KNNDA: A New Perspective of Alignment Recovery for Partially View-Aligned Clustering

Liang Zhao, Tianqi Yue, Shubin Ma, Ziyue Wang, ZhiYuan Liu, Bo Xu*

Dalian University of Technology, Dalian, Liaoning, China
BoXu@dlut.edu.cn*

Abstract

In multi-view clustering (MVC), complementary and consistent information from multiple views is integrated to improve clustering performance. However, inter-view sample correspondences may be partially missing in practice, making it difficult to learn cross-view consistency, which leads to the partially view-aligned problem (PVP). Most existing partially view-aligned clustering (PVC) methods first learn cross-view consistent representations based on known alignments, and then recover missing correspondences by measuring cross-view similarity between samples. However, such an indirect alignment recovery process depends on high-quality consistent representations and lacks effective utilization of known alignments, often resulting in sub-optimal outcomes. To address this, we propose a novel direct alignment recovery perspective, instantiated as K-Nearest Neighbors Direct Alignment (KNNDA). Specifically, we first construct an alignment domain by mapping the aligned neighbors of each unaligned sample into the aligned view. Then, we compute alignment confidence based on the similarity between known aligned pairs of neighbors. In particular, we use a dynamic threshold to filter out unreliable alignments. Finally, new alignments are generated within the high-confidence alignment domain. Contrastive loss is used to learn consistent representations for clustering. Comprehensive experiments on several real-world datasets show the effectiveness and superiority of our module in partially view-aligned clustering.

Code —
https://github.com/ICEY-Moon7/AAAI-26-KNNDA

Introduction

In real-world scenarios, the widespread adoption of multimodal technologies has led to data being no longer represented in a single form, but stored as multi-view data, composed of multiple distinct but interrelated modalities (Fig. 1a). As an important subfield of multi-view data processing (Trosten et al. 2023; Zhao and Xie 2024; Guo and Ye 2019; Wang et al. 2022a; Xu et al. 2022; Zhao et al. 2018; Zhang et al. 2021; Hu et al. 2025), Multi-View Clustering (MVC) eliminates the need for ground-truth labels, making it widely applicable in many real-world scenarios.

MVC aims to leverage complementary and consistent cross-view information for sample clustering. However, most existing MVC methods assume complete inter-view sample correspondences, which is rarely satisfied in practice. Imperfect data collection or storage errors often result in partially view-aligned data (Fig. 1b), leading to the Partially View-aligned Problem (PVP).

To tackle this issue, several traditional and deep learning-based Partially View-Aligned Clustering (PVC) methods have been proposed (Sun et al. 2024; Trosten et al. 2021; Wang et al. 2024b; Ma et al. 2025). MvCLN (Yang et al. 2021) relaxed alignment recovery to the class level and proposed a noise-robust contrastive loss that reduces the impact of false negative pairs. Undoubtedly, accurate alignment recovery plays a critical role in the success of PVC. However, most existing PVC approaches follow a two-stage paradigm: they first learn cross-view consistent representations based on the known alignments, and then compute sample-wise cross-view similarities to recover the unknown alignments. Such an indirect alignment recovery process heavily depends on high-quality representations. It depends on global similarity rather than local ground-truth correspondences, making it highly susceptible to noise and prone to suboptimal or even erroneous alignment.

Naturally, a new idea arises: Can we directly recover missing alignments based on the known ones? Intuitively, generating new alignments directly from existing ones enhances reliability. This direct recovery process no longer relies on high-quality representations, which allows alignment and representation learning to proceed simultaneously, eliminating the need for pretraining and reducing training costs.

Fortunately, our research demonstrates that this idea is feasible and effective. In this work, we propose a new paradigm for alignment recovery and instantiate it with a module named K-Nearest Neighbors Direct Alignment (KNNDA). Specifically, to restrict the alignment recovery process to local regions, we first construct an alignment domain for each unaligned sample by mapping its aligned neighbors into the corresponding aligned view. Then, we compute the confidence of each alignment domain based on the similarities among known aligned pairs. Domains with confidence below a dynamically adjusted threshold will be discarded. Both alignment confidence and dynamic threshold are designed to improve the reliability of alignment

*Corresponding author.

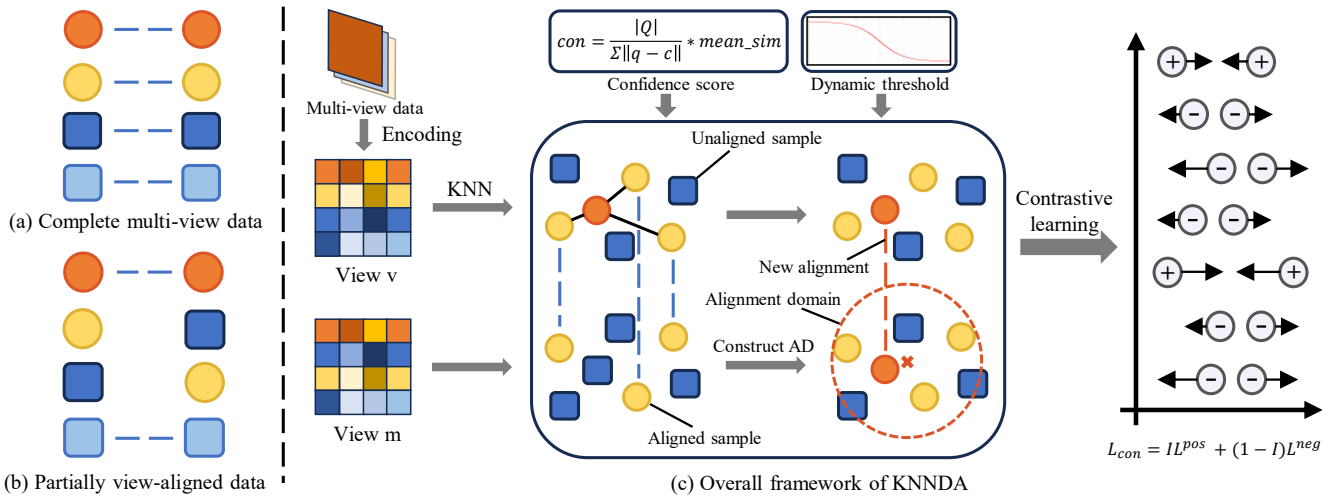


Figure 1: Schematic illustrations of multi-view data and KNNDA. The dashed lines indicate inter-view alignments.

recovery. Finally, the unaligned sample within the high-confidence domain which is most similar to the original unaligned sample is selected to form a new alignment pair.

