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

Supplementary Material

1. Appendix A: Related Work

1.1. Contrastive Learning for Consistency Learning

Exploring consistency information from complete instances across views is an effective way to alleviate instance observations missing and cluster distribution shifted in incomplete multi-view clustering (IMVC). Contrastive learning [2, 5, 21, 29], as an unsupervised representation learning [14, 22, 26], can learn the structural consistency information from multi-view data bring closer instances from positive samples and separate instances from negative samples [2, 5, 17, 21, 29], and has been successfully extended to multi-view clustering (MVC) task.

Specifically, the most widely applied contrastive learning paradigms construct positive and negative pairs at the instance-level. Despite instance-level paradigm have shown exceptional capability in consistency representation learning, two primary limits, false negative noise from intra-cluster observations for different instances and the local smoothness of instance representations, damage representation learning due to the loss of view-specific information.

After all, clustering is a one-to-many mapping. Recognizing false negative pairs (FPNs) causes detrimental impacts on clustering confidence and robustness, a cluster-level paradigm is proposed to discover cross-view cluster correspondences for intra-cluster but unpaired observations by reducing FPNs: TCL [11] selects pseudo-labels with confidence-based criteria to mitigate false negative impacts, while the noise-robust contrastive loss proposed by SURE [28] further discriminate false negative pairs by using a adaptive threshold calculated from distances of all positive and negative pairs. DIVIDE [15] utilizes an anchor-based approach to identify out-of-domain samples through high-order random walks to mitigate the issue of false negatives. They resolve the confusion of cross-view cluster correspondences caused by instance-level paradigms, but at the cost of cluster information within specific views, which hinders semantic consistency in representation learning.

1.2. Imputation and Alignment for IMVC

In IMVC, to preserve even recover relationships between data, imputation are supposed to handle missing data. Regarding the former, typical approaches include the cross-view transfer paradigm like neighborhood-based recovery, the cross-view interaction paradigm like adversarial generation or contrastive prediction. As members of transfer paradigm, the core idea of CRTC [19] and ICMVC [1] is to transfer the complete graph neighborhood relations from other views to missing views. However, neighborhood-based recovery, which uses cross-view neighbor information for imputation, overlooks complementary information specific to each view. To improve imputation performance, generative models such as autoencoders (AE) and generative adversarial networks (GAN), as well as discriminative models like contrastive learning, discover correlations across multi-view data to dynamically collaborate on both imputation and clustering. For examples, [20], [27] and [8] leverage the power of AEs in encoding latent representations to mine view-specific information for imputation; CPM-Nets [30] and GP-MVC [18] encode a common representation with consistency and complementarity information across views and employ adversarial strategies to reconstruct the common representation to approximate generated observations within views; COMPLETER [12] and DCP [13] unify cross-view consistency learning and missing prediction into a deep framework to constrain both complete paired observations and incomplete recovered observations by maximizing mutual information and minimizing conditional entropy across views. Although they successfully apply view-specific information in imputation, they lose the cluster structure information within the missing views. Thus, ProImp [10] proposed a novel paradigm based on within-view prototypes and cross-view observation-prototype relationships to further improve imputation performance.

However, the aforementioned imputation methods are limited by unsupervised learning and cannot restore the original distribution of view data. To achieve confident and robust clustering, a feasible solution is cross-view consistency alignment, generally categorized into cross-view cluster assignments-based, prototypes-based and distributions-based as the following works: To integrate soft labels from various views for decision fusion, DIMVC [23] aligns view-specific labels with a unified label using conditional entropy loss. DSIMVC [16] argues that multi-view data share common semantic information, so a contrastive loss is designed to align cluster assignments across views for consistency. CPSPAN [9] and ProImp [10] employ Hungarian algorithm and bounded contrastive loss [10] to calibrate prototype-shifted across views. To reduce cross-view distribution discrepancy arising from complete and incomplete data, APADC [25] minimize the mean discrepancy loss to align view distributions in a common representation space. SPCC [3] directly optimizes the distribution alignment loss

of K cluster across views.

Whether imputation or alignment, there is a deviation compared to the original data, and this deviation increases rapidly as the amount of available complete data decreases. To this end, different other IMVC methods, our FreeCSL, a novel consensus semantic-based paradigm, discover the shared semantic space through consensus prototype-based contrastive clustering, where all available observations are encoded as representations with consensus semantics for clustering. More specifically, during consensus learning, all observations can straightforwardly reach consensus on cluster semantic information without imputation and alignment.

2. Appendix B: Theorem Proof

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.*

2.1. Proof of Theorem 1

Theorem 1. *Consensus semantic learning yields more confident and robust cluster assignments than instance- and cluster-level paradigms.*

Case 1: Instance-level paradigm pull paired observations $(\overline{\mathbf{x}}_i^m, \overline{\mathbf{x}}_i^n)$ closer and push unpaired observations $(\mathbf{x}_i^m, \mathbf{x}_j^n)$ apart. However, if $C(\mathbf{x}_i^m, \mathbf{x}_j^n) = 1$, intra-cluster but unpaired observations are treated as negative pairs, introducing false negative noise into clustering.

Case 2: Cluster-level paradigm encourages the observation $\mathbf{x}_i^m$ to find its cluster-level counterparts $\mathbf{x}_j^n$ from different view n to mitigate false negative noisy. However, lacking within-view clustering mapping for view-specific cluster information, it explores cross-view cluster correspondences but fails to ensure cluster semantics consistency within views.

Case 3: Semantic-level paradigm construct a shared semantic space based on consensus prototypes $\mathbf{C}$ for all observations to eliminate semantic gaps and capture semantic relationships within clusters.

Proof. Define a general consistency learning objective as

$$\max \sum_{m \neq n}^{V} \sum_{i}^{N} \sum_{j \neq j'}^{N} \{Y \rho^+(\mathbf{x}_i^m, \mathbf{x}_j^n) + (Y-1)\rho^-(\mathbf{x}_i^m, \mathbf{x}_{j'}^n)\}, \tag{1}$$

where $Y = 1/0$ mean positive/negative pairs, and $\rho^{+/-}$ measure the similarity between positive/negative pairs.

