RELATIVE SEMANTIC REPRESENTATION LEARNING FOR PARTIALLY VIEW-ALIGNED CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Partially view-aligned clustering (PVC) aims to cluster multi-view data with partially known cross-view correspondences. However, existing PVC methods still face two limitations: i) view-specific representations may not sufficiently capture the relational semantics required for reliable cross-view correspondence estimation; and ii) a uniform association strategy may underutilize known correspondences and yield unreliable associations for unaligned samples. To address these issues, we propose Relative Semantic Representation Learning for Partially View-Aligned Clustering (RSR-PVC). Specifically, each sample is represented by a normalized distribution of its relations to the cross-view aligned samples, forming a relative semantic representation for cross-view comparison. Subsequently, a hard-to-soft relation-space association mechanism is developed to handle aligned and unaligned samples differently. Known aligned correspondences guide bidirectional matching, while associations for unaligned samples are modeled with soft distributions. Finally, the resulting associations are incorporated into relational consistency learning to enhance cross-view representation consistency. Extensive experiments on eight benchmark datasets against seven state-of-the-art methods demonstrate the effectiveness of RSR-PVC.

1 INTRODUCTION

Multi-view clustering (MVC) aims to uncover the underlying cluster structure by exploiting complementary and consistent information across multiple views (Fang et al., 2023; Liu et al., 2025c). It has been widely applied to computer vision (Wang et al., 2024a), multimedia analysis (Bian et al., 2024), and medical analysis (Kuang et al., 2024). Most conventional MVC methods rely on complete instance-level correspondence across views (Liu et al., 2025a; Li et al., 2026). However, this correspondence is difficult to maintain in real-world data collection (Guo et al., 2026). For example, cameras, LiDAR, and radar in autonomous-driving systems may operate asynchronously, making observations of dynamic scenes difficult to align precisely across modalities (Lin et al., 2025). These practical misalignments motivate Partially View-Aligned Clustering (PVC), which studies multi-view data with only partial cross-view correspondences (Huang et al., 2020; Xin et al., 2025).

Existing PVC studies mainly seek to extract useful cross-view information from the limited observed correspondences. Early studies usually established cross-view associations through learned latent representations. Huang et al. (2020) integrated differentiable matching into representation learning to establish correspondences in the latent space. Yang et al. (2021) relaxed strict instance-level matching by identifying category-level counterparts through noise-robust contrastive learning. Later studies broadened cross-view association by incorporating structural and semantic information. Wang et al. (2024c) introduced cross-view graph contrastive learning to achieve cluster-level alignment, while Qian et al. (2024) employed nonparametric clustering to discover additional semantic relations beyond the observed correspondences. However, the resulting associations are still largely inferred from similarities or relational structures constructed on view-dependent representations. Gao et al. (2026) recently showed that distance-based alignment can be sensitive to view-specific details and noise, which may reduce the reliability of cross-view matching. Therefore, establishing semantically comparable representations across heterogeneous views remains an important challenge.

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

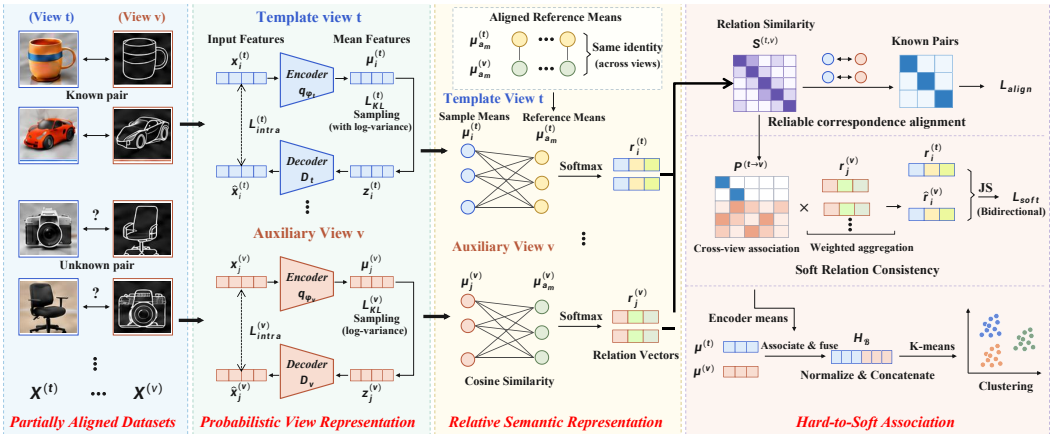


Figure 1: Overview of the proposed RSR-PVC framework. Probabilistic view representation learning captures view-specific latent information; relative semantic representation establishes a comparable relation space across views; hard-to-soft association exploits observed correspondences and models unknown correspondences with soft distributions; the learned associations are finally used for multi-view fusion and K-means clustering.

Beyond representation comparability, recent studies have increasingly examined the reliability of cross-view associations. Yan et al. (2025) introduced a coarse-to-fine re-alignment strategy that rematches unaligned samples according to their learned representation similarities. Zhao et al. (2026b) exploited known reliable alignments to propagate correspondence information under pseudo-label guidance. Xu et al. (2026) further modeled unknown correspondences with adaptive soft matching probabilities for contrastive representation learning. These studies suggest that known and unknown correspondences provide different forms of cross-view information. The former are explicitly observed, whereas the latter need to be inferred from the learned representations. How to effectively utilize both types of correspondence information while considering their different reliability remains an important issue.

To address these challenges, we propose Relative Semantic Representation Learning for Partially View-Aligned Clustering (RSR-PVC). RSR-PVC constructs a reference-relative semantic space from the observed aligned samples, enabling cross-view comparison through shared reference semantics. Based on this space, a hard-to-soft association mechanism exploits observed correspondences through bidirectional alignment and models unknown correspondences with soft association distributions and relational consistency. The learned associations are further used for multi-view representation fusion and clustering. The overall framework of RSR-PVC is illustrated in Figure 1. The main contributions are summarized as follows:

- We propose RSR-PVC, a unified framework for learning reliable cross-view associations from partially observed correspondences. It establishes cross-view comparability in a reference-relative semantic space and supports unified modeling of observed and unknown correspondences.
- We introduce a relative semantic representation that characterizes each sample by its normalized similarities to observed aligned references. The shared reference identities establish semantically comparable relation distributions across views.
- We develop a hard-to-soft relation-space association mechanism. Reliable observed correspondences are exploited through bidirectional alignment, while unknown correspondences are modeled with soft association distributions and cross-view relational consistency.
- Extensive experiments on eight benchmark datasets against seven representative PVC methods demonstrate the effectiveness of RSR-PVC. Ablation studies, robustness evaluations under different unaligned rates, and statistical analyses further validate the proposed framework.

2 Related Work

2.1 Deep Multi-View Clustering

Deep multi-view clustering employs neural networks to learn nonlinear representations and exploit complementary information across different views (Chowdhury et al., 2025). A representative line of research adopts autoencoders to learn view-specific embeddings while preserving cross-view consistency and complementarity (Du et al., 2021; Xu et al., 2021). With the development of self-supervised learning, contrastive learning has been introduced to enhance semantic consistency across views at both feature and semantic levels (Xu et al., 2022; Chen et al., 2023; Trosten et al., 2023). Recent studies further incorporate inter-sample relationships and structural information into representation learning (Fei et al., 2025). Wang & Feng (2024) jointly model contrastive semantics and cross-view structural relations, while Peng et al. (2026) align view-specific relationship matrices with a global relational structure. These studies show an increasing emphasis on semantic and relational consistency in deep MVC. However, most existing deep MVC methods rely on known instance correspondences across views, which limits their direct application to partially aligned data.