Once missing alignments are recovered, contrastive learning is applied to all aligned sample pairs to learn consistent representations for clustering. Unlike most prior PVC methods, our alignment recovery module does not require pretraining and is able to perform jointly with representation learning in a unified framework. This enables mutual enhancement between alignment recovery and representation learning, thereby improving the performance of downstream clustering tasks. Experiments on several real-world datasets show that our method outperforms SOTA PVC methods in terms of both alignment recovery and clustering performance. Our main contributions are summarized as follows:

- We propose a novel perspective for alignment recovery: direct alignment recovery, and present a concrete implementation of this idea. Our alignment recovery module requires no pretraining, and can be integrated with representation learning in an end-to-end fashion.
- Unlike the existing PVC paradigm, our alignment recovery module infers new alignments from local reliable correspondences, greatly improving the reliability of alignment recovery. Meanwhile, the scope of cross-view similarity computation is reduced from the dataset to local alignment domains, which reduces computational cost.
- To enhance the reliability of alignment recovery, we introduce an alignment confidence measure and design a dynamically adaptive thresholding mechanism. We further explore and analyze the interaction between alignment recovery and representation learning, and their influence on final clustering and alignment performance.

Related Works

Indirect Partially View-aligned Clustering

In real-world applications, it is common that the alignments between views in multi-view data are partially missing, leading to the problem of partially view-aligned clustering. Although some studies have considered this scenario (Jin et al. 2023; Eaton, desJardins, and Jacob 2010; Lu and Peng 2013), they almost all follow a paradigm that learns cross-view consistent representations and infers alignments based on cross-view similarities among samples. UPMGC-SM (Wen et al. 2024) proposed a framework based on graph reconstruction loss, which recovers unknown alignments through learnable alignment transformation matrices for each view. DGPPVC (Zhao et al. 2025) is the first to employ graph convolutional networks for alignment identification, progressively discovering cross-view alignments by dynamically updating adjacency graphs. NC^{3}L (Qian et al. 2024) further tackled the challenge of mitigating false negative pairs in contrastive learning by proposing a non-parametric, clustering-guided contrastive learning method.

Direct Partially View-aligned Clustering

Unfortunately, almost no studies have come close to the concept of direct alignment recovery. CGDAPVC (Wang et al. 2024a) calculates intra-view similarity graphs using semantic features and aligns distributions across views via Wasserstein distance to align views. However, this alignment recovery method still relies on a global similarity graph, rather than being directly guided by local ground-truth alignments. As the semantic complexity of the dataset increases, it becomes significantly more difficult to learn correct semantic correspondences, which in turn leads to a substantial decline in both alignment and clustering performance. Comparative experiments conducted on large-scale datasets with complex semantics validate the correctness of this hypothesis.

Methodology

In this section, we first provide a problem formulation of PVC. Then, we introduce the proposed Direct Alignment Recovery Paradigm, along with our module named KNNDA, and the overall framework of our model.

Problem Formulation

Given a multi-view dataset $\chi = \{X^1, X^2, \ldots, X^m\}$, where $X^v = \{x_1^v, x_2^v, ..., x_n^v\}\,(v = 1, 2, ..., m)$ represents the raw feature space of the v-th view, m denotes the number of views, and n is the number of samples. Under the assumption of partially-view alignment, only a subset of samples have known correspondences across different views. Thus, the data in each view can be divided into two disjoint subsets $\{A^v, U^v\}_{v=1}^m$, where $\{A^v\}_{v=1}^m$ denote the sets of aligned samples whose correspondences across views are known. Specifically, for any two views v and w, the samples x_i^v and $x_i^w (v \neq w)$ at the same index position are projections of the same instance. In contrast, the subset $\{U^v\}_{v=1}^m$ contains samples whose cross-view correspondences are missing.

The goal of partially view-aligned clustering is to infer the unknown alignments based on the known correspondences, and to assign all n samples into c clusters, where c is the predefined number of categories. To achieve improved clustering performance, we first use view-specific encoders to transform the original feature spaces $\{X^v\}_{v=1}^m$ into latent representations $\{Z^v\}_{v=1}^m$.

Accordingly, the alignment between view v and view w can be formalized as:

$$Z^v \sim P^{vw} Z^w \tag{1}$$

where P^{vw} is the permutation matrix representing the alignments between the two views. $P_{ij}^{vw} = 1$ if the i-th sample in view v is aligned with the j-th sample in view w, otherwise $P_{ij}^{vw} = 0$.

Without loss of generality, we use the two-view scenario as an example in the subsequent subsections. The extension to cases where the number of views is greater than 2 can be naturally achieved by selecting one view as the anchor view and conducting $m - 1$ pairwise alignment processes. The overall architecture of our KNNDA is illustrated in Fig. 1c.

