RELIABLE CORRESPONDENCE LEARNING FOR INCOMPLETE AND UNALIGNED MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Incomplete and unaligned multi-view clustering (IUMVC) has emerged as a challenging paradigm for discovering shared cluster structures from multiple views, where observations may be missing and cross-view sample correspondence is unknown. However, the unknown correspondence makes reliable cross-view matching difficult, as missing observations can result in multiple plausible matches, while erroneous correspondences may propagate errors to subsequent representation learning and multi-view consensus. To address these challenges, we propose a reliable correspondence learning framework, termed CAMVC, for robust clustering of unaligned incomplete multi-view data. Specifically, we first establish a canonical coordinate system without using paired samples and develop a cluster-routed local sparse transport mechanism to identify plausible correspondences, avoiding both premature hard matching and exhaustive global matching. Furthermore, we exploit correspondence-induced reconstruction consistency to quantify sample-level reliability and feature-dependent residual variation, allowing unreliable correspondences to exert less influence on subsequent representation learning. Finally, we integrate cross-view geometric consistency with reliability-aware consensus and high-order tensor regularization to capture complementary and higher-order dependencies across views, followed by reliability- and coverage-aware fusion for clustering. Extensive experiments on six benchmarks under varying levels of missingness and misalignment demonstrate that CAMVC improves clustering performance and effectively learns cross-view correspondences, even as the common sample-order assumption is progressively relaxed.

1 INTRODUCTION

Multi-view data describe the same objects through different modalities or feature representations, providing both shared semantic information and complementary cues (Zhao et al., 2020; Zhang et al., 2012; Wang et al., 2019). Multi-view clustering (MVC) exploits this consistency and complementarity to discover groups that are difficult to identify from a single view (Yang & Wang, 2018; Zhou et al., 2024; Lu et al., 2016; Kang et al., 2021). Conventional MVC typically assumes that every object is observed in every view and that cross-view sample correspondence is known. In practice, acquisition failures can leave views missing, while independent collection can obscure sample correspondence. Both violations undermine direct cross-view integration, motivating the study of MVC beyond the fully observed and aligned setting.

Existing methods address these challenges under different combinations of observation completeness and cross-view correspondence. For incomplete multi-view data with known correspondence, Hu & Chen (2019) used observation-weighted matrix factorization to learn shared sample representations and align view-specific bases. Wen et al. (2023a) estimated sample-relation confidence from shared nearest neighbors to guide consensus graph learning, while Wen et al. (2021) jointly inferred missing features and similarity graphs with a low-rank tensor constraint. For unaligned multi-view data with unavailable correspondence, Wen et al. (2023b) matched view-specific graphs based on relational structures to guide cross-view alignment before graph fusion. Li et al. (2024) learned view-specific anchors and matched the resulting bipartite graphs to a reference graph for common graph construction. More recent studies jointly consider missing observations and unavailable cor-

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

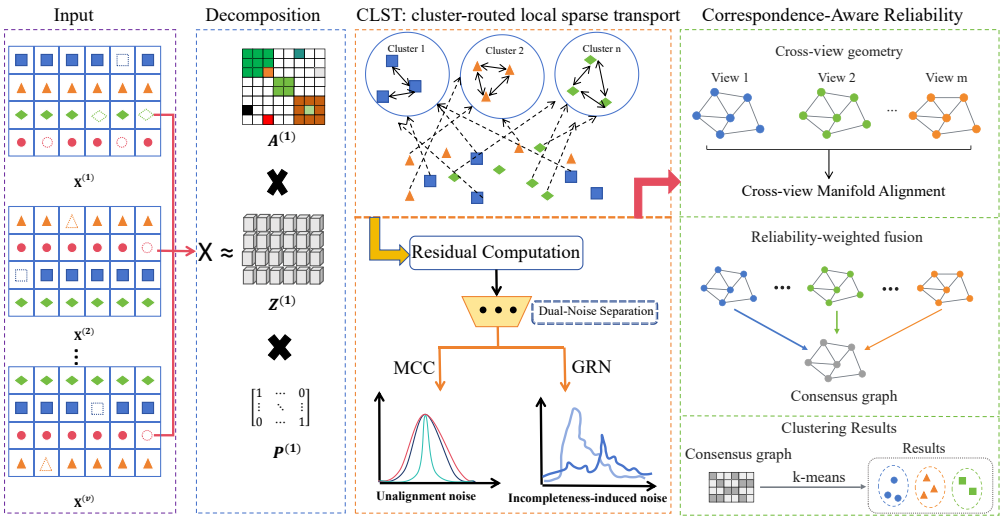


Figure 1: Overview of CAMVC. Cluster-routed local sparse transport retains plausible correspondence candidates, while correspondence-dependent reconstruction residuals provide sample-specific weights and observed-data reconstruction quality guides view-level graph fusion. Cross-view geometric consistency and high-order tensor regularization are integrated with reliability-aware consensus, followed by reliability and coverage-aware fusion for clustering.

respondence. Zeng et al. (2023) learned a shared semantic distribution without paired samples to support clustering. Liu et al. (2025) recovered missing samples in a low-dimensional space and aligned self-representation matrices to a reference view, coupling recovery and cross-view structure learning through tensor regularization and adaptive alignment.

Despite this progress, three difficulties remain when missingness and unalignment coexist. First, existing correspondence recovery methods often make hard assignment decisions over cross-view samples, committing each local sample to a single counterpart. However, under incomplete and unaligned settings, missing observations and view-specific variations may make multiple samples plausible counterparts for the same observation. Such hard assignments discard this correspondence ambiguity and may introduce erroneous matches, which can subsequently propagate errors to representation learning and structural integration. Second, existing methods typically quantify reliability at the view level and assign a uniform weight to all observations within a view. However, correspondence reliability may vary substantially across samples, rendering view-level weighting insufficient to distinguish unreliable matches. Third, correspondence alignment alone does not guarantee equally reliable neighborhood structures across views. Missing observations and correspondence errors can distort the local geometry of different views to different degrees, making direct fusion of the aligned graphs prone to propagating unreliable relations into the consensus structure.

To this end, we propose CAMVC, a correspondence-aware framework for incomplete and unaligned multi-view clustering, as illustrated in Figure 1. The framework progressively addresses correspondence uncertainty from local matching to sample- and view-level reliability. To preserve the ambiguity of cross-view correspondence, cluster-routed local sparse transport (CLST) confines matching to plausible candidate regions and assigns soft weights to multiple candidates within a sparse support, avoiding premature global hard assignments. This localized correspondence structure further enables sample-level reliability estimation: correspondence-dependent reconstruction residuals identify poorly reconstructed observations and reduce their influence during representation learning. At the view level, however, reliable sample correspondences do not necessarily translate into reliable view structures. Observed-data reconstruction errors are therefore used to quantify the reliability of each canonicalized view and regulate graph fusion accordingly. The representations and correspondences are updated alternately, allowing the two to mutually refine each other. Covariance-aware residual modeling and higher-order tensor regularization further enhance robustness (Zhong et al., 2026). The main contributions are summarized as follows:

- We introduce a correspondence-aware framework for incomplete and independently ordered multi-view clustering, incorporating correspondence uncertainty into representation learning without requiring known sample pairs or cross-view permutations.
- We develop a cluster-routed local correspondence mechanism that combines structural routing with sparse soft matching, retaining multiple plausible candidates while avoiding premature hard assignments and dense all-pairs matching.
- We couple correspondence learning with reliability-aware integration: correspondence-dependent reconstruction residuals downweight inconsistent samples, while observed-data reconstruction quality regulates the contribution of each canonicalized view to the consensus structure.

2 Related Work

2.1 Incomplete Multi-View Clustering

Incomplete multi-view clustering addresses missing observations, typically assuming known cross-view sample correspondence. For example, Hu & Chen (2019) proposed weighted semi-nonnegative matrix factorization to learn a shared representation, aligning view-specific bases through regularized regression to mitigate the effect of missing observations. Beyond feature-based factorization, graph-based methods exploit relationships among observed samples. Wen et al. (2023a) used shared-nearest-neighbor confidence to guide consensus graph learning, while Wang et al. (2022) jointly learned anchors and incomplete bipartite graphs to reduce the cost of graph integration. Tensor-based methods capture higher-order dependencies across views through low-rank tensor modeling, spanning representation recovery (Zhang et al., 2023; Liang et al., 2024), graph completion (Zhao et al., 2023), kernel learning (Wu et al., 2024), and missing-data recovery (Wen et al., 2021; Gu et al., 2024; Zhong et al., 2026). In deep IMVC, methods learn representations from observed views and address missingness through cross-view prediction, prototype-based imputation, or sample and prototype alignment (Lin et al., 2021; Li et al., 2023; Jin et al., 2023).

2.2 Unaligned Multi-View Clustering

Unaligned multi-view clustering addresses unavailable cross-view sample correspondence through structural alignment or consistency learning. Wen et al. (2023b) matched view-specific structures to estimate cross-view correspondence before graph fusion, while Li et al. (2024) combined anchor learning with bipartite graph matching for scalable graph integration. Extending structural alignment to multiple scales, Ji et al. (2024) learned and aligned multi-scale representations, using tensor regularization to integrate them and capture higher-order cross-view correlations. Related studies consider an unpaired setting in which each sample is observed in only one view, and establish cross-view connections through reliable cluster structures. Xin et al. (2024) proposed single and multiple reliable view guidance to improve the cluster structures of other views. To improve the reliability of sample relationships, Xin et al. (2025) selected sample pairs with consistent relationships across clustering levels for within-view contrastive learning, mitigating clustering errors associated with ambiguous boundary samples.

2.3 Joint Handling of Missingness and Misalignment

Incomplete and unaligned multi-view clustering considers the joint challenges of missing observations and unknown cross-view correspondences. Existing methods mainly establish cross-view consistency through shared semantics, explicit recovery and alignment, or partially observed correspondences. Zeng et al. (2023) learned invariant semantic distributions across views without paired samples, enabling clustering together with the realignment or imputation of defective instances. Similarly, Dai et al. (2025) learned shared prototypes and employed modularity-based graph clustering to capture within-view semantic relationships without explicit imputation or alignment. In contrast, Liu et al. (2025) explicitly recovered missing samples in a low-dimensional space and aligned view-specific self-representations to a reference view, followed by tensor regularization for integration. When partial cross-view correspondences are available, Yang et al. (2022) exploited them as positive pairs for noise-robust contrastive learning, while Wang et al. (2024) combined dual

prediction for missing-view recovery with instance- and class-level alignment. Zhao et al. (2025) further used aligned complete samples as a bridge to construct cross-view correspondences under partial unalignment and missingness, together with a dual noise-robust contrastive objective.

However, existing methods do not explicitly model the ambiguity of cross-view correspondence while accounting for sample- and view-level reliability in representation learning and structural integration. CAMVC addresses these issues through cluster-routed local sparse transport to retain multiple plausible correspondences, correspondence-aware reconstruction residuals to identify unreliable observations, and reconstruction-quality-driven graph fusion to regulate the integration of view-specific structures.

