

Imputation-free and Alignment-free: Incomplete Multi-view Clustering Driven by Consensus Semantic Learning

Yuzhuo Dai¹, Jiaqi Jin¹, Zhibin Dong¹, Siwei Wang^{2,*}, Xinwang Liu^{1,*}, En Zhu^{1,*},
 Xihong Yang¹, Xinbiao Gan¹, Yu Feng¹

¹National University of Defense Technology, Changsha, China

²Intelligent Game and Decision Lab, Beijing, China

{yzdai24, wangsiwei13}@nudt.edu.cn

Abstract

In incomplete multi-view clustering (IMVC), missing data induce prototype shifts within views and semantic inconsistencies across views. A feasible solution is to explore cross-view consistency in paired complete observations, further imputing and aligning the similarity relationships inherently shared across views. Nevertheless, existing methods are constrained by two-tiered limitations: (1) Neither instance- nor cluster-level consistency learning construct a semantic space shared across views to learn consensus semantics. The former enforces cross-view instances alignment, and wrongly regards unpaired observations with similar semantics as negative pairs; the latter focuses on cross-view cluster counterparts while coarsely handling fine-grained intra-cluster relationships within views. (2) Excessive reliance on consistency results in unreliable imputation and alignment without incorporating view-specific cluster information. Thus, we propose an IMVC framework, imputation- and alignment-free for consensus semantics learning (FreeCSL). To bridge semantic gaps across all observations, we learn consensus prototypes from available data to discover a shared space, where semantically similar observations are pulled closer for consensus semantics learning. To capture semantic relationships within specific views, we design a heuristic graph clustering based on modularity to recover cluster structure with intra-cluster compactness and inter-cluster separation for cluster semantics enhancement. Extensive experiments demonstrate, compared to state-of-the-art competitors, FreeCSL achieves more confident and robust assignments on IMVC task.

1. Introduction

Thanks to representation learning enhanced by data observed from different perspectives, multi-view clustering

*Corresponding author

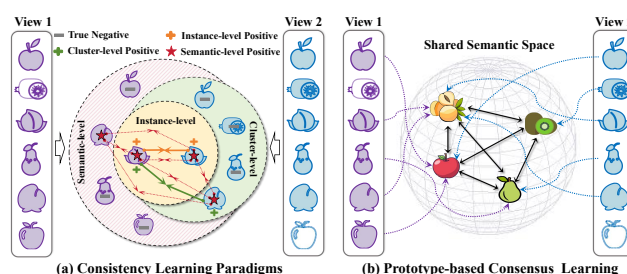


Figure 1. Research Motivation for Consensus Semantic Learning. (a) two existing paradigms, instance- and cluster-level, either treat cross-view unpaired observations with similar semantics as false negatives or neglect to treat within-view observations with similar semantics as true positives; (b) we propose a novel semantic-level paradigm based on contrastive clustering with a set of consensus prototypes to foster semantic consistency across all view data.

(MVC) has achieved significant breakthroughs in the field of unsupervised learning [3, 6, 11, 17, 21, 34, 46, 57]. However, in practical applications, the assumption of data completeness is often difficult to satisfy that incomplete multi-view clustering (IMVC) is introduced [7, 25, 29, 33, 35, 37, 54]. In IMVC, missing data causes prototypes shifts within views and semantic misalignment across views, due to the discrepancy between the distributions of complete and incomplete instances [14, 19]. More and more studies [4, 16] have noted that variations in complete instances across different views further exacerbate prototype misalignment. It is challenging to achieve semantic consistency on cluster assignments across all view data.

To alleviate prototype shifts and misalignment while promoting semantic consensus in cluster assignments, existing IMVC methods explore consistency information from complete instances for imputation and alignment [20, 24, 32]. Unfortunately, they still suffer from several significant drawbacks in practical applications. In terms of consistency learning, as shown in Fig. 1 (a), one widely used paradigm is instance-level [13, 36], which pulls paired ob-

servations (the same instance in different views) closer in the representation space by enforcing highly similar representations, but may inadvertently discard view-specific information [26, 31, 48]. More importantly, it tends to introduce false-negative noise, where unpaired observations with the same semantics are incorrectly treated as negative pairs [2, 8, 27]. [49],[52] and [27] attempted to optimize this issue by proposing a cluster-level paradigm [47] that encourages observations to find their cluster counterparts across different views. It learns a cluster space shared across views but not applicable within specific views, as there are no clustering interactions among intra-view observations.

Since the above consistency learning paradigms fail to account for semantic consistency across all view data, they cannot mitigate semantic gaps among intra-cluster observations [22, 30, 56]. Therefore, imputation of missing data is required to restore the original data distribution, including neighborhood-based recovery via cross-view graph structure transfer [39], adversarial generation or contrastive prediction through cross-view mutual information interaction [15, 37, 42, 53], and prototype-based imputation via cross-view sample-prototype relationship inheritance [19]. Meanwhile, cross-view alignment of assignments [22, 32], prototypes [16, 19, 38], or distributions [4, 45] is also a crucial approach for further enhancing consistent learning. More related works are enumerated in Appendix A. Both imputation and alignment are limited by the consistent information from cross-view paired data and cannot fully exploit within-view unpaired data to mine view-specific cluster information, *i.e.*, within-cluster compactness and between-cluster separation [9, 19]. Particularly, once the amount of missing data is too excessive to provide sufficient consistent information, model performance may even decline rapidly due to the noise introduced by improper imputation and alignment.

Realizing the above issues, we ponder: imputation and alignment aim to restore the similarity relationships inherited from other complete views for the clustering task. Can we directly bridge the semantic gaps while exploring the semantic relationships among all data in consistency learning, thereby avoiding imputation and alignment operations with uncertainty noise? Thus, we propose an IMVC framework, a novel semantic-level paradigm, as shown in Fig. 1 (b), that is driven by imputation- and alignment-free consensus semantic learning (FreeCSL). Notably, our proposed consensus learning involves concurrent interaction among all data, rather than being limited to within or across views. To bridge semantic gaps among view observations and learn cluster semantics information, FreeCSL employs contrastive clustering based on consensus prototypes to discover a shared semantic space, where observations converge toward their semantic prototypes respectively. In practice, we set missing statistical weights for observations to facilitate view collaboration in consensus representations that in-

tegrate detailed information of all views while adapting the impact of different views with different missing instances. Based on consensus representation, our model can construct robust consensus prototypes for semantic-level clustering without imputation and alignment. To discover view-specific semantic relationships and further enrich consensus representations, FreeCSL exploits graph clustering to capture the cluster structure for each view, which maximizes graph modularity within views to enhance intra-cluster connections and reduce inter-cluster interactions. In short, our model encodes data correlation, discovered by semantic learning, into a shared space to obtain consensus semantic representations for instance clustering. Our prominent contributions can be summarized as follows:

- In terms of bridging semantic gaps, we design the consensus semantic learning module, a novel semantic-level paradigm based on prototypical contrastive clustering, to discover a common semantic space where all observations are embedded as representations with consistent semantics, avoiding additional imputation and alignment.
- In terms of exploring semantic relationships, we employ the cluster semantics enhancement module, a heuristic graph clustering method with modularity-based learning objective, to mine the inherent cluster structures that reveal the semantic correlations within views.
- Extensive experiments show our model surpasses state-of-the-art (SOTA) competitors in complex tasks with high missing rates, multiple clusters and large-scale data.

2. Method

In this section, FreeCSL, a deep IMVC method without imputation or alignment, is proposed to learn consensus semantic representations for clustering. The framework in Fig. 2 coordinates reconstruction (REC) module, cross-view consensus semantic learning (CSL) module, and within-view cluster semantic enhancement (CSE) module.