Direct Alignment Recovery Paradigm

Unlike alignment recovery methods that rely on high-quality consistent representations, direct alignment recovery needs no pretraining and instead depends directly on known alignments. This leads to two key challenges: 1) The importance of each known alignment varies in the recovery process. If every generated new alignment is based on all known alignments, the computational cost becomes prohibitively high. 2) The newly generated alignments will be used in subsequent iterations to recover further alignments. Therefore, ensuring their reliability and accuracy is critical, otherwise the model could be misled by incorrect alignments.

To address Challenge 1, we construct a local alignment domain for each unaligned sample using its K-nearest aligned neighbors, thereby restricting alignment recovery to a local region and reducing computational overhead. To address Challenge 2, we introduce an alignment confidence metric and design a dynamically adaptive thresholding mechanism to ensure that only reliable and accurate alignments are retained throughout the iterative process. Furthermore, we argue that such domain restriction and filtering mechanisms are fundamental components that should be incorporated into all direct alignment recovery methods, and we formally refer to this design principle as the Direct Alignment Recovery Paradigm.

K-Nearest Neighbors Direct Alignment

After obtaining the unified latent representation $\{Z^v\}_{v=1}^m$, for a given unaligned sample z_i^v, we first construct its alignment domain (AD) in view w using its K-nearest aligned neighbors set N_i^v. The corresponding neighbors set in view w is denoted as N_i^w, which can be formally defined by:

$$\begin{aligned} N_i^w &= \{z_j^w \mid P_{kj}^{vw} = 1\} \\ \text{s.t.} \quad & z_k^v \in N_i^v,\ k, j = 1, 2, \ldots, n \end{aligned} \tag{2}$$

where P_{kj}^{vw}=1 indicates that the sample in N_i^w is aligned one-to-one with the corresponding sample in N_i^v.

After identifying the corresponding K-nearest neighbors N_i^w in view w, we use their centroid $c_i^w = mean(z_j^w)$ as the center of the alignment domain. The radius r_i^w of the alignment domain is defined as the distance from the center to the farthest neighbor in N_i^w. The unaligned samples within this domain constitute the set of alignment candidates Q_i^w, which can be formally described as follows:

$$\begin{aligned} Q_i^w &= \{z_j^w \mid \left\|z_j^w - c_i^w\right\| \leq r_i^w \wedge P_{kj}^{vw} = 0\} \\ \text{s.t.} \quad & k, j = 1, 2, ..., n \end{aligned} \tag{3}$$

where $\|\cdot\|$ denotes the Euclidean distance operator.

After obtaining Q_i^w, the new alignment sample for z_i^v will be selected from this set. However, to address Challenge 2, it is necessary to discard unreliable alignments. To assess the reliability of each alignment domain, we compute a confidence score based on two factors: the average similarity between known aligned sample pairs within the domain and the distribution of candidate samples in the domain, which is formally defined by:

$$con_i^w = \frac{|Q_i^w|}{\sum \|q_i^w - c_i^w\|} * mean\_sim \quad \text{s.t. } q_i^w \in Q_i^w \tag{4}$$

where the second term represents the average similarity of the known aligned sample pairs within the domain, and the first term serves as a correction factor. $|\cdot|$ denotes the operator used to compute the number of samples in a set.

We define the correction factor as the ratio between the number of candidate points within the alignment domain and their aggregated distance to the domain center. The correction factor evaluates the compactness of the alignment domain: the more candidate points it contains and the closer they are to the center, the denser the cluster is considered to be. This is based on the assumption that samples closer to the cluster center tend to have higher density, and are thus more likely to belong to the same class (Li et al. 2020). The second term reflects the similarity between the local structure of the alignment domain and the K-nearest neighbor of z_i^v, serving as the primary basis for alignment.

The average similarity of the known aligned sample pairs within the alignment domain can be computed by:

$$\begin{aligned} mean\_sim &= \frac{1}{\mathrm{K}} * \sum \cos\left(z_k^v, z_j^w\right) \\ \text{s.t. } & z_k^v \in N_i^v, z_j^w \in N_i^w, P_{kj}^{vw} = 1 \end{aligned} \tag{5}$$

where the operator $\cos(\cdot,\cdot)$ denotes the computation of cosine similarity between vectors.

After computing the confidence scores for all alignment domains, Min-Max normalization is applied to map the confidence values into the range $[0, 1]$. Subsequently, only the alignment domains with confidence scores above the current threshold th are retained. For a given unaligned sample z_i^v, the most similar sample within its candidate set Q_i^w is selected to form a new alignment. The formal expression of the index for newly aligned sample in Q_i^w is as follows:

$$j = argmax\ \cos(z_i^v, z_j^w) \quad \text{s.t. } z_j^w \in Q_i^w \tag{6}$$

After establishing all new alignments, the alignment permutation matrix P^{vw} is updated by setting $P_{ij}^{vw} = 1$.

Contrastive Consistent Representation Learning

After obtaining the updated alignment permutation matrix, we construct positive pairs from all aligned sample pairs. For each positive pair, we randomly select M unaligned samples to construct M negative pairs. We adopt the loss of SIAMESE (Hadsell, Chopra, and LeCun 2006) for contrastive consistency representation learning, defined as:

$$L_{con} = \frac{1}{2N}\sum_{i=1}^{N}(IL_i^{pos} + (1-I)L_i^{neg}) \tag{7}$$

where N denotes the number of sample pairs, and $I = 1/0$ for positive/negative pairs. When the pair is positive, term $L_i^{\text{pos}} = \|z_i^v - z_j^w\|^2$ will become effective, otherwise $L_i^{\text{neg}} = \left(\max\left(\text{mar} - \|z_i^v - z_j^w\|, 0\right)\right)^2$ takes effect.