3 METHOD

3.1 PROBLEM SETUP AND FRAMEWORK

Consider an incomplete and unaligned multi-view dataset consisting of m views and n underlying samples. View v is represented by $\mathbf{X}^{(v)} \in \mathbb{R}^{d_v \times n}$, whose columns follow an unknown local order. The diagonal mask $\mathbf{W}^{(v)} = \operatorname{diag}(w_1^{(v)}, \ldots, w_n^{(v)})$ indicates the availability of local observations, and $\mathcal{O}_v = \{i : w_i^{(v)} = 1\}$ denotes the observed local indices. During data construction, each view and its mask are permuted together, independently of the other views. The model therefore receives neither canonical sample identities nor known cross-view sample pairs. Before permutation, we ensure that every underlying sample is observed in at least one view.

The views differ in both feature space and sample order, preventing direct comparison of their observations. To account for feature heterogeneity, we introduce a view-specific orthogonal basis and latent representations. To accommodate unknown correspondence, these representations are indexed in a canonical sample order and mapped to the local observation order through a correspondence matrix. A residual term accounts for reconstruction errors. With reconstruction restricted to observed samples, the model is formulated as follows:

$$\begin{aligned} \mathbf{X}^{(v)}\mathbf{W}^{(v)} &= \mathbf{A}^{(v)}\mathbf{Z}^{(v)}\big(\mathbf{P}^{(v)}\big)^{\top}\mathbf{W}^{(v)} + \mathbf{E}^{(v)}\mathbf{W}^{(v)}, \\ \text{s.t.} \quad & \big(\mathbf{A}^{(v)}\big)^{\top}\mathbf{A}^{(v)} = \mathbf{I}_t. \end{aligned} \tag{1}$$

Here, $\mathbf{A}^{(v)} \in \mathbb{R}^{d_v \times t}$ is a view-specific orthogonal basis, and $\mathbf{Z}^{(v)} \in \mathbb{R}^{t \times n}$ contains representations in a canonical sample order. Each row of $\mathbf{P}^{(v)} \in \mathbb{R}^{n \times n}$ associates a local sample with canonical samples, allowing $\mathbf{A}^{(v)}\mathbf{Z}^{(v)}\big(\mathbf{P}^{(v)}\big)^{\top}$ to reconstruct the local observations; $\mathbf{E}^{(v)}$ accounts for the remaining residual. To establish the canonical coordinates without known pairs, a reference view is selected by evaluating the explained scatter and cluster balance of its observed SVD embedding. The selected view is reordered as view 1, and its truncated SVD defines a fixed guide $\mathbf{Z}_0 = \mathbf{D}_0\mathbf{V}_0^{\top}$, where $\mathbf{D}_0 \in \mathbb{R}^{t \times t}$ is the diagonal matrix of retained singular values. This guide fixes the coordinate system but contains no cross-view pair.

Concretely, if WSS_v and TSS_v are the within- and total-scatter of the independently clustered embedding and π_{vr} is its empirical cluster proportion, the selection score is

$$\operatorname{score}(v) = \left(1 - \frac{\mathrm{WSS}_v}{\max(\mathrm{TSS}_v, \epsilon)}\right)_+ \sqrt{\frac{\exp(-\sum_r \pi_{vr} \log \pi_{vr})}{c}}. \tag{2}$$

The balance factor penalizes nearly collapsed partitions that could otherwise receive high scores from low within-cluster scatter. Scoring uses only each view's own observed samples and does not require cross-view pairs.

With this coordinate system fixed, CAMVC must determine both which local samples correspond to canonical samples and how strongly their information should influence learning. CLST first restricts correspondence estimation to plausible local candidates. The resulting reconstruction then provides sample-level reliability weights and view-level weights for integrating canonicalized geometry. Covariance-aware residual modeling and a high-preservation tensor prior support feature-dependent variation and higher-order agreement, respectively. These components are coupled through alternating updates, followed by reliability-weighted representation fusion for clustering (Figure 1).

3.2 Cluster-Routed Local Sparse Transport

The canonical coordinates permit cross-view comparison, but missing observations may still leave multiple plausible correspondences. A hard permutation cannot retain these alternatives, whereas dense soft matching includes incompatible candidates and requires quadratic storage. CLST instead performs coarse cluster routing followed by local sparse matching. The reference transport is $\mathbf{P}^{(1)} = \mathbf{I}_n$; for $v \geq 2$,

$$\begin{cases} \mathbf{P}_{ij}^{(v)} \geq 0, & \sum_j \mathbf{P}_{ij}^{(v)} = 1, \quad w_i^{(v)} = 1, \\ \mathbf{P}_{i:}^{(v)} = \mathbf{0}, & \qquad\qquad\quad w_i^{(v)} = 0. \end{cases} \tag{3}$$

To construct the candidate support, we row-normalize $\mathbf{Z}_0^\top$ as $\widehat{\mathbf{Z}}_0 = \mathcal{N}_{\mathrm{row}}(\mathbf{Z}_0^\top)$ and partition it into $c_t = \min(n, \max(c, \mathrm{round}(\sqrt{n})))$ routing clusters, where c is the requested number of output clusters. Let a_j and $\boldsymbol{\xi}_r$ be the assignment of canonical sample j and the centroid of route r. The current anchor projects local samples as $\widehat{\mathbf{H}}^{(v)} = \mathcal{N}_{\mathrm{row}}((\mathbf{X}^{(v)}\mathbf{W}^{(v)})^\top \mathbf{A}^{(v)})$. For an observed sample i, CLST computes

$$\begin{aligned} D_{ir}^{(v)} &= \|\widehat{\mathbf{H}}_{i:}^{(v)} - \boldsymbol{\xi}_r\|_2^2, & \pi_{ir}^{(v)} &\propto \exp(-D_{ir}^{(v)}/\vartheta_{i,c}^{(v)}), \\ C_{ij}^{(v)} &= \|\widehat{\mathbf{H}}_{i:}^{(v)} - \widehat{\mathbf{Z}}_{0,j:}\|_2^2, & \psi_{ij|r}^{(v)} &\propto \exp(-C_{ij}^{(v)}/\vartheta_{i,w}^{(v)}), \\ \widetilde{\mathbf{P}}_{ij}^{(v)} &= \pi_{i,a_j}^{(v)} \psi_{ij|a_j}^{(v)}, & j &\in \bigcup_{r \in \mathcal{R}_i^{(v)}} \mathcal{K}_{ir}^{(v)}. \end{aligned} \tag{4}$$

Here, $\mathcal{R}_i^{(v)}$ contains the r_c most probable routes, and $\mathcal{K}_{ir}^{(v)}$ contains the k nearest canonical samples within retained route r (or all samples if fewer are available). Thus, the proposal combines route probability $\pi_{ir}^{(v)}$ with conditional matching probability $\psi_{ij|r}^{(v)}$ within that route. Matching-distance scales can vary across samples. Row-adaptive temperatures $\vartheta_{i,c}^{(v)}$ and $\vartheta_{i,w}^{(v)}$ account for these differences by scaling the respective softmax distributions to the row-wise distance dispersion. However, row normalization alone does not control the total assignment mass received by each canonical sample. A small number of clipped column-balancing steps therefore limits excessive concentration on a few canonical samples. Finally, because matching costs depend on the evolving anchors, the balanced proposal is combined with the previous transport:

$$\mathbf{P}^{(v)} \leftarrow \mathrm{RowNorm}\Big(\mathrm{Prune}_{r_c k}\Big[\alpha_p^{(s)} \mathbf{P}_{\mathrm{new}}^{(v)} + (1 - \alpha_p^{(s)}) \mathbf{P}^{(v)}\Big]\Big), \quad \alpha_p^{(s)} = \max\Big(\alpha_{p,\min}, \frac{\alpha_{p,0}}{s^{\gamma_p}}\Big). \tag{5}$$

This continuation avoids abrupt correspondence changes after the representation warm-up. Each row has at most $r_c k$ nonzeros; for balanced routes, CLST storage is $O(n r_c k)$ and its dominant matching cost is $O(n^2 t / c_t)$ rather than $O(n^2 t)$. Initialization, adaptive temperatures, balancing, and failure safeguards are detailed in Appendix B.1.

3.3 Correspondence-Aware Reliability and Geometry

CLST specifies which correspondences are admissible, but an admissible match may still yield an inconsistent reconstruction. To regulate its influence, we evaluate the observed reconstruction residual $\mathbf{R}^{(v)} = [\mathbf{X}^{(v)} - \mathbf{A}^{(v)}\mathbf{Z}^{(v)}(\mathbf{P}^{(v)})^\top]\mathbf{W}^{(v)}$. At iteration r, the maximum correntropy criterion (MCC) gives

$$q_{i,r}^{(v)} = \exp\left(-\frac{\|\mathbf{R}_{r,:i}^{(v)}\|_2^2}{2\sigma_{\mathrm{mcc},r-1}^2}\right), \qquad \mathbf{Q}_r^{(v)} = \mathrm{diag}(q_{1,r}^{(v)}, \ldots, q_{n,r}^{(v)}), \tag{6}$$

where $\sigma_{\mathrm{mcc},r-1}$ is the bandwidth from the preceding iteration. Larger residuals reduce a sample's influence on subsequent anchor and residual updates. These scalar weights control sample-level influence but cannot represent correlations across features. We therefore combine them with a feature covariance $\boldsymbol{\Sigma}^{(v)} \succ 0$:

$$\begin{aligned} \mathcal{N}^{(v)} = \|\mathbf{E}^{(v)}\mathbf{W}^{(v)}(\mathbf{Q}^{(v)})^{1/2}\|_F^2 + \Psi_{\sigma_{\mathrm{mcc}}}(\mathbf{Q}^{(v)}) \\ + \mathrm{Tr}\Big((\mathbf{E}^{(v)}\mathbf{W}^{(v)})^\top (\boldsymbol{\Sigma}^{(v)})^{-1} \mathbf{E}^{(v)}\mathbf{W}^{(v)}\Big) + n_v \log|\boldsymbol{\Sigma}^{(v)}|. \end{aligned} \tag{7}$$

Here, $n_v = |\mathcal{O}_v|$, and $\Psi_{\sigma_{\mathrm{mcc}}}$ is the auxiliary potential associated with MCC reweighting. The two lines address sample-level influence and feature-dependent variation, respectively; reconstruction consistency serves as a reliability proxy, not a guarantee of correct matching. Bandwidth updates and covariance variants are detailed in Appendix B.2.