2.2 Partially View-Aligned Clustering

Partially view-aligned clustering (PVC) considers multi-view data where only a subset of cross-view sample correspondences is known. Existing methods generally exploit the observed correspondences to establish associations for unaligned samples and learn consistent representations across views (Qian et al., 2024; Wang et al., 2024c; Liu et al., 2025b). Huang et al. (2020) integrated differentiable matching with representation learning to establish cross-view correspondences. Yang et al. (2021) introduced noise-robust contrastive learning to identify category-level counterparts while alleviating false negative pairs. He et al. (2024) further modeled latent representations probabilistically to preserve both cross-view commonality and view-specific information under partial alignment.

Recent methods further exploit structural and semantic cues for cross-view association. Wang et al. (2024b) represented inter-sample relationships with similarity graphs and estimated correspondences by matching graph distributions. Peng et al. (2025) reduced cross-view representation shifts and exploited semantic relations from both aligned and unaligned samples. Gao et al. (2025) jointly modeled one-to-many instance relations and cluster-level associations, while Zhao et al. (2026a) utilized known aligned neighborhoods and confidence estimation to recover reliable correspondences. More recently, Xu et al. (2026) explicitly modeled partial correspondence through adaptive soft matching probabilities for contrastive representation learning. Despite these advances, cross-view association can still be affected by representation discrepancies, while inferred correspondences may vary considerably in reliability. How to establish robust cross-view relations and effectively exploit both observed and inferred correspondence information therefore remains an important issue in PVC. RSR-PVC addresses this issue from a relation-space perspective. Cross-view associations are established in a reference-relative semantic space constructed from observed aligned samples, reducing the dependence on direct similarity comparison between view-specific representations. Furthermore, observed and unknown correspondences are handled according to their reliability through bidirectional alignment and soft association, respectively.

3 Methodology

3.1 Problem Formulation

Let $\mathcal{X} = \{\mathbf{X}^{(v)}\}_{v=1}^{V}$ denote a partially view-aligned multi-view dataset, where $\mathbf{X}^{(v)} \in \mathbb{R}^{N \times d_v}$ represents the v-th view with N samples and feature dimension d_v. Let $\mathcal{A} \subseteq \{1, \ldots, N\}$ denote the aligned subset with $|\mathcal{A}| = N_a$, whose cross-view correspondences are known across all views. The remaining samples form the unaligned subset $\mathcal{U}$ with $|\mathcal{U}| = N_u$, where $N = N_a + N_u$. For $i \in \mathcal{A}$, samples with the same index across different views correspond to the same underlying object, while the cross-view correspondences of samples in $\mathcal{U}$ are unknown. The objective is to exploit information from all views and partition the N samples into K clusters.

3.2 Probabilistic View Representation Learning

To characterize the latent distribution of each view, view-specific encoders are used to learn probabilistic latent representations (He et al., 2024). A probabilistic representation captures both the latent feature and its associated uncertainty, providing a flexible characterization of view-specific observations. Specifically, for a sample $\mathbf{x}_i^{(v)}$ from the v-th view, its latent representation is modeled by a Gaussian distribution:

$$q_{\phi_v}(\mathbf{z}_i^{(v)} \mid \mathbf{x}_i^{(v)}) = \mathcal{N}\Big(\boldsymbol{\mu}_i^{(v)}, \mathrm{diag}(\exp(\boldsymbol{\ell}_i^{(v)}))\Big), \tag{1}$$

where d is the latent dimension, and $\boldsymbol{\mu}_i^{(v)}, \boldsymbol{\ell}_i^{(v)} \in \mathbb{R}^d$ denote the mean and log-variance, respectively. Following the reparameterization strategy (Kingma & Welling, 2014), the latent representation is sampled as $\mathbf{z}_i^{(v)} = \boldsymbol{\mu}_i^{(v)} + \exp(\boldsymbol{\ell}_i^{(v)}/2) \odot \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The sampled latent representation is reconstructed by the corresponding decoder D_v as $\hat{\mathbf{x}}_i^{(v)} = D_v(\mathbf{z}_i^{(v)})$.

To retain the intrinsic information of each view, an intra-view contrastive objective is imposed between the reconstructed and original samples:

$$\mathcal{L}_{\mathrm{intra}}^{(v)} = -\frac{1}{B}\sum_{i=1}^{B} \log \frac{\exp(s(\hat{\mathbf{x}}_i^{(v)}, \mathbf{x}_i^{(v)})/\tau_{\mathrm{intra}})}{\sum_{j=1}^{B} \exp(s(\hat{\mathbf{x}}_i^{(v)}, \mathbf{x}_j^{(v)})/\tau_{\mathrm{intra}})}, \tag{2}$$

where B is the batch size, $s(\cdot,\cdot)$ denotes cosine similarity, and τ_{intra} is the temperature parameter. The intra-view contrastive loss preserves view-specific information and enhances sample discrimination within each view. Meanwhile, the approximate posterior is constrained by the standard Gaussian prior $p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I})$ through the KL divergence:

$$\mathcal{L}_{\mathrm{KL}}^{(v)} = \frac{1}{2Bd}\sum_{i=1}^{B}\sum_{q=1}^{d}\Big[(\mu_{iq}^{(v)})^2 + \exp(\ell_{iq}^{(v)}) - \ell_{iq}^{(v)} - 1\Big]. \tag{3}$$

This term regularizes the latent distribution by encouraging the approximate posterior to remain close to the standard Gaussian prior. The two objectives are averaged over all views to obtain $\mathcal{L}_{\mathrm{intra}}$ and $\mathcal{L}_{\mathrm{KL}}$.

3.3 Relative Semantic Representation with Aligned References

Cross-view association in PVC requires comparing samples across different views. However, representation discrepancies across views can make direct comparison of their latent features unreliable. To address this issue, the aligned samples with known cross-view correspondences are used as semantic references to construct relative semantic representations. For a mini-batch $\mathcal{B}$, denote the indices of its M aligned samples by $\mathcal{A}_{\mathcal{B}} = \{a_1, \ldots, a_M\}$. The reference indexed by a_m corresponds to the same underlying object across views, providing consistent reference semantics. The sample-reference relation is measured using cosine similarity:

$$c_{im}^{(v)} = \frac{(\boldsymbol{\mu}_i^{(v)})^\top \boldsymbol{\mu}_{a_m}^{(v)}}{\|\boldsymbol{\mu}_i^{(v)}\|_2 \|\boldsymbol{\mu}_{a_m}^{(v)}\|_2}, \tag{4}$$

where $c_{im}^{(v)}$ denotes the cosine similarity between the i-th sample and reference a_m in the v-th view. Both representations are taken from the same view, avoiding direct similarity computation across heterogeneous latent spaces. The sample-reference similarities are then normalized over the aligned references:

$$r_{im}^{(v)} = \frac{\exp(c_{im}^{(v)}/\tau_{\mathrm{ref}})}{\sum_{n=1}^{M} \exp(c_{in}^{(v)}/\tau_{\mathrm{ref}})}, \tag{5}$$

where τ_{ref} controls the sharpness of the distribution over the aligned references, and $r_{im}^{(v)}$ denotes the normalized relation between sample i and reference a_m. The resulting vector $\mathbf{r}_i^{(v)} = [r_{i1}^{(v)}, \ldots, r_{iM}^{(v)}] \in \mathbb{R}^M$ is the relative semantic representation of sample i in the v-th view. Its m-th element refers to the same aligned identity a_m across views, providing consistent reference semantics for cross-view comparison.

3.4 HARD-TO-SOFT RELATION-SPACE ASSOCIATION

The relative semantic representations enable samples from different views to be compared with respect to the same aligned references. However, cross-view correspondences are available only for the aligned samples, while the associations of unaligned samples remain unknown. Using the same association strategy for both types of samples may overlook reliable aligned correspondences and introduce unreliable associations for unaligned samples. To address this issue, a hard-to-soft relation-space association mechanism is developed. The t-th view is used as the template, and each remaining view $v \neq t$ serves as an auxiliary view. Both $\mathbf{r}_i^{(t)}$ and $\mathbf{r}_j^{(v)}$ are normalized relation distributions over the same set of aligned references, making them directly comparable in the relation space. Their similarity is measured using the Bhattacharyya coefficient (Pándy et al., 2022):

$$S_{ij}^{(t,v)} = \sum_{m=1}^{M} \sqrt{r_{im}^{(t)} r_{jm}^{(v)}}, \tag{6}$$

where $S_{ij}^{(t,v)}$ denotes the relation-space similarity between sample i in the template view and sample j in the v-th view, and $\mathbf{S}^{(t,v)} \in \mathbb{R}^{B\times B}$ collects all pairwise similarities. A larger value indicates more similar relation patterns over the aligned references.