In contrastive learning, we aim to minimize the distances between positive pairs while maximizing the distances between negative pairs. However, simply increasing the negative pair distance may lead to trivial solutions. Therefore, we introduce a margin value mar, which constrains the distances between negative pairs from growing excessively. mar is computed only once at the start of training by:

$$mar = \frac{1}{N_{pos}}\sum \|z_i^v - z_j^w\|^2 + \frac{1}{N_{neg}}\sum \|z_i^v - z_j^w\|^2 \tag{8}$$

where N_{pos} and N_{neg} represent the number of positive and negative pairs, respectively.

The overall loss function of our framework is defined as:

$$L = L_{rec} + L_{con} \tag{9}$$

where $L_{rec} = \frac{1}{n}\sum_{i=1}^{n}\|x_i^v - \hat{x}_i^v\|^2$ denotes the loss of reconstruction, and $\hat{x}_i^v = g^v(f^v(x_i^v))$ is the original data reconstructed by encoder-decoder network.

The whole process of our model is shown in Algorithm 1.

Algorithm 1 Optimization procedures for KNNDA

Input: Partially view-aligned dataset χ; parameters K, th; number of training epochs T; number of clusters c.
Output: Learned consistent representation.
1: Initialize the networks. Set epoch $t = 0$.
2: **while** $t < T$ **do**:
3: Learn the latent representations $\{Z^v\}_{v=1}^m$.
4: Input $\{Z^v\}_{v=1}^m$ into the KNNDA module to obtain updated $\{P^{1v}\}_{v=2}^m$.
5: Construct positive and negative pairs by updated $\{P^{1v}\}_{v=2}^m$ for contrastive learning .
6: Update $\{f^v\}_{v=1}^m$ and $\{g^v\}_{v=1}^m$ by Eq. (9).
7: $t = t + 1$
8: **end while**
Return: Consistent representation Z^* by concatenating aligned $\{Z^v\}_{v=1}^m$.

Experiments

Datasets, Metrics and Experimental Settings

Our experiments utilized six multi-view datasets, including Scene-15 (Fei-Fei and Perona 2005), HandWritten (Zhang et al. 2022), BDGP (Cai et al. 2012), ALOI-100[1], 100Leaves (Cope and Barman 2013), and BBCSports (Greene and Cunningham 2006). Details of these datasets are provided in Table 1. To evaluate the effectiveness of our framework when the number of views is greater than 2, we conducted experiments on the selected 2 and 3 views of 100Leaves and BBCSports datasets. Following all baseline methods, we use ACC, NMI, and ARI as evaluation metrics.

Dataset	Views	Samples	Clusters
Scene-15	59/20	4485	15
ALOI-100	64/77	10800	100
BDGP	79/1750	2500	5
HandWritten	240/216	2000	10
100Leaves-2V	64/64	1600	100
100Leaves-3V	64/64/64	1600	100
BBCSports-2V	2544/2465	282	5
BBCSports-3V	2465/2544/2582	282	5

Table 1: Datasets Summary.

Our encoders are implemented by multi-layer neural networks, with layer dimensions d_{in}-1024-1024-1024-d_{out}. Each linear layer is followed by a Batch Normalization layer (Ioffe and Szegedy 2015), a ReLU activation function (Nair and Hinton 2010), and a Dropout layer (Srivastava et al. 2014). The dropout ratio is set to 0.2.

We implement our framework on PyTorch 2.0.0 and run experiments on an NVIDIA GeForce 4090 GPU. The network is optimized by the Adam optimizer (Kingma and Ba 2015) with the learning rate of 0.001. The batch size is set to 1024 for all datasets, except BBCSports to 128.

The settings of our model-specific parameters, including the number of neighbors K and the threshold th, will be detailed in the sections on parameter sensitivity analysis and threshold selection strategy.

We compare KNNDA with 11 SOTA multi-view clustering baselines, including AE^2-Nets (Zhang, Liu, and Fu 2019), PVC (Huang et al. 2020), MVC-UM (Yu et al. 2021), MvCLN, FMVACC (Wang et al. 2022b), SURE (Yang et al. 2023), DIVIDE (Lu et al. 2024), DGPPVC, NC^{3}L, CGDAPVC, and TCLPVC (Gao et al. 2025). In particular, AE^2-