Reliability weighting alone does not specify how views should share neighborhood information. Such information must first be expressed in the canonical sample order; directly fusing locally ordered graphs would connect unrelated samples. For each $\mathbf{Z}^{(v)}$, we blend its dynamic K-nearest-neighbor graph with the fixed reference graph: $\mathbf{S}^{(v)} = (1-\eta_g)\mathbf{S}_0 + \eta_g \mathbf{S}_{\mathrm{dyn}}^{(v)}$. Let $\mathcal{L}$ denote the normalized Laplacian, $\mathbf{L}^{(v)} = \mathcal{L}(\mathbf{S}^{(v)})$, and $\mathbf{L}_{-v} = (m-1)^{-1}\sum_{u\neq v}\mathbf{L}^{(u)}$. This construction supports two complementary forms of geometric regularization:

$$\begin{aligned} \mathcal{G}_{\mathrm{cross}} &= \sum_v \mathrm{Tr}\Big(\mathbf{Z}^{(v)}\mathbf{L}_{-v}(\mathbf{Z}^{(v)})^\top\Big), \\ \mathcal{G}_{\mathrm{rel}} &= \sum_v \mathrm{Tr}\Big(\mathbf{Z}^{(v)}\mathbf{L}_{s}(\mathbf{Z}^{(v)})^\top\Big). \end{aligned} \tag{8}$$

Here, $\mathbf{L}_s = \mathcal{L}(\mathbf{S}_s)$ is the normalized Laplacian of the consensus graph $\mathbf{S}_s = \sum_v \zeta_v \mathbf{S}^{(v)}$, where $\zeta_v \propto (e_v+\epsilon)^{-1}$ and $\sum_v \zeta_v = 1$. The first term encourages each representation to preserve neighborhoods from the other views. The second encourages agreement with the reliability-weighted consensus graph. Its normalized inverse-error weights ζ_v reduce the contribution of views with large mean squared observed reconstruction error e_v. Thus, the same correspondence-dependent reconstruction informs local sample weighting and global structure integration, while the two geometry terms serve distinct cross-view and consensus roles.

3.4 Objective and Optimization

The preceding terms regulate reconstruction and neighborhood consistency. To additionally encourage higher-order agreement, we stack the canonical representations along the view mode into $\mathcal{Z}$ and introduce an auxiliary tensor $\mathcal{G}$ with the supporting high-preservation regularizer $\widetilde{\mathcal{R}}_\delta$. Let Θ collect the view-specific variables $\mathbf{A}^{(v)}, \mathbf{Z}^{(v)}, \mathbf{E}^{(v)}, \mathbf{Q}^{(v)}$, and $\mathbf{\Sigma}^{(v)}$. For fixed transports and graphs, the resulting subproblem is

$$\begin{aligned} \min_{\Theta,\mathcal{G}} \quad & \widetilde{\mathcal{R}}_\delta(\mathcal{G}) + \lambda_1 \sum_v \mathcal{N}^{(v)} + \lambda_2 \mathcal{G}_{\mathrm{cross}} + \lambda_3 \mathcal{G}_{\mathrm{rel}} \\ \text{s.t.} \quad & \mathbf{X}^{(v)}\mathbf{W}^{(v)} = \mathbf{A}^{(v)}\mathbf{Z}^{(v)}(\mathbf{P}^{(v)})^\top\mathbf{W}^{(v)} + \mathbf{E}^{(v)}\mathbf{W}^{(v)}, \\ & (\mathbf{A}^{(v)})^\top\mathbf{A}^{(v)} = \mathbf{I}_t, \qquad \mathcal{Z} = \mathcal{G}, \\ & \mathbf{P}^{(1)} = \mathbf{I}_n, \qquad \mathbf{P}^{(v)} \text{ satisfies Eq. (3)}, \quad v \geq 2. \end{aligned} \tag{9}$$

The parameters λ_1, λ_2, and λ_3 control residual modeling, cross-view geometry, and reliability-aware geometry, respectively; CLST, tensor, and stabilization constants are fixed. Optimization alternates between reliability and representation updates, with periodic transport and graph refreshes, as summarized in Algorithm 1 in Appendix B. Updated anchors refine matching costs, while updated transports change reconstruction and hence the reliability weights. Because these refreshes alter the subproblem, the solver is an alternating surrogate procedure rather than ADMM on one fixed objective. Geometry and tensor terms are gradually activated after warm-up.

For fixed transports and graphs, the continuous blocks reduce to orthogonal Procrustes, sparse linear systems, ridge updates, and tensor proximal mapping. The full update equations, stopping rule, and complexity derivation are given in Appendices B.4–B.5.

3.5 Consensus Readout

After optimization, the canonical representations share a sample order, but their usefulness for clustering depends on both reconstruction and transport quality. Unlike the graph weights ζ_v, the final fusion weights also account for transport concentration and coverage. Let p_v be the mean row maximum of $\mathbf{P}^{(v)}$ over observed samples, and let $n_{\mathrm{eff}}^{(v)}$ denote the entropy-based effective number of covered canonical columns. We measure transport concentration by $h_v =$

$\max\{0, (p_v - 1/n)/(1-1/n)\}$ and effective coverage by $g_v = \min\{1, n_{\text{eff}}^{(v)}/|\mathcal{O}_v|\}$. Their geometric mean, $u_v = \sqrt{h_v g_v}$ for $v \geq 2$, favors concentrated assignments with broad canonical coverage; we set $u_1 = 1$ for the reference view. Combining this measure with reconstruction quality gives the fusion weights and consensus representation:

$$\omega_v = \frac{u_v/(e_v + \epsilon)}{\sum_{q=1}^{m} u_q/(e_q + \epsilon)}, \qquad \bar{\mathbf{Z}} = \sum_{v=1}^{m} \omega_v \mathbf{Z}^{(v)}. \tag{10}$$

A minimum reference weight stabilizes the fusion. We column-normalize $\bar{\mathbf{Z}}$ and apply k-means to $\bar{\mathbf{Z}}^\top$. Detailed definitions of p_v and $n_{\text{eff}}^{(v)}$ are provided in Appendix B.6.

4 Experiments

4.1 Experimental Setup

Datasets and Experimental Protocol. To evaluate CAMVC under simultaneous missingness and unalignment, we use six datasets: Synthetic3d, Reuters, Mfeat, Handwritten, Hdigit, and Cifar10, covering 600–50,000 samples. For each prescribed missing rate $r \in \{0.1, 0.3, 0.5\}$, observations are independently removed while retaining at least one observation per underlying sample. The selected reference view retains its local order; each other view is independently permuted together with its mask. All methods receive identical incomplete and shuffled data without ground-truth permutations or known sample pairs. Class labels are not supplied for representation learning. Dataset statistics, parameter grids, and implementation details are provided in Appendix A.1.

Baselines. We compare CAMVC with three groups of methods: aligned incomplete-view methods, including ASCR Xu et al. (2025), FSIMVC Du et al. (2024), and a covariance-aware baseline; unpaired methods, including UPMGC Wen et al. (2023b), MVC-UM Yu et al. (2021), and PAVuC-ATS Dong et al. (2025); and semantic-based methods, represented by SMILE Zeng et al. (2023).

Evaluation Metrics. Performance is measured by ACC, NMI, purity, and F-score, with higher values indicating better clustering.

4.2 Clustering Performance

Table 1 compares CAMVC with the baselines on six datasets at 10% missingness and full cross-view unalignment, using ACC, NMI, purity, and F-score. CAMVC attains the highest value in all 24 dataset–metric comparisons. On Synthetic3d, its NMI increases from SMILE at 0.3631 to 0.4813, a gain of 0.1182. On Reuters, its ACC increases from PAVuC_ATS at 0.3942 to 0.4967, a gain of 0.1025. Handwritten shows a similar advantage in F-score, which rises from MVC-UM's 0.4182 to 0.5094, a gain of 0.0912. Together, these results show that CAMVC learns effective clustering representations in the evaluated incomplete and unaligned setting without supplied cross-view sample pairs. Results at the other missing rates are provided in Appendix A.2, and Partial-unalignment comparisons are provided in Appendix A.3.

4.3 Parameter Sensitivity and Convergence

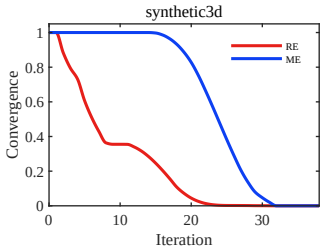
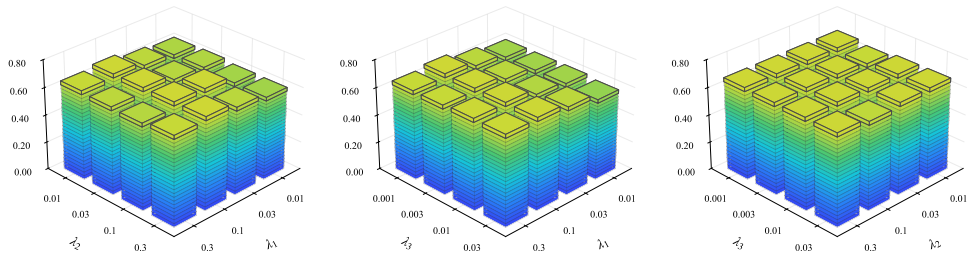


Figure 3: Convergence behavior

Figure 2 reports Mfeat ACC over the (λ_1, λ_2), (λ_1, λ_3), and (λ_2, λ_3) grids, with the remaining parameter fixed. For example, in the (λ_1, λ_2) panel, the bar at $\lambda_1 = \lambda_2 = 0.3$ is lower than several bars at intermediate settings, while neighboring combinations also retain relatively high ACC. This indicates that CAMVC is effective over more than one tested parameter combination, although the choice of regularization strengths still affects performance. Complete six-dataset plots are given in Appendices A.4.

Table 1: Performance comparison with state-of-the-art methods on six benchmark datasets under 10% missing rate. Best results are highlighted in **bold**.