Reliable Correspondence Alignment. The aligned samples provide reliable cross-view correspondences, which can be used to guide relation-space similarity learning. For each aligned sample, its known counterpart is expected to exhibit higher similarity than the remaining aligned candidates. To exploit this correspondence information, a bidirectional matching objective is introduced:

$$\mathcal{L}_{\text{align}}^{(t,v)} = -\frac{1}{2M} \sum_{m=1}^{M} \left[\log \frac{\exp(S_{a_m a_m}^{(t,v)}/\tau_{\text{match}})}{\sum_{n=1}^{M} \exp(S_{a_m a_n}^{(t,v)}/\tau_{\text{match}})} + \log \frac{\exp(S_{a_m a_m}^{(t,v)}/\tau_{\text{match}})}{\sum_{n=1}^{M} \exp(S_{a_n a_m}^{(t,v)}/\tau_{\text{match}})} \right], \tag{7}$$

where τ_{match} is the matching temperature. The first term encourages the known counterpart to have higher similarity than other aligned candidates in the template-to-auxiliary direction, while the second term applies the same constraint in the reverse direction. This bidirectional objective encourages the relation-space similarities to preserve the known cross-view correspondences.

Soft Correspondence Propagation. For unaligned samples, the true cross-view correspondences are unavailable, and several candidates may exhibit similar relative semantic patterns. Directly assigning a single counterpart can therefore introduce unreliable matches. To address this issue, the similarities in $\mathbf{S}^{(t,v)}$ are converted into soft association distributions:

$$\bar{P}_{ij}^{t\to v} = \frac{\exp(S_{ij}^{(t,v)}/\tau_{\text{match}})}{\sum_{k=1}^{B} \exp(S_{ik}^{(t,v)}/\tau_{\text{match}})}, \tag{8}$$

where $\bar{P}_{ij}^{t\to v}$ denotes the soft association weight from sample i in the template view t to sample j in the auxiliary view v. For aligned samples, the known same-index correspondences are retained as one-hot assignments, while the learned soft distributions are preserved for unaligned samples:

$$P_{ij}^{t\to v} = \begin{cases} 1, & i \in \mathcal{A}_{\mathcal{B}},\ j = i, \\ 0, & i \in \mathcal{A}_{\mathcal{B}},\ j \neq i, \\ \bar{P}_{ij}^{t\to v}, & i \notin \mathcal{A}_{\mathcal{B}}. \end{cases} \tag{9}$$

The reverse association $\mathbf{P}^{v\to t}$ is constructed in the same manner. The resulting association distribution is further used to aggregate relative semantic representations from the auxiliary view into an associated counterpart:

$$\hat{\mathbf{r}}_i^{(v)} = \sum_{j=1}^{B} \text{sg}(P_{ij}^{t\to v}) \mathbf{r}_j^{(v)}, \tag{10}$$

where $\text{sg}(\cdot)$ denotes the stop-gradient operation. The correspondence weights are detached when optimizing the consistency objective, preventing gradients from the consistency loss from directly updating the correspondence estimation.

Algorithm 1 Training Procedure of RSR-PVC

Input: Multi-view data $\mathcal{X} = \{\mathbf{X}^{(v)}\}_{v=1}^{V}$, aligned subset $\mathcal{A}$, cluster number K, and training epochs E.
Output: Cluster assignments $\mathbf{Y}$.

```
Initialize the probabilistic encoders and decoders.
for e = 1 to E do
    Encode and reconstruct all views, and compute L_intra and L_KL by Eqs. (2) and (3).
    Construct the relative semantic representations {r^(v)}_{v=1}^V by Eq. (5).
    for each auxiliary view v ≠ t do
        Compute S^(t,v) by Eq. (6), and construct P^{t→v} and P^{v→t} by Eq. (9).
        Compute L_align^(t,v) and L_soft^(t,v) by Eqs. (7) and (11).
    end for
    Update the network parameters by minimizing Eq. (12).
end for
Fuse the learned multi-view representations by Eq. (13).
Apply K-means to obtain Y.
```

To propagate the cross-view association information into representation learning, each relative semantic representation is encouraged to agree with the counterpart obtained from the other view. Both representations are probability distributions over the same aligned references, and their discrepancy is measured by the Jensen–Shannon divergence:

$$\mathcal{L}_{\text{soft}}^{(t,v)} = \frac{1}{2}\left[\frac{1}{B}\sum_{i=1}^{B} D_{\text{JS}}\left(\mathbf{r}_i^{(t)} \| \hat{\mathbf{r}}_i^{(v)}\right) + \frac{1}{B}\sum_{j=1}^{B} D_{\text{JS}}\left(\mathbf{r}_j^{(v)} \| \hat{\mathbf{r}}_j^{(t)}\right)\right], \tag{11}$$

where $D_{\text{JS}}(\cdot\|\cdot)$ denotes the Jensen–Shannon divergence. The bidirectional objective encourages each sample to maintain a consistent relative semantic pattern with its associated counterpart across views.

3.5 Overall Objective and Clustering

By integrating the objectives introduced above, the overall objective of the proposed method is formulated as:

$$\mathcal{L} = \mathcal{L}_{\text{intra}} + \beta_{\text{kl}}\mathcal{L}_{\text{KL}} + \lambda_{\text{align}}\mathcal{L}_{\text{align}} + \lambda_{\text{soft}}\mathcal{L}_{\text{soft}}, \tag{12}$$

where $\mathcal{L}_{\text{intra}}$ preserves the intrinsic information of each view, while $\mathcal{L}_{\text{KL}}$ regularizes the probabilistic latent distributions. The two cross-view objectives, $\mathcal{L}_{\text{align}}$ and $\mathcal{L}_{\text{soft}}$ are averaged over the corresponding objectives between the template view and each auxiliary view. β_{kl}, λ_{align}, and λ_{soft} are the corresponding trade-off coefficients.

Let $\mathbf{M}^{(v)} = [\boldsymbol{\mu}_1^{(v)}, \ldots, \boldsymbol{\mu}_B^{(v)}]^\top \in \mathbb{R}^{B\times d}$ denote the latent mean matrix of view v. For each auxiliary view, $\widehat{\mathbf{M}}^{(v)} = \mathbf{P}^{t\to v}\mathbf{M}^{(v)}$ is obtained according to the learned cross-view associations. The resulting representations are fused as

$$\mathbf{H}_{\mathcal{B}} = \text{Concat}\left(\text{Norm}(\mathbf{M}^{(t)}), \left\{\text{Norm}(\widehat{\mathbf{M}}^{(v)})\right\}_{v\neq t}\right), \tag{13}$$

where $\text{Norm}(\cdot)$ denotes row-wise ℓ_2 normalization. The fused representations are concatenated over all mini-batches, and K-means is applied to obtain the final clustering assignments. A detailed computational complexity analysis is provided in Appendix A. The overall training procedure of RSR-PVC is summarized in Algorithm 1.

4 Experiments

The experiments are organized around the following research questions: **Q1:** How does RSR-PVC compare with representative PVC methods? **Q2:** How does each core component affect the clustering performance? **Q3:** How robust is RSR-PVC under different unaligned rates? Due to space

constraints, additional analyses are provided in the **Appendix: Q4:** Are the performance improvements statistically consistent across datasets? **Q5:** How sensitive is RSR-PVC to the hyperparameter settings? **Q6:** Do the learned representations exhibit clear and discriminative cluster structures? **Q7:** Does the optimization exhibit stable convergence? **Q8:** How does the running time compare with the competing methods?

