specific information and serve as transport states and endpoints. Building on these semantic conditions, we introduce a target-excluded conditional flow matching strategy that temporarily treats an observed view as a pseudo-missing target and learns source-to-target transport routes in the frozen latent space without accessing target-view information during conditioning. Furthermore, TRACE employs an observed-priority multi-source recovery mechanism, in which each missing view is independently recovered from all authentic observed views and the resulting target candidates are aggregated, while generated views are prevented from participating in subsequent recovery. This design avoids recursive error propagation and preserves the information of genuinely observed views. Extensive experiments on five datasets demonstrate that TRACE achieves competitive or superior performance against recent state-of-the-art methods.

1 INTRODUCTION

Multi-view clustering (MVC) aims to discover meaningful groups by integrating complementary information from multiple heterogeneous views without class labels (Chao et al., 2021; Chowdhury et al., 2025; Liu et al., 2026). Most MVC methods, however, assume complete multi-view data. This assumption is frequently violated because of acquisition costs, sensor failures, communication constraints, and access restrictions. Incomplete multi-view clustering (IMVC) addresses clustering when some views are missing and seeks a coherent cluster structure by exploiting complementary information across views (Wen et al., 2019; Liu et al., 2019a; Lin et al., 2021).

Existing IMVC methods can be broadly categorized into two paradigms based on how they handle missing information: consensus-based methods and completion-based methods. Consensus-based methods learn a shared clustering structure directly from incomplete multi-view data (Liu et al., 2019a; Xu et al., 2024; Dai et al., 2025). Although these methods avoid errors caused by explicit imputation, they do not directly recover the complementary information associated with missing views. In contrast, completion-based methods first reconstruct missing features or latent representations through mechanisms such as cross-view prediction and prototype-based modeling, and then perform joint clustering using the completed representations (Wen et al., 2019; Lin et al., 2021; 2023; Li et al., 2023; Wang et al., 2021; Cai et al., 2024). More recently, generative models have been introduced for missing-view completion. For example, diffusion-based methods recover missing representations through iterative conditional denoising (Wen et al., 2024; Zhang et al., 2025), while OPTION employs conditional Flow Matching (FM) to construct continuous transport paths for missing-view completion while preserving cross-view structural information (Lipman et al., 2023; Zhou & Gu, 2026).

Despite these developments, existing generative approaches still face two challenges in recovering missing views. First, existing methods often use cross-view representations to guide recovery without explicitly disentangling shared semantic information from view-specific information. Reliable

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.

recovery, however, requires both: shared semantics should provide consistent guidance across views, while view-specific information should be preserved in the latent state being transported. Without distinguishing these roles, it is difficult to simultaneously maintain cross-view semantic consistency and preserve target-specific characteristics. Second, generative methods can exploit an observed target view as endpoint supervision during training, but the target is inherently unavailable when it is missing at inference. Although the target latent provides valid endpoint supervision during training, its semantic information cannot be accessed at inference. Conditioning on target-dependent semantics therefore introduces a train-test mismatch. Moreover, multiple observed views can provide complementary evidence for the same missing target, motivating direct multi-source recovery rather than recursively reusing generated views.

To address these challenges, we propose seman**T**ic-guided c**R**oss-view l**A**tent transport framework via **C**onditional flow matching for incompl**E**te multi-view clustering (**TRACE**). TRACE follows a two-stage design that first establishes a semantic-guided latent space and then learns cross-view transport within this space. In Stage I, we learn two complementary representations for each view: a view-specific latent representation that retains view-dependent information and a compact semantic code that captures shared clustering structure across views. We jointly optimize cross-view semantic consistency, prototype-based semantic organization, semantic anchoring, and variance regularization to obtain compact and stable semantic codes. In Stage II, we learn conditional cross-view latent transport based on the learned semantic codes. During training, an observed view is treated as a pseudo-missing target. Its latent representation provides the target endpoint for Flow Matching, whereas its semantic code is excluded from the transport condition, which is constructed solely from the other observed views. This separation allows target supervision to be retained without introducing target-dependent semantics into the conditioning process. At inference, each missing view is independently recovered from all authentic observed views. Each source view is transported to the same missing target to produce a candidate latent representation, and the candidates are subsequently aggregated. Observed representations are preserved, while generated representations are not reused as sources, yielding a non-recursive recovery process. The main contributions of this work are summarized as follows:

- We propose TRACE, a semantic-guided cross-view latent transport framework for incomplete multi-view clustering. TRACE learns a dual-level representation that separates compact semantic codes for cross-view semantic guidance from view-specific latent representations for cross-view transport.
- We develop a target-excluded conditional Flow Matching strategy for learning cross-view latent transport. By freezing the learned representation modules, TRACE trains a shared velocity field in the established latent space, using the latent representation of an observed target as endpoint supervision while excluding its semantic code from the transport condition.
- We introduce an observed-priority, non-recursive multi-source recovery mechanism. Each missing view is independently recovered from all authentic observed views, and the resulting candidate representations are aggregated while preserving observed representations and preventing generated views from being reused as sources.

2 Related Work

Incomplete multi-view clustering (IMVC) handles missing views mainly through consensus-based or completion-based methods (Liu et al., 2017; 2019b; 2024; Yu et al., 2024; Liu et al., 2025).

Consensus-based methods directly learn a common clustering structure from the available views without explicitly reconstructing the missing ones. Early approaches either align view-specific embeddings in a common space or directly optimize a consensus clustering matrix from incomplete observations (Wen et al., 2019; Liu et al., 2019a). With the development of deep learning, subsequent methods exploit complementary information across views through different forms of semantic and structural regularization. DIMVC mines cluster complementarity to learn discriminative representations from incomplete multi-view data (Xu et al., 2022). ICMVC further integrates view-specific information through instance-level attention fusion and high-confidence guidance (Chao et al., 2024), while variational approaches aggregate view-specific information to infer shared latent structures within a unified probabilistic framework (Xu et al., 2024). FreeCSL introduces consensus semantic learning to obtain more reliable cross-view clustering information (Dai et al., 2025). More recently,