2.1. Problem Statement

Notations. Given a multi-view dataset $\mathcal{X} = \{\mathbf{X}^v \in \mathbb{R}^{N \times D_v}\}_{v=1}^V$ with N instances across V views, $\tilde{\mathbf{X}}^v = \{\tilde{\mathbf{x}}_i^v \in \mathbb{R}^{D_v}\}_{i=1}^{N_v}$ is an incomplete subset of the v -th view with N_v observations, and $\bar{\mathbf{X}}^{m,n} = \{(\bar{\mathbf{x}}_i^m, \bar{\mathbf{x}}_i^n)\}_{i=1}^{N_{mn}}$ is a pair-observed subset with N_{mn} instances observed in both the m -th and n -th view. The task is to partition N instances into K clusters.

Definition 1. *Instance-level Consistency (IC):* $\forall m \neq n, \mathbf{x}_i^m$ and $\mathbf{x}_j^n$ are instance-level consistent across views if $i = j$ (they are cross-view observations of the same instance $\mathbf{x}$), expressed as $I(\mathbf{x}_i^m, \mathbf{x}_j^n) = 1$ and 0 otherwise.

Definition 2. *Cluster-level Consistency (CC):* $\forall m \neq n, \mathbf{x}_i^m$ and $\mathbf{x}_j^n$ are cluster-level consistent across views if they belong to the same cluster k , expressed as $C(\mathbf{x}_i^m, \mathbf{x}_j^n) = 1$ and 0 otherwise.

Definition 3. *Semantic-level Consensus (SC):* $\forall m$ and n , $\mathbf{x}_i^m$ and $\mathbf{x}_j^n$ achieve semantic-level consensus in MVC task if all observations share a set of cluster prototypes $\mathbf{C} = \{\mathbf{c}_k\}_{k=1}^K$ and $\arg \max_k \rho(\mathbf{x}_i^m, \mathbf{c}_k) = \arg \max_k \rho(\mathbf{x}_j^n, \mathbf{c}_k)$, expressed as $S(\mathbf{x}_i^m, \mathbf{x}_j^n) = 1$ and 0 otherwise.

Theorem 1. *Consensus semantic learning yields more confident and robust cluster assignments than instance- and cluster-level paradigms. (Proof is provided in Appendix B.)*

Theorem 2. *Since paired observations $(\bar{\mathbf{x}}_i^m, \bar{\mathbf{x}}_i^n)$ inherently satisfy instance- and cluster-level consistency, they can achieve semantic consensus via a shared set of prototypes $\mathbf{C}$. (Proof is provided in Appendix B.)*

2.2. Within-view Reconstruction

To avoid clustering instability from similarity measures on manifold structures in high-dimensional spaces, we use autoencoders to encode view observations $\tilde{\mathbf{X}}^v$ into clustering-friendly low-dimensional representations $\tilde{\mathbf{Z}}^v$. Considering multiple views are mostly heterogeneous and differently distributed, we provide an independent encoder-decoder $\{\mathcal{E}_v, \mathcal{D}_v\}_{v=1}^V$ for each view. The encoder embeds the latent representation and the decoder recovers the original data from it, which jointly minimize reconstruction loss:

$$\begin{aligned} \mathcal{L}_{rec} &= \sum_{v=1}^V \left\| \tilde{\mathbf{X}}^v - \mathcal{D}_v(\tilde{\mathbf{Z}}^v) \right\|_F^2 \\ &= \sum_{v=1}^V \sum_{i=1}^{N_v} \left\| \tilde{\mathbf{x}}_i^v - \mathcal{D}_v(\mathcal{E}_v(\tilde{\mathbf{x}}_i^v)) \right\|_2^2, \end{aligned} \quad (1)$$

where $\mathcal{E}_v(\tilde{\mathbf{X}}^v; \theta_v) : \tilde{\mathbf{X}}^v \in \mathbb{R}^{N_v \times D_v} \rightarrow \tilde{\mathbf{Z}}^v \in \mathbb{R}^{N_v \times d}$ and $\mathcal{D}_v(\tilde{\mathbf{Z}}^v; \phi_v) : \tilde{\mathbf{Z}}^v \in \mathbb{R}^{N_v \times d} \rightarrow \hat{\mathbf{X}}^v \in \mathbb{R}^{N_v \times D_v}$.

2.3. Cross-view Consensus Semantic Learning

Based on Theorem 1, we design prototypical contrastive clustering for consensus semantic learning. To reach semantic-level consensus on assignments, consensus prototypes with all view information, are introduced into contrastive clustering to explore a shared semantic space, where all observations with similar semantics are pulled closer.

Consensus Representations and Consensus Prototypes. Due to the assumptions of consistency and complementarity in MVC, inseparable clusters in one view will become linearly separable by introducing complementary information from other views [10, 43, 55]. We utilize a fusion manner, denoted as $\mathbb{T}(\cdot)$, to map the latent representations $\mathbf{Z}^v$ into a linearly weighted representation space for consensus representations $\mathbf{Z} \in \mathbb{R}^{N \times d}$ that integrates consistency and complementary information from multiple views:

$$\mathbf{Z} = \mathbb{T}(\{\mathbf{Z}^v\}_{v=1}^V) = \sum_{v=1}^V \mathbf{w}^v \mathbf{Z}^v = \left\{ \sum_{v=1}^V w_i^v \mathbf{z}_i^v \right\}_{i=1}^N, \quad (2)$$

where w_i^v is the instance-level fine-grained fusion weight:

$$w_i^v = \frac{\mathbb{I}(\mathbf{z}_i^v \neq NaN)}{\sum_{v'=1}^V \mathbb{I}(\mathbf{z}_i^{v'} \neq NaN)}, \quad (3)$$

where $\mathbb{I}(\cdot)$ is the indicative function that takes 1 when $\mathbf{z}_i^v$ is the representation of i -th instance $\mathbf{X}_i$ observed in view v , and 0 otherwise. We set a completeness statistical weight w_i^v based on the number of observations in V views for instance $\mathbf{X}_i$ as the fusion weight, which makes use of view-specific complementary information without discarding unpaired observations and mitigating the negative impact of missing noise by adapting to the differences in missing instances across different views.

We derive a set of consensus semantic prototypes $\mathbf{C} = \{\mathbf{c}_k \in \mathbb{R}^d\}_{k=1}^K$ via k -means on consensus representations $\mathbf{Z}$. As “a representative embedding for semantically similar observations”, $\mathbf{C}$ comprehensively captures the semantic information of all data [50] and is continuously refined throughout consensus representation learning.

Prototype-based Contrastive Clustering. We perform semantic-level contrastive clustering in a shared vector space spanned by consensus prototypes, where the latent representation $\mathbf{z}_i^v$ encoded as the semantic representation $\mathbf{h}_i^v$ and $\mathbf{h}_i^v$ is assigned to the prototype $\mathbf{c}_k$ with the longest projected distance to obtain optimal assignments, in two steps:

- Semantic similarity measure: Encode the semantic representation $\mathbf{h}_i^v$ and project it onto each prototype, then calculate the probability $p_{i,k}^v$ of belonging to cluster k :

$$p_{i,k}^v = \frac{\exp(\mathbf{h}_i^{v\top} \mathbf{c}_k / \tau)}{\sum_{k'=1}^K \exp(\mathbf{h}_i^{v\top} \mathbf{c}_{k'} / \tau)}, \quad (4)$$

where τ is a temperature parameter [40]. $\mathbf{p}_i^v = \{p_{i,k}^v\}_{k=1}^K$ is the soft assignments of observation $\mathbf{x}_i^v$.

