SPORT: STRUCTURE-AWARE PROTOTYPE DISENTANGLEMENT FOR INCOMPLETE MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Prototype-based incomplete multi-view clustering (IMVC) has recently gained attention for leveraging prototypes to recover missing views. However, existing methods often overlook view-specific semantics in prototypes, fail to preserve cluster-level relational structures, and rely on global prototypes without exploiting local geometric information, resulting in limited multi-view expressiveness and inaccurate imputation. To address these limitations, we propose a novel framework termed **S**tructure-aware **PrO**totype disentanglement fo**R** incomplete multi-view clus**T**ering (SPORT), which explicitly disentangles shared and view-specific components of prototypes while preserving cluster-level relational structures. Specifically, we decouple prototypes into orthogonal shared and view-specific components, aligning only shared components to capture consensus semantics while decorrelating view-specific components to preserve complementary information. Meanwhile, a structure-aware contrastive learning mechanism is incorporated to explicitly model cluster-level relationships during cross-view representation learning. Furthermore, a hybrid imputation strategy integrates global prototype matching with local neighborhood matching, enabling joint exploitation of semantic prototypes and manifold structures for missing-view recovery. Extensive experiments on six benchmark datasets show that SPORT achieves superior performance over state-of-the-art methods under various missing rates.

1 INTRODUCTION

With the rapid development of knowledge discovery and data mining techniques, data are increasingly collected and stored in multiple modalities. Such multi-source data are referred to as multi-view data. Multi-View Clustering (MVC) seeks to partition unlabeled multi-view data by identifying shared clustering structures across multiple views, thereby offering a more comprehensive characterization of the underlying relationships among samples Qiang et al. (2026); Qin et al. (2026); Xu et al. (2016). Conventional MVC frameworks typically assume that all samples are fully observed across all views Wang (2025); Gu & Wang (2025); Trosten et al. (2023). However, in real-world scenarios, incomplete multi-view data are ubiquitous due to sensor limitations, hardware failures, or storage constraints, with some samples missing one or more views (Liu et al., 2026; Ding et al., 2025; Liu et al., 2025). This give rise to the emergence of numerous Incomplete Multi-View Clustering (IMVC) approaches Wang et al. (2026); Zhang et al. (2026); Wu et al. (2025b), which aims to mine information from available views to uncover semantically consistent clustering structures.

According to the learning paradigm, existing IMVC methods can be broadly categorized into traditional methods Zhao et al. (2017); Zhou et al. (2019); Wen et al. (2020); Guo & Ye (2019) and deep learning-based methods (Ding et al., 2025; Zhang et al., 2024; Xu et al., 2023b). Owing to their superior representation capacity and scalability, deep learning-based methods have garnered increasing attention in addressing IMVC tasks. For instance, Lin et al. (2021) proposed an information-theoretic framework that jointly performs data recovery and consistency learning, achieving cross-view consistency and reconstructing missing data via dual prediction and contrastive losses. Tang & Liu (2022) designed a neighborhood-based IMVC framework that imputed missing views from dynamically mined k-nearest neighbors and introduced a weighting function to suppress low-quality imputations. Moreover, a growing line of literature focuses on prototype-based methods Yuan et al.

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

(2025); Jin et al. (2023); Wang et al. (2025); Zhu et al. (2026), which exploit global semantics via representative prototypes to better handle missing information and improve clustering robustness. For example, Wang et al. (2025) aligned the prototypes via contrastive learning and recovered the missing data through an explicit prototype graph matching strategy. Li et al. (2025) designed a bi-alignment guidance module to ensure discriminative representations and employed an attention layer to enhance the compactness of the view-specific sample-prototype relationship.

Despite the notable advancement in prototype-based IMVC methods, three key limitations persist. First, existing methods typically enforce strict cross-view prototype alignment under the assumption that corresponding prototypes should share identical semantic representations across views. However, each view is generated from a distinct observation process, and its prototype inherently encodes both shared semantics and view-specific characteristics. Such overly strict alignment suppresses view-specific information, limiting the expressive capacity of multi-view representations, a phenomenon we term the Prototype Over-alignment Problem (POP). Second, most methods employ hard contrastive learning that treats only same-sample pairs as positive pairs, discarding cluster-level semantic similarity among neighboring instances (i.e., samples belonging to the same cluster generally exhibit closely related features which should also be harnessed), causing semantically similar samples to be pushed apart in the feature space. Furthermore, existing prototype-based imputation methods rely exclusively on prototypes as semantic anchors for missing-view imputation. Due to the coarse-grained nature of prototypes, they primarily capture global cluster-level semantics while overlooking local neighborhood relationships, resulting in structurally over-smoothed reconstructions that fail to preserve fine-grained instance-level consistency.