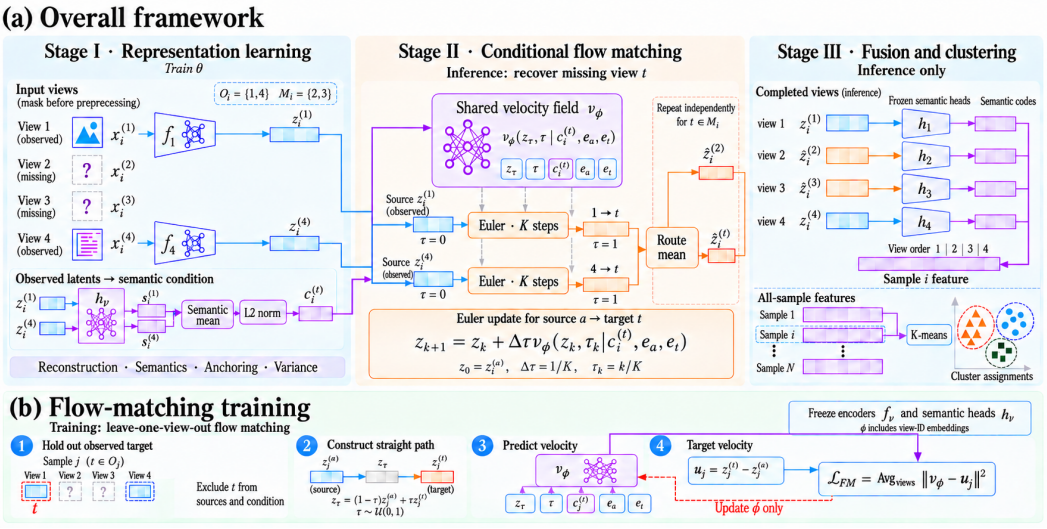


Figure 1: Overview of TRACE. Stage I learns compact semantic codes and view-specific latent representations. Stage II learns target-excluded conditional Flow Matching in the frozen latent space. At inference, multi-source recovery aggregates candidates from authentic observed views, preserves observed representations, and excludes generated views from subsequent recovery.

ACOVA explicitly models cross-view correlations to further exploit dependencies among incomplete views (Xu et al., 2026).

Completion-based methods reconstruct missing features or latent representations to recover complementary information unavailable from observed views. Early approaches primarily rely on cross-view prediction or prototype-based modeling. For example, COMPLETER (Lin et al., 2021) and Dual-Completer (Lin et al., 2023) perform cross-view prediction in a shared latent space and couple completion with contrastive learning to preserve cross-view consistency. CPSPAN (Jin et al., 2023) and ProImp (Li et al., 2023) leverage prototype information for incomplete multi-view representation learning and prototype-based imputation, respectively. DSIMVC dynamically imputes missing views from semantic neighbors and selectively uses imputed samples to mitigate the degradation caused by semantically inconsistent completion (Tang & Liu, 2022). HSACC (Ding et al., 2025) further combines hierarchical semantic alignment with cooperative completion. Other methods impose distribution-level constraints, coupling feature projection with distribution alignment to reduce the discrepancy between observed and generated views (Xu et al., 2023). Beyond deterministic prediction, generative approaches model the distribution of missing views more explicitly. For example, generative MVC (Wang et al., 2021) employs adversarial generation and cycle consistency to synthesize missing views, while CMVAE (Cai et al., 2024) adopts probabilistic latent modeling to learn complete generative representations and infer missing views. More recently, diffusion-based methods (Wen et al., 2024; Zhang et al., 2025) formulate missing-view completion as an iterative conditional denoising process, progressively transforming noisy representations toward the missing-view distribution. OPTION (Zhou & Gu, 2026) further explores conditional Flow Matching for missing-view imputation by learning continuous transport paths to recover missing latent representations.

3 Method

3.1 Problem Formulation and Overview

Let $\mathcal{X} = \{X^{(1)}, \ldots, X^{(V)}\}$ be an aligned multi-view dataset with N samples and V views, indexed by $\mathcal{V} = \{1, \ldots, V\}$. Each view $X^{(v)} \in \mathbb{R}^{N \times D_v}$ contains D_v-dimensional features, with row $x_i^{(v)}$ representing sample i. Incomplete multi-view clustering aims to partition these samples into C clusters using available views without ground-truth labels. A binary mask $M = [m_i^{(v)}] \in \{0,1\}^{N \times V}$ indicates view availability. The observed and missing view sets are $\mathcal{O}_i = \{v \in \mathcal{V} : m_i^{(v)} = 1\}$

and $\mathcal{M}_i = \mathcal{V} \setminus \mathcal{O}_i$, respectively, with $|\mathcal{O}_i| \geq 1$ for every sample. The mask is constructed before preprocessing, and each view-specific preprocessor is fit only to that view's observed rows.

Figure 1 summarizes TRACE. Stage I learns dual-level representations from observed views. Stage II freezes these representations and trains a shared velocity field using target-excluded pseudo-missing supervision. At inference, each missing view is recovered from all authentic observed views by aggregating source-to-target candidates. Observed representations are preserved, and generated views are never reused as sources.

3.2 Dual-Level Representation Learning

TRACE separates compact semantic codes for clustering-oriented conditioning from full latent representations that retain view-specific information for transport. Stage I learns both using only observed views.

Dual-level representations. For every observed cell, TRACE computes

$$z_i^{(v)} = E_v\left(x_i^{(v)}\right), \qquad s_i^{(v)} = H_v\left(z_i^{(v)}\right), \qquad \widehat{x}_i^{(v)} = G_v\left(z_i^{(v)}\right), \tag{1}$$

where E_v and H_v are the view-specific encoder and semantic head, $z_i^{(v)} \in \mathbb{R}^d$ is the full latent representation, $s_i^{(v)} \in \mathbb{R}^p$ is its compact semantic code, and G_v is a view-specific decoder. Here, "complete" denotes the full encoder output, not exact recovery of unobserved information. Let $\mathcal{B} \subseteq \{1, \ldots, N\}$ denote the current minibatch, $\mathcal{I}_v = \{i \in \mathcal{B} : m_i^{(v)} = 1\}$ its observed rows for view v, and $\mathcal{V}_B = \{v \in \mathcal{V} : |\mathcal{I}_v| > 0\}$ the views represented in the minibatch. To retain view-specific information in the complete latent representations, the reconstruction objective is evaluated only on authentic observations:

$$\mathcal{L}_{\text{rec}} = \frac{1}{|\mathcal{V}_B|} \sum_{v \in \mathcal{V}_B} \frac{1}{|\mathcal{I}_v| D_v} \sum_{i \in \mathcal{I}_v} \left\| G_v(z_i^{(v)}) - x_i^{(v)} \right\|_2^2. \tag{2}$$