- Swapped knowledge distillation: Based on Theorem 2, $(\bar{\mathbf{x}}_i^m, \bar{\mathbf{x}}_i^n)$ share the same cluster semantics on consensus prototypes $\mathbf{C}$, thus their true labels $\mathbf{y}_i^m, \mathbf{y}_i^n \in \mathbb{R}^K$ should be are distributionally consistent. To this end, “swapped” knowledge distillation (KD) utilizes pseudo-labels $\mathbf{q}_i^m, \mathbf{q}_i^n$ as mutual supervised signals to prompt their cluster assignments $\mathbf{p}_i^m, \mathbf{p}_i^n$ as identical as possible:

$$\ell_{cc}^{m,n} = \ell_{kd}(\mathbf{H}^m, \mathbf{Q}^n) + \ell_{kd}(\mathbf{H}^n, \mathbf{Q}^m), \quad (5)$$

where $\ell_{kd}(\mathbf{H}^m, \mathbf{Q}^n) = -\frac{1}{N_{mn}} \sum_{i=1}^{N_{mn}} \sum_{k=1}^K \mathbf{q}_i^n \log \mathbf{p}_i^m$. We

extend Eq. (5) to more than two views, fostering greater view collaboration to enhance semantic learning:

$$\mathcal{L}_{cc} = \sum_{m=1}^V \sum_{\substack{n=1 \\ n \neq m}}^V \ell_{cc}^{m,n}. \quad (6)$$

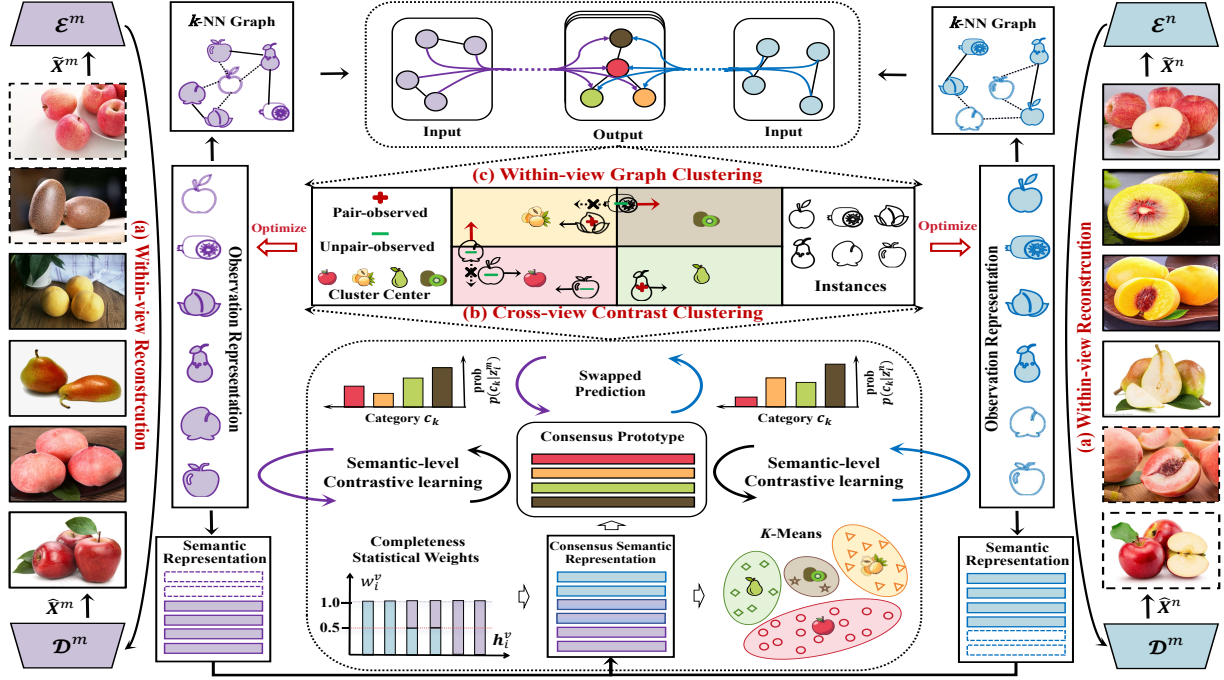


Figure 2. The framework of FreeCSL. (a) Reconstruction module, encodes observations into clustering-friendly representations for each view; (b) Consensus semantic learning module, learns semantic representations through cross-view contrastive clustering in a shared semantic space, where paired observations are assigned to their nearest semantic prototype for consistent assignments. (c) Cluster semantic enhancement module, enriches semantic representations with cluster structure information through within-view graph clustering, which applies GCN to aggregate view-specific semantic information and maximizes spectral modularity to recover cluster structure with greater separation and lower entropy. Ultimately, perform k -means on the consensus semantic representations to predict cluster labels.

- Consensus label solution: To obtain pseudo-labels $\mathbf{Q}^m = \{\mathbf{q}_i^m\}_{i=1}^{N_{mn}}$, project semantic representations $\mathbf{H}^m$ onto consensus prototypes $\mathbf{C}$ and maximize similarity between the representation $\mathbf{h}_i^m$ and its assigned prototype $\mathbf{c}_k$:

$$\mathbf{Q}^* = \arg \max_{\mathbf{Q}^m \in \mathcal{Q}} \text{Tr}\{(\mathbf{C}\mathbf{Q}^m)^\top \mathbf{H}^m\} + \alpha \mathcal{R}(\mathbf{Q}^m), \quad (7)$$

where $\mathcal{R}(\mathbf{Q}^m) = -\sum_i^{N_{mn}} \mathbf{q}_i^m \log \mathbf{q}_i^m$, an entropy regularizer prevents trivial solutions via the smoothness α .

2.4. Within-view Cluster Semantic Enhancement

Considering that graph pooling can discover data implicit correlation and reduce missing noise via aggregating the information of neighboring nodes, we design a graph clustering to enhance representations with cluster semantic information. It combines a heuristic learning objective based on modularity, an evaluation metric for cluster structure quality, to effectively recover the cluster structure[51].

Modularity. Modularity [28] quantifies the deviation between the actual cluster structure and the expected structure generated by random combinations in the null model. Modularity matrix $\mathbf{B}$ is defined as $\mathbf{B} = \mathbf{A} - \frac{\mathbf{d}\mathbf{d}^\top}{2m}$, where $\mathbf{A}$ is the adjacency matrix, $\mathbf{d}$ is the degree vector and m is the total number of edges in the graph. The entry $b_{ij} \in \mathbf{B}$ measures connection strength between nodes i and j .

Modularity-based Graph clustering. Graph clustering relies on graph pooling and node assignment. For each view, we construct graph structures with k -nearest neighbors (KNN), then deploy graph convolutional network (GCN) $\mathcal{G}_v(\cdot)$ to aggregate graph embeddings:

$$\mathcal{G}_v(\mathbf{Z}^v; \mathbf{A}^v) = \sigma(\mathbf{D}^{v-\frac{1}{2}} \mathbf{A}^v \mathbf{D}^{v-\frac{1}{2}} \mathbf{Z}^v \mathbf{W}^v + \mathbf{Z}^v \mathbf{W}_s^v), \quad (8)$$

where $\sigma(\cdot)$ is an activation function used to reinforce the nonlinear aggregation capability of GCN. We replace a skip connection matrix $\mathbf{W}_s^v$ with the self-loop $\mathbf{A}^v + \mathbf{I}$ to alleviate the over-smoothing of node embeddings. Graph pooling operates on the latent representations $\mathbf{Z}^v$, while the adjacency matrix $\mathbf{A}^v$ is constructed from the original observations $\mathbf{X}^v$:

$$a_{i,j}^v = \begin{cases} 1, & \text{if } (\mathbf{x}_i^v \text{ and } \mathbf{x}_j^v \neq \text{NaN}) \& \\ & (\mathbf{x}_i^v \in \mathcal{N}_k(\mathbf{x}_j^v) \text{ or } \mathbf{x}_j^v \in \mathcal{N}_k(\mathbf{x}_i^v)), \\ 0, & \text{otherwise,} \end{cases} \quad (9)$$

where $\mathcal{N}_k(\mathbf{x}_i^v)$ is a set including the k -nearest neighbors of $\mathbf{x}_i^v$. It brings neighbors closer in the embedding space without introducing missing noise into the graph structure.