4.1 Datasets and Baselines

Experiments are conducted on eight multi-view benchmark datasets: Wikipedia (Rasiwasia et al., 2010), WebKB (Wen et al., 2024), Handwritten (Wen et al., 2024), UCI3view (Xia et al., 2014), MNIST-USPS (Peng et al., 2019), Hdigit (Wen et al., 2024), ReutersEF (Huang et al., 2019), and Cifar10 (Zhang et al., 2018). Detailed dataset statistics are provided in Appendix B. To construct the partially view-aligned setting, N_a samples retain their known cross-view correspondences, while the remaining $N - N_a$ samples are randomly shuffled across views. The unaligned rate is defined as $\rho = (N - N_a)/N$.

We compare RSR-PVC against seven representative methods for partially view-aligned clustering, including VITAL (He et al., 2024), NC3L (Qian et al., 2024), CANDY (Guo et al., 2024), AF-UMC (Sun et al., 2025), DW-InfoNCE (Chen et al., 2026), KNNDA (Zhao et al., 2026a), and BIN (Deng et al., 2026).

Implementation Details. The proposed method is implemented in Python 3.10 with PyTorch. All experiments are conducted on an Ubuntu system equipped with an NVIDIA GB10 GPU and 128 GB unified system memory. We employ three widely used clustering metrics for evaluation, including clustering accuracy (ACC), normalized mutual information (NMI), and adjusted rand index (ARI). All results are reported as percentages (%), with higher values indicating better clustering performance.

For each view, a probabilistic autoencoder with three 2000-dimensional hidden layers is employed. The latent dimension d is selected from $\{10, 30, 50, 100\}$ according to the dataset. The model is trained for 200 epochs using the Adam optimizer (Kingma & Ba, 2014), with a learning rate of 1×10^{-4} and a weight decay of 0.01. The batch size is adjusted according to the dataset scale, and the first view is used as the template view. The temperature parameters are set to $\tau_{\text{intra}} = 0.4$, $\tau_{\text{ref}} = 0.3$, and $\tau_{\text{match}} = 0.2$. The trade-off coefficients are selected from $\beta_{\text{kl}} \in \{0.01, 0.03, 0.1, 0.3, 1\}$, $\lambda_{\text{align}} \in \{0.02, 0.04, 0.08, 0.16\}$, and $\lambda_{\text{soft}} \in \{0.04, 0.08, 0.16, 0.32, 0.64\}$. After training, K-means is applied to the final fused representation to obtain the clustering results.

4.2 Experimental Results and Analysis (Q1)

Tables 1 and 2 report the clustering results under unaligned rates of 0.4 and 0.8. Overall, RSR-PVC achieves strong performance across the eight benchmark datasets and three evaluation metrics. At an unaligned rate of 0.4, RSR-PVC obtains the best results on all datasets. In particular, its ACC exceeds the second-best method by 12.84% on WebKB and 21.82% on ReutersEF. The relatively large gap between ACC and NMI on WebKB is associated with its highly imbalanced cluster distribution, which affects the two metrics differently. Despite this imbalance, RSR-PVC still achieves clear improvements over the baseline methods on WebKB.

For the more challenging setting with an unaligned rate of 0.8, RSR-PVC continues to achieve the best or competitive performance across all benchmark datasets. Clear advantages are retained on WebKB, Handwritten, UCI3view, and ReutersEF. For example, its ACC exceeds the second-best result by 25.69% on WebKB and 8.21% on ReutersEF. Even with substantially fewer observed correspondences, RSR-PVC maintains strong overall clustering performance.

4.3 Ablation Study (Q2)

To examine the contribution of the main components in RSR-PVC, four variants are constructed. **Direct Feature Matching** replaces the relative semantic representation with direct cross-view matching of latent mean representations. **w/o Reliable Alignment** removes $\mathcal{L}_{\text{align}}$. **w/o Soft Propagation** removes $\mathcal{L}_{\text{soft}}$. **Hard Association** replaces the soft association distribution of each unaligned sample with a one-hot assignment to its most similar candidate.

Table 1: Clustering performance (%) on eight benchmark datasets under an unaligned rate of 0.4. The best results are highlighted in bold, and the second-best results are underlined.

Method	Wikipedia			WebKB			Handwritten			UCI3view		
	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
VITAL	41.70	27.17	20.08	51.53	11.78	-4.70	90.38	84.90	83.14	84.82	81.93	76.86
NC3L	43.43	27.77	21.58	80.02	25.09	29.86	83.66	80.28	71.67	82.79	82.58	74.59
CANDY	22.80	7.49	2.87	69.86	0.64	-3.73	85.26	76.59	71.24	78.83	72.31	64.38
AF-UMC	21.21	8.02	2.93	72.55	5.85	11.62	67.53	64.80	53.10	62.40	61.31	48.30
DW-InfoNCE	38.96	23.04	15.60	51.93	10.55	-4.08	83.41	79.16	74.03	84.52	81.51	76.62
KNNDA	39.25	22.65	14.83	59.42	3.98	4.05	87.12	77.36	74.14	83.92	69.28	67.84
BIN	36.80	20.65	13.92	51.95	4.48	-4.24	89.70	81.36	78.93	87.80	79.03	75.58
Ours	**45.17**	**32.39**	**24.46**	**92.86**	**54.72**	**70.85**	**96.80**	**93.26**	**92.99**	**93.55**	**86.50**	**86.12**
Method	**MNIST-USPS**			**Hdigit**			**ReutersEF**			**Cifar10**		
	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
VITAL	95.51	89.26	90.34	97.41	92.82	94.33	45.61	27.89	17.46	85.73	73.40	72.24
NC3L	96.33	90.84	92.03	97.56	93.26	94.66	47.56	29.41	19.61	32.23	46.59	17.23
CANDY	41.41	33.67	23.32	56.49	47.55	33.54	30.93	7.95	3.07	63.86	55.87	38.51
AF-UMC	47.72	41.45	28.33	39.43	33.12	19.91	37.51	14.28	11.14	83.97	70.05	67.87
DW-InfoNCE	96.43	90.87	92.23	97.44	93.00	94.39	33.80	16.64	3.85	94.26	86.76	87.86
KNNDA	44.43	32.06	22.75	61.72	52.68	44.44	48.19	29.78	24.93	95.30	88.78	89.98
BIN	75.82	54.03	53.58	83.10	65.34	66.08	46.04	19.74	17.32	84.01	66.68	67.84
Ours	**96.92**	**92.31**	**93.30**	**98.29**	**95.19**	**96.24**	**70.01**	**49.19**	**47.06**	**95.88**	**90.18**	**91.19**

Table 2: Clustering performance (%) on eight benchmark datasets under an unaligned rate of 0.8. The best results are highlighted in bold, and the second-best results are underlined.

Method	Wikipedia			WebKB			Handwritten			UCI3view		
	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
VITAL	25.25	10.66	5.30	50.52	8.38	-4.75	90.63	81.21	80.35	81.98	75.26	70.56
NC3L	25.69	8.60	4.69	53.76	3.41	-4.45	83.00	74.55	66.30	85.45	79.88	68.66
CANDY	19.77	7.51	2.20	67.55	0.32	-2.84	65.45	59.21	48.76	54.00	45.71	33.43
AF-UMC	21.50	7.29	2.76	60.32	0.11	1.06	71.55	69.41	59.36	73.85	71.40	62.92
DW-InfoNCE	26.12	9.47	4.34	51.47	6.10	-4.85	84.90	78.62	72.94	83.75	74.75	69.05
KNNDA	20.92	5.00	1.69	57.09	0.35	-1.76	55.60	52.83	40.42	71.05	57.08	50.23
BIN	20.49	5.46	1.77	58.04	0.34	-1.89	82.40	71.08	65.73	77.70	69.59	62.82
Ours	**28.86**	**10.91**	**6.48**	**93.24**	**55.83**	**72.09**	**94.45**	**88.24**	**88.05**	**91.80**	**82.93**	**82.74**
Method	**MNIST-USPS**			**Hdigit**			**ReutersEF**			**Cifar10**		
	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
VITAL	83.73	70.09	68.06	90.84	79.68	80.85	55.06	34.63	27.11	90.75	79.66	80.76
NC3L	90.06	**80.00**	79.51	93.17	84.10	85.51	38.88	29.59	12.03	92.13	82.72	83.60
CANDY	26.02	21.40	10.50	30.76	24.69	10.60	27.71	3.55	3.74	79.31	65.86	59.65
AF-UMC	44.42	42.75	26.87	44.10	43.99	28.34	39.01	17.80	12.22	79.65	63.29	60.04
DW-InfoNCE	85.46	72.51	70.65	91.94	81.51	83.01	35.28	20.97	5.30	91.96	82.33	83.24
KNNDA	26.10	17.43	10.10	44.96	35.75	23.16	39.32	23.30	16.78	92.23	82.87	83.75
BIN	47.12	24.96	19.39	50.33	31.61	26.39	30.65	7.96	6.64	65.17	40.98	38.90
Ours	**90.40**	79.86	**80.03**	**94.93**	**87.76**	**89.11**	**63.27**	**44.41**	**38.44**	**92.36**	**83.44**	**84.06**