[1]https://aloi.science.uva.nl/

Methods	Scene-15			BDGP			HandWritten		
Metrics(%)	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
AE^2-Nets (CVPR'19)	26.25	24.04	12.03	39.16	17.77	6.15	69.10	66.45	56.08
PVC (NeurIPS'20)	35.85	39.43	19.83	77.40	53.30	51.29	71.65	74.64	63.90
MVC-UM (KDD'21)	21.11	22.20	15.39	32.16	7.53	24.00	46.25	48.58	29.52
MvCLN (CVPR'21)	38.17	39.29	24.47	65.14	45.01	43.82	59.75	53.74	42.26
FMVACC (NeurIPS'22)	35.05	33.89	18.28	52.20	33.78	27.33	78.41	73.77	66.69
SURE (TPAMI'23)	36.91	39.28	21.87	44.96	25.70	14.85	59.08	54.36	41.90
DIVIDE (AAAI'24)	34.85	31.79	17.23	44.88	14.39	10.25	64.90	52.72	43.34
DGPPVC (TNNLS'24)	39.94	37.88	21.88	82.93	60.71	62.14	83.15	76.53	69.97
NC^{3}L (TIP'24)	44.57	**43.93**	26.46	81.08	71.01	56.37	83.80	78.02	70.22
CGDAPVC (MM'24)	41.63	42.05	22.57	90.74	77.40	78.84	83.16	79.24	73.01
TCLPVC (TCSVT'25)	36.23	40.10	20.92	91.92	77.87	80.94	90.45	83.41	79.84
KNNDA (ours)	**45.08**	42.55	**26.75**	**93.32**	**81.80**	**83.92**	**91.20**	**84.63**	**81.79**
Methods	ALOI-100			100Leaves-2V			BBCSports-2V		
Metrics(%)	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
AE^2-Nets (CVPR'19)	3.76	9.96	0.43	36.24	64.93	22.38	33.07	4.23	3.67
PVC (NeurIPS'20)	42.35	67.16	30.49	39.56	63.76	22.22	37.87	8.43	4.19
MVC-UM (KDD'21)	17.03	43.46	11.04	39.93	69.31	27.14	26.29	4.34	0.81
MvCLN (CVPR'21)	44.40	64.07	34.31	40.19	70.39	26.49	40.43	9.36	6.02
FMVACC (NeurIPS'22)	30.10	49.54	15.75	35.81	73.85	26.99	36.24	8.91	6.18
SURE (TPAMI'23)	39.46	62.14	31.77	65.00	82.32	52.11	37.59	7.31	2.15
DIVIDE (AAAI'24)	46.13	62.41	30.12	36.88	66.68	22.58	30.85	4.30	3.24
DGPPVC (TNNLS'24)	52.47	66.32	38.18	48.19	69.83	29.72	37.23	7.83	3.23
NC^{3}L (TIP'24)	37.36	64.38	25.91	38.25	69.84	24.33	43.26	9.67	4.64
CGDAPVC (MM'24)	22.28	47.69	13.12	28.70	65.93	17.80	34.75	4.97	4.42
TCLPVC (TCSVT'25)	49.88	74.06	41.76	25.19	60.35	15.54	34.40	3.32	3.14
KNNDA (ours)	**85.05**	**88.62**	**75.80**	**70.38**	**84.44**	**58.11**	**47.52**	**16.35**	**16.99**

Table 2: Clustering performance on six two-view datasets. The optimal results are marked in bold, and the suboptimal scores are underlined.

Nets, FMVACC, and DIVIDE were not designed for partially view-aligned clustering. Therefore, to simulate partially aligned scenarios for those methods that are not specifically designed for partial alignment, we first apply Principal Component Analysis (PCA) to reduce the dimensionality of all views after shuffling the datasets, and then use the Hungarian algorithm to establish cross-view alignments. For all compared methods, we report the results under the same experimental settings as described in their original papers, or follow the best configurations provided in their available code implementations. In the comparison analysis section, the initial alignment ratio ar is set to 0.5.

Clustering Comparative Analysis

Table 2 reports the clustering results of all methods on six 2-view datasets. Our KNNDA almost outperforms all baselines across all datasets and metrics. Especially, on ALOI-100, KNNDA achieves improvements of 62.1% (ACC), 19.66% (NMI), and 81.51% (ARI) over the second-best scores, indicating the capability of our framework to recover missing alignments with high quality while learning discriminative latent representations. Secondly, partially view-aligned clustering methods generally outperform non-partially aligned methods, highlighting the importance of a high-quality alignment recovery module.

Notably, CGDAPVC is the only method that learns alignments without relying on representation similarity. However, it still does not recover the missing alignments directly from the known alignments and shows significant performance drops on the ALOI-100 and 100Leaves datasets, both of which have complex clusters. In contrast, our KNNDA is based on known alignments to recover unknown alignments, achieves stable performance across all datasets.

In addition, we evaluate the performance of KNNDA under three-view setting. The results are presented in Table 3, together with those of the baseline methods. Since some baselines are not applicable to three-views setting, we only report results of part of baselines. It can be observed that KNNDA still maintains a performance advantage when handling three views. Furthermore, compared to the two-view setting, our framework achieves slightly better clustering performance in the three-view case, indicating that KNNDA can effectively leverage the complementary information provided by additional views to enhance the quality of representation learning.

Ablation Study

To verify the effectiveness of our module, we conduct an ablation study on the Scene-15 and HandWritten datasets. We design two vanilla versions of KNNDA: 1) one that com-

Methods	100Leaves-3V			100Leaves-2V			BBCSports-3V			BBCSports-2V		
Metrics(%)	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
AE^2-Nets	49.81	73.85	34.43	36.24	64.93	22.38	36.17	4.84	0.61	33.07	4.23	3.67
MVC-UM	54.88	79.43	41.75	39.93	69.31	27.14	28.75	2.80	0.67	26.29	4.34	0.81
FMVACC	40.30	70.82	28.46	35.81	73.85	26.99	37.70	12.28	5.34	36.24	8.91	6.18
DIVIDE	37.69	66.84	22.45	36.88	66.68	22.58	27.66	3.76	1.48	30.85	4.30	3.24
DGPPVC	35.06	60.93	19.32	48.19	69.83	29.72	34.04	1.88	0.40	37.23	7.83	3.23
TCLPVC	36.06	66.64	24.13	25.19	60.35	15.54	35.11	2.97	2.78	34.40	3.32	3.14
KNNDA(ours)	**83.06**	**89.34**	**72.14**	**70.38**	**84.44**	**58.11**	**49.65**	**21.46**	**16.05**	**47.52**	**16.35**	**16.99**

Table 3: Three-view clustering performance on 100Leaves and BBCSports datasets.

Version	Scene-15			HandWritten		
	ACC	NMI	ARI	ACC	NMI	ARI
V1	27.25	26.96	12.71	73.15	71.91	60.57
V2	31.22	27.23	13.86	82.05	75.35	67.92
KNNDA	**45.08**	**42.55**	**26.75**	**91.20**	**84.63**	**81.79**

Table 4: Ablation study on Scene-15 and HandWritten datasets with vanilla versions of KNNDA.

pletely removes our alignment module and directly applies the Hungarian algorithm for alignment; 2) one that retains the KNNDA alignment process but removes the confidence computation and threshold-based filtering mechanisms. The second version corresponds to Challenge 2 introduced in the Direct Alignment Recovery Paradigm section. Regarding Challenge 1, although removing domain constraints from the model is theoretically feasible for ablation, we observe out-of-memory errors during experiments. This further confirms the necessity and correctness of our Direct Alignment Recovery Paradigm.