The softmax classifier offers the soft assignment $\mathbf{P}^v$ for node embeddings encoded by $\mathcal{G}_v(\mathbf{Z}^v; \mathbf{A}^v)$:

$$\mathbf{P}^v = \text{Softmax}(\mathcal{G}_v(\mathbf{Z}^v; \mathbf{A}^v)). \quad (10)$$

To obtain a well-separated cluster assignment $\mathbf{P}^v$, we maximize the spectral modularity $\text{Tr}((\mathbf{P}^v)^\top \mathbf{B}^v \mathbf{P}^v)$ [1, 18], which captures intra-cluster connections and inter-cluster margins by mapping $\mathbf{P}^v$ to modularity $\mathbf{B}^v$; To further improve the confidence and robustness of $\mathbf{P}^v$, we apply a self-supervised signal $\mathbf{L}^v$, learned from contrastive clustering as detailed in Sec. 2.3, to guide $\mathbf{P}^v$ in chasing $\mathbf{L}^v$ via self-knowledge distillation. Specifically, the following objective is designed to optimize $\mathbf{P}^v$:

$$\ell_m^v(\mathbf{P}^v; \mathbf{L}^v) = -\frac{1}{2m} \text{Tr}((\mathbf{P}^v)^\top \mathbf{B}^v \mathbf{P}^v) + \lambda \text{KL}(\mathbf{L}^v \parallel \mathbf{P}^v), \quad (11)$$

where Kullback-Leibler divergence $\text{KL}(\cdot)$ serves as a robust regularizer, leveraging λ to regulate the information flow of $\mathbf{L}^v \in \mathbb{R}^{N \times K}$ to ensure stable cluster performance. We project semantic representations $\mathbf{H}^v$ learned through contrastive clustering onto their k -means prototypes and introduce Student's t -distribution [41] as kernel to predict high-confidence labels $\mathbf{L}^v$ with a nearly uniform distribution:

$$l_{i,k}^v = \frac{\left(1 + \frac{\|\mathbf{h}_i^v - \mathbf{c}_k^v\|^2}{\gamma}\right)^{-\frac{\gamma+1}{2}}}{\sum_{k'} \left(1 + \frac{\|\mathbf{h}_i^v - \mathbf{c}_{k'}^v\|^2}{\gamma}\right)^{-\frac{\gamma+1}{2}}} \in \mathbf{L}^v. \quad (12)$$

We conduct graph clustering on each view separately, then optimize assignments across all views concurrently:

$$\mathcal{L}_{gc} = \sum_{v=1}^V \ell_m^v(\mathbf{P}^v, \mathbf{L}^v). \quad (13)$$

2.5. Clustering Driven by Consensus Semantics

Our FreeCSL learns the semantic representations $\mathbf{H}^v$ end-to-end by jointly minimizing the reconstruction loss, contrastive clustering loss and graph clustering loss:

$$\mathcal{L} = \mathcal{L}_{rec} + \mathcal{L}_{cc} + \mathcal{L}_{gc}. \quad (14)$$

Without searching for the optimal balancing weights of three loss terms, outstanding performance is easily achieved by applying k -means to the consensus semantic representations $\mathbf{H}$ learned via the fusion manner $\mathbb{T}(\{\mathbf{H}^v\}_{v=1}^V)$.

The implementation for FreeCSL consists of two stages: warm-up training for the encoder-decoder, followed by fine-tuning for semantic representation learning and clustering. For details, refer to Algorithm 1.

3. Experiment

3.1. Experimental Settings

We select four representative datasets namely Caltech-5V[44], ALOI-100[5], YoutubeFace10[12] and NoisyMNIST[23] to demonstrate our model's performance

Algorithm 1 FreeCSL for Learning and Clustering

Input: Complete, incomplete and pair-observed multi-view dataset $\{\mathbf{X}^v\}_{v=1}^V$, $\{\tilde{\mathbf{X}}^v\}_{v=1}^V$, $\{\bar{\mathbf{X}}^{m,n}\}_{m \neq n}^V$; networks $\{\mathcal{E}_v, \mathcal{D}_v, \mathcal{G}_v\}_{v=1}^V$; warm-up and fine-tuning epochs e, E .

for $t = 1$ **to** e **do**

 On $\{\tilde{\mathbf{X}}^v\}_{v=1}^V$, warming up $\{\mathcal{E}_v, \mathcal{D}_v\}_{v=1}^V$ with Eq.(1).

for $t = 1$ **to** E **do**

 Obtain latent representations $\{\mathbf{Z}^v | \mathcal{E}_v : \mathbf{X}^v \rightarrow \mathbf{Z}^v\}_{v=1}^V$;
 Construct consensus prototypes $\mathbf{C} = \{\mathbf{c}_k\}_{k=1}^K$ as elaborated in Sec. 2.3.

 Compute the reconstruction loss $\mathcal{L}_{rec}$ with Eq.(1).

 // Cross-view Contrastive Clustering

for $m, n = 1$ **to** V ($m \neq n$) **do**

 Compute the cluster probability $\mathbf{p}_i^m, \mathbf{p}_i^n$ with Eq.(4);

 Solve consensus pseudo-label $\mathbf{q}_i^m$ with Eq.(7);

 Get swapped loss $\ell_{cc}^{m,n}$ with Eq.(5) on $\bar{\mathbf{X}}^{m,n}$.

 Calculate contrastive clustering loss $\mathcal{L}_{cc}$ with Eq.(6)

 // Within-view Graph Clustering

for $v = 1$ **to** V **do**

 Construct adjacency matrix $\mathbf{A}^v$ with Eq.(9).

 Solve node assignment $\mathbf{P}^v$ with Eq.(10)

 Solve pseudo-labels $\mathbf{L}^v$ with Eq.(12);

 Compute KL-modularity loss ℓ_m^v with Eq.(11)

 Calculate graph clustering loss $\mathcal{L}_{gc}$ with Eq.(13)

 // Semantic Representation Learning

 Calculate the overall loss $\mathcal{L}$ with Eq.(14);

 Optimize $\{\mathcal{E}_v, \mathcal{D}_v, \mathcal{G}_v\}_{v=1}^V$ to minimize $\mathcal{L}$;

 Learn latent and consensus representations $\{\mathbf{Z}^v\}_{v=1}^V$ and $\mathbf{Z}$, semantic representations $\{\mathbf{H}^v\}_{v=1}^V$, consensus semantic representations $\mathbf{H}$, consensus prototypes $\mathbf{C}$.

Output: $\hat{\mathbf{Y}} = \{\hat{y}_i\}_{i=1}^N$ predicted by k -means on $\mathbf{H}$.

in comparison with seven SOTA methods summarized in Table 1. To comprehensively evaluate experimental results, three metrics are adopted: accuracy (ACC), normalized mutual information (NMI), and adjusted rand index (ARI).

Table 1. SOTA methods categorized by the types of techniques for consistency, imputation, and alignment.

Competitors	Consistency	Imputation	Alignment
CPM-Nets (TPAMI'20)	instance-level	mutual information interaction	\
COMPLETER (CVPR'21)	instance-level	mutual information interaction	
DIMVC (AAAI'22)	instance-level	\	assignment-based
SURE (TPAMI'23)	cluster-level		
Prolmp (IJCAI'23)	instance-level	sample-prototype relationship inheritance	prototype-based assignment-based
ICMVC (AAAI'24)	instance-level	graph structure transfer	
DIVIDE (AAAI'24)	cluster-level	mutual information interaction	

3.2. Implementation details

Models are trained on an NVIDIA 3090 GPU using Adam optimizer in PyTorch 2.1.0. The warming up and fine-tuning learning rates are 0.0003 and 0.0005, with a batch size of 512. The hyperparameters are set as follows for

different datasets: smoothness $\alpha = 0.5$, temperature $\tau \in \{0.1, 0.2\}$, neighbors $\zeta = 3$, and regularizer weight $\lambda \in \{0.05, 0.1, 0.2, 0.3\}$. For all datasets, 4-layer autoencoders and 2-layer GCNs with same MLP structures are used for each view. The layers of contrastive clustering and graph clustering are shared across all views via a single FC layer.