Instance-level paradigms: When $Y = I(\mathbf{x}_i^m, \mathbf{x}_j^n)$, the objective of instance-level paradigms f_{ic} is formulated as:

$$f_{ic} = \sum_{m \neq n}^{V} \sum_{i}^{N} [\rho^+(\mathbf{x}_i^m, \mathbf{x}_i^n) - \sum_{j \neq i}^{N} \rho^-(\mathbf{x}_i^m, \mathbf{x}_j^n)]. \tag{2}$$

When $C(\mathbf{x}_i^m, \mathbf{x}_j^n) = 1$, the instance-level paradigm incorrectly treats them as negative pairs, introducing false negative noise $\epsilon = \mathbb{P}(C = 1 | I = 0)$. $\mathbb{P}(C = 1 | I = 0)$ is false negative probability that is determined by the cross-view same-cluster probability and the quality of the cluster structure. It is defined as $\mathbb{P}(C = 1 | I = 0) = \frac{1}{K} + \beta r$, where β quantifies the negative impact of missing rate r on cluster structure quality.

Define the number of instance-level positive pairs N_{ip}, the number of instance-level negative pairs N_{in}, the number of false negative pairs in unpaired observations N_{fn} in views m, n as:

$$\begin{aligned}
N_{ip} &= \mathbb{E}[\sum_{i=j}^{N} I(\mathbf{x}_i^m, \mathbf{x}_j^n) = 1] = (1-r)^2 N, \\
N_{in} &= \mathbb{E}[\sum_{i \neq j}^{N} I(\mathbf{x}_i^m, \mathbf{x}_j^n) = 0] = 2r(1-r)N(N-1), \\
N_{fn} &= \mathbb{E}[\sum_{i \neq j}^{N} \{C(\mathbf{x}_i^m, \mathbf{x}_j^n) \cdot \mathbb{I}(I(\mathbf{x}_i^m, \mathbf{x}_j^n) = 0)\} = 1] \\
&= 2r(1-r)N(N-1) \cdot \mathbb{P}(C = 1 | I = 0) \\
&= 2r(1-r)N(N-1) \cdot \epsilon,
\end{aligned} \tag{3}$$

The objective function f_{ic} is further revised, and its expectation is as follows:

$$\begin{aligned}
f_{ic} &= \sum_{m \neq n}^{V} \sum_{i}^{N_{ip}} [\rho^+(\mathbf{x}_i^m, \mathbf{x}_i^n) - (1+\epsilon) \sum_{j \neq i}^{N_{in}} \rho^-(\mathbf{x}_i^m, \mathbf{x}_j^n)], \\
\mathbb{E}[f_{ic}] &= V(V-1)\{N_{ip} \cdot \mathbb{E}[\rho^+] - (1+\epsilon) \cdot N_{in} \cdot \mathbb{E}[\rho^-]\}
\end{aligned} \tag{4}$$

- When maximizing f_{ic}, the noise term amplifies ϵ the penalty for negative pairs by $(1+\epsilon)$, which suppresses intra-cluster similarity and undermines clustering performance.

- Furthermore, since $N_{ip} \propto \frac{1}{r^2}$, $N_{in} \propto r^2$ and $\rho \propto r$, as r increases, the impact of false negative noise ρ on model performance will also increase.

Cluster-level paradigms: When $Y = C(\mathbf{x}_i^m, \mathbf{x}_j^n)$, the objective of cluster-level paradigms f_{cc} is formulated as:

$$f_{cc} = \sum_{m \neq n}^{V} \sum_{i}^{N} [\rho^+(\mathbf{x}_i^m, \mathbf{x}_j^n) - \sum_{j \neq i}^{N} \rho^-(\mathbf{x}_i^m, \mathbf{x}_j^n)]. \quad (5)$$

Due to the different data distribution across views caused by varying missing observations in each view, as well as the lack of clustering interaction among instances within views, there may be inconsistencies in cluster semantics and cluster distributions between view m and n, introducing cluster consistency errors $\delta^{m,n}$.

Define $\mathbf{C}^v = \{\mathbf{c}_k^v\}_{k=1}^K$ as a set of cluster prototypes for v-th view data $\mathbf{X}^v$ and $p(\mathbf{X}^v|\mathbf{c}_k^v)$ as the probability distribution of $\mathbf{X}^v$ in the k-th cluster. $\delta^{m,n}$ include the following two errors:

- Cluster semantic error $\delta_{se}^{m,n}$: two observations $\mathbf{x}_i^m, \mathbf{x}_j^n$ from the same semantic cluster may be assigned to different clusters across views. Formally, when $S(\mathbf{x}_i^m, \mathbf{x}_j^n) = 1$, Cluster-level paradigm mistakes $\arg\max_k \rho(\mathbf{x}_i^m, \mathbf{c}_k^v) \neq \arg\max_{k'} \rho(\mathbf{x}_j^n, \mathbf{c}_{k'}^v)$) and can be quantified as:

$$\delta_{se}^{m,n} = \mathbb{A}(\mathbf{C}^m, \mathbf{C}^n), \quad (6)$$

  where $\mathbb{A}(\cdot)$ is the cost function for optimally matching the prototypes between views, like cost matrix in Hungarian Algorithm, Optimal transport distance in Optimal Transport and contrastive loss in Contrastive Learning.
- Cluster distribution error $\delta_{st}^{m,n}$: the data distribution of the same semantic cluster k may vary across views. It means $p(\mathbf{X}^m|\mathbf{c}_k^m) \neq p(\mathbf{X}^n|\mathbf{c}_k^n)$ and and can be quantified as:

$$\delta_{st}^{m,n} = \sum_{k}^{K} \mathbb{D}(p(\mathbf{X}^m|\mathbf{c}_k^m)||p(\mathbf{X}^n|\mathbf{c}_k^n)), \quad (7)$$

where $\mathbb{D}(\cdot)$ quantifies the difference between the two distributions, like Kullback-Leibler Divergence, Total Variation Distance and Maximum Mean Discrepancy Distance.