Cross-view semantic consistency. While reconstruction preserves view-specific information in $z_i^{(v)}$, the semantic codes $s_i^{(v)}$ should capture information that is shared across views. For each sample i, we therefore construct an observed-set semantic reference by averaging its available semantic codes: $\bar{s}_i = |\mathcal{O}_i|^{-1} \sum_{v \in \mathcal{O}_i} s_i^{(v)}$. We first align each observed code with a stop-gradient copy of this mean:

$$\mathcal{L}_{\text{agr}} = \frac{1}{|\mathcal{V}_B|} \sum_{v \in \mathcal{V}_B} \frac{1}{|\mathcal{I}_v| p} \sum_{i \in \mathcal{I}_v} \left\| s_i^{(v)} - \text{sg}(\bar{s}_i) \right\|_2^2, \tag{3}$$

where $\text{sg}(\cdot)$ stops gradients. To preserve sample-level correspondence across views, we also use symmetric cross-view contrastive learning. For a view pair (a, b), let $\mathcal{I}_{ab} = \mathcal{I}_a \cap \mathcal{I}_b$ and $\ell_{ij}^{(a,b)} = \langle \widetilde{s}_i^{(a)}, \widetilde{s}_j^{(b)} \rangle / T_c$, with $\widetilde{s} = s / \max(\|s\|_2, \epsilon)$, where $T_c > 0$ is the contrastive temperature and $\epsilon > 0$ is a numerical stability constant. The pairwise loss is

$$\mathcal{L}_{\text{inst}}^{(a,b)} = -\frac{1}{2|\mathcal{I}_{ab}|} \sum_{i \in \mathcal{I}_{ab}} \left[\log \frac{\exp(\ell_{ii}^{(a,b)})}{\sum_{j \in \mathcal{I}_{ab}} \exp(\ell_{ij}^{(a,b)})} + \log \frac{\exp(\ell_{ii}^{(a,b)})}{\sum_{j \in \mathcal{I}_{ab}} \exp(\ell_{ji}^{(a,b)})} \right]. \tag{4}$$

Let $\mathcal{P}_B = \{(a, b) \in \mathcal{V}^2 : a < b, |\mathcal{I}_{ab}| \geq 2\}$ denote the valid view pairs. We define $\mathcal{L}_{\text{inst}} = |\mathcal{P}_B|^{-1} \sum_{(a,b) \in \mathcal{P}_B} \mathcal{L}_{\text{inst}}^{(a,b)}$ when $\mathcal{P}_B$ is nonempty, and set it to zero otherwise.

Balanced prototype organization. Agreement alone does not prevent semantically collapsed or highly imbalanced groups. TRACE therefore introduces C shared learnable prototypes $\{\omega_k \in \mathbb{R}^p\}_{k=1}^C$. Their prediction distribution for an observed code is $p_{ik}^{(v)} = \text{softmax}_k(\langle \widetilde{s}_i^{(v)}, \widetilde{\omega}_k \rangle / T_p)$, where $\widetilde{\omega}_k = \omega_k / \max(\|\omega_k\|_2, \epsilon)$ and $T_p > 0$ is the prototype temperature. For each valid pair, Sinkhorn normalization over its co-observed samples produces the balanced stop-gradient assignments $Q^{(a|ab)} = [q_{ik}^{(a|ab)}]$ and $Q^{(b|ab)} = [q_{ik}^{(b|ab)}]$ in $\mathbb{R}^{|\mathcal{I}_{ab}| \times C}$. TRACE swaps these assignments across views:

$$\mathcal{L}_{\text{proto}}^{(a,b)} = -\frac{1}{2|\mathcal{I}_{ab}|} \sum_{i \in \mathcal{I}_{ab}} \sum_{k=1}^{C} \left(q_{ik}^{(a|ab)} \log p_{ik}^{(b)} + q_{ik}^{(b|ab)} \log p_{ik}^{(a)} \right). \tag{5}$$

We define $\mathcal{L}_{\text{proto}} = |\mathcal{P}_B|^{-1} \sum_{(a,b)\in\mathcal{P}_B} \mathcal{L}_{\text{proto}}^{(a,b)}$ when $\mathcal{P}_B$ is nonempty, and set it to zero otherwise. The unified semantic objective is $\mathcal{L}_{\text{sem}} = (\mathcal{L}_{\text{agr}} + \mathcal{L}_{\text{inst}} + \mathcal{L}_{\text{proto}})/3$. The agreement and instance terms promote view-compatible sample semantics, whereas the prototype term organizes non-degenerate group structure.

Semantic anchoring and variance regularization. Cross-view consistency alone does not ensure that compact semantic codes remain compatible with the view-specific complete latent representations used for transport. TRACE therefore introduces a view-specific anchor $A_v : \mathbb{R}^p \to \mathbb{R}^d$ that maps the observed-set semantic mean back to each view-specific complete latent representation:

$$\mathcal{L}_{\text{anchor}} = \frac{1}{|\mathcal{V}_B|} \sum_{v\in\mathcal{V}_B} \frac{1}{|\mathcal{I}_v| d} \sum_{i\in\mathcal{I}_v} \left\| A_v(\bar{s}_i) - z_i^{(v)} \right\|_2^2. \tag{6}$$

A variance penalty $\mathcal{L}_{\text{std}}$ discourages constant semantic dimensions. For each view in $\mathcal{V}_B' = \{v \in \mathcal{V}_B : |\mathcal{I}_v| > 1\}$, we compute the coordinate-wise population standard deviation of its observed semantic codes. The penalty is the mean squared deviation of these standard deviations from one, averaged over all p coordinates and all views in $\mathcal{V}_B'$. It is zero when $\mathcal{V}_B'$ is empty.

Representation-learning objective. The representation-learning objective is

$$\mathcal{L}_{\text{rep}} = \mathcal{L}_{\text{rec}} + \lambda_{\text{sem}}\mathcal{L}_{\text{sem}} + \lambda_{\text{anchor}}\mathcal{L}_{\text{anchor}} + \lambda_{\text{std}}\mathcal{L}_{\text{std}}. \tag{7}$$

Here, $\lambda_{\text{sem}}, \lambda_{\text{anchor}}, \lambda_{\text{std}} \geq 0$. Their values are reported in Section 4.1. After optimizing Equation 7, TRACE freezes the encoders, decoders, semantic heads, anchors, and prototypes. The semantic codes then condition transport between full view-specific latent states.