Dataset	Metric	ASCR	FSIMVC	GUITAR	UPMGC	MVC-UM	PAVuC_ATS	SMILE	CAMVC
Synthetic3d	ACC	0.4558	0.6037	0.3800	0.4338	0.5478	0.3817	0.7034	**0.7633**
	NMI	0.1063	0.1637	0.0075	0.0477	0.2638	0.0488	0.3631	**0.4813**
	PUR	0.4558	0.6040	0.3800	0.4390	0.5483	0.4167	0.7234	**0.7893**
	F-score	0.4245	0.4546	0.3395	0.3629	0.5726	0.4516	0.5833	**0.6494**
Reuters	ACC	0.3677	0.1865	0.2608	0.2225	0.3703	0.3942	0.3485	**0.4967**
	NMI	0.2540	0.0137	0.0367	0.0212	0.3448	0.2186	0.1136	**0.4351**
	PUR	0.3702	0.2070	0.2625	0.2333	0.3897	0.4050	0.3588	**0.4967**
	F-score	0.3036	0.2766	0.1895	0.1843	0.2992	0.2908	0.2411	**0.3310**
Mfeat	ACC	0.1338	0.2471	0.1340	0.4826	0.6272	0.3620	0.2265	**0.6695**
	NMI	0.0117	0.0852	0.0102	0.4019	0.5574	0.4126	0.1276	**0.6032**
	PUR	0.1374	0.2561	0.1390	0.5147	0.6312	0.4085	0.2355	**0.6965**
	F-score	0.1026	0.1423	0.1012	0.3382	0.5055	0.2796	0.1546	**0.5317**
Handwritten	ACC	0.1719	0.2499	0.2030	0.4644	0.6110	0.3555	0.2289	**0.6570**
	NMI	0.0526	0.0852	0.0945	0.4531	0.5071	0.3816	0.1186	**0.5814**
	PUR	0.1742	0.2630	0.2085	0.5196	0.6218	0.3833	0.2351	**0.6590**
	F-score	0.1224	0.1429	0.1434	0.3560	0.4182	0.2660	0.1500	**0.5094**
Hdigit	ACC	0.1139	0.2940	0.3725	0.4355	0.4557	0.1788	0.3397	**0.5226**
	NMI	0.0019	0.1200	0.2310	0.3257	0.4072	0.0879	0.2703	**0.4508**
	PUR	0.1149	0.3322	0.3770	0.4660	0.4855	0.1866	0.3406	**0.5580**
	F-score	0.1077	0.1866	0.2570	0.2824	0.3300	0.1755	0.2367	**0.4199**
Cifar10	ACC	0.1959	0.4752	N/A	0.3090	0.7490	0.7375	0.3557	**0.7607**
	NMI	0.1886	0.2188	N/A	0.1740	0.6603	0.6299	0.3901	**0.6943**
	PUR	0.2099	0.4771	N/A	0.3209	0.7490	0.7375	0.3622	**0.7607**
	F-score	0.1819	0.2725	N/A	0.1872	0.6225	0.6330	0.3028	**0.6779**



Mfeat-ACC

Figure 2: Representative parameter sensitivity on Mfeat.

Figure 3 shows RE and ME curves on Synthetic3d. RE denotes the relative reconstruction residual, and ME denotes the tensor equality error. The displayed RE curve falls close to its terminal level by about 20 iterations, while ME remains higher and approaches its displayed terminal level later. Both curves approach stable values across the six datasets, with Cifar10 requiring more iterations than the smaller benchmarks. The plotted RE and ME curves approach stable values over the observed iterations, providing empirical evidence of numerical stabilization. Complete curves and metric definitions are provided in Appendix A.5.

4.4 Ablation Study

To evaluate correspondence learning and reliability weighting, we compare four variants at 10% missingness. *Global Sparse Matching* replaces cluster routing with global matching under the same candidate budget $r_c k$; *Hard Local Matching* uses one-hot assignments within routed candidate regions; *Uniform Sample Weights* sets $q_i^{(v)} = 1$; and *Uniform Graph-Fusion Weights* sets $\zeta_v = 1/m$ while retaining the graph regularizer and final representation-fusion rule. All variants use the same data realizations, initialization seeds, hyperparameters, and stopping criteria.

Table 2: Ablation study results on six benchmark datasets under 10% missing rate. The best results are highlighted in bold.

Variant	Synthetic3d				Mfeat				Handwritten				Reuters				Hdigit				Cifar10			
	ACC	NMI	Pur.	F-score	ACC	NMI	Pur.	F-score	ACC	NMI	Pur.	F-score	ACC	NMI	Pur.	F-score	ACC	NMI	Pur.	F-score	ACC	NMI	Pur.	F-score
Ours (Full)	**0.7633**	**0.4813**	**0.7893**	**0.6494**	**0.6695**	**0.6032**	**0.6965**	**0.5317**	**0.6570**	**0.5814**	**0.6590**	**0.5094**	**0.4967**	**0.4351**	**0.4967**	**0.3310**	**0.5226**	**0.4508**	**0.5580**	**0.4199**	**0.7607**	**0.6943**	**0.7607**	**0.6779**
Global sparse matching	0.6650	0.3184	0.6650	0.5404	0.4835	0.4947	0.5180	0.4152	0.5445	0.5097	0.5800	0.4530	0.3783	0.1376	0.3808	0.2689	0.3974	0.2618	0.4003	0.2828	0.6710	0.6331	0.6873	0.6231
Hard local matching†	0.6767	0.3168	0.6767	0.5452	0.5405	0.4982	0.5760	0.4420	0.5285	0.5104	0.5840	0.4439	0.4375	0.2066	0.4567	0.2125	0.4199	0.3455	0.4636	0.3187	0.6692	0.6376	0.6775	0.6155
Uniform sample weights	0.6800	0.3203	0.6800	0.5482	0.5500	0.5010	0.5850	0.4408	0.5310	0.5097	0.5855	0.4471	0.4450	0.2222	0.4450	0.2228	0.4685	0.3827	0.5112	0.3569	0.6964	0.5952	0.7029	0.5890
Uniform graph-fusion weight	0.6650	0.3232	0.6650	0.5432	0.5385	0.5043	0.5775	0.4552	0.5490	0.5247	0.5840	0.4704	0.4500	0.2253	0.4500	0.2263	0.4753	0.3989	0.5164	0.3708	0.6651	0.6219	0.7086	0.5996

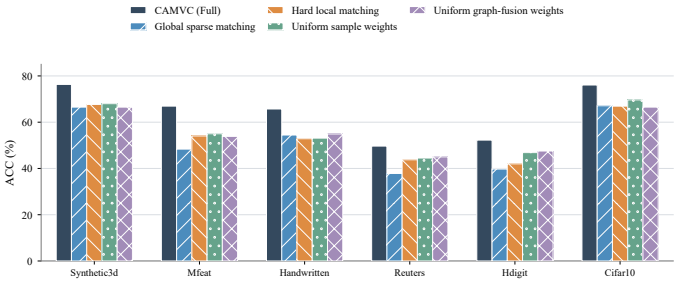


Figure 4: ACC of CAMVC and its ablation variants on six datasets at 10% missingness.

Figure 4 compares the ACC of CAMVC and the four variants across six datasets, while Table 2 reports all four clustering metrics. The full model obtains the highest scores in all 24 dataset–metric comparisons. For example, on Mfeat its ACC of 0.6695 falls to 0.4835 with global sparse matching and 0.5405 with hard local matching, reductions of 0.1860 and 0.1290, respectively. On Reuters, its NMI of 0.4351 falls to 0.2222 with uniform sample weights and 0.2253 with uniform graph-fusion weights. These results show that restricting candidate matches while retaining soft alternatives and weighting samples and graphs according to reconstruction reliability each contribute to CAMVC's clustering performance. A visualization of all four ablation metrics is provided in Appendix A.6.

4.5 Qualitative Correspondence and Structure Recovery

To examine cross-view structure recovery, Figure 5 visualizes the raw-view, aligned, and consensus similarity matrices on Synthetic3d, together with the matched class-level transport. In the aligned and consensus panels, within-class blocks remain visible; in the transport panel, the diagonal class associations are stronger than most off-diagonal associations. This class-level pattern indicates that the learned correspondences provide a common order in which cross-view structure can be integrated, complementing the clustering results in Table 1. Additional representation visualizations are given in Appendices A.7.

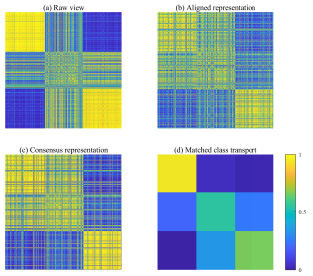


Figure 5: Structure recovery on Synthetic3d through alignment, consensus learning, and matched class-level transport.

5 Conclusion

In this work, we proposed CAMVC, a correspondence-aware framework for incomplete and unaligned multi-view clustering. CAMVC constructs a canonical coordinate system without paired samples and learns sparse soft correspondence through cluster-routed local sparse transport. Correspondence-dependent reconstruction residuals provide sample-specific reliability for representation learning, while observed-data reconstruction quality regulates graph fusion. Covariance-aware residual modeling and higher-order tensor regularization provide additional robustness. Experiments on six benchmarks demonstrate the effectiveness of CAMVC under varying levels of missingness and misalignment.

The current implementation requires repeated transport and graph updates and uses one automatically selected reference view. Future work will study scalable transport, multi-reference canonicalization, and severe missingness.

AI USE STATEMENT

Generative AI tools were used during manuscript preparation for language editing, LaTeX formatting, and organizational assistance. The authors reviewed and verified the scientific content, mathematical formulations, experimental design, reported results, and final manuscript, and take full responsibility for the submitted work.

REPRODUCIBILITY STATEMENT

The manuscript specifies the data construction protocol, compared methods, evaluation metrics, principal hyperparameter search space, optimization schedule, and the main implementation settings. The source code and experimental scripts will be released upon publication.

REFERENCES

Yuzhuo Dai, Jiaqi Jin, Zhibin Dong, Siwei Wang, Xinwang Liu, En Zhu, Xihong Yang, Xinbiao Gan, and Yu Feng. Imputation-free and alignment-free: Incomplete multi-view clustering driven by consensus semantic learning. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 5071–5081, 2025.

Wenhua Dong, Xiao-Jun Wu, Zhenhua Feng, Sara Atito, Muhammad Awais, and Josef Kittler. Probabilistically aligned view-unaligned clustering with adaptive template selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Liang Du, Yukai Shi, Yan Chen, Peng Zhou, and Yuhua Qian. Fast and scalable incomplete multi-view clustering with duality optimal graph filtering. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 8893–8902, 2024.

Zhibin Gu, Zhendong Li, and Songhe Feng. Edison: enhanced dictionary-induced tensorized incomplete multi-view clustering with gaussian error rank minimization. In *Forty-first international conference on machine learning*, 2024.

Menglei Hu and Songcan Chen. Doubly aligned incomplete multi-view clustering. *arXiv preprint arXiv:1903.02785*, 2019.

Jintian Ji, Songhe Feng, and Yidong Li. Tensorized unaligned multi-view clustering with multi-scale representation learning. In *Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining*, pp. 1246–1256, 2024.

Jiaqi Jin, Siwei Wang, Zhibin Dong, Xinwang Liu, and En Zhu. Deep incomplete multi-view clustering with cross-view partial sample and prototype alignment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11600–11609, 2023.

Zhao Kang, Zhiping Lin, Xiaofeng Zhu, and Wenbo Xu. Structured graph learning for scalable subspace clustering: From single view to multiview. *IEEE Transactions on Cybernetics*, 52(9): 8976–8986, 2021.

Haobin Li, Yunfan Li, Mouxing Yang, Peng Hu, Dezhong Peng, and Xi Peng. Incomplete multi-view clustering via prototype-based imputation. *arXiv preprint arXiv:2301.11045*, 2023.

Xingfeng Li, Yuangang Pan, Yinghui Sun, Quansen Sun, Ivor W Tsang, and Zhenwen Ren. Fast unpaired multi-view clustering. In *IJCAI*, pp. 4488–4496, 2024.

Cheng Liang, Daoyuan Wang, Huaxiang Zhang, Shichao Zhang, and Fei Guo. Robust tensor subspace learning for incomplete multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 36(11):6934–6948, 2024.

Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. Completer: Incomplete multi-view clustering via contrastive prediction. In *2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 11169–11178. IEEE, 2021.

Yapeng Liu, Wei Guo, Weiyu Li, Jingfeng Su, Qianlong Zhou, and Shanshan Yu. Tensor-based uncoupled and incomplete multi-view clustering. *Mathematics*, 13(9):1516, 2025.

Canyi Lu, Shuicheng Yan, and Zhouchen Lin. Convex sparse spectral clustering: Single-view to multi-view. *IEEE Transactions on Image Processing*, 25(6):2833–2843, 2016.

Jiatai Wang, Zhiwei Xu, Xuewen Yang, Xin Wang, and Li Tao. Hierarchical mutual information analysis: towards multi-view clustering in the wild. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. IEEE, 2024.

Lichen Wang, Zhengming Ding, Zhiqiang Tao, Yunyu Liu, and Yun Fu. Generative multi-view human action recognition. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6212–6221, 2019.

Siwei Wang, Xinwang Liu, Li Liu, Wenxuan Tu, Xinzhong Zhu, Jiyuan Liu, Sihang Zhou, and En Zhu. Highly-efficient incomplete large-scale multi-view clustering with consensus bipartite graph. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9776–9785, 2022.

Jie Wen, Zheng Zhang, Zhao Zhang, Lei Zhu, Lunke Fei, Bob Zhang, and Yong Xu. Unified tensor framework for incomplete multi-view clustering and missing-view inferring. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 10273–10281, 2021.

Jie Wen, Chengliang Liu, Gehui Xu, Zhihao Wu, Chao Huang, Lunke Fei, and Yong Xu. Highly confident local structure based consensus graph learning for incomplete multi-view clustering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 15712–15721, 2023a.

Yi Wen, Siwei Wang, Qing Liao, Weixuan Liang, Ke Liang, Xinhang Wan, and Xinwang Liu. Unpaired multi-view graph clustering with cross-view structure matching. *IEEE Transactions on Neural Networks and Learning Systems*, 35(11):16049–16063, 2023b.

Tingting Wu, Songhe Feng, and Jiazheng Yuan. Low-rank kernel tensor learning for incomplete multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 15952–15960, 2024.

Like Xin, Wanqi Yang, Lei Wang, and Ming Yang. Unpaired multiview clustering via reliable view guidance. *IEEE Transactions on Neural Networks and Learning Systems*, 36(6):11443–11455, 2024.

Like Xin, Wanqi Yang, Lei Wang, and Ming Yang. Multilevel reliable guidance for unpaired multi-view clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

Deng Xu, Chao Zhang, Cong Guo, Chunlin Chen, and Huaxiong Li. Fast incomplete multi-view clustering with adaptive similarity completion and reconstruction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 39, pp. 21734–21742, 2025.

Mouxing Yang, Yunfan Li, Peng Hu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Robust multi-view clustering with incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):1055–1069, 2022.

Yan Yang and Hao Wang. Multi-view clustering: A survey. *Big data mining and analytics*, 1(2): 83–107, 2018.

Hong Yu, Jia Tang, Guoyin Wang, and Xinbo Gao. A novel multi-view clustering method for unknown mapping relationships between cross-view samples. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pp. 2075–2083, 2021.

Pengxin Zeng, Mouxing Yang, Yiding Lu, Changqing Zhang, Peng Hu, and Xi Peng. Semantic invariant multi-view clustering with fully incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(4):2139–2150, 2023.

Chao Zhang, Huaxiong Li, Wei Lv, Zizheng Huang, Yang Gao, and Chunlin Chen. Enhanced tensor low-rank and sparse representation recovery for incomplete multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 11174–11182, 2023.

Haichao Zhang, Nasser M Nasrabadi, Yanning Zhang, and Thomas S Huang. Joint dynamic sparse representation for multi-view face recognition. *Pattern Recognition*, 45(4):1290–1298, 2012.

Feng Zhao, Jing Li, Lu Zhang, Zhe Li, and Sang-Gyun Na. Multi-view face recognition using deep neural networks. *Future Generation Computer Systems*, 111:375–380, 2020.

Liang Zhao, Ziyue Wang, Chuanye He, Qingchen Zhang, and Bo Xu. Dual robust unbiased multi-view clustering for incomplete and unpaired information. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 7092–7100, 2025.

Shuping Zhao, Jie Wen, Lunke Fei, and Bob Zhang. Tensorized incomplete multi-view clustering with intrinsic graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 11327–11335, 2023.

Zhenhao Zhong, Zhibin Gu, Pengpeng Yang, Ruiqiang Guo, et al. Gaussian regression-driven tensorized incomplete multi-view clustering with dual manifold regularization. *Advances in Neural Information Processing Systems*, 38:72102–72131, 2026.

Lihua Zhou, Guowang Du, Kevin Lü, Lizheng Wang, and Jingwei Du. A survey and an empirical evaluation of multi-view clustering approaches. *ACM Computing Surveys*, 56(7):1–38, 2024.

Appendix organization. Appendix A supplements the experiments in the main paper with detailed settings, numerical comparisons, and additional analyses. Appendix B provides the implementation and update details for the framework in Section 3. The notation, component definitions, and experimental settings follow the main paper.

A Supplementary Experiments and Reproducibility Details

A.1 Experimental Setup and Reproducibility

A.1.1 Datasets

Table 3 summarizes the numbers of samples, views, and classes.

Table 3: Statistics of the benchmark datasets

Dataset	#Samples	#Views	#Classes
Synthetic3d	600	3	3
Reuters	1200	5	6
Mfeat	2000	6	10
Handwritten	2000	6	10
Hdigit	10000	2	10
Cifar10	50000	3	10

A.1.2 Construction of Incomplete and Fully Unaligned Views

Since the original datasets are complete and aligned, we independently introduce missing observations and view-specific permutations. For a nominal missing rate r, each observation is removed independently with probability r, subject to each underlying instance remaining available in at least one view. The coverage constraint may make the realized missing fraction slightly smaller than r; throughout the paper, the reported rate denotes the prescribed missing probability.

For complete unalignment, the label-free reference-view selection strategy of CAMVC first identifies a reference view. This view retains its local order, whereas every non-reference view is independently permuted. Each permutation is applied consistently to the observed data and its observation mask. Neither the generated permutations nor the underlying cross-view correspondence is provided to any method. Consequently, every method receives the same incomplete and independently ordered views without predefined sample matches.

A.1.3 Compared Methods

We use the publicly released implementation of each method and set method-specific parameters according to the corresponding paper and source code. All methods receive identical incomplete and shuffled data; Ground-truth permutations and cross-view correspondences are never provided.

A.1.4 Evaluation Metrics

We report four standard clustering metrics: clustering accuracy (ACC), normalized mutual information (NMI), purity (PUR), and F-score. For ACC, the Hungarian algorithm resolves the permutation ambiguity between predicted cluster indices and ground-truth classes. Higher values indicate better performance for all metrics.

A.1.5 Implementation Details

CAMVC is implemented in MATLAB. We select $\lambda_1, \lambda_2 \in \{0.01, 0.03, 0.1, 0.3\}$ and $\lambda_3 \in \{0.001, 0.003, 0.01, 0.03\}$. Optimization runs for at most 300 iterations. The correspondence variables are fixed during the first 10 iterations for representation warm-up, after which CLST updates are activated. Final assignments are obtained by applying k-means with 20 replicates to the learned consensus representation.

Table 4: Performance comparison with state-of-the-art methods on six benchmark datasets under 30% missing rate. Best results are highlighted in **bold**.

Dataset	Metric	ASCR	FSIMVC	GUITAR	UPMGC	MVC-UM	PAVuC_ATS	SMILE	CAMVC
Synthetic3d	ACC	0.4910	0.5667	0.3517	0.4014	0.4656	0.3750	0.7026	**0.7183**
	NMI	0.1229	0.1261	0.0014	0.0212	0.1580	0.0280	**0.4072**	0.2387
	PUR	0.4910	0.5667	0.3517	0.4014	0.4661	0.3817	0.7026	**0.7183**
	F-score	0.4128	0.4239	0.3352	0.3510	0.5028	0.4017	0.4041	**0.5071**
Reuters	ACC	0.2626	0.1915	0.2317	0.2106	0.3267	0.3608	0.3533	**0.4423**
	NMI	0.3679	0.0095	0.0166	0.0177	0.1675	0.1272	0.1200	**0.3937**
	PUR	0.3685	**0.7587**	0.2333	0.2220	0.3418	0.3725	0.3652	0.4742
	F-score	**0.4141**	0.2616	0.1818	0.1756	0.2738	0.2569	0.2433	0.2873
Mfeat	ACC	0.1327	0.2388	0.1375	0.3778	0.4936	0.3150	0.2500	**0.5580**
	NMI	0.0103	0.0689	0.0098	0.3290	0.4428	0.3065	0.1568	**0.4459**
	PUR	0.1354	0.2502	0.1395	0.3976	0.5002	0.3395	0.2663	**0.5980**
	F-score	0.1052	0.1353	0.1008	0.2396	0.3143	0.2251	0.1738	**0.3467**
Handwritten	ACC	0.1872	0.2370	0.2130	0.3867	0.4926	0.3025	0.2273	**0.5375**
	NMI	0.0796	0.0698	0.0869	0.3555	0.4394	0.2878	0.1241	**0.5805**
	PUR	0.1914	0.2447	0.2155	0.4129	0.4975	0.3210	0.2442	**0.5375**
	F-score	0.1431	0.1347	0.1402	0.2594	0.3121	0.2195	0.1548	**0.3347**
Hdigit	ACC	0.1117	0.2727	0.2525	0.2488	0.3722	0.1701	0.3227	**0.4691**
	NMI	0.0014	0.1017	0.1414	0.1110	0.3258	0.0782	0.2935	**0.3936**
	PUR	0.1127	0.3117	0.2690	0.2625	0.3957	0.1768	0.3483	**0.4913**
	F-score	0.1183	0.1690	0.1728	0.1433	0.2319	0.1553	0.2429	**0.3423**
Cifar10	ACC	0.1858	0.4040	N/A	0.1680	0.5920	0.6169	0.4306	**0.6354**
	NMI	0.1398	0.1570	N/A	0.0381	0.5261	0.5203	0.4441	**0.5595**
	PUR	0.1893	0.4261	N/A	0.1790	0.5963	**0.6830**	0.4381	0.6478
	F-score	0.1686	0.2237	N/A	0.1045	0.3807	0.3867	0.3565	0.5155

Table 5: Performance comparison with state-of-the-art methods on six benchmark datasets under 50% missing rate. Best results are highlighted in **bold**.