To address the aforementioned issues, we propose a novel Structure-aware PrOtotype disentanglement foR incompleTe multi-view clustering framework (SPORT), which jointly exploits cross-view consistency, view-specific semantics, and structural relationships to learn robust and discriminative clustering representations. Specifically, as illustrated in Fig. 1, we explicitly decompose the multi-view prototypes obtained via k-means into complementary shared and view-specific components. The shared components are aligned across views to capture view-invariant semantics, while the view-specific components are regularized via a decorrelation constraint to preserve discriminative view-dependent information. This prototype decomposition strategy effectively disentangles cross-view consensus from view-specific semantics, leading to more discriminative clustering representations. In addition, a weighted contrastive learning mechanism is designed, where similarity-aware weights are introduced to characterize the structural relationships among samples, enabling cross-view feature alignment while preserving the underlying geometric structure of the data. Moreover, SPORT employs a hybrid imputation scheme that reconstructs missing samples by jointly exploiting information from matched prototypes and neighboring instances. By integrating cluster-level semantics with local structural information, the proposed strategy improves the quality of missing-data recovery and further enhances clustering performance. In summary, the contributions of our work can be summarized as follows:

- We propose a unified prototype alignment framework that decomposes prototypes into shared and view-specific components. The shared components are aligned across views, while the view-specific components are decorrelated to preserve complementary information, thereby mitigating prototype misalignment.
- We develop a structure-aware feature alignment mechanism by introducing instance-wise similarity weights to softly encode inter-sample relationships. This enables the contrastive objective to enforce cross-view consistency while preserving cluster-level relational structures, yielding more discriminative representations.
- We design a hybrid prototype imputation strategy that imputes missing data by integrating information from matched prototypes and neighboring samples, thereby jointly exploiting global cluster semantics and local structural information to improve reconstruction quality.

2 Related Work

Deep learning-based incomplete multi-view clustering (DIMVC) methods Zhang et al. (2026); Wang et al. (2026); Wu et al. (2025a) can be broadly categorized into four classes: generator-based, predictor-based, neighborhood-based, and prototype-based methods.

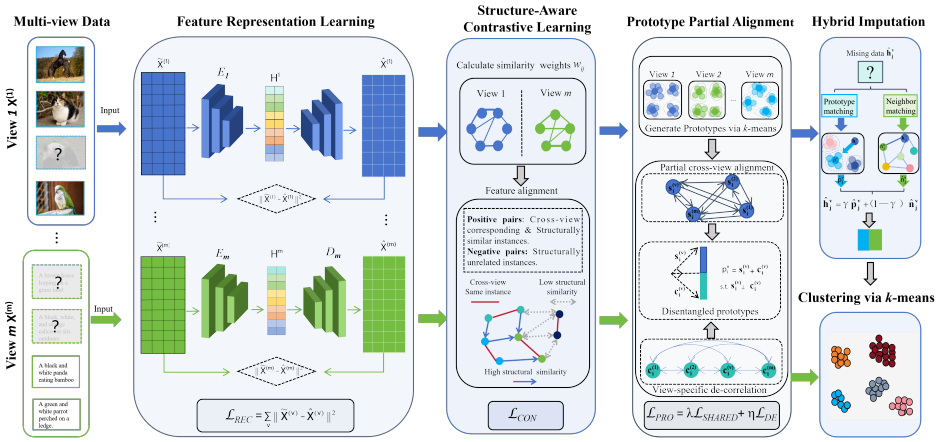


Figure 1: Framework of the proposed SPORT. Our method captures cross-view consistency and cluster-level structure via structure-aware relational contrastive learning, preserves complementary information through prototype partial alignment, and imputes missing views using a hybrid prototype imputation strategy.