3.3 Target-Excluded Conditional Latent Transport

Stage II trains only a shared velocity field and view-identity embeddings in the frozen latent space. Observed targets provide endpoint supervision but are excluded from the semantic condition to match missing-view inference.

Target-excluded semantic condition. For a target view t, TRACE summarizes the clustering-oriented semantics of the authentic observed views without using the target itself. With $\text{Norm}_2(q) = q/\max(\|q\|_2, \epsilon)$, the condition is

$$c_i^{(t)} = \text{Norm}_2\left(\frac{1}{|\mathcal{O}_i \setminus \{t\}|} \sum_{v\in\mathcal{O}_i\setminus\{t\}} s_i^{(v)} \right). \tag{8}$$

At inference, $t \in \mathcal{M}_i$ is already absent; during training, its semantic code is explicitly excluded. The normalized set mean in Equation 8 is invariant to observed-view order and shared across all routes for the same sample and target.

Pseudo-missing target construction. To obtain endpoint supervision without exposing target-view information to the condition, TRACE temporarily treats an observed view as a pseudo-missing target. For target $t \in \mathcal{V}$, the eligible samples are $\mathcal{E}_t = \{i \in \mathcal{B} : m_i^{(t)} = 1,\ |\mathcal{O}_i| \geq 2\}$. For each $i \in \mathcal{E}_t$, view t is used only as the supervised endpoint, while every $a \in \mathcal{O}_i \setminus \{t\}$ supplies an authentic observed source. Thus, neither the condition nor the source set accesses the pseudo-missing target, and genuinely missing views never enter the Flow Matching objective.

Conditional Flow Matching. For a source-to-target transport route $a \to t$, the frozen view-specific complete latent representations of the source and target define the straight probability path

$$z_{\tau,i}^{(a\to t)} = (1-\tau) z_i^{(a)} + \tau z_i^{(t)}, \qquad u_i^{(a\to t)} = z_i^{(t)} - z_i^{(a)}, \quad \tau \sim \mathcal{U}(0,1). \tag{9}$$

A shared velocity field v_ϕ receives the interpolated state, the target-excluded condition, a scalar coverage input, continuous time, and learned role-specific view embeddings $e_a^{\text{src}}, e_t^{\text{tgt}} \in \mathbb{R}^h$. The reported protocol uses $\rho_i^{(t)} = 1$ for every route:

$$\widehat{u}_{\phi,i}^{(a\to t)}(\tau) = v_\phi\left(z_{\tau,i}^{(a\to t)}, c_i^{(t)}, \rho_i^{(t)}, \tau, e_a^{\text{src}}, e_t^{\text{tgt}} \right). \tag{10}$$

The field outputs a d-dimensional velocity. Let $\mathcal{E}_{a,t} = \{i \in \mathcal{E}_t : a \in \mathcal{O}_i\}$ and $\mathcal{A}_t = \{a \in \mathcal{V} \setminus \{t\} : |\mathcal{E}_{a,t}| > 0\}$. The target-specific objective is

$$\mathcal{L}_{\text{flow}}^{(t)} = \frac{1}{|\mathcal{A}_t|} \sum_{a \in \mathcal{A}_t} \frac{1}{|\mathcal{E}_{a,t}| d} \sum_{i \in \mathcal{E}_{a,t}} \mathbb{E}_{\tau \sim \mathcal{U}(0,1)} \left\| \widehat{u}_{\phi,i}^{(a \to t)}(\tau) - u_i^{(a \to t)} \right\|_2^2. \tag{11}$$

Let $\mathcal{T}_B = \{t \in \mathcal{V} : |\mathcal{A}_t| > 0\}$ denote the active targets in the current minibatch. For a minibatch with $\mathcal{T}_B \neq \varnothing$, we summarize the target losses as $\mathcal{L}_{\text{flow}} = |\mathcal{T}_B|^{-1} \sum_{t \in \mathcal{T}_B} \mathcal{L}_{\text{flow}}^{(t)}$. Only minibatches containing at least one active target contribute a Stage II update. View-identity embeddings allow the shared velocity field to represent all source-to-target transport routes without allocating a separate network to each view pair.

Two-stage optimization and frozen latent space. Let θ collect the representation parameters (encoders, decoders, semantic heads, anchors, and prototypes), and ϕ the velocity-field and view-identity parameters. Stage I minimizes $\mathcal{L}_{\text{rep}}$ over θ. Stage II fixes the learned θ^* and minimizes $\mathcal{L}_{\text{flow}}$ over ϕ, with all representation modules frozen and in evaluation mode. Thus, $\partial \mathcal{L}_{\text{flow}} / \partial \theta^* = 0$: transport is learned without changing the representation geometry to reduce endpoint prediction error.

3.4 Multi-Source Nonrecursive Recovery

Target-excluded transport produces one target candidate from each authentic observed view.TRACE recovers each missing view independently by transporting and aggregating representations from authentic observed sources.

Source-to-target candidate generation. At inference, consider an authentic observed source $a \in \mathcal{O}_i$ and a missing target $t \in \mathcal{M}_i$. Starting from the source representation $\widehat{z}_{i,0}^{(a \to t)} = z_i^{(a)}$, TRACE solves the learned ODE with $K \in \mathbb{N}_+$ Euler steps:

$$\widehat{z}_{i,k+1}^{(a \to t)} = \widehat{z}_{i,k}^{(a \to t)} + \frac{1}{K} v_\phi \left(\widehat{z}_{i,k}^{(a \to t)}, c_i^{(t)}, \rho_i^{(t)}, \frac{k}{K}, e_a^{\text{src}}, e_t^{\text{tgt}} \right), \quad k = 0, \ldots, K-1. \tag{12}$$

This process produces one target candidate $\widehat{z}_{i,K}^{(a \to t)}$ from each authentic observed view.

Multi-source candidate aggregation. TRACE aggregates all target candidates for missing view t by an equal mean:

$$\widehat{z}_i^{(t)} = \frac{1}{|\mathcal{O}_i|} \sum_{a \in \mathcal{O}_i} \widehat{z}_{i,K}^{(a \to t)}. \tag{13}$$

This multi-source aggregation differs from the semantic-condition mean in Equation 8. The former combines target candidates in the view-specific complete latent space, whereas the latter summarizes compact semantic codes to condition transport.