Dataset	Metric	ASCR	FSIMVC	GUITAR	UPMGC	MVC-UM	PAVuC_ATS	SMILE	CAMVC
Synthetic3d	ACC	0.4647	0.5012	0.3583	0.3726	0.4289	0.3850	**0.7261**	0.5867
	NMI	0.1018	0.0747	0.0037	0.0063	0.1061	0.0193	0.3907	**0.4535**
	PUR	0.4717	0.5307	0.3667	0.3766	0.4294	0.4017	**0.7261**	0.5883
	F-score	0.4083	0.3942	0.3541	0.3557	0.4823	0.3715	**0.5905**	0.5096
Reuters	ACC	0.2468	0.1920	0.2242	0.2037	0.2225	0.2250	0.3377	**0.3817**
	NMI	**0.3784**	0.0085	0.0275	0.0155	0.0178	0.0174	0.0975	0.2888
	PUR	0.3485	**0.5764**	0.2267	0.2158	0.2267	0.2275	0.3404	0.4950
	F-score	0.2785	0.2426	0.1889	0.1687	0.1736	0.1819	0.2309	**0.3712**
Mfeat	ACC	0.1356	0.2207	0.1335	0.2676	0.3681	0.2525	0.2538	**0.4330**
	NMI	0.0112	0.0516	0.0093	0.1893	0.3254	0.2332	0.1355	**0.4179**
	PUR	0.1383	0.2460	0.1360	0.2964	0.3709	0.2675	0.2711	**0.4330**
	F-score	0.1114	0.1293	0.1005	0.1766	0.2066	0.1891	0.1640	**0.2651**
Handwritten	ACC	0.1814	0.2284	0.1740	0.2895	0.4355	0.2610	0.2561	**0.4845**
	NMI	0.0905	0.0529	0.0682	0.2011	0.3998	0.2503	0.1476	**0.4671**
	PUR	0.1850	0.2506	0.1840	0.3017	0.4355	0.2840	0.2698	**0.4845**
	F-score	0.1554	0.1308	0.1275	0.1822	**0.3161**	0.1901	0.1682	0.2272
Hdigit	ACC	0.1121	0.2728	0.2320	0.2009	0.2980	0.1304	0.3521	**0.3836**
	NMI	0.0017	0.1021	0.1170	0.0765	0.2555	0.0392	0.2907	**0.2974**
	PUR	0.1136	0.3018	0.2395	0.2085	0.3180	0.1350	0.3791	**0.3907**
	F-score	0.1338	0.1651	0.1572	0.1207	0.1878	0.1638	0.2531	**0.3076**
Cifar10	ACC	0.1685	0.3654	N/A	0.1800	0.4527	0.4550	0.3874	**0.5028**
	NMI	0.0913	0.1310	N/A	0.0601	0.4087	0.3616	0.4069	**0.4866**
	PUR	0.1749	0.3830	N/A	0.1912	0.4583	0.4573	0.4014	**0.5028**
	F-score	0.1427	0.1948	N/A	0.1145	0.2344	0.2427	0.3244	**0.4644**

A.2 Complete Results across Missing Rates

Tables 4 and 5 extend the comparison in Table 1 to prescribed missing rates of 30% and 50% under complete cross-view unalignment. All methods use the protocol in Appendix A.1.

CAMVC achieves the highest ACC on all six datasets at 30% missingness and on five datasets at 50%. On Mfeat, its ACC is 0.5580 and 0.4330 at these two rates, compared with 0.4936 and 0.3681 for MVC-UM, respectively. The advantage is not uniform: on Synthetic3d at 50% missingness, SMILE achieves 0.7261 ACC, compared with 0.5867 for CAMVC. These results support the effectiveness of CAMVC under the evaluated settings while identifying dataset-dependent limitations.

Table 6: Performance comparison with state-of-the-art methods on six benchmark datasets under 10% missing rate and 30% unaligned rate. Best available results are highlighted in **bold**.

Dataset	Metric	ASCR	FSIMVC	GUITAR	UPMGC	MVC-UM	PAVuC_ATS	SMILE	CAMVC
Synthetic3d	ACC	0.5957	0.7090	0.5233	0.4443	0.5548	0.4450	0.6158	**0.7867**
	NMI	0.2126	0.4547	0.0788	0.0538	0.2701	0.1082	0.2910	**0.4724**
	PUR	0.5957	0.7390	0.5233	0.4541	0.5556	0.4667	0.6158	**0.7967**
	F-score	0.4826	0.6325	0.3901	0.3664	0.5682	0.4303	0.5158	**0.6924**
Reuters	ACC	0.5005	0.1896	0.3642	0.2666	0.3821	0.3082	0.3232	**0.5442**
	NMI	0.3701	0.0194	0.1182	0.0534	0.2314	0.1739	0.1090	**0.4406**
	PUR	0.5031	**0.9120**	0.3642	0.2802	0.3979	0.3498	0.3494	0.5658
	F-score	0.3946	0.2785	0.2418	0.2115	0.3037	0.2812	0.2337	**0.4699**
Mfeat	ACC	0.5795	0.5720	0.5250	0.4875	0.6036	0.3625	0.2481	**0.6420**
	NMI	0.6054	0.5452	0.3683	0.4124	0.5613	0.4111	0.1468	**0.6975**
	PUR	0.5795	0.5720	0.5250	0.5212	0.6175	0.4075	0.2636	**0.6775**
	F-score	0.5081	0.4645	0.3626	0.3497	0.5017	0.2797	0.1670	**0.6448**
Handwritten	ACC	0.5970	0.5651	0.5315	0.4586	0.6043	0.3625	0.2450	**0.6290**
	NMI	0.5352	0.5390	0.3944	0.4303	0.5599	0.3793	0.1495	**0.6118**
	PUR	0.5970	0.6651	0.5635	0.5178	0.6157	0.3830	0.2557	**0.6845**
	F-score	0.4348	0.4549	0.3887	0.3502	0.4940	0.2617	0.1665	**0.5432**
Hdigit	ACC	0.3845	0.5143	0.4380	0.4408	0.4418	0.1838	0.3451	**0.7854**
	NMI	0.2433	0.3468	0.3424	0.3297	0.3958	0.1348	0.2669	**0.6406**
	PUR	0.4128	0.5404	0.4775	0.4690	0.5054	0.2014	0.3629	**0.7854**
	F-score	0.2626	0.3593	0.3201	0.2909	0.3404	0.1856	0.2391	**0.6378**
Cifar10	ACC	0.2866	0.5618	N/A	0.3024	0.6874	0.5438	0.3105	**0.7412**
	NMI	0.2447	0.5164	N/A	0.1993	0.6294	0.4364	0.3789	**0.6931**
	PUR	0.3030	0.5618	N/A	0.3272	0.6936	0.5705	0.3181	**0.7412**
	F-score	0.2225	0.4564	N/A	0.1999	0.5991	0.4277	0.2830	**0.6378**

Table 7: Performance comparison with state-of-the-art methods on six benchmark datasets under 10% missing rate and 50% unaligned rate. Best available results are highlighted in **bold**.

Dataset	Metric	ASCR	FSIMVC	GUITAR	UPMGC	MVC-UM	PAVuC_ATS	SMILE	CAMVC
Synthetic3d	ACC	0.5083	0.5015	0.4300	0.4248	0.5600	0.4167	0.6146	**0.7817**
	NMI	0.1434	0.3080	0.0420	0.0513	0.2607	0.0511	0.3491	**0.4211**
	PUR	0.5083	0.7027	0.4483	0.4468	0.5611	0.4383	0.7146	**0.7817**
	F-score	0.4274	0.5625	0.3656	0.3661	0.5663	0.4179	0.5678	**0.6493**
Reuters	ACC	0.4093	0.1864	0.3808	0.2478	0.3687	0.3267	0.3640	**0.4925**
	NMI	0.2642	0.0143	0.1262	0.0358	0.2270	0.0964	0.1274	**0.3494**
	PUR	0.4093	**0.8944**	0.3933	0.2579	0.3903	0.3308	0.3713	0.4925
	F-score	0.3279	0.2756	0.2631	0.2110	0.3001	0.2507	0.2491	**0.5190**
Mfeat	ACC	0.5399	0.5333	0.3750	0.4766	0.6082	0.3750	0.2547	**0.6400**
	NMI	0.3260	0.4719	0.1927	0.4137	**0.5548**	0.3839	0.1542	0.4932
	PUR	0.5415	0.5568	0.3950	0.5086	0.6082	0.4000	0.2743	**0.6755**
	F-score	0.3368	0.4893	0.2256	0.3431	0.4904	0.2666	0.1718	**0.5415**
Handwritten	ACC	0.5011	0.4708	0.4540	0.4713	0.4053	0.3741	0.2927	**0.5300**
	NMI	0.3620	0.4954	0.2675	0.4541	0.4517	0.3839	0.1661	**0.5117**
	PUR	0.5524	0.4708	0.4800	0.5198	0.4138	0.3910	0.2955	**0.5855**
	F-score	0.3596	0.3148	0.2883	0.3589	0.3876	0.2679	0.1822	**0.4459**
Hdigit	ACC	0.2608	0.4392	0.3760	0.4391	0.4570	0.1985	0.3260	**0.6405**
	NMI	0.0992	0.2536	0.2776	0.3227	0.4203	0.1290	0.2804	**0.5114**
	PUR	0.2678	0.4579	0.4045	0.4613	0.5425	0.2276	0.3600	**0.6444**
	F-score	0.1556	0.2850	0.2659	0.2870	0.3955	0.1826	0.2433	**0.4968**
Cifar10	ACC	0.2522	0.4669	N/A	0.3031	0.6427	0.3683	0.4334	**0.6971**
	NMI	0.2170	0.4604	N/A	0.1859	0.6008	0.3137	0.4357	**0.7314**
	PUR	0.2737	0.4731	N/A	0.3197	0.6563	0.3816	0.4372	**0.7077**
	F-score	0.2048	0.3068	N/A	0.1880	0.5845	0.4190	0.3530	**0.6390**

A.3 Robustness to Partial Unalignment

To complement the completely unaligned setting, we evaluate partial unalignment at a fixed missing rate of 10%. The unaligned rate denotes the fraction of samples whose cross-view ordering is disrupted, while the remaining samples retain their original correspondence. We consider unaligned rates of 30% and 50% and use the same data-construction and parameter-selection protocols for all methods.