Generator-based methods leverage the strong generative capability of generative models to reconstruct missing information from observed representations Shang et al. (2017); Wen et al. (2024); Fang et al. (2024); Zhang et al. (2025). Two predominant generative paradigms applied to IMVC tasks are generative adversarial networks (GANs) Goodfellow et al. (2020) and diffusion models Ho et al. (2020). Predictor-based methods estimate missing features from observed ones from an information-theoretic perspective Lin et al. (2021); Xu et al. (2023a); Lu et al. (2024); Ding et al. (2025). Neighborhood-based methods exploit structural relationships to address data incompleteness by leveraging local topology within or across views Yang et al. (2022); Tang & Liu (2022); Teng et al. (2024); Yan et al. (2025). K-nearest neighbor (kNN) Peterson (2009) and graph neural networks (GNNs) Corso et al. (2024) are two widely used paradigms. Prototype-based methods capture global semantics by learning view-wise prototypes as representative anchors and leveraging them for missing-data imputation Yuan et al. (2025); Jin et al. (2023); Wang et al. (2025); Li et al. (2023a); Du et al. (2026); Zhu et al. (2026). These prototypes are often misaligned across views, which has motivated extensive research on cross-view prototype alignment.

Detailed examples of these methods are presented in Appendix A.

3 Method

3.1 Notation and Problem Statement

In this paper, $\{\mathbf{X}^{(v)} \in \mathbb{R}^{n\times d_v}\}_{v=1}^{m}$ denotes the multi-view incomplete dataset consisting of n samples across m views with k clusters, where d_v is the feature dimension of the v-th view. Due to incomplete observations, we define the observed view set of sample i as :

$$\mathcal{V}_i = \{v \mid \mathbf{x}_i^{(v)} \text{ is observed}\} \subseteq \{1, \ldots, m\} \tag{1}$$

Let $\tilde{\mathbf{X}}^{(v)} \in \mathbb{R}^{n_v\times d_v}$ denote the observed instances in view v, where $n_v = |\{i \mid v \in \mathcal{V}_i\}|$ is the number of observed instances in view v. E_v and D_v are the respective encoder and decoder for view v, $\mathbf{H}^v \in \mathbb{R}^{n_v\times d}$ and $\hat{\mathbf{X}}^{(v)} \in \mathbb{R}^{n_v\times d_v}$ are the unified and reconstructed representations of $\tilde{\mathbf{X}}^{(v)}$. Let $\mathbf{h}_i^v$ to denote the ith row of $\mathbf{H}^v$. $\mathbf{P}_v \in \mathbb{R}^{k\times d}$ refers to the representation of k prototypes in view v.

Incomplete multi-view clustering aims to partition n unlabeled samples into k disjoint clusters through unsupervised learning, where each sample is only observed in certain views. To achieve this objective, we propose a structure-aware prototype disentanglement framework (SPORT), whose overall architecture is illustrated in Fig. 1.

3.2 Feature Representation Learning

Given the observed dataset $\tilde{\mathbf{X}}^{(v)} \in \mathbb{R}^{n_v \times d_v}$ of view v, to capture the relationships among samples, we utilize view-specific autoencoders Li et al. (2023b) to project all features into a common latent space. Specifically, using an encoder $E_v : \mathbb{R}^{n_v \times d_v} \to \mathbb{R}^{n_v \times d}$ and a decoder $D_v : \mathbb{R}^{n_v \times d} \to \mathbb{R}^{n_v \times d_v}$, the latent representation and the subsequent reconstruction process are formulated as $\mathbf{H}^v = E_v(\tilde{\mathbf{X}}^{(v)})$, $\hat{\mathbf{X}}^{(v)} = D_v(\mathbf{H}^v)$, where $\mathbf{H}^v \in \mathbb{R}^{n_v \times d}$ serves as the latent representation, and $\hat{\mathbf{X}}^{(v)} \in \mathbb{R}^{n_v \times d_v}$ is the reconstructed feature corresponding to view v.

To ensure the mapping fidelity of the autoencoders and preserve inherent structural characteristics, the overall reconstruction loss is defined as the Frobenius norm between the input and the reconstructed output across all views:

$$\mathcal{L}_{\text{REC}} = \sum_{v=1}^{m} \left\| \tilde{\mathbf{X}}^{(v)} - \hat{\mathbf{X}}^{(v)} \right\|_F^2, \tag{2}$$

where $\| \cdot \|_F$ denotes the Frobenius norm.