The results of ablation study are presented in Table 4. As observed, when the alignments are recovered solely using the Hungarian algorithm (V1), the model performance drops significantly. This is because removing all alignment modules degenerates the model into a conventional partially view-aligned contrastive clustering framework. In this setting, alignments are established purely based on similarity in the latent space, which may lead to incorrect pairings. Meanwhile, model performance remains suboptimal when the confidence and threshold filtering mechanisms are removed (V2), as unreliable or incorrect alignments introduce noise to the representation learning process, ultimately degrading clustering performance. These results demonstrate the effectiveness of both the proposed alignment module and the confidence-based thresholding strategy.

Robustness Study

To verify the robustness of the model under different initial alignment ratios $ar \in [0.1, 0.9]$, especially at the low alignment rate, we conducted comparative experiments on the BDGP dataset with three PVC baselines. In addition to used three existing metrics, we introduce the Class Alignment Rate (CAR) to evaluate the alignment quality of methods.

The results are illustrated in Fig. 2. It is evident from the results that our KNNDA consistently outperforms the other three PVC methods across all initial alignment ratios. Furthermore, as the initial alignment ratio increases, our clustering performance steadily improves, indicating that our KNNDA effectively leverages the available alignments to capture both consistency and complementarity across views. In particular, KNNDA demonstrates strong performance even when the alignment ratio is low, whereas other baselines suffer a significant drop in performance due to their limited ability to utilize sparse alignments.

Regarding the alignment recovery metric (CAR), KNNDA is able to recover nearly all alignments as the initial alignment ratio increases and still achieves substantial alignment recovery under low alignment ratios. This significantly facilitates subsequent representation learning. These results collectively demonstrate the robustness of KNNDA to varying degrees of initial alignment and validate the superiority of directly recovering missing alignments from known ones.

Parameter Sensitivity Analysis

The choice of parameter M follows the setting in MvCLN and is not further discussed here. For the number of neighbors K, we investigated its impact on KNNDA on the BDGP and Scene-15 datasets. Specifically, we report the clustering performance of our framework under different settings of K, with results shown in Fig. 3.

For the BDGP dataset, when K $\in [5, 20]$, KNNDA is relatively insensitive to its variation. And for the Scene-15 dataset, KNNDA consistently achieves superior performance across all settings of K. Therefore, we set K to 10 for all small-scale datasets and 35 for large-scale datasets.

Threshold Selection Strategy

In our KNNDA, only alignment pairs with confidence scores above a threshold are retained. Therefore, the choice of threshold is crucial: the threshold that is too low retains incorrect alignments, while an excessively high threshold slows down the recovery process. To determine an optimal threshold selection strategy, we conducted experiments on the BDGP and ALOI-100 datasets.

We first fixed the threshold th throughout the entire training process and investigated how different thresholds affect clustering performance. The results are shown in Fig. 4a. As the threshold increases, CAR gradually improves, indicating that higher thresholds are beneficial for alignment recovery. However, on the BDGP dataset, we observed an un-

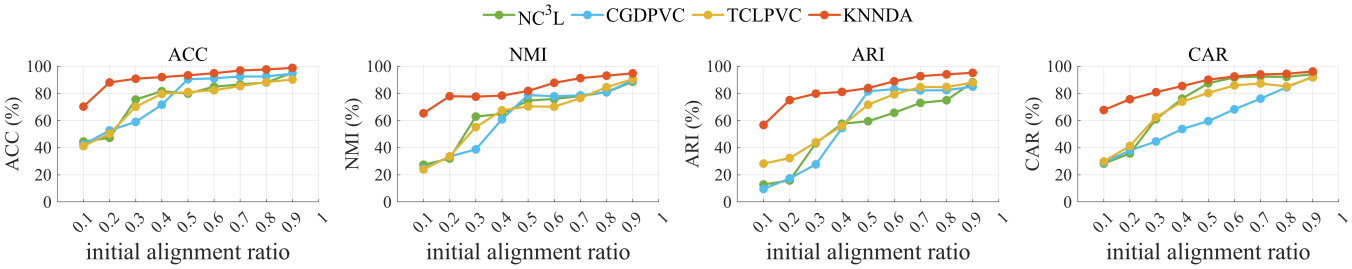
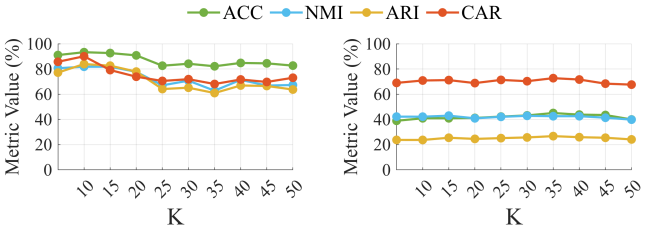
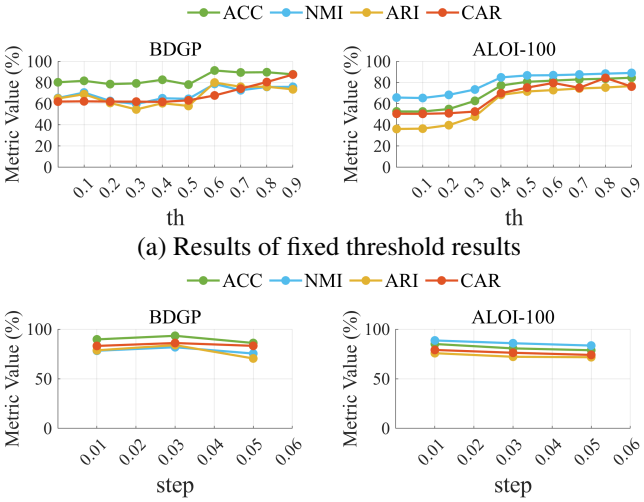


Figure 2: Clustering performance of our method is compared with three baselines on the BDGP dataset, considering varying initial alignment ratios.