As shown in Table 3, the full model consistently achieves the best performance on all three datasets. Direct feature matching causes clear performance degradation, particularly on WebKB, demonstrat-

Table 3: Ablation study on WebKB, UCI3view, and MNIST-USPS. The best results are highlighted in bold, and the second-best results are underlined.

Variant	WebKB			UCI3view			MNIST-USPS		
	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
Direct Feature Matching	73.17	4.36	13.22	84.10	76.34	71.26	93.80	86.13	86.79
w/o Reliable Alignment	<u>88.68</u>	<u>45.12</u>	<u>57.58</u>	91.05	82.58	81.29	86.26	76.53	73.65
w/o Soft Propagation	69.74	22.12	15.54	<u>92.60</u>	<u>84.92</u>	<u>84.28</u>	95.00	88.94	89.32
Hard Association	87.92	42.54	55.16	85.30	82.62	78.21	<u>96.54</u>	<u>91.58</u>	<u>92.51</u>
Full	**92.86**	**54.72**	**70.85**	**93.55**	**86.50**	**86.12**	**96.92**	**92.31**	**93.30**

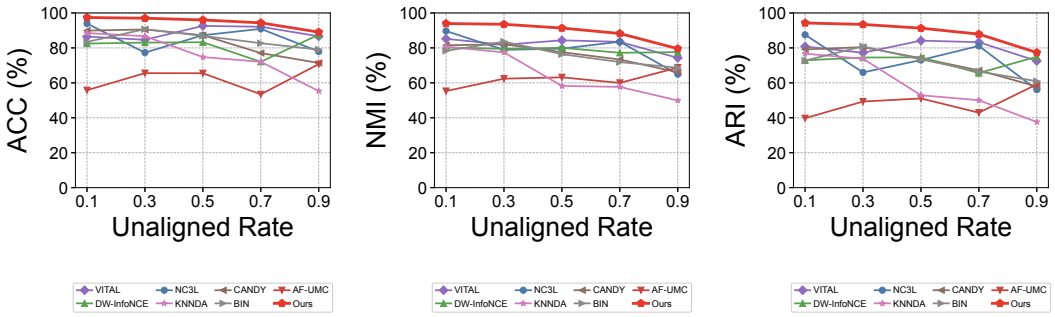


Figure 2: Clustering performance on Handwritten under different unaligned rates.

ing the importance of comparing cross-view samples through their relative relations to aligned references. Removing $\mathcal{L}_{\text{align}}$ also reduces the clustering performance, especially on MNIST-USPS, showing the importance of exploiting observed correspondences for reliable cross-view alignment. For the soft association mechanism, removing $\mathcal{L}_{\text{soft}}$ degrades the performance, indicating the benefit of cross-view relational consistency. Replacing the soft associations with hard assignments further results in inferior performance to the full model, supporting the use of soft associations for unaligned samples. These results collectively validate the relative semantic representation and the hard-to-soft association mechanism in RSR-PVC.

4.4 Effect of Unaligned Rate (Q3)

To investigate the influence of the unaligned rate, we conduct experiments on Handwritten with $\rho \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. As shown in Figure 2, RSR-PVC consistently achieves the best ACC, NMI, and ARI under all five settings. Although the clustering performance gradually decreases as fewer cross-view correspondences are available, RSR-PVC maintains a clear advantage over the compared methods even at $\rho = 0.9$. These results demonstrate that RSR-PVC remains effective under different degrees of partial view alignment.

5 Conclusion

This paper presents Relative Semantic Representation Learning for Partially View-Aligned Clustering (RSR-PVC), which learns cross-view comparable representations and reliable sample associations from partially observed correspondences. RSR-PVC characterizes each sample by its normalized similarities to aligned semantic references for cross-view comparison. Based on the resulting relation space, a hard-to-soft association mechanism exploits observed correspondences through bidirectional alignment and models unaligned samples with soft associations. The inferred associations are further used for relational consistency learning and multi-view representation fusion. Extensive experiments on eight benchmark datasets demonstrate the effectiveness of RSR-PVC under different unaligned rates. Future work will explore more flexible template-view selection strategies and extend the framework to settings with incomplete views and partial correspondences.

AI USE STATEMENT

In this work, we used generative AI tools (specifically ChatGPT) to assist with language editing of the manuscript, including improving readability, grammar, wording, and sentence-level clarity. Generative AI was not used for research ideation, methodology or experimental design, implementation, data processing, or result analysis and interpretation. All scientific content, technical claims, experimental results, and conclusions were developed and verified by the authors. We carefully reviewed and manually checked all AI-assisted edits to ensure that they preserved the original scientific meaning and did not introduce inaccurate, unsupported, or fabricated statements. We take responsibility for the final content of this work, including text, claims, or other materials produced with the assistance of generative AI.

REPRODUCIBILITY STATEMENT

To facilitate the reproducibility of this work, the complete source code, data processing scripts, and detailed execution instructions are publicly available through an anonymous repository: `https://anonymous.4open.science/r/RSR-PVC-75E2/`.

REFERENCES

Jintang Bian, Xiaohua Xie, Jian-Huang Lai, and Feiping Nie. Multi-view contrastive clustering via integrating graph aggregation and confidence enhancement. *Information Fusion*, 108:102393, 2024.

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

Qiuyu Chen, Xihong Yang, Zhibing Dong, Qian Qu, En Zhu, and Xinwang Liu. Dynamic weight-enhanced contrastive learning for partially view-aligned clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 36(8):12426–12438, 2026.

Anal Roy Chowdhury, Avisek Gupta, and Swagatam Das. Deep multi-view clustering: A comprehensive survey of the contemporary techniques. *Information Fusion*, 119:103012, 2025.

Shanghui Deng, Xiao Zheng, Chang Tang, Lianbo Guo, and Xinwang Liu. Game theory inspired cross-view interaction alignment for partially view-aligned clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–15, 2026.

Guowang Du, Lihua Zhou, Yudi Yang, Kevin Lü, and Lizhen Wang. Deep multiple auto-encoder-based multi-view clustering. *Data Science and Engineering*, 6(3):323–338, 2021.

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

Lunke Fei, Junlin He, Qi Zhu, Shuping Zhao, Jie Wen, and Yong Xu. Deep multi-view contrastive clustering via graph structure awareness. *IEEE Transactions on Image Processing*, 34:3805–3816, 2025.

Hang Gao, Cheng Liu, Ying Li, You Zhou, and Wei Du. Multi-level cross-view feature embedding for partial view-aligned clustering. *Knowledge-Based Systems*, pp. 115230, 2025.

Hang Gao, Zuosong Cai, Yuze Li, Cheng Liu, Gaoyang Li, Ying Li, Wei Du, and You Zhou. Learning whom to align with: Progressive anomaly combination detection for partially view-aligned clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 21171–21179, 2026.