3.3 Structure-Aware Relational Contrastive Learning

Although the feature representation learning module maps all features into a unified latent subspace with high reconstruction fidelity, it overlooks the dual consistency inherent in multi-view data: (1) cross-view consistency, where different views of the same sample share common semantics, and (2) structural consistency, where samples from the same cluster exhibit similar feature characteristics. Existing contrastive learning-based IMVC methods mainly exploit cross-view consistency by aligning representations of the same sample across views Li et al. (2023a); Jin et al. (2023); Yuan et al. (2025), while largely ignoring structural consistency among semantically related instances. By treating only cross-view pairs of the same instance as positives and all others as negatives, these methods may separate semantically similar samples, thereby disrupting cluster-level structures and degrading clustering performance.

To address this limitation, we propose a structure-aware relational contrastive learning module to jointly model cross-view and structural consistency. Specifically, the contrastive objective takes cross-view features of the same sample as positive pairs, and further uses pairwise similarity weights to characterize inter-instance affinities. As a result, similar samples are softly involved as positives, whereas unrelated samples serve as negatives. The pairwise similarity weight w_{ij} is defined as:

$$w_{ij} = \begin{cases} 1, & i = j, \\ \dfrac{1}{|\mathcal{V}_i \cap \mathcal{V}_j|} \displaystyle\sum_{v \in \mathcal{V}_i \cap \mathcal{V}_j} \exp\big(\big(S(\mathbf{h}_i^v, \mathbf{h}_j^v) - 1\big)/\tau_w\big), & |\mathcal{V}_i \cap \mathcal{V}_j| \geq 1, \\ 0, & |\mathcal{V}_i \cap \mathcal{V}_j| = 0. \end{cases} \tag{3}$$

where w_{ij} denotes the inter-instance similarity weight between sample i and j with $w_{ij} = w_{ji}$ and $w_{ii} = 1$. $\mathcal{V}_i$ is the set of views where sample i has complete data defined in Eq.(1). $|\mathcal{V}_i \cap \mathcal{V}_j|$ is the cardinality of $\mathcal{V}_i \cap \mathcal{V}_j$ which represents the set of views where sample i and j both have complete data. τ_w is a temperature hyperparameter, and $S(\mathbf{h}_i^v, \mathbf{h}_j^v)$ is the cosine similarity between $\mathbf{h}_i^v$ and $\mathbf{h}_j^v$.

These weights characterize the sample-wise similarity between samples, which captures the relationship between samples. Furthermore, for a view pair (v, u), we define the set of instances simultaneously observed in both views as:

$$\mathcal{N}_{v,u} = \{i \mid v \in \mathcal{V}_i \text{ and } u \in \mathcal{V}_i\} \tag{4}$$

To achieve the objective of aligning the closely related features of samples while deviating from the uncorrelated ones, the objective loss for feature alignment is defined as Eq. (5):

$$\mathcal{L}_{\text{CON}} = -\sum_{\substack{v,u=1 \\ u \neq v}}^{m} \frac{1}{n_v} \sum_{i=1}^{n_v} \mathbf{1}[i \in \mathcal{N}_{v,u}] \log \frac{\displaystyle\sum_{j=1}^{n_u} w_{ij} \exp\big(S(\mathbf{h}_i^v, \mathbf{h}_j^u)/\tau\big)}{\displaystyle\sum_{j=1}^{n_v} \exp\big(S(\mathbf{h}_i^v, \mathbf{h}_j^v)/\tau\big) + \sum_{j=1}^{n_u} \exp\big(S(\mathbf{h}_i^v, \mathbf{h}_j^u)/\tau\big)}. \tag{5}$$

where τ is the temperature hyperparameter, and $\mathbf{1}[\cdot]$ is the indicator function.