Observed-priority nonrecursive recovery and clustering. Observed representations are preserved, and only missing positions are filled:

$$z_{i,\text{final}}^{(v)} = \begin{cases} z_i^{(v)}, & v \in \mathcal{O}_i, \\ \widehat{z}_i^{(v)}, & v \in \mathcal{M}_i. \end{cases} \tag{14}$$

All missing targets use the same authentic observed sources; generated views are never added to this set. Recovery is therefore independent of target-processing order and avoids recursive error propagation. The frozen semantic heads produce $H_v(z_{i,\text{final}}^{(v)})$ for each completed latent. A fixed fusion operator $\Phi : (\mathbb{R}^d \times \mathbb{R}^p)^V \to \mathbb{R}^q$ combines the latent–semantic pairs across all views into $r_i \in \mathbb{R}^q$. Finally, K-means partitions the fused representations $\{r_i\}_{i=1}^N$ into C clusters, where C is the known number of classes. This step uses the completed latent–semantic representations without accessing the raw features of missing views or updating the frozen representation modules. The same fusion rule is applied to complete and incomplete samples.Each sample is assigned to the cluster whose centroid is nearest to its fused representation.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETTINGS

Datasets. Experiments are conducted on five widely used multi-view clustering benchmarks: BDGP, BBC, OutdoorScene, HandWritten, and ALOI-100.Detailed dataset statistics are provided in Appendix A,Table 3.

Baselines and Evaluation. TRACE is compared with six representative incomplete multi-view clustering methods, including DIMVC (Xu et al., 2022), DSIMVC (Tang & Liu, 2022), DVIMVC (Xu et al., 2024), ICMVC (Chao et al., 2024), DCG (Zhang et al., 2025), and OPTION (Zhou & Gu, 2026). Clustering performance is evaluated using ACC, NMI, and ARI, where higher values indicate better performance.

Implementation and training. The framework is implemented in PyTorch with view-specific MLP encoders and decoders and a shared conditional Flow Matching velocity field. In Stage I, the representation modules are optimized with Adam using a learning rate of 1×10^{-3} and a batch size of 256. In Stage II, the representation modules are frozen, and the velocity field is optimized with AdamW using a learning rate of 1×10^{-3}, a weight decay of 1×10^{-4}, and a batch size of 512. Missing representations are recovered with a fixed-step Euler solver using target-excluded semantic conditioning and observed-priority multi-source aggregation. Dataset-specific model dimensions and training budgets are selected from predefined bounded candidate sets.

4.2 COMPARISON RESULTS

Table 1 compares TRACE with six baselines across five datasets under different missing rates. The best and second-best results are highlighted in bold and underlined, respectively. On ALOI-100 with a 70% missing rate,TRACE surpasses the strongest competing results by 5.99%, 3.06%, and 20.36% in ACC, NMI, and ARI, respectively. TRACE also outperforms the generative baselines DCG and OPTION in a range of settings. For example, on BDGP with a 10% missing rate, TRACE improves ACC by 2.16% and 6.54% over DCG and OPTION, respectively. These results are consistent with the design of TRACE: dual-level representations separate shared semantics from view-specific information, target-excluded conditioning avoids dependence on unavailable target semantics, and multi-source transport exploits complementary information from authentic observed views. Together, these designs support robust missing-view recovery under varying degrees of view incompleteness.

4.3 ABLATION STUDY

We conducted four component ablations on HandWritten at a missing rate of 70%. Stage-I variants retrained the representations, whereas Stage-II variants retained the full Stage-I representations. The *w/o Flow Matching* variant used the frozen target-view anchor for recovery, while *w/o Observed Priority* changed only the completion rule using the same Flow checkpoint. Table 2 reports means and sample standard deviations; all four ablations reduced ACC, NMI, and ARI. Removing the unified semantic objective or semantic guidance reduced ACC by 1.50% and 0.62% , respectively, supporting semantic representation learning and conditioning. Removing Flow Matching reduced ACC by 8.81%, supporting cross-view latent transport. Removing observed priority caused the largest ACC decrease of 17.15%, highlighting the importance of preserving genuinely observed representations and recovering only missing views rather than regenerating observed representations.

Table 2: Ablation study on HandWritten. Best results are in bold.

Variant	ACC	NMI	ARI
Full TRACE	$\mathbf{90.16 \pm 0.92}$	$\mathbf{82.53 \pm 1.26}$	$\mathbf{79.95 \pm 1.66}$
w/o Unified Semantic Objective	88.66 ± 1.41	80.97 ± 1.70	77.49 ± 2.36
w/o Semantic Guidance	89.54 ± 1.20	81.55 ± 1.40	78.81 ± 1.96
w/o Flow Matching	81.35 ± 5.93	74.37 ± 2.13	68.10 ± 4.48
w/o Observed Priority	73.01 ± 4.22	65.27 ± 2.25	56.17 ± 3.50

4.4 SENSITIVITY ANALYSIS

We investigate the sensitivity of TRACE to three Stage-I loss weights and the inference-time Euler integration budget on HandWritten with a missing rate of 0.5. Specifically, we vary one factor at a time while fixing the others: $\lambda_{\text{sem}} \in \{0.025, 0.05, 0.1, 0.2, 0.4\}$, $\lambda_{\text{anchor}} \in$

Table 1: Clustering performance (ACC, NMI, ARI) under varying missing rates. Best results are bolded and second best are underlined.