Tables 6 and 7 report the comparisons under partial unalignment. CAMVC achieves the highest ACC on all six datasets at both rates. Across the four metrics, it obtains the best reported values in 23 of 24

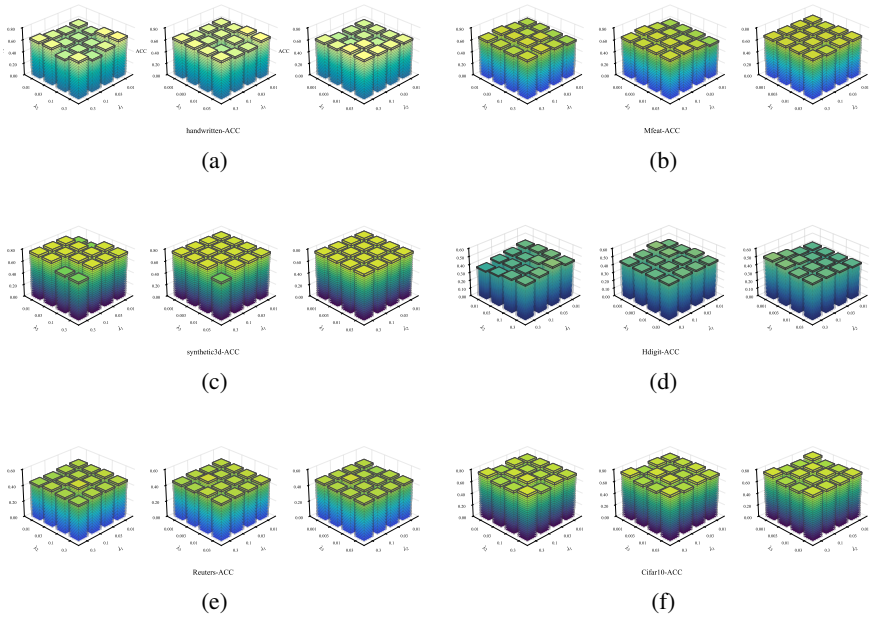


Figure 6: Parameter sensitivity of CAMVC (ACC) on six benchmark datasets over the $(\lambda_1, \lambda_2, \lambda_3)$ search grid.

dataset–metric combinations at 30% unalignment and in 22 of 24 at 50%. This behavior is consistent with the sparse soft correspondence formulation, which can exploit the preserved structure while retaining uncertainty for the disrupted portion. Together with the completely unaligned experiments, these results indicate that CAMVC is applicable across a continuum from partially preserved to fully unknown cross-view correspondence.

A.4 Parameter Sensitivity

We extend the parameter-sensitivity analysis to all six datasets. The parameters λ_1, λ_2, and λ_3 control correspondence-aware residual modeling, cross-view geometry transfer, and reliability-aware consensus geometry, respectively. For each parameter pair, we keep the remaining parameter at its selected value and report ACC.

Figure 6 shows that multiple parameter combinations yield relatively high ACC, although the performance variation differs across datasets. Overall, the model is not sensitive to a single finely tuned parameter combination over the investigated ranges.

A.5 Convergence Analysis

We examine numerical stabilization on all six datasets at a 10% missing rate. We monitor the relative reconstruction error (RE) and the tensor equality error (ME) associated with $\mathcal{Z} = \mathcal{G}$.

As shown in Fig. 7, both quantities approach stable values across the datasets. Reconstruction typically stabilizes before the tensor equality error, which is consistent with the optimization schedule: correspondence is held fixed during the initial warm-up and is subsequently refreshed at prescribed intervals. The larger CIFAR-10 dataset requires more iterations, but both curves stabilize within the maximum iteration budget.

A.6 Four-Metric Ablation Visualization

Figure 8 visualizes the four metrics reported in Table 2 across the six datasets, allowing the effects of the ablation variants to be compared across metrics.

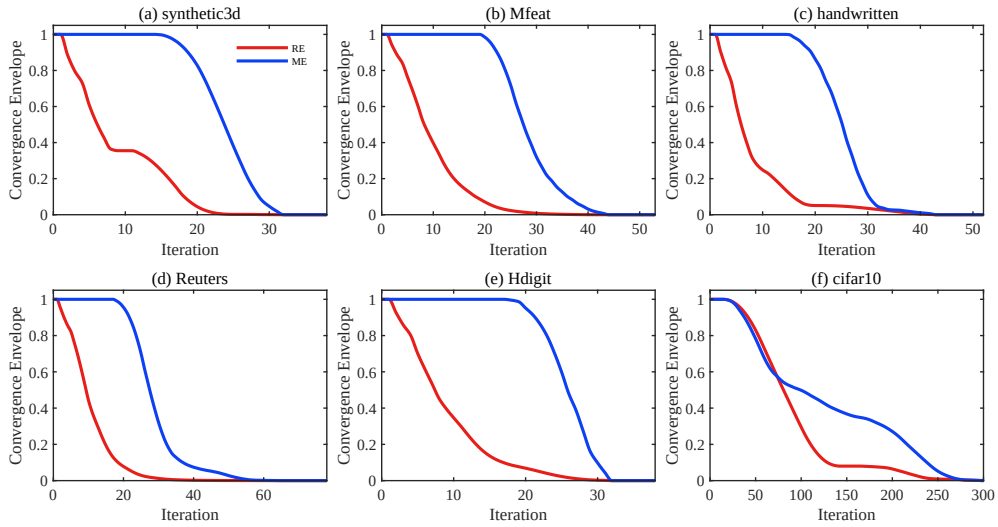


Figure 7: Convergence behavior of CAMVC on six datasets.

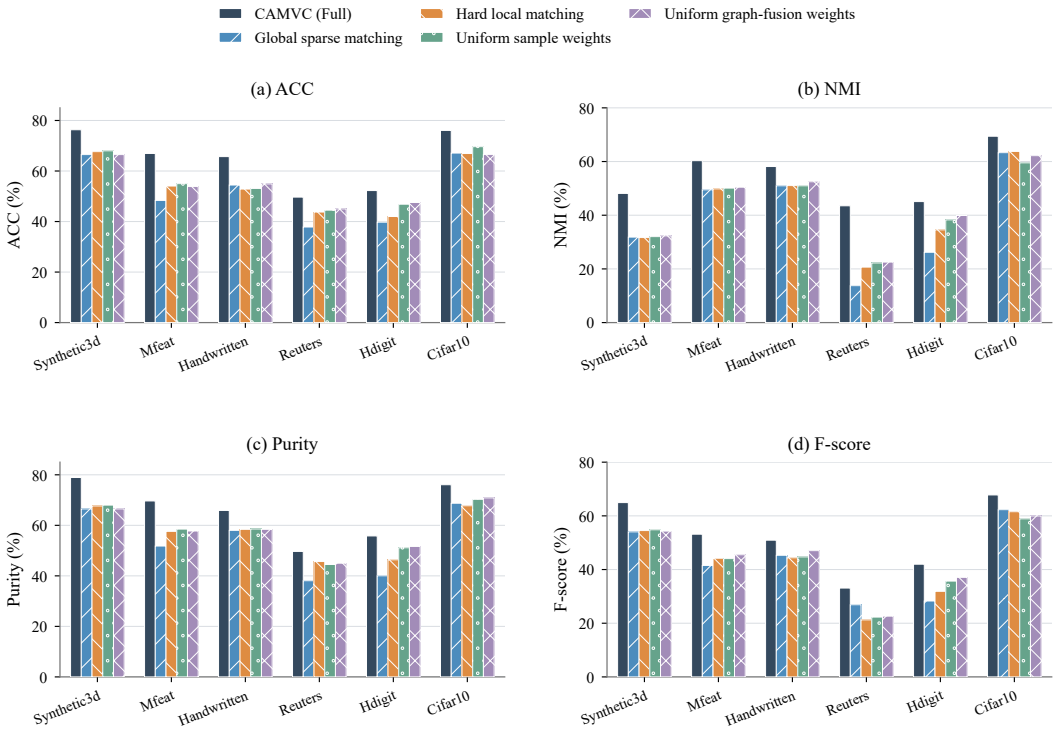
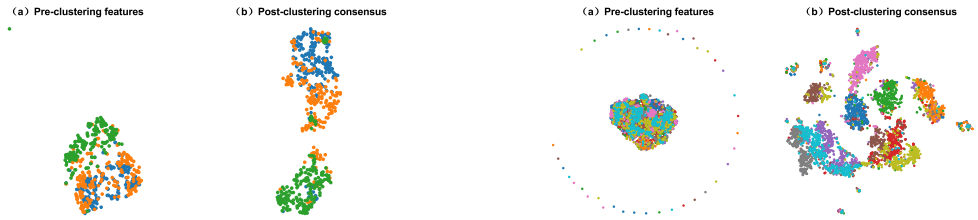


Figure 8: Ablation results on six datasets at 10% missingness. ACC, NMI, purity, and F-score are shown for CAMVC and its four variants.

A.7 Qualitative Representation Analysis

We use t-SNE to visualize the learned consensus representation at a missing rate of 10% on the Synthetic3d and CIFAR-10 datasets.

In Fig. 9, samples from the same class tend to occupy common regions, although overlap remains between some classes. Synthetic3d forms three relatively compact class regions, while CIFAR-10 exhibits more overlap because of its larger intra-class variation. The remaining local ambiguity is



Synthetic3d CIFAR-10

Figure 9: t-SNE visualizations of the consensus representations learned by CAMVC on Synthetic3d and CIFAR-10 at a missing rate of 10%.

Algorithm 1 Optimization of CAMVC

Require: Incomplete unaligned views $\{\mathbf{X}^{(v)}, \mathbf{W}^{(v)}\}$, cluster number c, and $\lambda_1, \lambda_2, \lambda_3$.
Ensure: Cluster labels.
1: Select the reference view and initialize $\mathbf{Z}_0$, anchors, residuals, and sparse transports.
2: **for** $r = 1, 2, \ldots$ **do**
3: Compute MCC weights and update orthogonal anchors.
4: **if** warm-up is finished and a transport update is due **then**
5: Refresh non-reference transports by Eqs. (4)–(5).
6: **end if**
7: Refresh canonical graphs when scheduled.
8: Update latent representations, the tensor auxiliary variable, residuals, covariances, and multipliers.
9: Stop when the feasibility and change criteria remain satisfied for the required patience.
10: **end for**
11: Fuse the canonical representations using the readout in Eq. (10), then apply k-means.

expected under simultaneous missingness and correspondence uncertainty. Since t-SNE can distort global distances, these plots are used as supplementary evidence for the clustering results.

B Algorithm and Implementation Details

This appendix specifies the update order and the implementation details needed to reproduce the method in Section 3. The notation and objective follow the main paper.

B.1 Initialization and CLST Implementation

To select the reference view, we standardize the observed features of each view, compute a truncated-SVD embedding, and cluster that embedding with k-means. The score in Eq. (2) selects the reference, which is reordered as view 1. Its rank-t SVD defines the fixed guide $\mathbf{Z}_0$ in Section 3, and all $\mathbf{Z}^{(v)}$ are initialized by this guide. Non-reference transports are initialized by matching independently obtained view clusters to the reference clusters, followed by sparse row normalization and two unsupervised CLST refinement steps. An invalid initialization falls back to a balanced sparse transport.

For the distances in Eq. (4), the row-adaptive temperatures are

$$\begin{aligned} \vartheta_{i,c}^{(v)} &= \max\Big(\tau_c \operatorname{std}_r(D_{ir}^{(v)}), \epsilon_c\Big), \\ \vartheta_{i,w}^{(v)} &= \max\Big(\tau_w \operatorname{std}_j(C_{ij}^{(v)}), \epsilon_w\Big). \end{aligned} \tag{11}$$

The r_c largest route probabilities are retained and renormalized. Within each retained route, at most k nearest canonical samples are used. After matching, clipped column-balancing steps limit excessive mass concentration. Missing rows are set to zero, observed rows are normalized, and at most $r_c k$ entries per row are retained. An update is accepted only if it is finite and satisfies the row-mass constraints.