3.4 Prototype Partial Alignment

To tackle POP, we design a prototype partial alignment module, which jointly utilizes cross-view consistency and view-specific semantics. Specifically, the ith prototype $\mathbf{p}_i^v \in \mathbb{R}^d$ of view v is decomposed each into two orthogonal complementary parts, namely the shared part $\mathbf{s}_i^v$ and view-specific $\mathbf{c}_i^v$ part. We separate these two components through a learnable weight matrix $\mathbf{U} \in \mathbb{R}^{d\times d_c}$ satisfying $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_{d_c}$, where $\mathbf{I}_{d_c}$ is the identity matrix $\in \mathbb{R}^{d_c\times d_c}$, d_c is a dimension smaller than d. Each prototype is decomposed into:

$$\mathbf{p}_i^v = \mathbf{s}_i^v + \mathbf{c}_i^v \tag{6}$$

where $\mathbf{s}_i^v$ and $\mathbf{c}_i^v$ are derived following:

$$\mathbf{s}_i^v = \mathbf{U}\mathbf{U}^\top\mathbf{p}_i^v \in \mathbb{R}^d, \quad \mathbf{c}_i^v = (\mathbf{I}_d - \mathbf{U}\mathbf{U}^\top)\mathbf{p}_i^v \in \mathbb{R}^d \tag{7}$$

Here $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_{d_c}$ preserves the orthogonality between $\mathbf{s}_i^v$ and $\mathbf{c}_i^v$ by:

$$\begin{aligned}(\mathbf{s}_i^v)^\top\,\mathbf{c}_i^v &= \left(\mathbf{U}\mathbf{U}^\top\,\mathbf{p}_i^v\right)^\top\left((\mathbf{I}_d - \mathbf{U}\mathbf{U}^\top)\,\mathbf{p}_i^v\right)\\ &= (\mathbf{p}_i^v)^\top(\mathbf{U}\mathbf{U}^\top)^\top(\mathbf{I}_d - \mathbf{U}\mathbf{U}^\top)\,\mathbf{p}_i^v\\ &= 0\end{aligned} \tag{8}$$

To ensure that the above decomposition is well-defined, the matrix $\mathbf{U}$ should remain column-orthonormal during optimization, such that $\mathbf{U}\mathbf{U}^\top$ forms a valid orthogonal projection onto the shared subspace. Instead of imposing an regularization term, we adopt a re-parameterization strategy based on the Cayley transform Jauch et al. (2018), which maps a skew-symmetric matrix to an orthogonal matrix. Specifically, we introduce a free matrix $\mathbf{W} \in \mathbb{R}^{d\times d}$ and construct a skew-symmetric matrix $\mathbf{A} = \mathbf{W} - \mathbf{W}^\top$. Then, an auxiliary orthogonal matrix $\mathbf{U}_{\text{full}} \in \mathbb{R}^{d\times d}$ is obtained at each forward pass as:

$$\mathbf{U}_{\text{full}} = (\mathbf{I}_d - \mathbf{A})(\mathbf{I}_d + \mathbf{A})^{-1} \tag{9}$$

Since $\mathbf{A}$ is skew-symmetric, i.e., $\mathbf{A}^\top = -\mathbf{A}$, the orthogonality of $\mathbf{U}_{\text{full}}$ holds by:

$$\mathbf{U}_{\text{full}}^\top\mathbf{U}_{\text{full}} = \left((\mathbf{I}_d - \mathbf{A})^{-1}(\mathbf{I}_d + \mathbf{A})\right)\left((\mathbf{I}_d - \mathbf{A})(\mathbf{I}_d + \mathbf{A})^{-1}\right) = \mathbf{I}_d$$

Finally $\mathbf{U}$ is derived by taking the first d_c columns of $\mathbf{U}_{\text{full}}$, and $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_{d_c}$ holds immediately.

After the prototypes are decomposed, to simultaneously capture cross-view consistency and view-specific semantics, we align the shared parts of the prototypes across different views while de-correlating the view-specific parts. This objective is achieved through a combination of the shared alignment loss and view-specific de-correlation loss.

Shared Alignment Loss To enforce the shared parts $\mathbf{s}_k^v$ is consistent across views for the same prototype-index g, we adopt a prototype contrastive objective. The shared alignment loss is defined as:

$$\mathcal{L}_{\text{SHARED}} = \sum_{\substack{v,u=1\\u\neq v}}^{m}\frac{1}{k}\sum_{r=1}^{t}\sum_{g=1}^{k}\frac{1}{r}\left(1 - \frac{\exp\left(S(\mathbf{s}_g^v, \mathbf{s}_g^u)/\tau\right)}{\sum_{j=1}^{k}\exp\left(S(\mathbf{s}_g^v, \mathbf{s}_j^u)/\tau\right)}\right)^r. \tag{10}$$

where τ is the temperature hyperparameter, t is a hyperparameter, k is the number of prototypes for each view. This loss simultaneously pulls together shared components of the same class across views and pushes apart those of different classes.