Dataset	Method	10%			30%			50%			70%		
		ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
BDGP	DIMVC (AAAI'22)	32.16	7.00	5.65	30.62	6.01	5.21	31.54	6.76	5.18	32.17	7.94	5.78
	DSIMVC (ICML'22)	37.18	13.60	9.30	36.82	14.69	9.70	38.69	**14.30**	10.01	37.77	**13.99**	9.91
	DVIMC (AAAI'24)	20.26	0.06	0.01	20.04	0.01	0.00	20.00	0.00	0.00	20.01	0.02	0.00
	ICMVC (AAAI'24)	32.58	7.18	4.88	29.28	5.94	3.18	28.18	4.90	2.82	27.82	3.84	2.19
	DCG (AAAI'25)	34.56	9.08	6.40	33.26	7.63	5.62	31.36	6.62	4.78	31.35	6.36	4.79
	OPTION (ICML'26)	38.94	11.85	9.56	35.94	10.52	7.70	33.39	9.55	6.73	34.45	9.22	6.63
	SCOPE (Ours)	**41.10**	**16.40**	**13.30**	**39.15**	**15.29**	**12.08**	**39.02**	13.62	**11.52**	**39.00**	13.04	**11.06**
BBC	DIMVC (AAAI'22)	69.99	55.11	53.89	**72.44**	58.18	**56.49**	53.43	37.57	28.83	50.72	33.96	20.56
	DSIMVC (ICML'22)	57.40	41.83	35.61	56.38	37.89	33.01	53.99	35.89	30.76	43.74	21.45	17.47
	DVIMC (AAAI'24)	50.86	33.16	21.79	55.21	37.95	27.96	53.40	34.61	25.19	51.80	35.33	22.33
	ICMVC (AAAI'24)	33.87	4.79	1.38	32.09	3.06	1.32	34.13	3.40	1.68	31.12	2.95	1.47
	DCG (AAAI'25)	42.72	16.12	13.01	41.37	13.72	11.06	36.29	9.16	7.09	31.80	6.20	3.48
	OPTION (ICML'26)	69.20	42.36	42.48	56.55	32.55	32.69	58.48	31.29	33.25	42.80	16.70	12.83
	SCOPE (Ours)	**71.42**	**63.17**	**55.17**	71.30	**63.24**	56.17	**74.72**	**66.62**	**62.16**	**80.70**	**68.80**	**66.51**
OutdoorScene	DIMVC (AAAI'22)	35.39	21.86	13.40	32.51	19.67	10.57	32.46	17.35	9.43	34.90	21.99	12.03
	DSIMVC (ICML'22)	54.80	50.38	35.11	56.32	50.66	35.27	58.11	50.13	35.48	57.27	49.19	34.74
	DVIMC (AAAI'24)	63.39	57.68	44.97	63.13	57.75	45.12	62.72	57.22	44.30	62.35	56.44	43.82
	ICMVC (AAAI'24)	53.18	49.58	31.40	51.41	46.09	24.30	48.40	41.41	19.85	45.65	40.15	20.05
	DCG (AAAI'25)	66.00	56.11	46.06	61.73	51.83	41.54	63.28	52.18	42.53	59.26	47.56	37.55
	OPTION (ICML'26)	63.47	57.62	43.67	64.55	57.23	44.79	65.27	56.87	44.34	63.92	56.87	43.60
	SCOPE (Ours)	**77.72**	**62.86**	**57.28**	**77.46**	**62.34**	**56.74**	**76.96**	**61.83**	**55.94**	**74.88**	**60.47**	**53.81**
HandWritten	DIMVC (AAAI'22)	41.48	48.97	32.84	39.67	46.26	30.82	45.99	50.66	33.49	44.68	50.39	33.91
	DSIMVC (ICML'22)	74.72	78.54	68.93	77.09	81.67	72.90	73.03	78.37	68.43	69.04	74.57	63.55
	DVIMC (AAAI'24)	83.12	80.44	74.53	**89.86**	84.45	81.43	80.48	79.24	71.80	82.25	81.41	75.40
	ICMVC (AAAI'24)	39.73	39.46	21.10	34.68	34.35	19.14	28.76	28.56	11.11	29.19	27.58	8.46
	DCG (AAAI'25)	88.76	85.46	80.73	88.03	82.69	78.64	85.81	80.04	75.17	68.86	63.31	50.38
	OPTION (ICML'26)	89.45	**87.05**	82.79	85.08	**86.52**	80.08	84.88	84.50	78.05	81.60	82.48	74.47
	SCOPE (Ours)	**91.92**	85.51	**83.24**	89.80	85.13	**82.16**	**93.37**	**87.40**	**86.01**	**91.88**	**85.41**	**83.08**
ALOI-100	DIMVC (AAAI'22)	75.97	88.51	67.05	75.61	87.96	65.87	72.66	85.67	53.77	74.63	84.91	45.07
	DSIMVC (ICML'22)	49.28	70.54	36.35	50.61	71.08	37.86	45.91	67.42	33.14	42.44	65.82	30.37
	DVIMC (AAAI'24)	3.53	9.33	0.00	3.74	10.55	0.00	3.82	10.73	0.01	3.82	10.95	0.01
	ICMVC (AAAI'24)	54.81	74.05	42.27	54.45	73.56	41.50	50.92	70.63	37.49	33.82	60.42	22.72
	DCG (AAAI'25)	66.23	81.45	49.16	56.06	75.31	39.84	43.24	66.40	30.13	29.50	56.85	19.71
	OPTION (ICML'26)	66.34	83.14	56.04	66.77	81.85	55.69	68.01	81.36	56.24	63.85	79.10	52.24
	SCOPE (Ours)	**86.77**	**90.85**	**79.30**	**86.00**	**90.30**	**78.42**	**84.72**	**89.65**	**77.02**	**80.62**	**87.97**	**72.60**

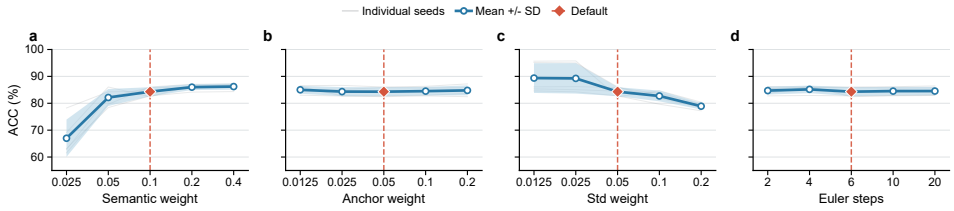


Figure 2: Sensitivity analysis of four key hyperparameters on HandWritten (mean ± std, 5 runs).

$\{0.0125, 0.025, 0.05, 0.1, 0.2\}$, $\lambda_{\text{std}} \in \{0.0125, 0.025, 0.05, 0.1, 0.2\}$, and $K \in \{2, 4, 6, 10, 20\}$. We quantify sensitivity to a parameter p by $S_p = (\text{ACC}_{\max} - \text{ACC}_{\min})/\text{ACC}_{\max}$, where smaller values indicate greater robustness. As shown in Figure 2, λ_{anchor} and the Euler budget exhibit limited sensitivity, with ACC ranges of 0.70% and 0.86% , respectively. In contrast, λ_{sem} and λ_{std} show larger variations of 19.22% and 10.46%, mainly at extreme settings. These results indicate that TRACE is robust to variations in the anchor weight and Euler integration budget, while appropriate semantic and standard-deviation regularization is important for stable clustering performance.