3.3. Competitiveness of FreeCSL

FreeCSL is compared with seven SOTA methods across three metrics in Table 2. The results indicate that FreeCSL excellently handles challenges of high missing rates, multiple clusters, and large-scale issues in IMVC:

- From the perspective of effectiveness, FreeCSL surpasses most SOTA models, particularly on challenging datasets like ALOI-100 (100 clusters) and YouTubeFace10 (30,000 samples). It achieves significant improvements in ACC, with gains of 15.12%, 26.38%, 23.91%, 26.21% for ALOI-100, and 1.16%, 8.81%, 6.32%, 1.68% for YouTubeFace10, at missing rate $r = 0.1, 0.3, 0.5, 0.7$.
- From the perspective of robustness, Fig. 3 (visualized from Table 2) shows our model’s stability as r increases. FreeCSL declines gradually, unlike other models with sharp ACC drops, highlighting its robustness in IMVC task. On the small dataset Caltech-5V with $r = 0.7$, its Acc remains at 84.56%, while on the larger dataset NoisyMNIST, also achieves 92.19%.

The above phenomenon can be explained as follows: As r increases, paired observations become scarce. Methods like instance-level consistency learning or cross-view imputation and alignment (e.g., CPM-Net, COMPLETER, ICMVC), constrained by their dependence on consistency information, will introduce biases such as false negatives and semantic inconsistencies, resulting in degraded performance. For the complex task ALOI-100, cluster-level methods like SURE and DIVIDE surpass instance-level by exploiting cross-view cluster consistency. But, their reliance on cross-view graphs or mutual information for imputation, while neglecting within-view cluster affiliations, restricts performance. ProImp acknowledges the need to integrate within-view and cross-view information for semantically consistent imputation. However, its prototype alignment accumulates errors, particularly with a large number of clusters, achieving only 20% ACC on ALOI-100.

The superiority of FreeCSL stems from the synergistic effect of the CSL and CSE modules, which eliminate semantic gaps and ensure consistent assignments for all observations in a shared semantic space.

3.4. Understanding FreeCSL

Ablation Study. As shown in Table 3, ablation studies on FreeCSL’s three components reveal that the CSL module contributes the most. On ALOI-100, it demonstrates exceptional performance and strong stability, with ACC drop-

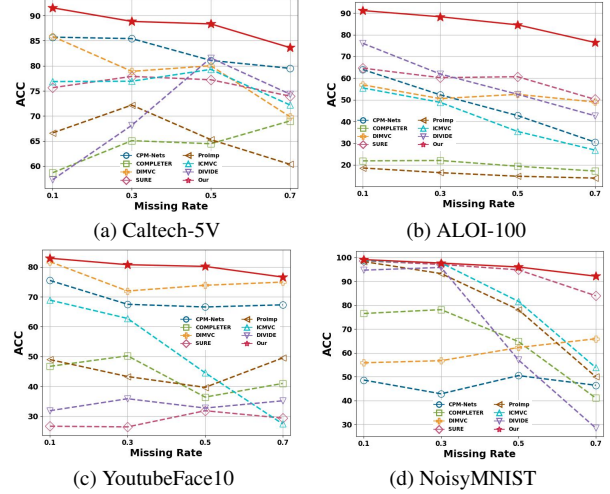


Figure 3. Visualization for Table 2 based on metric ACC.

ping only 12.4% as r rises from 0.1 to 0.7. This is attributed to CSL’s establishment of a shared semantic space via prototype-based semantic contrast, reducing semantic gaps within clusters. The CSE module also plays a positive role in discovering tighter cluster structures by optimizing modularity. Thanks to its enhanced semantics, both the REC and CSL modules achieve notable improvements.

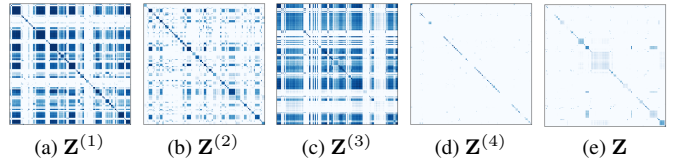


Figure 4. Similarity matrices of $\{Z^v\}_{v=1}^4, Z$ without consensus semantic learning on ALOI-100 with $r = 0.5$.

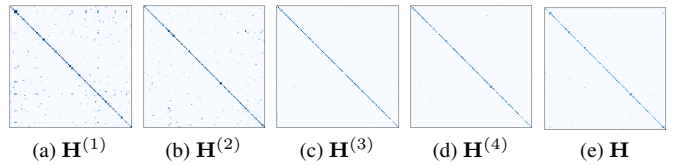


Figure 5. Similarity matrices of $\{H^v\}_{v=1}^4, H$ with consensus semantic learning on ALOI-100 with $r = 0.5$.

Imputation- and Alignment-free CSL. To verify FreeCSL can reduce semantic gaps and capture semantic relationships without requiring imputation or alignment, we set up two imputation experiments: one imputation for latent representations Z^v without consensus semantic learning named ILR, and the second for semantic representations H^v learned from consensus semantic learning named ISR. They transfer complete k -nn graph structures from other views to incomplete views and utilize corresponding k neighbors for imputation. The control groups perform k -means on the sum of all view matrices to predict cluster

Table 2. Performance comparison on four multi-view benchmarks. The best and second-best results are highlighted in red and blue.

Missing rates		$r = 0.1$			$r = 0.3$			$r = 0.5$			$r = 0.7$		
Metrics		ACC (%)	NMI (%)	ARI (%)	ACC (%)	NMI (%)	ARI (%)	ACC (%)	NMI (%)	ARI (%)	ACC (%)	NMI (%)	ARI (%)
Caltech-5V	CPM-Nets	85.72	74.62	71.92	85.43	72.90	70.84	81.08	68.61	66.08	79.50	65.52	62.70
	COMPLETER	58.68	67.69	42.98	65.07	68.75	47.10	64.46	66.09	42.80	69.00	67.46	51.49
	DIMVC	85.86	80.01	74.63	78.90	78.76	69.55	80.01	70.08	66.35	69.86	60.16	48.70
	SURE	75.64	69.12	62.83	77.89	67.89	61.15	77.22	65.99	60.99	73.96	60.89	53.66
	ProImp	66.62	57.66	47.54	72.17	61.70	52.49	65.28	55.77	46.22	60.36	49.23	39.63
	ICMVC	76.86	73.92	67.37	76.93	73.39	66.30	79.29	72.67	66.21	72.21	66.63	58.18
	DIVIDE	57.29	47.00	33.23	68.14	58.93	44.16	81.57	71.45	59.20	74.36	68.62	60.14
	Ours	91.57	85.32	83.14	88.86	81.26	78.68	88.36	80.01	78.02	83.64	71.79	68.06
ALOI-100	CPM-Nets	63.95	79.14	51.26	52.30	70.74	39.79	42.74	62.46	28.40	30.52	54.03	17.18
	COMPLETER	21.86	46.94	12.10	22.04	46.45	11.62	19.38	44.81	10.78	17.18	43.68	9.60
	DIMVC	56.90	75.00	35.59	50.55	72.51	28.10	52.54	71.45	32.56	48.97	70.80	26.16
	SURE	64.59	77.74	52.61	60.22	74.76	47.21	60.65	74.84	47.39	50.23	68.91	37.48
	ProImp	26.15	52.94	31.40	24.52	53.07	28.18	27.97	58.10	33.42	20.33	50.10	24.28
	ICMVC	55.48	74.04	42.91	48.83	67.46	36.50	35.42	56.98	24.19	26.84	47.43	16.63
	DIVIDE	76.01	88.59	70.59	61.90	80.53	53.29	52.56	73.17	40.57	42.66	67.82	30.93
	Ours	91.13	95.22	87.59	88.28	93.34	83.95	84.56	90.53	78.90	76.44	85.00	67.68
YouTubeFace10	CPM-Nets	75.48	78.56	66.50	67.50	75.82	61.15	66.60	75.19	62.01	67.32	74.10	62.54
	COMPLETER	46.72	52.04	21.03	50.24	52.81	27.77	36.41	35.24	15.18	40.98	39.92	20.82
	DIMVC	81.77	81.32	74.59	71.96	76.42	62.29	73.87	75.44	63.97	74.94	79.34	68.96
	SURE	26.65	23.60	11.41	26.43	22.40	10.55	31.84	28.77	15.27	29.37	27.48	12.85
	ProImp	48.97	53.20	33.83	43.30	49.04	26.53	39.67	50.18	26.68	49.52	51.47	29.12
	ICMVC	68.90	74.17	60.06	62.76	65.44	53.03	44.56	52.51	32.51	27.44	27.09	15.33
	DIVIDE	31.85	29.64	11.35	35.86	30.81	12.31	32.74	30.82	12.74	35.17	33.47	15.93
	Ours	82.93	83.55	74.76	80.77	81.46	71.62	80.19	81.07	71.37	76.62	81.31	73.22
NoisyMNIST	CPM-Nets	48.64	44.08	32.00	42.88	37.01	26.92	50.52	43.54	33.56	46.46	37.18	27.04
	COMPLETER	76.55	87.95	77.18	78.16	86.11	76.83	64.79	65.77	54.68	41.15	40.60	27.58
	DIMVC	55.88	54.78	41.29	56.77	58.29	43.51	62.28	62.66	48.01	65.94	67.43	54.58
	SURE	98.61	95.89	96.96	97.11	93.75	91.64	94.85	89.06	87.89	84.00	78.25	76.02
	ProImp	98.30	95.31	96.35	93.29	86.05	85.38	78.16	69.45	64.37	50.04	41.33	33.34
	ICMVC	98.78	96.36	97.35	97.75	93.71	95.11	81.64	79.24	75.47	53.91	50.91	44.56
	DIVIDE	94.72	91.44	89.43	95.85	89.71	91.06	57.09	57.29	46.28	28.57	25.61	11.14
	Ours	99.13	97.23	98.10	97.68	93.94	94.94	96.04	89.81	91.48	92.19	82.50	83.56