Ruiming Guo, Mouxing Yang, Yijie Lin, Xi Peng, and Peng Hu. Robust contrastive multi-view clustering against dual noisy correspondence. *Advances in Neural Information Processing Systems*, 37:121401–121421, 2024.

Wei Guo, Yinghui Sun, Xingfeng Li, Zhenwen Ren, and Quansen Sun. Collaborative anchor graph learning with similarity-guided alignment for multi-view clustering. *IEEE Transactions on Multimedia*, pp. 1–14, 2026. doi: 10.1109/TMM.2026.3704856.

Changhao He, Hongyuan Zhu, Peng Hu, and Xi Peng. Robust variational contrastive learning for partially view-unaligned clustering. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 4167–4176, 2024.

Zhenyu Huang, Joey Tianyi Zhou, Xi Peng, Changqing Zhang, Hongyuan Zhu, and Jiancheng Lv. Multi-view spectral clustering network. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pp. 2563–2569. International Joint Conferences on Artificial Intelligence Organization, 7 2019.

Zhenyu Huang, Peng Hu, Joey Tianyi Zhou, Jiancheng Lv, and Xi Peng. Partially view-aligned clustering. *Advances in Neural Information Processing Systems*, 33:2892–2902, 2020.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014.

Yabin Kuang, Minzhu Xie, Zhanhong Zhao, Dongze Deng, and Ergude Bao. Multi-view contrastive clustering for cancer subtyping using fully and weakly paired multi-omics data. *Methods*, 232: 1–8, 2024.

Zhenglai Li, Jun Wang, Chang Tang, Xinzhong Zhu, Wei Zhang, and Xinwang Liu. Balanced multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026.

Chen Lin, Zhicheng He, Yu Qiu, and Yuanyi Huang. A method of time alignment in bev features for multimodal fusion object detection of intelligent vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 26(8):12432–12445, 2025.

Ruimeng Liu, Xin Zou, Chang Tang, Xiao Zheng, Xingchen Hu, Kun Sun, and Xinwang Liu. Sparsemvc: probing cross-view sparsity variations for multi-view clustering. *Advances in Neural Information Processing Systems*, 38:154594–154622, 2025a.

Wenzhe Liu, Jiongcheng Zhu, Jingbo Tan, Huibing Wang, and Yong Zhang. Partially view-aligned clustering via data recoupling and elastic bi-consistency learning. *Expert Systems with Applications*, pp. 129741, 2025b.

Xinwang Liu, Ke Liang, Jun Wang, Suyuan Liu, Xiangke Wang, and Huaimin Wang. Two decades of multi-view clustering: Taxonomy, application, and challenge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025c.

Liang Peng, Yixuan Ye, Cheng Liu, Hangjun Che, Fei Wang, Zhiwen Yu, Si Wu, and Hau-San Wong. Smart: Semantic matching contrastive learning for partially view-aligned clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025.

Shuangmei Peng, Zhe Chen, Tianyang Xu, and Xiaojun Wu. Relationship alignment for view-aware multi-view clustering. In *International Conference on Learning Representations*, volume 2026, pp. 147335–147347, 2026.

Xi Peng, Zhenyu Huang, Jiancheng Lv, Hongyuan Zhu, and Joey Tianyi Zhou. Comic: Multi-view clustering without parameter selection. In *International Conference on Machine Learning*, pp. 5092–5101. PMLR, 2019.

Michal Pándy, Andrea Agostinelli, Jasper Uijlings, Vittorio Ferrari, and Thomas Mensink. Transferability estimation using bhattacharyya class separability. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9162–9172, 2022.

Shengsheng Qian, Dizhan Xue, Jun Hu, Huaiwen Zhang, and Changsheng Xu. Nonparametric clustering-guided cross-view contrastive learning for partially view-aligned representation learning. *IEEE Transactions on Image Processing*, 33:6158–6172, 2024.

Nikhil Rasiwasia, Jose Costa Pereira, Emanuele Coviello, Gabriel Doyle, Gert RG Lanckriet, Roger Levy, and Nuno Vasconcelos. A new approach to cross-modal multimedia retrieval. In *Proceedings of the 18th ACM international conference on Multimedia*, pp. 251–260, 2010.

Bohang Sun, Yuena Lin, Tao Yang, Zhen Zhu, Zhen Yang, and Gengyu Lyu. AF-UMC: An alignment-free fusion framework for unaligned multi-view clustering. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Daniel J Trosten, Sigurd Løkse, Robert Jenssen, and Michael C Kampffmeyer. On the effects of self-supervision and contrastive alignment in deep multi-view clustering. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23976–23985. IEEE, 2023.

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

Jing Wang and Songhe Feng. Contrastive and view-interaction structure learning for multi-view clustering. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, pp. 5055–5063, 8 2024.

Xibiao Wang, Hang Gao, Xindian Wei, Liang Peng, Rui Li, Cheng Liu, Si Wu, and Hau-San Wong. Contrastive graph distribution alignment for partially view-aligned clustering. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 5240–5249, 2024b.

Yiming Wang, Dongxia Chang, Zhiqiang Fu, Jie Wen, and Yao Zhao. Partially view-aligned representation learning via cross-view graph contrastive network. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(8):7272–7283, 2024c.

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, 2024.

Rongkai Xia, Yan Pan, Lei Du, and Jian Yin. Robust multi-view spectral clustering via low-rank and sparse decomposition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28, pp. 2149–2155, 2014.

Like Xin, Wanqi Yang, Lei Wang, and Ming Yang. Multilevel reliable guidance for unpaired multiview clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 36(10):18968–18982, 2025.

Gehui Xu, Chengliang Liu, Yabo Liu, Yong Xu, and Jie Wen. Partial contrastive learning for partially view-aligned multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(10):12068–12081, 2026.

Jie Xu, Yazhou Ren, Guofeng Li, Lili Pan, Ce Zhu, and Zenglin Xu. Deep embedded multi-view clustering with collaborative training. *Information Sciences*, 573:279–290, 2021.

Jie Xu, Huayi Tang, Yazhou Ren, Liang Peng, Xiaofeng Zhu, and Lifang He. Multi-level feature learning for contrastive multi-view clustering. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16030–16039. IEEE, 2022.

Wenbiao Yan, Jihua Zhu, Jinqian Chen, Haozhe Cheng, Shunshun Bai, Liang Duan, and Qinghai Zheng. Partially multi-view clustering via re-alignment. *Neural Networks*, 182:106884, 2025.

Mouxing Yang, Yunfan Li, Zhenyu Huang, Zitao Liu, Peng Hu, and Xi Peng. Partially view-aligned representation learning with noise-robust contrastive loss. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1134–1143, 2021.

Zheng Zhang, Li Liu, Fumin Shen, Heng Tao Shen, and Ling Shao. Binary multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7):1774–1782, 2018.

Liang Zhao, Tianqi Yue, Shubin Ma, Ziyue Wang, ZhiYuan Liu, and Bo Xu. Knnda: A new perspective of alignment recovery for partially view-aligned clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 28706–28714, 2026a.

Liang Zhao, Tianqi Yue, Shubin Ma, Ziyue Wang, and Lin Yuanbo Wu. Pseudo-label guided alignment propagation for partially view-aligned clustering. *IEEE Transactions on Multimedia*, pp. 1–12, 2026b. doi: 10.1109/TMM.2026.3727807.

APPENDIX

A COMPUTATIONAL COMPLEXITY ANALYSIS

Let N, V, d_v, h, d, B, M, and E denote the numbers of samples, views, the input dimension of the v-th view, hidden dimension, latent dimension, batch size, aligned references in a mini-batch, and training epochs, respectively. In each epoch, probabilistic representation learning requires $\mathcal{O}(Nh\sum_{v=1}^{V} d_v + VNhd + NB\sum_{v=1}^{V} d_v)$. Constructing the relative semantic representations costs $\mathcal{O}(VNMd)$, while the hard-to-soft association requires $\mathcal{O}((V-1)NBM)$. Accordingly, the overall training complexity can be written as $\mathcal{O}\Big(EN[(h+B)\sum_{v=1}^{V} d_v + Vhd + VM(d+B)]\Big)$. The space complexity is mainly determined by the batch-wise relation and association matrices. Sequential processing of the auxiliary views requires a space complexity of $\mathcal{O}(B^2)$.