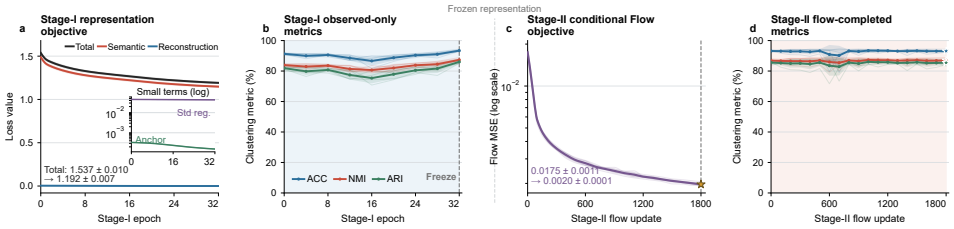


Figure 3: Two-stage training of TRACE on HandWritten (50% missing). Thin lines: individual runs; solid lines: means; shaded bands: standard deviations across five seeds.

4.5 Convergence Analysis

We track TRACE's clustering metrics and stage-specific objectives on HandWritten at a missing rate of 0.5 to assess two-stage optimization stability.Figure 3 shows the mean and standard deviation over five fixed seeds. During Stage I, the clustering metrics remain stable, with ACC, NMI, and ARI improving slightly from 87.11%, 80.42%, and 76.45% to 87.69%, 80.96%, and 77.45%, respectively, while the representation objective decreases from 1.6761 to 1.4082. After freezing the representation networks, Stage II yields a rapid improvement during the first 100 Flow updates, followed by a stable plateau. By iteration 900, ACC, NMI, and ARI reach 89.67%, 81.97%, and 79.34%, while the normalized Flow Matching objective decreases from 0.02983 to 0.00548. Narrow standard-deviation bands, decreasing objectives, and stable clustering metrics indicate reproducible convergence under two-stage training.

4.6 Visualization

Figure 4 compares t-SNE visualizations before and after training on BBC and HandWritten at a missing rate of 0.5. Compared with the raw inputs, TRACE produces more compact clusters and clearer class separation, supporting its clustering effectiveness.

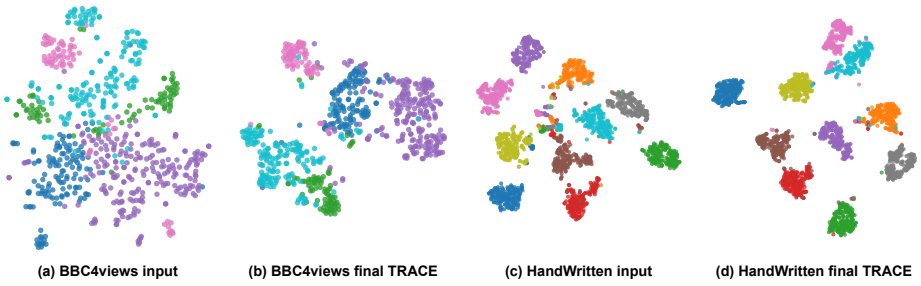


Figure 4: t-SNE visualization of latent representations before and after training.

5 Conclusion

In this work, we present TRACE, a semantic-guided cross-view latent transport framework via conditional Flow Matching for incomplete multi-view clustering. TRACE separates compact, clustering-oriented semantic codes from full view-specific latent representations. With the representation networks frozen, target-excluded conditional Flow Matching learns source-to-target transport through pseudo-missing supervision without using target-view information in the condition. At inference, each missing view is independently recovered from all authentic observed views, and the resulting candidates are aggregated. This observed-priority recovery preserves observed representations and prevents recursive error propagation by excluding generated views from subsequent recovery. Experiments on five benchmark datasets across multiple missing rates demonstrate competitive or superior performance against recent methods.

AI USE STATEMENT

In this work, we used large language model-based generative AI tools, including OpenAI ChatGPT/Codex , to assist with Chinese-to-English translation where needed and with the wording and presentation of interpretations based on author-provided experimental results. We did not use generative AI to generate synthetic datasets, develop theoretical or conceptual models, formulate mathematical claims or proofs, propose research hypotheses, design the research methodology or experiments, implement the core method, clean or reformat datasets, or conduct qualitative or thematic data analysis.

Additionally, generative AI tools were used to draft and revise portions of the manuscript from author-provided technical content, improve readability and conciseness, suggest paragraph and section organization, check terminology consistency, assist with LaTeX and reference formatting, and support code organization, documentation, and software checks. The authors manually reviewed all AI-assisted text against the method definitions, implementation, experimental records, and cited sources. All numerical results and references were independently verified by the authors, and any AI-assisted software changes were inspected and tested before use. The research ideas, methodological decisions, experimental design, reported results, and final scientific conclusions remain those of the authors. We take responsibility for the final content of this work, including all text, claims, code, and other artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work uses public benchmark datasets and does not involve human participant recruitment, intervention, or the collection of new personal data.

REPRODUCIBILITY STATEMENT

The model formulation, training objectives, and inference procedure are described in Section 3, while the datasets and evaluation protocol are specified in Section 4.1. To support reproducibility, we provide an anonymized implementation in the supplementary material, including environment specifications, dataset-specific configurations, and scripts for the main experiments, ablation studies, and sensitivity analyses. The accompanying documentation describes data preparation, preprocessing, and checkpoint selection, together with instructions for running experiments and aggregating results. The code package also includes checks for source and dataset integrity.

REFERENCES

H. Cai, W. Huang, S. Yang, S. Ding, Y. Zhang, B. Hu, F. Zhang, and Y.-M. Cheung. Realize generative yet complete latent representation for incomplete multi-view learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(5):3637–3652, 2024. doi: 10.1109/TPAMI.2023.3346869.

G. Chao, S. Sun, and J. Bi. A survey on multiview clustering. *IEEE Transactions on Artificial Intelligence*, 2(2):146–168, 2021. doi: 10.1109/tai.2021.3065894.

G. Chao, Y. Jiang, and D. Chu. Incomplete contrastive multi-view clustering with high-confidence guiding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 11221–11229, 2024. doi: 10.1609/aaai.v38i10.29000.

A. R. Chowdhury, A. Gupta, and S. Das. Deep multi-view clustering: A comprehensive survey of the contemporary techniques. *Information Fusion*, 119:103012, 2025. doi: 10.1016/j.inffus.2025.103012.

Y. Dai, J. Jin, Z. Dong, S. Wang, X. Liu, E. Zhu, X. Yang, X. Gan, and Y. Feng. Imputation-free and alignment-free: Incomplete multi-view clustering driven by consensus semantic learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5071–5081, 2025. doi: 10.1109/cvpr52734.2025.00478.