Define the number of cluster-level positive pairs N_{cp} and cluster-level negative pairs N_{cp} as:

$$\begin{aligned} N_{cp} &= \mathbb{E}[\sum_{i,j}^{N} C(\mathbf{x}_i^m, \mathbf{x}_j^n) = 1] = (1-r)^2 N^2 \cdot \mathbb{P}(y_i^m = y_j^n), \\ N_{cn} &= \mathbb{E}[\sum_{i \neq j}^{N} C(\mathbf{x}_i^m, \mathbf{x}_j^n) = 0] \\ &= (1-r)^2 N^2 \cdot (1 - \mathbb{P}(y_i^m = y_j^n)), \end{aligned} \quad (8)$$

where $\mathbb{P}(y_i^m = y_j^n)$ represents the probability that $\mathbf{x}_i^m$ and $\mathbf{x}_j^n$ belong to the same semantic cluster. If instances are uniformly distributed across K clusters, $\mathbb{P}(y_i^m = y_j^n) = \frac{1}{K}$.

The objective function f_{cc} is further revised, and its expectation is as follows:

$$\begin{aligned} f_{cc} &= \sum_{m \neq n}^{V} \{\sum_{i}^{N_{cp}} [\rho^+(\mathbf{x}_i^m, \mathbf{x}_j^n) - \sum_{j \neq i}^{N_{cn}} \rho^-(\mathbf{x}_i^m, \mathbf{x}_j^n)] - \delta^{m,n}\}, \\ \mathbb{E}[f_{cc}] &= V(V-1)\{N_{cp} \cdot \mathbb{E}[\rho^+] - N_{cn} \cdot \mathbb{E}[\rho^-] - \mathbb{E}[\delta^{m,n}]\}, \end{aligned} \quad (9)$$

- To ensure cluster semantic and distributions consistency, the cluster-level paradigm needs to optimize error term $\mathbb{E}[\delta^{m,n}]$. However, $\mathbb{E}[\delta^{m,n}]$ cannot be entirely eliminated and can only be minimized, which inevitably degrades the model's performance.
- Furthermore, the missing rate r disrupts the uniformity of the original cluster distribution ($\mathbb{P}(y_i^m = y_j^n)$ is no longer equal to $\frac{1}{K}$), thereby introducing both false negative and false positive noise in N_{cp} and N_{cn}. This perturbation consequently exacerbates the degree of prototype and distribution shifts. As a result, $\mathbb{E}[\delta^{m,n}]$ will increase with r.
- Meanwhile, due to $\delta^{m,n} \propto K$ and $\mathbb{E}[f_{cc}] \propto V(V-1)$, an excessive number of clusters and views can cause $\mathbb{E}[\delta^{m,n}]$ to surge, significantly increasing the difficulty of optimization.

Semantic-level paradigms: Define the quantities of semantic-level positive and negative pairs:

$$\begin{aligned} N_{sp} &= \mathbb{E}\left[\sum_{i,j}^{N} S(\mathbf{x}_i^m, \mathbf{x}_j^n) = 1\right] \\ &= (1-r)^2 N^2 \cdot \mathbb{P}(y_i^m = y_j^n), \\ N_{sn} &= \mathbb{E}\left[\sum_{i \neq j}^{N} S(\mathbf{x}_i^m, \mathbf{x}_j^n) = 0\right] \\ &= (1-r)^2 N^2 \cdot \left(1 - \mathbb{P}(y_i^m = y_j^n)\right), \end{aligned} \quad (10)$$

where $\mathbb{P}(y_i^m = y_j^n)$ still represents the same-cluster probability of cross-view observations.

When $Y = S(\mathbf{x}_i^m, \mathbf{x}_j^n)$, the objective of semantic-level paradigms f_{sc} is formulated as:

$$f_{sc} = \sum_{m \neq n}^{V} \sum_{i}^{N_{sp}} [\rho^+(\mathbf{x}_i^m, \mathbf{x}_j^n) - \sum_{j \neq i}^{N_{sn}} \rho^-(\mathbf{x}_i^m, \mathbf{x}_j^n)]. \quad (11)$$

Compared with IC in False negative noise mitigation: semantic-level positive pairs N_{sp} are defined as $S(\mathbf{x}_i^m, \mathbf{x}_j^n) = 1$, and its false negative noise ϵ_{sc} is quantified as:

$$\epsilon_{\text{sc}} = \mathbb{P}(\arg\max_k \rho(\mathbf{x}_i^m, \mathbf{c}_k) \neq \arg\max_k \rho(\mathbf{x}_j^n, \mathbf{c}_k) \mid C = 1) \quad (12)$$

- N_{sp} are constructed through consensus prototypes $\mathbf{C}$, avoiding cross-view matching:

$$\epsilon_{\text{sc}} = \mathbb{P}(S(\mathbf{x}_i^m, \mathbf{x}_j^n) = 0 \mid C(\mathbf{x}_i^m, \mathbf{x}_j^n) = 1) \approx 0$$

  Therefore, $N_{fn}^{sc} \propto \epsilon_{\text{sc}} \approx 0$,
- As the prototypes $\mathbf{C}$ are optimized, the distance between different cluster prototypes $\|\mathbf{c}_k - \mathbf{c}_{k'}\|$ increases, causing $\mathbb{P}(\cdot) \propto \exp(-\|\mathbf{c}_k - \mathbf{c}_{k'}\|^2/\sigma^2)$ to decay exponentially. This drives $N_{fn}^{sc} \to 0$.

Compared with CC in Cluster Consistency Errors optimization: According to Definition 3, semantic-level paradigms enforces all views to share the same set of cluster prototypes $\mathbf{C}$, fundamentally eliminating cross-view cluster semantic ambiguity. This is specifically manifested as:

- Cross-view Semantic Consistency of shared Prototypes: $\forall m, n, \mathbf{c}_k^m = \mathbf{c}_k^n = \mathbf{c}_k$, directly eliminating cluster semantic error $\delta_{se}^{m,n}$ (i.e., $\delta_{se}^{m,n} = 0$).
- Implicit Constraint on Distribution Discrepancy: The shared prototypes project data from each view into a common space through the mapping function $\psi(\cdot)$, causing the distribution discrepancy $\delta_{st}^{m,n}$ to be constrained by the embedding distance $\delta_{st}^{m,n} \propto \|\psi(\mathbf{X}^m) - \psi(\mathbf{X}^n)\|^2 \to 0$, which is automatically minimized during optimization.
- False Positive/Negative Suppression: Due to sharing a set of semantic prototypes, the estimation of $\mathbb{P}(y_i^m = y_j^n)$ remains $\frac{1}{K}$ unaffected by the view missing rate r (compared to $\mathbb{P}(y_i^m = y_j^n) \neq 1/K$ in Cluster-level paradigms), thereby avoiding false negatives and false positives.