B DATASET STATISTICS

Table 4 summarizes the number of samples, views, clusters, and feature dimensions of the eight benchmark datasets.

Table 4: Statistics of the benchmark datasets.

Dataset	Samples	Views	Clusters	Dimensions
Wikipedia	693	2	10	128/10
WebKB	1,051	2	2	2949/334
Handwritten	2,000	6	10	216/76/64/6/240/47
UCI3view	2,000	3	10	240/76/6
MNIST-USPS	5,000	2	10	784/256
Hdigit	10,000	2	10	784/256
ReutersEF	18,758	2	6	100/100
Cifar10	50,000	3	10	512/2048/1024

C STATISTICAL VALIDATION OF THE MAIN RESULTS (Q4)

C.1 CROSS-DATASET RADAR-CHART ANALYSIS

Figure 3 provides a unified comparison of the clustering performance across eight datasets under $\rho \in \{0.4, 0.8\}$. Each radar chart corresponds to one evaluation metric. RSR-PVC forms the outer performance profile on nearly all axes under both unaligned rates, with particularly clear advantages on several challenging datasets such as WebKB and ReutersEF. The consistently broad radar profiles demonstrate the strong cross-dataset performance of RSR-PVC under different degrees of partial view alignment.

C.2 PAIRWISE STATISTICAL SIGNIFICANCE

To further evaluate the statistical reliability of the performance improvements, one-sided exact Wilcoxon signed-rank tests are conducted between RSR-PVC and each baseline, using the eight datasets as paired observations for each metric and unaligned rate. As shown in Table 5, all comparisons reach the $p < 0.01$ significance level, and nearly all comparisons further achieve $p < 0.005$. These results provide statistical evidence that the performance advantages of RSR-PVC are consistent across different datasets and unaligned rates.

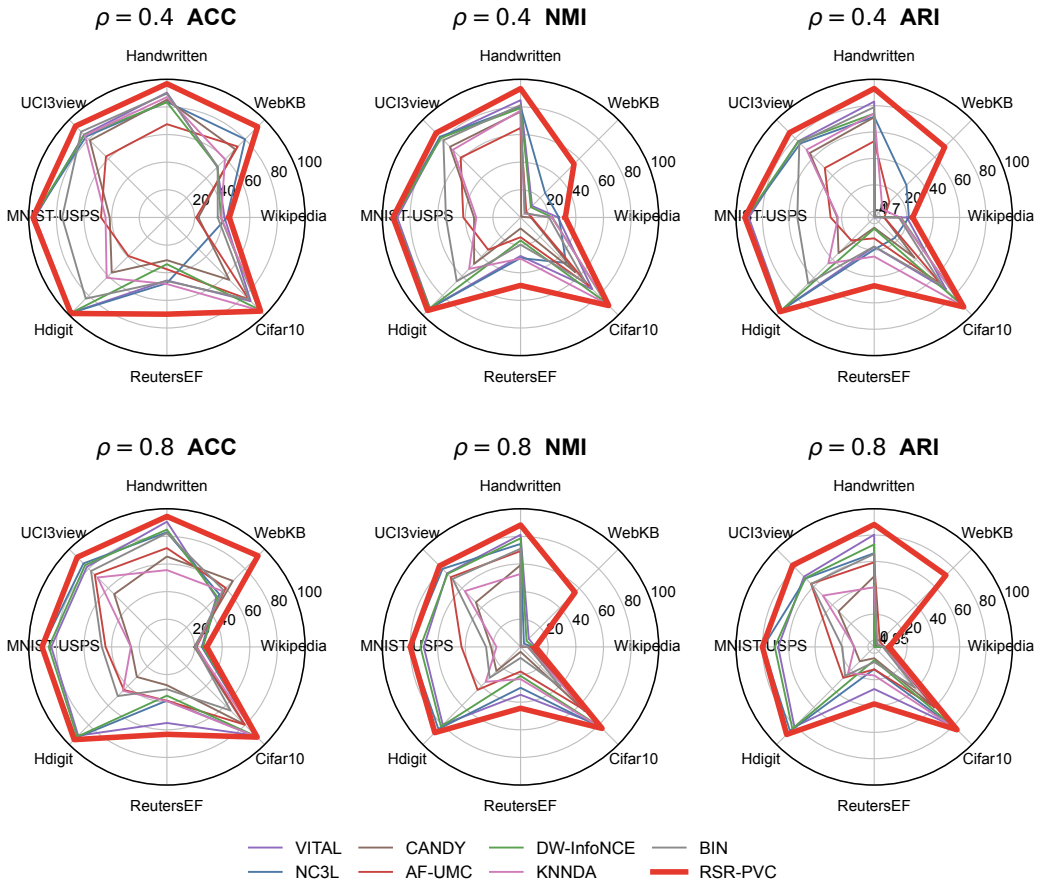


Figure 3: Cross-dataset radar-chart comparison of ACC, NMI, and ARI under two unaligned rates.

Table 5: One-sided exact Wilcoxon signed-rank tests comparing RSR-PVC with each baseline across eight datasets. Significance levels are denoted by $\circ$ for $p < 0.05$, $\triangle$ for $p < 0.01$, and $\star$ for $p < 0.005$.

Method	$\rho = 0.4$			$\rho = 0.8$		
	ACC	NMI	ARI	ACC	NMI	ARI
VITAL	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
NC3L	$\star$	$\star$	$\star$	$\star$	$\triangle$	$\star$
CANDY	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
AF-UMC	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
DW-InfoNCE	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
KNNDA	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$
BIN	$\star$	$\star$	$\star$	$\star$	$\star$	$\star$

D Parameter Sensitivity Analysis (Q5)

In this section, the sensitivity of the main trade-off parameters is investigated. First, λ_{align} and λ_{soft} are varied over $\{0.02, 0.04, 0.08, 0.16\}$ and $\{0.04, 0.08, 0.16, 0.32, 0.64\}$, respectively. Figure 4 reports the ACC, NMI, and ARI results on UCI3view. The three metrics show similar trends under different parameter combinations, with relatively small λ_{align} and small-to-moderate λ_{soft} generally

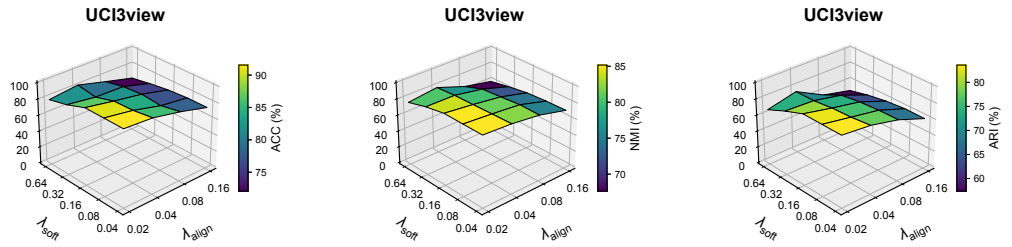


Figure 4: Parameter sensitivity of λ_{align} and λ_{soft} on UCI3view in terms of ACC, NMI, and ARI.

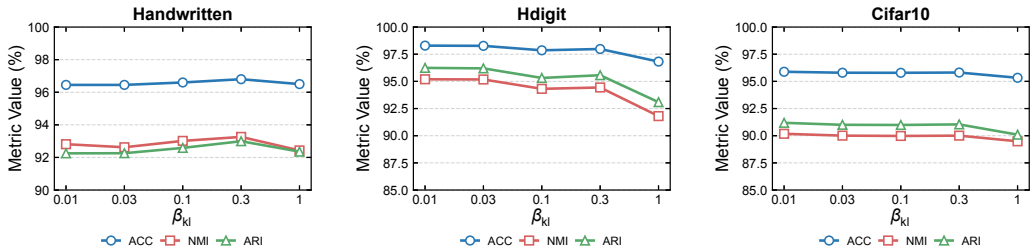


Figure 5: Parameter sensitivity of β_{kl} on Handwritten, Hdigit, and Cifar10.

yielding better performance. Across different datasets, the preferred value of λ_{align} varies, whereas λ_{soft} is mainly concentrated at 0.04 and 0.08.