Table 3. Ablation study on Caltech-5V and ALOI-100. $\checkmark$ denotes FreeCSL with the component and the best results are highlighted in red.

Components			$r = 0.1$			$r = 0.3$			$r = 0.5$			$r = 0.7$		
$\mathcal{L}_{rec}$ $\mathcal{L}_{cc}$ $\mathcal{L}_{gc}$			ACC (%)	NMI (%)	ARI (%)	ACC (%)	NMI (%)	ARI (%)	ACC (%)	NMI (%)	ARI (%)	ACC (%)	NMI (%)	ARI (%)
Caltech-5V	$\checkmark$		75.00	65.64	58.89	69.00	57.84	51.49	52.21	44.18	36.71	51.93	43.42	35.75
	$\checkmark$	$\checkmark$	85.50	78.07	72.28	82.00	74.53	67.67	85.36	74.32	71.10	79.14	67.81	62.78
	$\checkmark$		81.07	70.45	66.13	80.86	69.61	67.49	74.14	59.39	55.90	70.71	53.25	49.95
	$\checkmark$	$\checkmark$	91.57	85.32	83.14	88.86	81.26	78.68	88.36	80.01	78.02	83.64	71.79	68.06
ALOI-100	$\checkmark$		63.81	77.92	49.38	44.69	65.64	31.11	32.20	55.83	18.94	23.78	48.90	11.31
	$\checkmark$	$\checkmark$	87.69	93.64	84.12	85.25	91.02	79.45	80.23	88.04	74.15	75.29	83.38	65.72
	$\checkmark$		81.06	85.96	71.21	64.13	75.87	51.49	47.52	65.59	34.29	32.73	56.35	19.23
	$\checkmark$	$\checkmark$	91.13	95.22	87.59	88.28	93.34	83.95	84.56	90.53	78.90	76.44	85.00	67.68

Table 4. Imputation- and alignment-free study on Caltech-5V and ALOI-100. ILR and ISR are filled with K-NN imputation via cross-view graph for latent representations and semantic representations $\mathbf{Z}^{(v)}$, $\mathbf{H}^{(v)}$. The best results are highlighted in red.

Missing rates		$r = 0.1$			$r = 0.3$			$r = 0.5$			$r = 0.7$		
Metrics		ACC (%)	NMI (%)	ARI (%)	ACC (%)	NMI (%)	ARI (%)	ACC (%)	NMI (%)	ARI (%)	ACC (%)	NMI (%)	ARI (%)
Caltech-5V	ILR	91.71	85.76	83.44	89.14	81.69	79.07	88.79	80.79	78.37	77.86	65.64	55.12
	ISR	91.64	85.55	83.30	89.43	82.26	79.62	88.86	80.78	78.74	81.29	66.68	63.59
	FreeCSL	91.57	85.32	83.14	88.86	81.26	78.68	88.36	80.01	78.02	83.64	71.79	68.06
ALOI-100	ILR	58.92	79.67	41.54	53.46	75.88	36.89	45.83	69.47	27.41	38.08	58.96	21.35
	ISR	90.16	95.38	87.86	88.91	94.08	85.48	82.53	88.73	71.54	65.36	77.85	46.06
	FreeCSL	91.13	95.22	87.59	88.28	93.34	83.95	84.56	90.53	78.90	76.44	85.00	67.68

labels. Table 4 shows FreeCSL’s robustness advantage at high r on Caltech-5V, with notable performance disparity on ALOI-100. ILR, lacking consensus semantic learning, struggles with multi-cluster tasks due to cross-view semantic gaps. ISR avoids semantic confusion via consensus semantic learning but underperforms FreeCSL due to biased

imputation as consistency information decreases at higher r . FreeCSL, by contrast, integrates consistency and complementary information in consensus semantic representations for confident decisions.

To further illustrate how consensus semantic learning “free up” imputation and alignment, we construct cosine

similarity matrices on ALOI-100 for view-specific latent representations $\{\mathbf{Z}^v\}_{v=1}^4$ and their consensus $\mathbf{Z}$, as well as view-specific semantic representations $\{\mathbf{H}^v\}_{v=1}^4$ and their consensus $\mathbf{H}$ in Fig. 5, then analyze their information entropy. The similarity matrices of $\{\mathbf{Z}^v\}_{v=1}^4$ are chaotic, with high uncertainty in intra- and inter- cluster relationships. The fusion manner $\mathbb{T}(\cdot)$ with Eq. (2) alleviates high entropy by incorporating view-specific complementary information. Compared to Fig. 4, the similarity matrices of $\{\mathbf{H}^v\}_{v=1}^4$ shows a clear block structure, with high intra-cluster similarity and distinct inter-cluster differences. This confirms $\mathbf{H}$, enhanced by view-specific information, reduces cross-view semantic gaps, while consensus semantic learning, capturing cluster semantic relationship, promotes high-confidence assignments as stated in Theorem 1.

3.5. Analysis on FreeCSL

Convergence and Robustness Analysis. In Fig. 6, we plot metrics and losses over training iterations, with error bands from 5 random Caltech-5V experiments to assess robustness. Both converge to stable values with minimal fluctuations and reach stability simultaneously. The gradual decrease in reconstruction loss indicates that the CSL and CSE modules are reasonable and beneficial, as they don't introduce significant discrepancies between semantic and latent representations, further underscoring the benefits of pre-training. Thanks to the harmonious collaboration of the three modules, our FreeCSL achieved satisfactory results within 30 iterations at minimal computational cost.