B.2 Reliability, Graphs, and Tensor Prior

The MCC weights use the observed residual in Eq. (6). After the weights and anchors are updated, the bandwidth for the next iteration is

$$\sigma_{\mathrm{mcc},r} = \mathrm{clip}\left(\sqrt{\mathrm{median}_{v,i:w_i^{(v)}=1} \|\mathbf{R}_{r,:i}^{(v)}\|_2^2 + \epsilon}, \sigma_{\min}, \sigma_{\max}\right). \tag{12}$$

The residual model is given in Eq. (7). Full feature covariance is used for moderate dimensions, while the scalable variant retains only feature-wise residual variances.

The reference graph $\mathbf{S}_0$ is a symmetrized binary K-nearest-neighbor graph constructed within routing clusters. The graph blend is defined in Section 3; its implementation uses $\eta_g = 0.25$ and refreshes the dynamic graphs every ten iterations. For a symmetric nonnegative graph $\mathbf{S}$, the normalized Laplacian is

$$\mathcal{L}(\mathbf{S}) = \mathbf{I} - \frac{1}{2}\left(\mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{-1/2} + (\mathbf{D}^{-1/2}\mathbf{S}\mathbf{D}^{-1/2})^\top\right), \quad \mathbf{D} = \mathrm{diag}(\mathbf{S}\mathbf{1}). \tag{13}$$

The observed-data reconstruction error and graph-fusion weight are

$$e_v = \frac{\|[\mathbf{X}^{(v)} - \mathbf{A}^{(v)}\mathbf{Z}^{(v)}(\mathbf{P}^{(v)})^\top]\mathbf{W}^{(v)}\|_F^2}{d_v n_v}, \qquad \zeta_v = \frac{1/(e_v+\epsilon)}{\sum_u 1/(e_u+\epsilon)}. \tag{14}$$

B.3 Higher-Order Consensus Prior

For the stacked canonical representations $\mathcal{Z} = \Phi(\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(m)})$, the auxiliary tensor satisfies $\mathcal{Z} = \mathcal{G}$. The regularizer used in Eq. (9) is

$$\mathcal{R}_\delta(\mathcal{G}) = \frac{1}{m}\sum_{k=1}^{m}\sum_i \frac{(1+\delta)\tanh(\bar{s}_i^{(k)})}{\delta + \tanh(\bar{s}_i^{(k)})}, \qquad \bar{s}_i^{(k)} = \frac{s_i^{(k)}}{\max_j s_j^{(k)} + \epsilon}. \tag{15}$$

Here $s_i^{(k)}$ denotes a singular value of the k-th Fourier slice. The implemented penalty is $\widetilde{\mathcal{R}}_\delta = \kappa_T \mathcal{R}_\delta$, with $\kappa_T = 2 \times 10^{-3}$.

B.4 Alternating Updates

Let $\mathbf{Y}^{(v)}$ and $\mathcal{M}$ denote the reconstruction and tensor multipliers. After T_w warm-up iterations, geometry and tensor regularization are activated according to $\chi_r = \min\{1, \max\{0, (r - T_w)/T_r\}\}$.

Anchors and transports. For each view, define

$$\mathbf{M}_{A,r}^{(v)} = \mathbf{X}^{(v)}\mathbf{W}^{(v)}\mathbf{Q}_r^{(v)}\mathbf{W}^{(v)}\mathbf{P}_{r-1}^{(v)}(\mathbf{Z}_{r-1}^{(v)})^\top. \tag{16}$$

If $\mathbf{M}_{A,r}^{(v)} = \mathbf{U}_A\mathbf{\Lambda}_A\mathbf{V}_A^\top$, the orthogonal anchor update is $\mathbf{A}_r^{(v)} = \mathbf{U}_A\mathbf{V}_A^\top$. After warm-up, non-reference transports are refreshed periodically using Eqs. (4)–(5); $\mathbf{P}^{(1)} = \mathbf{I}_n$ remains fixed. Graphs are refreshed on their separate schedule using the reconstruction errors in Eq. (14).

Latent representations. Define $\widetilde{\mathbf{X}}^{(v)} = \mathbf{X}^{(v)}\mathbf{W}^{(v)} - \mathbf{E}^{(v)}\mathbf{W}^{(v)} + \mathbf{Y}^{(v)}/\mu$ and $\mathbf{C}_v = (\mathbf{P}^{(v)})^\top\mathbf{W}^{(v)}\mathbf{P}^{(v)}$. To account for sparse-transport curvature, set

$$\bar{c}_v = \mathrm{Tr}(\mathbf{C}_v)/n, \qquad r_v = \mathrm{Tr}(\mathbf{W}^{(v)})/n, \qquad s_v = \mathrm{clip}\left(\frac{r_v}{\max(\bar{c}_v, \epsilon)}, 1, s_{\max}\right). \tag{17}$$

For weakly covered canonical columns, let $\mathbf{K}_v^{(r)} = \mathrm{diag}(\kappa_1^{(v)}, \ldots, \kappa_n^{(v)})$, where $c_j^{(v)} = (\mathbf{C}_v)_{jj}$, $a_v = \max(\mu s_v \bar{c}_v, \epsilon_a)$, $\widetilde{c}_v$ is the median positive entry of $s_v c_j^{(v)}$, and

$$\kappa_j^{(v)} = a_v\left[\gamma_0 + \gamma_u \max\left(0, 1 - \frac{s_v c_j^{(v)}}{\max(\widetilde{c}_v, \epsilon)}\right)\right]. \tag{18}$$

The undamped update solves $\mathbf{Z}_\star^{(v)}\mathcal{R}^{(v)} = \mathcal{B}^{(v)}$, with

$$\begin{aligned}\mathcal{R}^{(v)} &= \mu s_v \mathbf{C}_v + 2\chi_r\lambda_2\mathbf{L}_{-v} + 2\chi_r\lambda_3\mathbf{L}_s + \rho\mathbf{I}_n + \mathbf{K}_v^{(r)} + \mathbb{I}[v=1]\gamma_A a_v \mathbf{I}_n,\\ \mathcal{B}^{(v)} &= \mu s_v (\mathbf{A}^{(v)})^\top \widetilde{\mathbf{X}}^{(v)}\mathbf{P}^{(v)} + \rho\mathbf{G}^{(v)} - \mathbf{M}^{(v)} + \mathbf{Z}_{\mathrm{old}}^{(v)}\mathbf{K}_v^{(r)} + \mathbb{I}[v=1]\gamma_A a_v \mathbf{Z}_0.\end{aligned} \tag{19}$$

The last terms in these two lines are an update-level reference-guide stabilizer. The symmetric systems are solved row-wise by PCG, and the accepted solution is damped: $\mathbf{Z}^{(v)} \leftarrow \beta_z^{(r)}\mathbf{Z}_\star^{(v)} + (1-\beta_z^{(r)})\mathbf{Z}_{\mathrm{old}}^{(v)}$.

Tensor, residual, and covariance variables. During warm-up, $\mathcal{G}$ tracks $\mathcal{Z}$ and $\mathcal{M}$ remains zero. Subsequently,

$$\mathcal{G} = \mathrm{prox}_{\chi_r\widetilde{\mathcal{R}}_\delta/\rho}(\mathcal{Z} + \mathcal{M}/\rho). \tag{20}$$

The Fourier slices use iteratively reweighted singular-value shrinkage for Eq. (15).

For an observed local column i, let $\mathbf{r}_i^{(v)}$ be the corresponding column of $\mathbf{X}^{(v)}\mathbf{W}^{(v)} - \mathbf{A}^{(v)}\mathbf{Z}^{(v)}(\mathbf{P}^{(v)})^\top\mathbf{W}^{(v)} + \mathbf{Y}^{(v)}/\mu$. The full-covariance residual update solves

$$\left[2\lambda_1(\mathbf{\Sigma}^{(v)})^{-1} + (2\lambda_1 q_i^{(v)} + \mu + \epsilon)\mathbf{I}_{d_v}\right]\widehat{\mathbf{e}}_i^{(v)} = \mu\mathbf{r}_i^{(v)}. \tag{21}$$

The diagonal-covariance version is solved elementwise. Missing residual columns are reset to zero. The full covariance candidate is

$$\widehat{\mathbf{\Sigma}}^{(v)} = \frac{\mathbf{E}^{(v)}\mathbf{W}^{(v)}(\mathbf{E}^{(v)}\mathbf{W}^{(v)})^\top}{\max(n_v, 1)} + \epsilon\mathbf{I}_{d_v}, \tag{22}$$

followed by damping; the scalable variant retains feature-wise mean squared residuals.

Multipliers. Writing $\mathbf{R}^{(v)}$ for the observed reconstruction residual defined in Section 3, the reconstruction multiplier is updated by $\mathbf{Y}^{(v)} \leftarrow \mathbf{Y}^{(v)} + \beta_Y\mu(\mathbf{R}^{(v)} - \mathbf{E}^{(v)}\mathbf{W}^{(v)})$. The tensor multiplier follows $\mathcal{M} \leftarrow \mathcal{M} + \rho(\mathcal{Z} - \mathcal{G})$. The implementation increases μ up to a fixed cap and adapts ρ using tensor primal–dual residual balancing.

B.5 Stopping Rule

The solver monitors normalized tensor primal and dual residuals, reconstruction change, changes in $\mathbf{Z}$, $\mathbf{P}$, and $\mathbf{E}$, the CLST row-sum error, and the PCG residual. Stopping requires these criteria to remain below their tolerances for several checkpoints after the regularization ramp. Because transports and graphs are refreshed periodically, this procedure is treated as an alternating surrogate solver rather than an optimization method for one unchanged subproblem.

B.6 Consensus Readout

The main paper defines the final weights through transport concentration, effective coverage, and reconstruction error. For observed local rows $\mathcal{O}_v$, the concentration statistic is $p_v = |\mathcal{O}_v|^{-1}\sum_{i\in\mathcal{O}_v}\max_j P_{ij}^{(v)}$. The coverage statistic uses

$$b_j^{(v)} = \sum_{i\in\mathcal{O}_v} P_{ij}^{(v)}, \qquad \bar{b}_j^{(v)} = \frac{b_j^{(v)}}{\sum_\ell b_\ell^{(v)}}, \qquad n_{\mathrm{eff}}^{(v)} = \exp\left(-\sum_{j:\bar{b}_j^{(v)}>0}\bar{b}_j^{(v)}\log\bar{b}_j^{(v)}\right). \tag{23}$$

If the reference weight computed from Eq. (10) is below the prescribed minimum, it is clipped and the remaining mass is redistributed proportionally.