X. Ding, L. Zhao, X. Li, and X. Zhu. Incomplete multi-view clustering via hierarchical semantic alignment and cooperative completion. In *Advances in Neural Information Processing Systems*, volume 38, pp. 52175–52195, 2025. doi: 10.52202/085713-1739.

J. Jin, S. Wang, Z. Dong, X. Liu, and E. 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*, pp. 11600–11609, 2023. doi: 10.1109/cvpr52729.2023.01116.

H. Li, Y. Li, M. Yang, P. Hu, D. Peng, and X. Peng. Incomplete multi-view clustering via prototype-based imputation. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pp. 3911–3919, 2023. doi: 10.24963/ijcai.2023/435.

Y. Lin, Y. Gou, Z. Liu, B. Li, J. Lv, and X. Peng. COMPLETER: Incomplete multi-view clustering via contrastive prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11169–11178, 2021. doi: 10.1109/cvpr46437.2021.01102.

Y. Lin, Y. Gou, X. Liu, J. Bai, J. Lv, and X. Peng. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45 (4):4447–4461, 2023. doi: 10.1109/TPAMI.2022.3197238.

Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Le. Flow matching for generative modeling. In *International Conference on Learning Representations*, 2023.

J. Liu, X. Liu, X. Wan, K. Liang, W. Liang, S. Zhou, H. Wu, and K. Guo. Incomplete multi-view deep clustering with data imputation and alignment. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-2981.

S. Liu, J. Zhang, Y. Wen, X. Yang, S. Wang, Y. Zhang, E. Zhu, C. Tang, L. Zhao, and X. Liu. Sample-level cross-view similarity learning for incomplete multi-view clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(12):14017–14025, 2024. doi: 10.1609/aaai.v38i12.29310.

X. Liu, M. Li, L. Wang, Y. Dou, J. Yin, and E. Zhu. Multiple kernel k-means with incomplete kernels. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017. doi: 10.1609/aaai.v31i1.10893.

X. Liu, X. Zhu, M. Li, C. Tang, E. Zhu, J. Yin, and W. Gao. Efficient and effective incomplete multi-view clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1):4392–4399, 2019a. doi: 10.1609/aaai.v33i01.33014392.

X. Liu, X. Zhu, M. Li, L. Wang, C. Tang, J. Yin, D. Shen, H. Wang, and W. Gao. Late fusion incomplete multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(10):2410–2423, 2019b. doi: 10.1109/TPAMI.2018.2879108.

X. Liu, K. Liang, J. Wang, S. Liu, X. Wang, and H. Wang. Two decades of multi-view clustering: Taxonomy, application, and challenge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(3):3744–3764, 2026. doi: 10.1109/TPAMI.2025.3640109.

H. Tang and Y. Liu. Deep safe incomplete multi-view clustering: Theorem and algorithm. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 21090–21110. PMLR, 2022.

Q. Wang, Z. Ding, Z. Tao, Q. Gao, and Y. Fu. Generative partial multi-view clustering with adaptive fusion and cycle consistency. *IEEE Transactions on Image Processing*, 30:1771–1783, 2021. doi: 10.1109/TIP.2020.3048626.

J. Wen, Z. Zhang, Y. Xu, B. Zhang, L. Fei, and H. Liu. Unified embedding alignment with missing views inferring for incomplete multi-view clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1):5393–5400, 2019. doi: 10.1609/aaai.v33i01.33015393.

J. Wen, S. Deng, W. Wong, G. Chao, C. Huang, L. Fei, and Y. Xu. Diffusion-based missing-view generation with the application on incomplete multi-view clustering. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 52762–52778. PMLR, 2024.

G. Xu, J. Wen, C. Liu, B. Hu, Y. Liu, L. Fei, and W. Wang. Deep variational incomplete multi-view clustering: Exploring shared clustering structures. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(14):16147–16155, 2024. doi: 10.1609/aaai.v38i14.29548.

J. Xu, C. Li, Y. Ren, L. Peng, Y. Mo, X. Shi, and X. Zhu. Deep incomplete multi-view clustering via mining cluster complementarity. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 8761–8769, 2022. doi: 10.1609/aaai.v36i8.20856.

J. Xu, C. Li, L. Peng, Y. Ren, X. Shi, H. T. Shen, and X. Zhu. Adaptive feature projection with distribution alignment for deep incomplete multi-view clustering. *IEEE Transactions on Image Processing*, 32:1354–1366, 2023. doi: 10.1109/TIP.2023.3243521.

Z. Xu, A. Tang, S. Chen, X. Zou, C. Lang, R. A. Mancisidor, and M. Kampffmeyer. Beyond independence: Learning correlated views for variational incomplete multi-view clustering. In *Proceedings of the Forty-third International Conference on Machine Learning*, 2026.

S. Yu, S. Wang, P. Zhang, M. Wang, Z. Wang, Z. Liu, L. Fang, E. Zhu, and X. Liu. DVSAI: Diverse view-shared anchors based incomplete multi-view clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(15):16568–16577, 2024. doi: 10.1609/aaai.v38i15.29595.

Y. Zhang, Y. Lin, W. Yan, L. Yao, X. Wan, G. Li, C. Zhang, G. Ke, and J. Xu. Incomplete multi-view clustering via diffusion contrastive generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(21):22650–22658, 2025. doi: 10.1609/aaai.v39i21.34424.

S. Zhou and Z. Gu. OPTION: Optimal transport-guided flow matching for incomplete and unaligned multi-view clustering. In *Forty-third International Conference on Machine Learning*, 2026.

A Dataset Details

We evaluate TRACE on five multi-view clustering benchmarks: BDGP, BBC, OutdoorScene, Hand-Written, and ALOI-100. Table 3 summarizes their sample counts, numbers of views and classes, and feature dimensions.

Table 3: Statistics of the five benchmark datasets.

Dataset	Samples	Views	Feature dimensions	Classes
BDGP	2,500	3	$(1000, 500, 250)$	5
BBC	685	4	$(4659, 4633, 4665, 4684)$	5
OutdoorScene	2,688	4	$(512, 432, 256, 48)$	8
HandWritten	2,000	6	$(216, 76, 64, 6, 240, 47)$	10
ALOI-100	10,800	4	$(77, 13, 64, 125)$	100