Additionally, the influence of β_{kl} is evaluated on Handwritten, Hdigit, and Cifar10 over $\{0.01, 0.03, 0.1, 0.3, 1\}$. As shown in Figure 5, the clustering performance remains relatively stable when β_{kl} varies from 0.01 to 0.3. A larger value of 1 causes a moderate decrease on Hdigit and Cifar10, while only small fluctuations are observed on Handwritten. These results show that RSR-PVC is relatively insensitive to β_{kl} within a broad range.

E T-SNE VISUALIZATION ANALYSIS (Q6)

To qualitatively evaluate the discriminative ability of the learned representations, t-SNE is employed to project the high-dimensional features into a two-dimensional space. Figure 6 presents the visualization results of RSR-PVC and five representative deep PVC methods on Handwritten with an unaligned rate of 0.4. Several comparison methods exhibit overlapping clusters or scattered samples between different categories. In contrast, RSR-PVC produces relatively compact clusters with clear inter-cluster separation and less overlap among different clusters. The visualization is consistent with the quantitative results, providing additional qualitative evidence for the discriminative quality of the learned representations.

F OPTIMIZATION BEHAVIOR AND COMPUTATIONAL EFFICIENCY (Q7 & Q8)

F.1 CONVERGENCE ANALYSIS (Q7)

We further investigate the convergence behavior of the proposed method on all eight benchmark datasets. Figure 7 presents the variation of the overall training objective with respect to the number of training epochs. The objective values gradually decrease and become stable during optimization across different datasets, indicating that the proposed method exhibits stable convergence behavior.

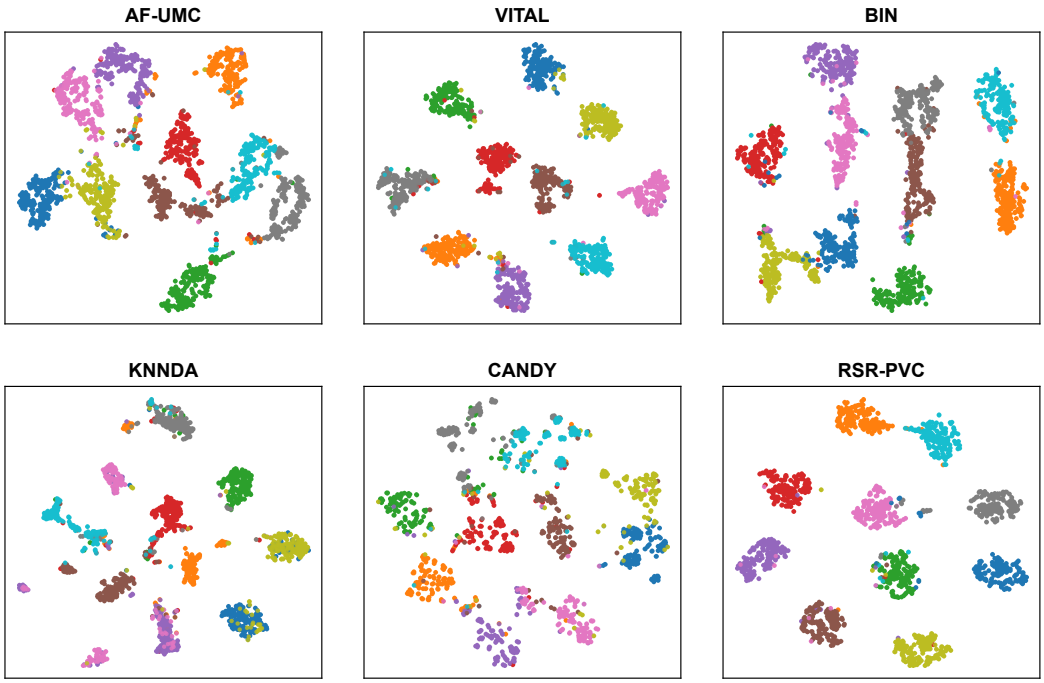


Figure 6: t-SNE visualizations of the learned representations on Handwritten under an unaligned rate of 0.4.

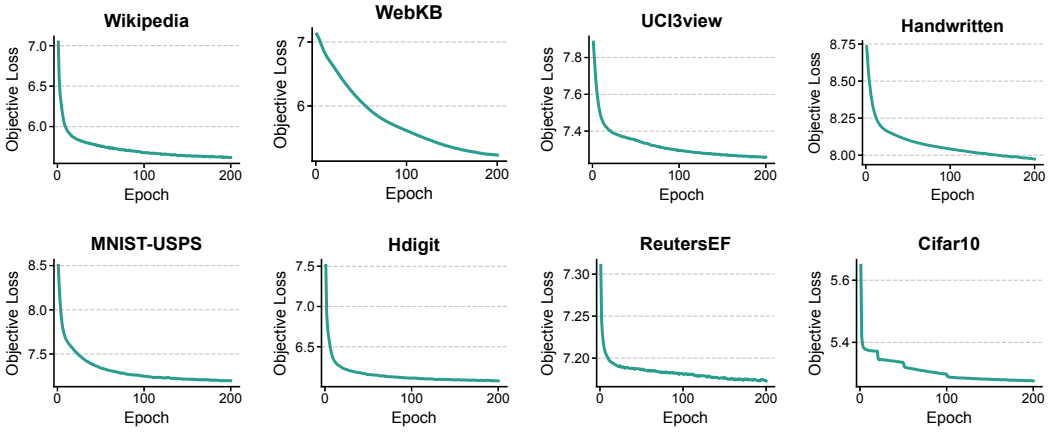


Figure 7: Convergence curves of the overall objective loss on eight benchmark datasets.

F.2 Running Time Analysis (Q8)

Figure 8 presents the running-time comparison of RSR-PVC and the compared methods on eight benchmark datasets, where Time is measured in seconds and reported as $\log_{10}(\text{Time}+1)$. RSR-PVC shows moderate running time on most datasets compared with the other methods. The additional cost mainly arises from the in-batch sample-reference relation modeling and pairwise cross-view association. The running time generally increases on larger datasets, while RSR-PVC maintains strong clustering performance, showing a reasonable trade-off between clustering effectiveness and computational cost.

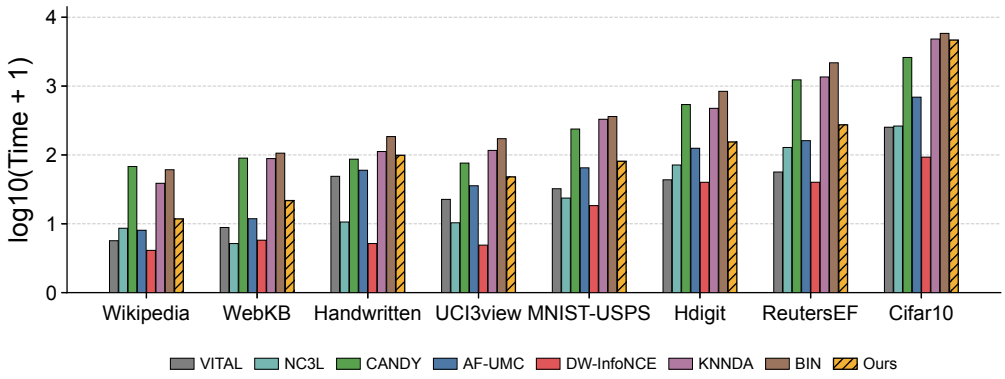


Figure 8: Running time comparison on eight benchmark datasets.