The objective function f_{sc} can be formally expressed in expectation form as:

$$\mathbb{E}[f_{sc}] = V(V-1)\{N_{sp} \cdot \mathbb{E}[\rho^+] - N_{sn} \cdot \mathbb{E}[\rho^-]\}. \quad (13)$$

- Confidence and Robustness for Noise ϵ and Error δ: Compared to instance-level paradigms (containing explicit noise term $(1+\epsilon)\mathbb{E}[\rho^-]$) and cluster-level paradigms (containing non-eliminable $\mathbb{E}[\delta^{m,n}]$), the semantic-level objective has no additional noise and error terms, and N_{fn}^{sc}, N_{fp}^{sc} decays during optimization.
- Confidence and Robustness for Missing Rate r: Due to the shared prototype constraint, the ratio between N_{sp} and N_{sn} remains stable ($\mathbb{P}(y_i^m = y_j^n) = 1/K$). Even with high r, the objective function can still accurately model the cluster structure.

□

2.2. Proof of Theorem 2

Theorem 2. *Since Paired observations $(\overline{\mathbf{x}}_i^m, \overline{\mathbf{x}}_i^n)$ inherently satisfy instance- and cluster-level consistency, they can achieve semantic consensus via a shared set of prototypes* $\mathbf{C}$.

Proof. **Instance-level Consistency:** According to Definition 1, paired observations $(\overline{\mathbf{x}}_i^m, \overline{\mathbf{x}}_i^n)$ satisfy the condition that both are cross-view observations of the same instance $\mathbf{x}_i$, thus they are instance-level consistency $I\big(\overline{\mathbf{x}}_i^m, \overline{\mathbf{x}}_i^n\big) = 1$. Two observations essentially belong to the same underlying instance, with only view-specific noise or modality discrepancies causing observational differences.

Cluster-level Consistency: $\overline{\mathbf{x}}_i^m$ and $\overline{\mathbf{x}}_i^n$ are cross-view observations of the same instance, they must belong to the same cluster. According to Definition 2, paired observations $(\overline{\mathbf{x}}_i^m, \overline{\mathbf{x}}_i^n)$ are instance-level consistency $C\big(\overline{\mathbf{x}}_i^m, \overline{\mathbf{x}}_i^n\big) = 1$. This further ensures that, in addition to being similar in features, these two observations are also consistent in their cluster structure, indicating that both are grouped into the same semantic cluster across different views.

Semantic-level Consistency: According to Definition 3, semantic-level consensus requires:

- Shared Cluster Prototypes: All observations share the same set of prototypes $\mathbf{C} = \{c_k\}_{k=1}^K$.
- Consistent Prototype Assignment: $\arg\min_k \rho(\overline{x}_i^m, c_k) = \arg\min_k \rho(\overline{x}_i^n, c_k)$.

Paired observations $(\overline{\mathbf{x}}_i^m, \overline{\mathbf{x}}_i^n)$ satisfy the following conditions:

- Condition 1: Since $\mathbf{C}$ is globally shared, observations from all views are assigned based on the same set of prototypes.
- Condition 2: Assume the nearest prototype for $\overline{\mathbf{x}}_i^m$ is $\mathbf{c}_k$:

$$\arg\min_k d(\overline{\mathbf{x}}_i^m, \mathbf{c}_k) = k.$$

  Since $\overline{\mathbf{x}}_i^m$ and $\overline{\mathbf{x}}_i^n$ belong to the same cluster $\mathbf{c}_k$ (CC), and prototype $\mathbf{c}_k$ is the central representation of this cluster, the nearest prototype for $\overline{\mathbf{x}}_i^n$ should also be $\mathbf{c}_k$. Otherwise, if the nearest prototype for $\overline{\mathbf{x}}_i^n$ is $\mathbf{c}_{k'}$ ($k' \neq k$), it would contradict the cluster consistency (CC). Therefore, it must satisfy:

$$\arg\min_k \rho(\overline{\mathbf{x}}_i^m, \mathbf{c}_k) = \arg\min_k \rho(\overline{\mathbf{x}}_i^n, \mathbf{c}_k) = k.$$

The conditions of SC all hold. According to Definition 3, the paired observations $(\overline{\mathbf{x}}_i^m, \overline{\mathbf{x}}_i^n)$ have reached semantic-level consensus $S(\mathbf{x}_i^m, \mathbf{x}_i^n) = 1$.

□

3. Appendix C: Experiments

3.1. Experimental Settings

Datasets. From the perspective of clustering task complexity in the number of clusters, views, feature dimensions, and samples, six widely applied public datasets are selected for experiments:

Competitors. To validate the effectiveness of our model from the perspective of consistency learning, imputation and alignment, we select seven state-of-the-art methods as competitors and summarize them in Table **??** according to

Table 1. Multi-view benchmark datasets in experiments.

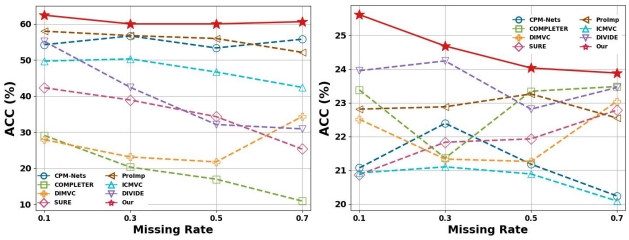
Dataset	Samples	Clusters	Views	Dimensionality
Yale [31]	165	15	3	3304/6750/4096
Caltech-5V[24]	1400	7	5	1984/512/928/254/40
NUSWIDEOBJ10[7]	6251	10	5	129/74/145/226/65
ALOI-100[4]	10800	100	4	77/13/64/125
YouTubeFace10[6]	38654	10	4	944/576/512/640
NoisyMNIST[12]	70000	10	2	784/784

the consistency, imputation and alignment techniques they employ.