(a) BDGP (b) Scene-15

Figure 3: Clustering performance of KNNDA on BDGP and Scene-15 datasets under different K.



(a) Results of fixed threshold results

(b) Results of different steps for dynamic threshold

Figure 4: Clustering performance of KNNDA on BDGP and ALOI-100 datasets by different threshold selection strategy.

expected phenomenon: when the threshold exceeds 0.5, although CAR continues to increase with the threshold, clustering metrics begin to drop. This inconsistency between alignment quality and clustering performance raised an important question: what causes this divergence?

Through extensive experiments across different datasets, we found that this inconsistency arises in semantically simple datasets (i.e., with fewer categories) such as BDGP. In such cases, the representation learning module can quickly learn cluster-friendly features in the early training stages, leading to improved clustering metrics. However, when the threshold is too high, alignment recovery lags behind. By the time alignment improves, the model is already overfitted, causing the final clustering performance to decline. In contrast, on semantically complex datasets such as ALOI-100, representation learning is more difficult. This naturally synchronizes the pace of alignment recovery and representation learning, mitigating the negative impact.

These findings offer a new suggestion: fast alignment recovery is preferred for simpler datasets to match the learning progress of representations, whereas slower recovery can be beneficial for complex datasets.

According to the above analysis, we propose a dynamic threshold strategy: a high threshold is used at the beginning of training to prevent the model from being affected by unreliable alignments. Then, the threshold is gradually decreased each iteration until it reaches a predefined lower bound. For different datasets, the speed of threshold decay (i.e., the step) can be adjusted to control the recovery pace.

Based on the fixed-threshold experiments, we selected 0.8 as the initial threshold and 0.65 as the lower bound. We then evaluated model performance under different step sizes, as shown in Fig. 4b. The results indicate that the optimal step size for the BDGP dataset is 0.03, whereas for ALOI-100, the optimal value is 0.01. This aligns with our hypothesis: for semantically simple datasets, the threshold should be reduced more quickly to accelerate alignment recovery. Moreover, too large step size leads to overly aggressive threshold decay, allowing more incorrect alignments to be retained, thereby harming clustering performance across all datasets.

Conclusion

In this paper, we propose a new paradigm of alignment recovery for PVC and instantiate it with a module named K-Nearest Neighbors Direct Alignment (KNNDA). KNNDA directly recovers missing alignments by constructing alignment domains based on K-nearest neighbors and filters out incorrect alignments by confidence scores derived from known alignments. In addition, our dynamic thresholding strategy ensures strong performance across datasets with varying characteristics. Extensive experiments demonstrate the superiority and effectiveness of our model in partial alignment clustering.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (62476038), the Science and Technology Project of Liaoning Province (2024JH2/102600027), the Science and Technology Project of Dalian City (2024JJ12GX025, 2023JJ12SN029, and 2023JJ11CG005), and the Fundamental Research Funds for the Central Universities (DUT25YG246 and DUTZD25216).

References

Cai, X.; Wang, H.; Huang, H.; and Ding, C. 2012. Joint stage recognition and anatomical annotation of drosophila gene expression patterns. *Bioinformatics*, 28(12): i16–i24.

Cope, B. T. R. P., James; and Barman, S. 2013. One-hundred plant species leaves data set. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5RG76.

Eaton, E.; desJardins, M.; and Jacob, S. 2010. Multi-view clustering with constraint propagation for learning with an incomplete mapping between views. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, 389–398. New York, NY, USA: Association for Computing Machinery. ISBN 9781450300995.

Fei-Fei, L.; and Perona, P. 2005. A Bayesian hierarchical model for learning natural scene categories. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, 524–531 vol. 2.

Gao, H.; Liu, C.; Cai, Z.; Sun, H.; Li, G.; Li, Y.; and Du, W. 2025. A Novel Approach for Effective Partially View-Aligned Clustering with Triple-Consistency. *IEEE Transactions on Circuits and Systems for Video Technology*, 1–1.

Greene, D.; and Cunningham, P. 2006. Practical solutions to the problem of diagonal dominance in kernel document clustering. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, 377–384. New York, NY, USA: Association for Computing Machinery. ISBN 1595933832.

Guo, J.; and Ye, J. 2019. Anchors Bring Ease: An Embarrassingly Simple Approach to Partial Multi-View Clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01): 118–125.

Hadsell, R.; Chopra, S.; and LeCun, Y. 2006. Dimensionality Reduction by Learning an Invariant Mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, 1735–1742.

Hu, S.; Zhang, C.; Zou, G.; Lou, Z.; and Ye, Y. 2025. Deep Multiview Clustering by Pseudo-Label Guided Contrastive Learning and Dual Correlation Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 36(2): 3646–3658.

Huang, Z.; Hu, P.; Zhou, J. T.; Lv, J.; and Peng, X. 2020. Partially View-aligned Clustering. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems*, volume 33, 2892–2902. Curran Associates, Inc.

Ioffe, S.; and Szegedy, C. 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, 448–456. JMLR.org.

Jin, J.; Wang, S.; Dong, Z.; Liu, X.; and Zhu, E. 2023. Deep Incomplete Multi-View Clustering with Cross-View Partial Sample and Prototype Alignment. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11600–11609.