Parameter Sensitivity Analysis. Two hyperparameters in FreeCSL warrant investigation, namely, graph neighbors ζ in Eq. (9) and regularizer coefficient λ in Eq. (11). We expanded ζ and λ to the range of 0.05 to 0.5 and 3 to 32 respectively. In Fig. 7, our model is insensitive to ζ but more affected by variations in λ . This is because the robust regularizer leverages semantic knowledge learned from the CSL module to guide the CSE model toward faster and more stable learning. A small λ cannot provide effective consistency constraints while a large one makes CSL to dominate the learn direction, hindering CSE from exploring cluster structure information. Therefore, we set $\lambda \in [0.05, 0.2]$ to balance cooperative relationship between CSE and CSL. Without sacrificing ACC and NMI, we prefer to set the minimum $\zeta = 3$ to reduce computational complexity.

3.6. Visualization of Consensus Semantic Clusters

To further examine the quality of cluster structures formed by consensus semantic representations, t-SNE visualization based on true labels are plotted in Fig. 8 and clearly show minimal misclassification and clusters with strong intra-cluster cohesion while distinct inter-cluster separation. This indicates FreeCSL nearly recovers true cluster structures by learning consensus semantic information from all data.

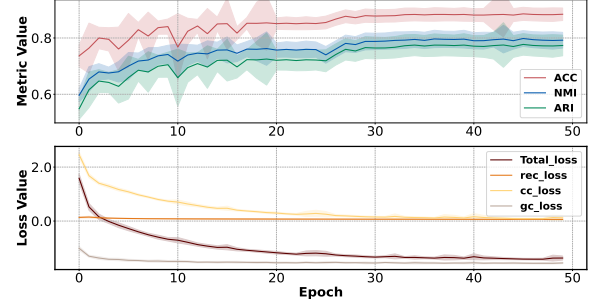


Figure 6. Convergence analysis on Caltech-5V with $r = 0.5$.

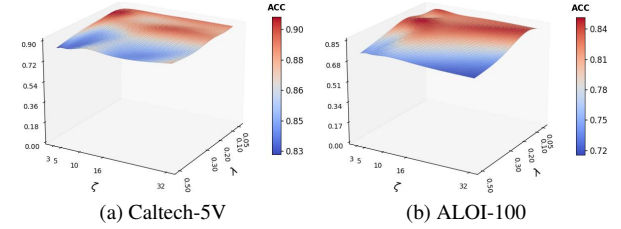


Figure 7. Parameter analyses for ζ and λ with $r = 0.5$.

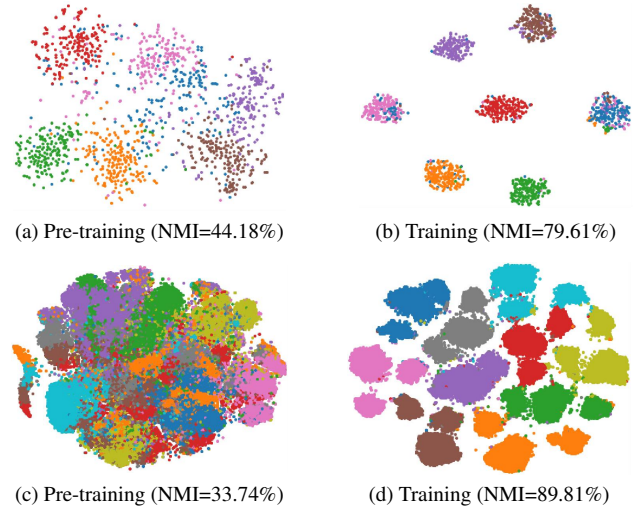


Figure 8. Visualization for Caltech-5V and NoisyMNIST.

4. Conclusion

In this paper, we propose FreeCSL, a novel semantic learning paradigm free from imputation and alignment compared to existing consistency learning. We design prototype-based contrastive clustering to discover a shared semantic space, where observations converge toward their respective semantic prototype and are encoded as consensus semantics representations for clustering. Furthermore, we employ modularity-inspired graph clustering to enrich semantic representation with view-specific cluster information. The effective synergy of consensus semantic learning and cluster semantic enhancement makes FreeCSL excel in most complex IMVC tasks.

Acknowledgments

This work is supported by the National Key R&D Program of China under Grant No.2022ZD0209103, the National Natural Science Foundation of China (project no. 62325604, 62276271, 62406329, 62476281, 62441618), National Natural Science Foundation of China Joint Found under Grant No. U24A20323.

References

- [1] Ulrik Brandes, Daniel Delling, Marco Gaertler, Robert Görke, Martin Hoefer, Zoran Nikoloski, and Dorothea Wagner. Maximizing modularity is hard. *arXiv preprint physics/0608255*, 2006. **5**
- [2] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *Advances in neural information processing systems*, 33:9912–9924, 2020. **2**
- [3] Chenhang Cui, Yazhou Ren, Jingyu Pu, Jiawei Li, Xiaorong Pu, Tianyi Wu, Yutao Shi, and Lifang He. A novel approach for effective multi-view clustering with information-theoretic perspective. *Advances in Neural Information Processing Systems*, 36, 2024. **1**
- [4] Zhibin Dong, Jiaqi Jin, Yuyang Xiao, Bin Xiao, Siwei Wang, Xinwang Liu, and En Zhu. Subgraph propagation and contrastive calibration for incomplete multiview data clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2024. **1, 2**
- [5] 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. **5**
- [6] 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. **1**
- [7] Wei Feng, Guoshuai Sheng, Qianqian Wang, Quanyue Gao, Zhiqiang Tao, and Bo Dong. Partial multi-view clustering via self-supervised network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 11988–11995, 2024. **1**
- [8] Ruiming Guo, Mouxiang 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. **2**
- [9] 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*, pages 4167–4176, 2024. **2**
- [10] Junlin Hu, Jiwen Lu, and Yap-Peng Tan. Sharable and individual multi-view metric learning. *IEEE transactions on pattern analysis and machine intelligence*, 40(9):2281–2288, 2017. **3**
- [11] Peng Hu, Liangli Zhen, Xi Peng, Hongyuan Zhu, Jie Lin, Xu Wang, and Dezhong Peng. Deep supervised multi-view learning with graph priors. *IEEE Transactions on Image Processing*, 33:123–133, 2023. **1**
- [12] Dong Huang, Chang-Dong Wang, and Jian-Huang Lai. Fast multi-view clustering via ensembles: Towards scalability, superiority, and simplicity. *IEEE Transactions on Knowledge and Data Engineering*, 35(11):11388–11402, 2023. **5**
- [13] Jiabo Huang, Shaogang Gong, and Xiatian Zhu. Deep semantic clustering by partition confidence maximisation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8849–8858, 2020. **1**
- [14] 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. **1**
- [15] Yangbangyan Jiang, Qianqian Xu, Zhiyong Yang, Xiaochun Cao, and Qingming Huang. Dm2c: Deep mixed-modal clustering. *Advances in Neural Information Processing Systems*, 32, 2019. **2**
- [16] Jiaqi Jin, Siwei Wang, Zhibin Dong, Xinwang Liu, and En Zhu. Deep incomplete multi-view clustering with cross-view partial sample and prototype alignment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11600–11609, 2023. **1, 2**
- [17] Guanzhou Ke, Bo Wang, Xiaoli Wang, and Shengfeng He. Rethinking multi-view representation learning via distilled disentangling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26774–26783, 2024. **1**
- [18] Brian W Kernighan and Shen Lin. An efficient heuristic procedure for partitioning graphs. *The Bell system technical journal*, 49(2):291–307, 1970. **5**
- [19] Haobin Li, Yunfan Li, Mouxiang Yang, Peng Hu, Dezhong Peng, and Xi Peng. Incomplete multi-view clustering via prototype-based imputation. *arXiv preprint arXiv:2301.11045*, 2023. **1, 2**
- [20] Xingfeng Li, Yinghui Sun, Quansen Sun, Zhenwen Ren, and Yuan Sun. Cross-view graph matching guided anchor alignment for incomplete multi-view clustering. *Information Fusion*, 100:101941, 2023. **1**
- [21] Yingming Li, Ming Yang, and Zhongfei Zhang. A survey of multi-view representation learning. *IEEE transactions on knowledge and data engineering*, 31(10):1863–1883, 2018. **1**
- [22] Yunfan Li, Peng Hu, Zitao Liu, Dezhong Peng, Joey Tianyi Zhou, and Xi Peng. Contrastive clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8547–8555, 2021. **2**
- [23] Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. Completer: Incomplete multi-view clustering via contrastive prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11174–11183, 2021. **5**
- [24] Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4447–4461, 2022. **1**
- [25] Xinwang Liu, Xinzhong Zhu, Miaomiao Li, Lei Wang, Chang Tang, Jianping Yin, Dinggang Shen, Huaimin Wang,