- CPM-Net [30], encodes view-specific information into a common representation based on instance-level consistency and employs GANs to impute missing data across views.
- COMPLETER [12], maximize mutual information and minimize conditional entropy across views based on instance-level consistency to achieve contrastive representation learning and duel missing prediction.
- DIMVC [23], performs instance-level contrastive learning to construct a common representation, while aligns view-specific cluster assignments with the common assignment for decision fusion.
- SURE [28], introduces an adaptive distance threshold for positive-negative pairs to identify and penalize false negative pairs, enabling cluster-level contrastive learning. Additionally, it transfers the cluster relationships from other complete views to the missing views for imputation.
- ProImp [10], conducts instance-level contrastive learning and prototypes alignment to ensure consistency across views, then fills in missing observations by referring to prototypes in the missing views and the observation-prototype relationships in other complete views.
- ICMVC [1], transfers graph relationships from complete views to missing views for imputation based on instance-level consistency. To further enhance consistency in cluster assignments, it constrains view-specific assignments to align with the high-confidence common representation.
- DIVIDE [15], leverages random walks to progressively discover positive and negative pairs for cross-view cluster alignment. Through cluster-level contrastive learning, it explores cross-view consistency information to recover missing views.

Table 2. 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	graph structure transfer	\
ProImp (IJCAI'23)	instance-level	sample-prototype relationship inheritance	prototype-based
ICMVC (AAAI'24)	instance-level	graph structure transfer	assignment-based
DIVIDE (AAAI'24)	cluster-level	mutual information interaction	\



(a) Yale (b) NUSWIDEOBJ10

Figure 1. Visualization for Table 4 based on metric ACC.

3.2. Implementation details

Our model consists of three modules: reconstruction (REC) module, consistency semantic learning (CSL) module and cluster semantic enhancement (CSE) module, as well as four components: encoder, decoder, contrastive clustering and graph clustering. The implementation details are as follows:

Table 3. FreeCSL architecture details.

Component	Layer	Dimension
Encoder	4-layer MLPs	$view\_dim \to 500 \to 500 \to 2000 \to 64$
Decoder	4-layer MLPs	$64 \to 2000 \to 500 \to 500 \to view\_dim$
contrastive clustering	1-layer FC	$64 \to 64$
graph clustering	2-layer GCNs and 1-layer FC	$64 \to 128 \to 64 \to cluster\_num$

3.3. Competitiveness of FreeCSL

To further enhance the credibility of our model, we supply a comparative experiment on Yale and NUSWIDEOBJ10 dataset, and present the comparison results, along with the visualizations based on ACC and NMI metric, in Table 4 and Fig. 1. As mentioned in our main text, FreeCSL surpasses all competitors and demonstrates more stable performance in various missing rates even on the small-sample dataset Yale, as FreeCSL avoids the errors associated with imputation and alignment.

3.4. Understanding FreeCSL

Ablation Study. The proposed FreeCSL contains three modules: reconstruction (REC) module, cross-view consistency semantic learning (CSL) module, and within-view cluster semantic enhancement (CSE) module. To further verify the importance of each module, we conducted extra ablation experiments on YoutubeFace10, NoisyMNIST, Yale and NUSWIDEOBJ10 datasets as shown in Table 5. With the REC module as the baseline, both CSE module and CSE module contribute significantly to the improved performance of all datasets. Furthermore, due to the synergistic effect of the three modules, our model exhibits more confident and stable performance across different missing rates compared to the ablation group.

Imputation- and Alignment-free CSL. To demonstrate our model can learn semantic knowledge from view data

Table 4. Clustering performance comparisons on Yale and NUSWIDEOBJ10. 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 (%)
Yale	CPM-Nets	54.24	60.82	37.55	56.66	63.25	40.22	53.34	59.58	34.22	55.76	58.20	33.10
	COMPLETER	29.09	37.10	2.36	20.30	29.61	1.20	16.97	26.08	0.97	10.91	16.88	0.32
	DIMVC	27.91	32.27	7.94	23.12	26.79	2.85	21.76	26.92	3.46	34.32	39.47	11.21
	SURE	42.30	49.57	22.12	38.91	43.90	13.61	34.30	39.07	9.08	25.33	33.79	3.92
	ProImp	57.98	63.37	38.95	56.77	60.43	35.54	55.96	58.47	32.91	52.12	56.11	30.19
	ICMVC	49.70	61.52	30.64	50.30	61.62	31.49	46.67	58.91	27.84	42.42	54.43	23.21
	DIVIDE	55.15	56.37	28.97	42.42	45.38	18.25	32.12	35.80	8.40	30.91	32.03	6.48
	Our	62.42	65.87	45.71	60.00	64.73	41.33	60.00	63.14	40.85	60.61	60.30	37.69
NUSWIDEOBJ10	CPM-Nets	21.07	7.76	3.93	22.39	6.88	3.97	21.18	5.97	3.06	20.24	4.60	1.86
	COMPLETER	23.38	8.16	2.58	21.36	9.90	4.61	23.34	9.94	4.60	23.48	10.96	5.37
	DIMVC	22.51	11.46	6.61	21.33	11.89	5.43	21.26	10.64	5.03	23.04	10.40	5.68
	SURE	20.87	10.90	5.39	21.83	11.24	6.07	21.93	11.14	5.92	22.78	10.54	6.16
	ProImp	22.81	11.31	5.85	22.88	11.40	6.11	23.26	11.20	6.20	22.55	11.24	5.94
	ICMVC	20.92	10.15	5.06	21.10	10.59	5.19	20.89	10.20	5.04	20.09	9.58	5.06
	DIVIDE	23.95	12.97	7.75	24.24	13.22	7.67	22.81	12.90	7.43	23.45	10.78	6.05
	Ours	25.61	16.31	8.75	24.68	15.14	7.95	24.03	14.10	7.69	23.88	12.67	6.82