Kingma, D. P.; and Ba, J. 2015. Adam: A Method for Stochastic Optimization. In Bengio, Y.; and LeCun, Y., eds., *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Li, Q.; Yue, S.; Wang, Y.; Ding, M.; Li, J.; and Wang, Z. 2020. Boundary Matching and Interior Connectivity-Based Cluster Validity Anlysis. *Applied Sciences*, 10(4).

Lu, Y.; Lin, Y.; Yang, M.; Peng, D.; Hu, P.; and Peng, X. 2024. Decoupled Contrastive Multi-View Clustering with High-Order Random Walks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(13): 14193–14201.

Lu, Z.; and Peng, Y. 2013. Unified Constraint Propagation on Multi-View Data. *Proceedings of the AAAI Conference on Artificial Intelligence*, 27(1): 640–646.

Ma, S.; Zhao, L.; Lu, M.; Guo, Y.; and Xu, B. 2025. Consistency-Aware Padding for Incomplete Multi-Modal Alignment Clustering Based on Self-Repellent Greedy Anchor Search. *arXiv preprint arXiv:2507.03917*.

Nair, V.; and Hinton, G. E. 2010. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, 807–814. Madison, WI, USA: Omnipress. ISBN 9781605589077.

Qian, S.; Xue, D.; Hu, J.; Zhang, H.; and Xu, C. 2024. Nonparametric Clustering-Guided Cross-View Contrastive Learning for Partially View-Aligned Representation Learning. *IEEE Transactions on Image Processing*, 33: 6158–6172.

Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; and Salakhutdinov, R. 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1): 1929–1958.

Sun, Y.; Qin, Y.; Li, Y.; Peng, D.; Peng, X.; and Hu, P. 2024. Robust Multi-View Clustering With Noisy Correspondence. *IEEE Transactions on Knowledge and Data Engineering*, 36(12): 9150–9162.

Trosten, D. J.; Lokse, S.; Jenssen, R.; and Kampffmeyer, M. 2021. Reconsidering Representation Alignment for Multi-View Clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1255–1265.

Trosten, D. J.; Løkse, S.; Jenssen, R.; and Kampffmeyer, M. C. 2023. 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)*, 23976–23985.

Wang, R.; Li, L.; Tao, X.; Wang, P.; and Liu, P. 2022a. Contrastive and attentive graph learning for multi-view clustering. *Information Processing Management*, 59(4): 102967.

Wang, S.; Liu, X.; Liu, S.; Jin, J.; Tu, W.; Zhu, X.; and Zhu, E. 2022b. Align then fusion: generalized large-scale multi-view clustering with anchor matching correspondences. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781713871088.

Wang, X.; Gao, H.; Wei, X.; Peng, L.; Li, R.; Liu, C.; Wu, S.; and Wong, H.-S. 2024a. Contrastive Graph Distribution Alignment for Partially View-Aligned Clustering. In *Proceedings of the 32nd ACM International Conference on Multimedia*, MM '24, 5240–5249. New York, NY, USA: Association for Computing Machinery. ISBN 9798400706868.

Wang, Y.; Chang, D.; Fu, Z.; Wen, J.; and Zhao, Y. 2024b. Partially View-Aligned Representation Learning via Cross-View Graph Contrastive Network. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(8): 7272–7283.

Wen, Y.; Wang, S.; Liao, Q.; Liang, W.; Liang, K.; Wan, X.; and Liu, X. 2024. Unpaired Multi-View Graph Clustering With Cross-View Structure Matching. *IEEE Transactions on Neural Networks and Learning Systems*, 35(11): 16049–16063.

Xu, J.; Tang, H.; Ren, Y.; Peng, L.; Zhu, X.; and He, L. 2022. Multi-Level Feature Learning for Contrastive Multi-View Clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16051–16060.

Yang, M.; Li, Y.; Hu, P.; Bai, J.; Lv, J.; and Peng, X. 2023. Robust Multi-View Clustering With Incomplete Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1): 1055–1069.

Yang, M.; Li, Y.; Huang, Z.; Liu, Z.; Hu, P.; and Peng, X. 2021. Partially View-aligned Representation Learning with Noise-robust Contrastive Loss. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1134–1143.

Yu, H.; Tang, J.; Wang, G.; and Gao, X. 2021. 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*, KDD '21, 2075–2083. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383325.

Zhang, B.; Qiang, Q.; Wang, F.; and Nie, F. 2022. Fast Multi-View Semi-Supervised Learning With Learned Graph. *IEEE Transactions on Knowledge and Data Engineering*, 34(1): 286–299.

Zhang, C.; Liu, Y.; and Fu, H. 2019. AE2-Nets: Autoencoder in Autoencoder Networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2572–2580.

Zhang, C.; Wang, S.; Liu, J.; Zhou, S.; Zhang, P.; Liu, X.; Zhu, E.; and Zhang, C. 2021. Multi-view Clustering via Deep Matrix Factorization and Partition Alignment. In *Proceedings of the 29th ACM International Conference on Multimedia*, MM '21, 4156–4164. New York, NY, USA: Association for Computing Machinery. ISBN 9781450386517.

Zhao, L.; and Xie, Q. 2024. Distribution-Level Multi-View Clustering for Unaligned Data. *IEEE Signal Processing Letters*, 31: 2330–2334.

Zhao, L.; Xie, Q.; Li, Z.; Wu, S.; and Yang, Y. 2025. Dynamic Graph Guided Progressive Partial View-Aligned Clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 36(5): 9370–9382.

Zhao, Y.; You, X.; Yu, S.; Xu, C.; Yuan, W.; Jing, X.-Y.; Zhang, T.; and Tao, D. 2018. Multi-view manifold learning with locality alignment. *Pattern Recognition*, 78: 154–166.