- and Wen Gao. Late fusion incomplete multi-view clustering. *IEEE transactions on pattern analysis and machine intelligence*, 41(10):2410–2423, 2018. [1](#)
- [26] Yiding Lu, Haobin Li, Yunfan Li, Yijie Lin, and Xi Peng. A survey on deep clustering: from the prior perspective. *Vicinityearth*, 1(1):4, 2024. [2](#)
- [27] Yiding Lu, Yijie Lin, Mouxing Yang, Dezhong Peng, Peng Hu, and Xi Peng. Decoupled contrastive multi-view clustering with high-order random walks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 14193–14201, 2024. [2](#)
- [28] Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582, 2006. [4](#)
- [29] Jingyu Pu, Chenhang Cui, Xinyue Chen, Yazhou Ren, Xiaorong Pu, Zhifeng Hao, S Yu Philip, and Lifang He. Adaptive feature imputation with latent graph for deep incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 14633–14641, 2024. [1](#)
- [30] Yuming Shen, Ziyi Shen, Menghan Wang, Jie Qin, Philip Torr, and Ling Shao. You never cluster alone. *Advances in Neural Information Processing Systems*, 34:27734–27746, 2021. [2](#)
- [31] Yuan Sun, Yang Qin, Yongxiang Li, Dezhong Peng, Xi Peng, and Peng Hu. Robust multi-view clustering with noisy correspondence. *IEEE Transactions on Knowledge and Data Engineering*, 2024. [2](#)
- [32] Huayi Tang and Yong Liu. Deep safe multi-view clustering: Reducing the risk of clustering performance degradation caused by view increase. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 202–211, 2022. [1](#), [2](#)
- [33] Jingjing Tang, Qingqing Yi, Saiji Fu, and Yingjie Tian. Incomplete multi-view learning: Review, analysis, and prospects. *Applied Soft Computing*, page 111278, 2024. [1](#)
- [34] Xinhang Wan, Jiyuan Liu, Xinbiao Gan, Xinwang Liu, Siwei Wang, Yi Wen, Tianjiao Wan, and En Zhu. One-step multi-view clustering with diverse representation. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–13, 2024. [1](#)
- [35] Xinhang Wan, Bin Xiao, Xinwang Liu, Jiyuan Liu, Weixuan Liang, and En Zhu. Fast continual multi-view clustering with incomplete views. *IEEE Transactions on Image Processing*, 33:2995–3008, 2024. [1](#)
- [36] Haoqing Wang, Xun Guo, Zhi-Hong Deng, and Yan Lu. Rethinking minimal sufficient representation in contrastive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16041–16050, 2022. [1](#)
- [37] Qianqian Wang, Zhengming Ding, Zhiqiang Tao, Quanxue Gao, and Yun Fu. Generative partial multi-view clustering with adaptive fusion and cycle consistency. *IEEE Transactions on Image Processing*, 30:1771–1783, 2021. [1](#), [2](#)
- [38] Siwei Wang, Xinwang Liu, Suyuan Liu, Jiaqi Jin, Wenxuan Tu, Xinzhong Zhu, and En Zhu. Align then fusion: Generalized large-scale multi-view clustering with anchor matching correspondences. *Advances in Neural Information Processing Systems*, 35:5882–5895, 2022. [2](#)
- [39] Yiming Wang, Dongxia Chang, Zhiqiang Fu, Jie Wen, and Yao Zhao. Incomplete multi-view clustering via cross-view relation transfer. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022. [2](#)
- [40] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3733–3742, 2018. [3](#)
- [41] Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*, pages 478–487. PMLR, 2016. [5](#)
- [42] Cai Xu, Ziyu Guan, Wei Zhao, Hongchang Wu, Yunfei Niu, and Beilei Ling. Adversarial incomplete multi-view clustering. In *IJCAI*, pages 3933–3939, 2019. [2](#)
- [43] Jie Xu, Chao Li, Yazhou Ren, Liang Peng, Yujie Mo, Xiaoshuang Shi, and Xiaofeng Zhu. Deep incomplete multi-view clustering via mining cluster complementarity. In *Proceedings of the AAAI conference on artificial intelligence*, pages 8761–8769, 2022. [3](#)
- [44] Jie Xu, Huayi Tang, Yazhou Ren, Liang Peng, Xiaofeng Zhu, and Lifang He. Multi-level feature learning for contrastive multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16051–16060, 2022. [5](#)
- [45] Jie Xu, Chao Li, Liang Peng, Yazhou Ren, Xiaoshuang Shi, Heng Tao Shen, and Xiaofeng Zhu. Adaptive feature projection with distribution alignment for deep incomplete multi-view clustering. *IEEE Transactions on Image Processing*, 32:1354–1366, 2023. [2](#)
- [46] Xiaoqiang Yan, Zhixiang Jin, Fengshou Han, and Yangdong Ye. Differentiable information bottleneck for deterministic multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27435–27444, 2024. [1](#)
- [47] 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*, pages 1134–1143, 2021. [2](#)
- [48] 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*, pages 1134–1143, 2021. [2](#)
- [49] Mouxing Yang, Yunfan Li, Peng Hu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Robust multi-view clustering with incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):1055–1069, 2022. [2](#)
- [50] Shengju Yu, Suyuan Liu, Siwei Wang, et al. Sparse low-rank multi-view subspace clustering with consensus anchors and unified bipartite graph. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. [3](#)
- [51] Shengju Yu, Siwei Wang, Zhibin Dong, et al. A non-parametric graph clustering framework for multi-view data. In *Proceedings of the AAAI conference on artificial intelligence*, pages 16558–16567, 2024. [4](#)

- [52] Pengxin Zeng, Mouxing Yang, Yiding Lu, Changqing Zhang, Peng Hu, and Xi Peng. Semantic invariant multi-view clustering with fully incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(4):2139–2150, 2023. [2](#)
- [53] Changqing Zhang, Yajie Cui, Zongbo Han, Joey Tianyi Zhou, Huazhu Fu, and Qinghua Hu. Deep partial multi-view learning. *IEEE transactions on pattern analysis and machine intelligence*, 44(5):2402–2415, 2020. [2](#)
- [54] Yi Zhang, Xinwang Liu, Siwei Wang, Jiyuan Liu, Sisi Dai, and En Zhu. One-stage incomplete multi-view clustering via late fusion. In *Proceedings of the 29th ACM international conference on multimedia*, pages 2717–2725, 2021. [1](#)
- [55] Yi Zhang, Fengyu Tian, Chuan Ma, Miaomiao Li, Hengfu Yang, Zhe Liu, En Zhu, and Xinwang Liu. Regularized instance weighting multiview clustering via late fusion alignment. *IEEE Transactions on Neural Networks and Learning Systems*, 2024. [3](#)
- [56] Huasong Zhong, Chong Chen, Zhongming Jin, and Xian-Sheng Hua. Deep robust clustering by contrastive learning. *arXiv preprint arXiv:2008.03030*, 2020. [2](#)
- [57] Lihua Zhou, Guowang Du, Kevin Lü, Lizheng Wang, and Jingwei Du. A survey and an empirical evaluation of multi-view clustering approaches. *ACM Computing Surveys*, 56(7):1–38, 2024. [1](#)