Table 5. Ablation study on YoutubeFace10, NoisyMNIST, Yale and NUSWIDEOBJ10. ✓ 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 (%)
YouTube10	✓			71.57	75.65	64.43	68.37	75.63	63.87	65.28	69.41	55.66	59.63	62.36	46.51
	✓	✓		79.20	81.58	71.60	77.13	80.07	68.86	75.13	79.17	65.38	71.80	75.95	65.24
	✓		✓	76.55	80.13	69.21	72.89	70.94	63.68	72.08	68.06	61.62	68.88	64.71	56.58
	✓	✓	✓	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	✓			33.45	26.44	16.81	25.25	14.80	8.01	23.52	15.46	7.46	24.17	15.46	6.68
	✓	✓		98.17	97.04	96.97	96.27	93.69	93.91	95.25	89.08	90.38	90.96	81.15	82.37
	✓		✓	53.38	50.60	37.16	39.24	37.59	20.84	33.89	29.02	14.19	33.08	26.70	14.23
	✓	✓	✓	99.13	97.23	98.10	97.68	93.94	94.94	96.04	89.81	91.48	92.19	82.50	83.56
Yale	✓			50.91	60.79	36.40	44.85	50.61	25.49	34.55	44.81	17.03	33.33	40.25	12.86
	✓	✓		55.15	58.99	34.75	56.97	59.11	36.01	56.97	61.71	61.71	56.36	59.57	35.19
	✓		✓	54.55	58.72	33.98	46.06	46.06	23.87	36.36	47.59	19.45	35.15	42.16	12.20
	✓	✓	✓	62.42	65.87	45.71	60.00	64.73	42.14	60.00	63.14	40.85	60.61	60.30	37.69
NUSWIDEOBJ	✓			19.32	5.69	2.70	18.00	3.80	1.57	17.65	2.96	0.57	19.20	3.57	0.45
	✓	✓		23.68	16.13	8.50	23.56	14.84	7.62	23.36	13.63	7.29	22.86	12.60	6.50
	✓		✓	23.23	9.67	4.98	23.63	8.21	4.95	20.83	5.67	2.96	20.22	6.10	2.42
	✓	✓	✓	25.61	16.31	8.75	24.68	15.14	7.95	24.03	14.10	7.69	23.88	12.67	6.82

and achieve consistent and reliable clustering assignments without imputation or alignment, we make efforts in two aspects: conducting imputation experiments and visualizing similarity matrices, both based on latent and semantic representations learned from YoutubeFace10, NoisyMNIST, and NUSWIDEOBJ10 datasets.

Notably, both the latent and semantic representations $\{\mathbf{Z}^v\}_{v=1}^V, \{\mathbf{H}^v\}_{v=1}^V$ are outputs of our model after training. The latent representation $\mathbf{Z}^v$ refers to the output after the decoder but before the CSL module, while the semantic representation $\mathbf{H}^v$ has undergone nonlinear mapping through the CSL module. We impute the missing views for two sets $\{\mathbf{Z}^v\}_{v=1}^V$ and $\{\mathbf{H}^v\}_{v=1}^V$, with mean values based on the neighborhood relationships observed in complete view data. Finally, we perform K-means on consensus representations $\mathbf{Z}$ and $\mathbf{H}$ fused by the representation fusion manner $\mathbb{T}(\{\mathbf{Z}^v\}_{v=1}^V$, $\mathbb{T}(\{\mathbf{H}^v\}_{v=1}^V$ described in Section 2.3 of our main text.

In Table 6, at small missing rates, our model performs comparably regardless of whether the missing data are imputed or not. As the missing rate increases and the available information for imputation decreases, our model without imputation exhibit superior robustness. Improper imputation introduces noise, while our model, combining the CSL and CSE modules, successfully captures semantic knowledge from view data (embedded in both latent and semantic representations) and leveraging the fusion method $\mathcal{T}(\cdot)$, effectively integrate the consistency and complementary information across views. Thus, our FreeCSL achieves excellent performance without incurring extra computational cost or suffering clustering accuracy loss arising from imputation.

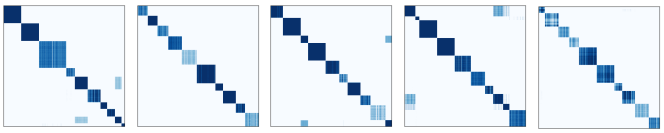
We visualize the cosine similarity matrices of the latent representations $\{\mathbf{Z}^v\}_{v=1}^V$, semantic representations $\{\mathbf{H}^v\}_{v=1}^V$, and their consensus representations $\mathbf{Z}$, $\mathbf{H}$ learned from YoutubeFace10, NoisyMNIST, Yale and NUSWIDEOBJ10 datasets in Fig. 2-8, further confirming the advantages of our model in consensus semantic learn-

Table 6. Imputation- and alignment-free study on YoutubeFace10, NoisyMNIST, Yale and NUSWIDEOBJ10. ILR and ISR are filled by K-NN imputation via cross-view graph for 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 (%)
YouTube10	ILR	82.66	82.79	74.20	80.46	81.18	71.54	80.37	81.28	71.58	73.68	75.70	63.39
	ISR	82.72	82.86	72.69	81.07	82.63	72.82	80.63	81.67	72.00	73.91	75.84	63.81
	FreeCSL	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	ILR	99.12	97.21	98.08	98.06	94.50	95.83	95.98	89.74	91.10	90.99	80.19	81.16
	ISR	99.15	97.31	98.15	97.83	93.86	95.28	95.80	89.23	90.98	90.69	79.76	80.57
	FreeCSL	99.13	97.23	98.10	97.68	93.94	94.94	96.04	89.81	91.48	92.19	82.50	83.56
Yale	ILR	55.15	61.63	37.54	56.36	62.72	40.53	53.33	59.89	35.05	50.30	55.21	29.07
	ISR	58.18	60.84	37.37	60.00	64.26	42.48	56.97	60.33	36.39	52.73	55.56	29.91
	FreeCSL	62.42	65.87	45.71	60.00	64.73	42.14	60.00	63.14	40.85	60.61	60.30	37.69
NUSWIDEOBJ	ILR	24.09	15.29	7.48	24.50	14.06	7.39	22.32	12.67	5.79	22.78	11.39	5.32
	ISR	24.22	16.44	8.53	25.07	15.26	8.33	23.93	13.79	7.13	22.45	11.83	6.16
	FreeCSL	25.61	16.31	8.75	24.68	15.14	7.95	24.03	14.10	7.69	23.88	12.67	6.82

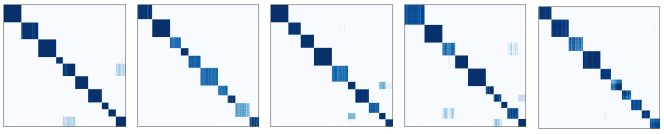
ing. The experimental results on Four datasets commonly reflect two findings:

- The similarity matrices of semantic representations, compared to latent ones, show a clearer and more uniform block structure along the diagonal. This indicates that the semantic representations, jointly optimized by the CSL and CSE modules, are well-suited for clustering task.
- Our consensus prototype-based semantic learning and consensus representation-based semantic clustering, effectively reduces entropy within clusters and enhances more confident assignments by integrating view-specific information.



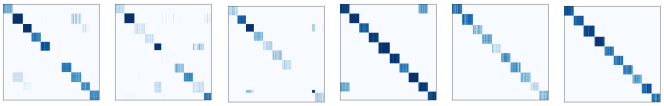
(a) $\mathbf{Z}^{(1)}$ (b) $\mathbf{Z}^{(2)}$ (c) $\mathbf{Z}^{(3)}$ (d) $\mathbf{Z}^{(4)}$ (e) $\mathbf{Z}$

Figure 2. Similarity matrices of $\{\mathbf{Z}^v\}_{v=1}^4$, $\mathbf{Z}$ on YouTubeFace10 with $r = 0.5$.



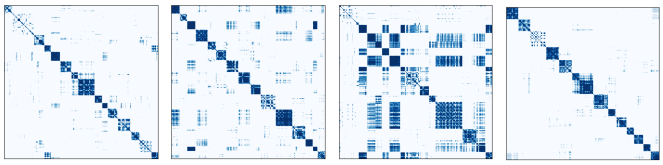
(a) $\mathbf{H}^{(1)}$ (b) $\mathbf{H}^{(2)}$ (c) $\mathbf{H}^{(3)}$ (d) $\mathbf{H}^{(4)}$ (e) $\mathbf{H}$

Figure 3. Similarity matrices of $\{\mathbf{H}^v\}_{v=1}^4$, $\mathbf{H}$ on YouTubeFace10 with $r = 0.5$.



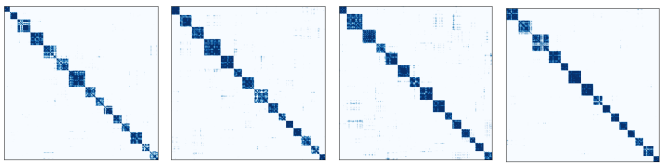
(a) $\mathbf{Z}^{(1)}$ (b) $\mathbf{Z}^{(2)}$ (c) $\mathbf{Z}$ (d) $\mathbf{H}^{(1)}$ (e) $\mathbf{H}^{(2)}$ (f) $\mathbf{H}$

Figure 4. Similarity matrices of $\{\mathbf{Z}^v\}_{v=1}^2$ and $\mathbf{Z}$, $\{\mathbf{H}^v\}_{v=1}^2$ and $\mathbf{H}$ on NoisyMNIST with $r = 0.5$.



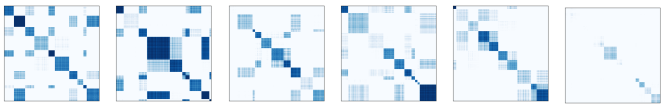
(a) $\mathbf{Z}^{(1)}$ (b) $\mathbf{Z}^{(2)}$ (c) $\mathbf{Z}^{(3)}$ (d) $\mathbf{Z}$

Figure 5. Similarity matrices of $\{\mathbf{Z}^v\}_{v=1}^3$, $\mathbf{Z}$ on Yale with $r = 0.5$.



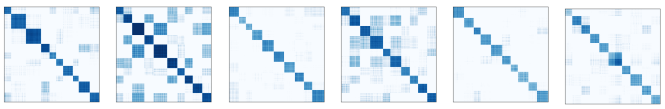
(a) $\mathbf{H}^{(1)}$ (b) $\mathbf{H}^{(2)}$ (c) $\mathbf{H}^{(3)}$ (d) $\mathbf{H}$

Figure 6. Similarity matrices of $\{\mathbf{H}^v\}_{v=1}^3$, $\mathbf{H}$ on Yale with $r = 0.5$.



(a) $\mathbf{Z}^{(1)}$ (b) $\mathbf{Z}^{(2)}$ (c) $\mathbf{Z}^{(3)}$ (d) $\mathbf{Z}^{(4)}$ (e) $\mathbf{Z}^{(5)}$ (f) $\mathbf{Z}$

Figure 7. Similarity matrices of $\{\mathbf{Z}^v\}_{v=1}^5$ and $\mathbf{Z}$ on NUSWIDEOBJECT10 with $r = 0.5$.



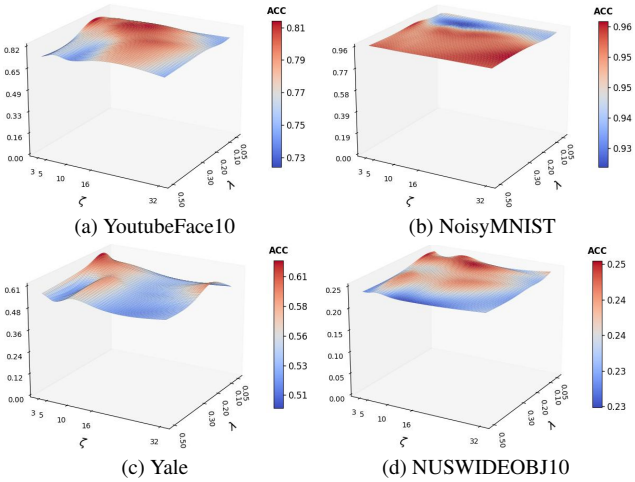
(a) $\mathbf{H}^{(1)}$ (b) $\mathbf{H}^{(2)}$ (c) $\mathbf{H}^{(3)}$ (d) $\mathbf{H}^{(4)}$ (e) $\mathbf{H}^{(5)}$ (f) $\mathbf{H}$

Figure 8. Similarity matrices of $\{\mathbf{H}^v\}_{v=1}^5$ and $\mathbf{H}$ on NUSWIDEOBJECT10 with $r = 0.5$.

3.5. Analysis on FreeCSL

Parameter Sensitivity Analysis. As in Section 3.5, we perform a parameter sensitivity analysis on the number of neighbors λ and the regularization coefficient ζ in graph clustering, on YoutubeFace10, NoisyMNIST, Yale

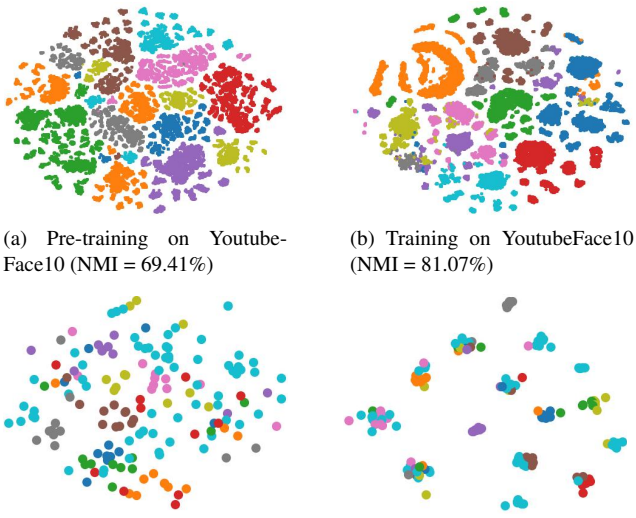
and NUSWIDEOBJ10 datasets. Fig. 9 shows our model is highly stable, with minimal performance fluctuation even when λ and ζ are adjusted to ranges of 3 to 32 and 0.05 to 0.5, respectively. A smaller number of neighbors λ and more relaxed regularization constraints ζ, will yield higher clustering accuracy (ACC). Except for the large-scale NoisyMNIST dataset, where a larger number of neighbors effectively enhance model performance by aggregating more useful neighbor information to discover cluster structures. In conclusion, our model present outstanding performance in complex clustering tasks without sacrificing computational resources for clustering accuracy or relying on strict regularization constraints.



(a) YoutubeFace10 (b) NoisyMNIST

(c) Yale (d) NUSWIDEOBJ10

Figure 9. Parameter analyses for ζ and λ with $r = 0.5$.



(a) Pre-training on YoutubeFace10 (NMI = 69.41%)

(b) Training on YoutubeFace10 (NMI = 81.07%)

(c) Pre-training on Yale (NMI = 44.81%)

(d) Training on Yale (NMI = 63.14%)

Figure 10. Visualization on YoutubeFace10 and Yale with $r = 0.5$.

3.6. Visualization for Consensus Semantic Clusters

Referring to true labels, we visualize the clustering effect of consensus semantic representations on YoutubeFace10 and Yale with the setting of missing rate $r = 0.5$, shown in Fig. 10 respectively. We can observe that after the training of our model, all instances converge toward their respective clusters, where instances within the same cluster become more compact, and instances from different clusters are separated far away. In addition, the visualization results of the prototypes of each cluster further confirm that through consensus prototype-based semantic learning, the shifted prototypes are re-estimated and accurately calibrated without the need for extra alignment processes.

References

[1] Guoqing Chao, Yi Jiang, and Dianhui Chu. Incomplete contrastive multi-view clustering with high-confidence guiding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 11221–11229, 2024. 1, 5

[2] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020. 1

[3] 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

[4] 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

[5] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020. 1

[6] 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

[7] Zhenyu Huang, Joey Tianyi Zhou, Xi Peng, Changqing Zhang, Hongyuan Zhu, and Jiancheng Lv. Multi-view spectral clustering network. In *IJCAI*, page 4, 2019. 5

[8] Qiang Ji, Yanfeng Sun, Junbin Gao, Yongli Hu, and Baocai Yin. A decoder-free variational deep embedding for unsupervised clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 33(10):5681–5693, 2021. 1

[9] 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

[10] Haobin Li, Yunfan Li, Mouxing Yang, Peng Hu, Dezhong Peng, and Xi Peng. Incomplete multi-view clustering via prototype-based imputation. *arXiv preprint arXiv:2301.11045*, 2023. 1, 5

[11] Yunfan Li, Mouxing Yang, Dezhong Peng, Taihao Li, Jiantao Huang, and Xi Peng. Twin contrastive learning for online clustering. *International Journal of Computer Vision*, 130 (9):2205–2221, 2022. 1

[12] 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. 1, 5

[13] 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

[14] Chong Liu, Yuqi Zhang, Hongsong Wang, Weihua Chen, Fan Wang, Yan Huang, Yi-Dong Shen, and Liang Wang. Efficient token-guided image-text retrieval with consistent multimodal contrastive training. *IEEE Transactions on Image Processing*, 32:3622–3633, 2023. 1

[15] 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. 1, 5

[16] 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

[17] Dong Wang, Ning Ding, Piji Li, and Haitao Zheng. Cline: Contrastive learning with semantic negative examples for natural language understanding. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2332–2342, 2021. 1

[18] 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

[19] 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. 1

[20] Shaowei Wei, Jun Wang, Guoxian Yu, Carlotta Domeniconi, and Xiangliang Zhang. Deep incomplete multi-view multiple clusterings. In *2020 IEEE International Conference on Data Mining (ICDM)*, pages 651–660. IEEE, 2020. 1

[21] 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. 1

[22] Enze Xie, Jian Ding, Wenhai Wang, Xiaohang Zhan, Hang Xu, Peize Sun, Zhenguo Li, and Ping Luo. Detco: Unsupervised contrastive learning for object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8392–8401, 2021. 1

[23] 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. 1, 5

[24] 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

[25] 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. 1

[26] Xiaolong Xu, Hongsheng Dong, Lianyong Qi, Xuyun Zhang, Haolong Xiang, Xiaoyu Xia, Yanwei Xu, and Wanchun Dou. Cmclrec: Cross-modal contrastive learning for user cold-start sequential recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1589–1598, 2024. 1

[27] Lin Yang, Wentao Fan, and Nizar Bouguila. Deep clustering analysis via dual variational autoencoder with spherical latent embeddings. *IEEE Transactions on Neural Networks and Learning Systems*, 34(9):6303–6312, 2021. 1

[28] Mouxing Yang, Yunfan Li, Peng Hu, Jinfeng Bai, Jian Cheng Lv, and Xi Peng. Robust multi-view clustering with incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 1, 5

[29] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *International conference on machine learning*, pages 12310–12320. PMLR, 2021. 1

[30] 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. 1, 5

[31] Pengfei Zhu, Xinjie Yao, Yu Wang, Binyuan Hui, Dawei Du, and Qinghua Hu. Multi-view deep subspace clustering networks. *arXiv preprint arXiv:1908.01978*, 2019. 5