View-specific De-correlation Loss To allow the view-specific parts $\mathbf{c}_k^v$ capture genuinely distinct information across views rather than redundant features, we minimize the cosine alignment between view-specific parts of different views:

$$\mathcal{L}_{\text{DE}} = \frac{1}{k}\sum_{g=1}^{k}\sum_{\substack{v,u=1\\u\neq v}}^{m}\left|\left\langle\frac{\mathbf{c}_g^v}{\|\mathbf{c}_g^v\|}, \frac{\mathbf{c}_g^{(v')}}{\|\mathbf{c}_g^{(v')}\|}\right\rangle\right| \tag{11}$$

Minimizing this term drives the view-specific parts of different views toward orthogonal directions, ensuring each view retains its own distinctive representation.

The overall combinatorial loss is defined as:

$$\mathcal{L}_{\text{PRO}} = \lambda \mathcal{L}_{\text{SHARED}} + \eta \mathcal{L}_{\text{DE}} \tag{12}$$

where λ and η are hyperparameters used to balance the shared and view-specific loss.

3.5 Overall objective function

To achieve our IMVC objective, we combine the losses defined above, and the overall loss function for our SPORT model is formulated as:

$$\mathcal{L} = \mathcal{L}_{\text{REC}} + \alpha \mathcal{L}_{\text{CON}} + \beta \mathcal{L}_{\text{PRO}} \tag{13}$$

where α and β are hyperparameters to reconcile the contrastive loss and the prototype partial alignment loss. Due to space limits, the training procedure our SPORT model is differed Appendix B.

3.6 Hybrid Prototype Imputation Strategy

After prototype alignment, these prototypes are employed as semantic anchors for missing-view reconstruction. Existing methods Yuan et al. (2025); Jin et al. (2023); Li et al. (2023a) typically rely solely on the most relevant prototypes to recover missing features. Such approaches mainly exploit global shared information encoded in prototypes, while neglecting the local semantics captured by neighboring instances. Consequently, their performance often degrades under high missing rates or when unreliable prototypes are generated.

To mitigate this drawback, we propose a novel hybrid prototype imputation strategy which utilizes both prototypes and neighboring information. Specifically, for a missing data in view v of sample j, we first calculate the average cosine similarity between sample j and all prototypes in views where sample j is complete. Then, for all other samples who have complete data in view v, we calculate the average similarity between them and sample j in views where both samples have complete features. In addition following Eq.(14), we match the prototype and sample with the largest average cosine similarity with sample j, and denote their features in view v as $\hat{\mathbf{p}}_j^v$ and $\hat{\mathbf{n}}_j^v$ respectively. Lastly, the missing information of sample j in view v is imputed by a weighted sum of $\hat{\mathbf{p}}_j^v$ and $\hat{\mathbf{n}}_j^v$ as Eq.(15), where the imputed feature is denoted by $\hat{\mathbf{h}}_j^v$, and the weights are γ and $1-\gamma$ with $\gamma \in [0,1]$.

$$\hat{\mathbf{p}}_j^v = \arg\max_{\mathbf{p}_k^v \in \mathbf{P}_v} \frac{1}{|\mathcal{V}_j|} \sum_{v \in \mathcal{V}_j} S(\mathbf{p}_k^v, \mathbf{h}_j^v), \quad \hat{\mathbf{n}}_j^v = \arg\max_{\mathbf{h}_i^v \in \tilde{\mathbf{H}}^v} \frac{1}{|\mathcal{V}_j \cap \mathcal{V}_i|} \sum_{u \in \mathcal{V}_j \cap \mathcal{V}_i} S(\mathbf{h}_i^u, \mathbf{h}_j^u) \tag{14}$$

$$\hat{\mathbf{h}}_j^v = \gamma \cdot \hat{\mathbf{p}}_j^v + (1-\gamma) \cdot \hat{\mathbf{n}}_j^v, \quad \gamma \in [0,1] \tag{15}$$

By exploiting both prototype- and neighborhood-based information for missing data imputation, our method achieves stable clustering performance even under increasing missing rates, as demonstrated in the following experiments.

4 Experiments

In this section, we demonstrate the superior performance of our proposed SPORT model through through extensive evaluations, including experimental settings, performance comparisons, ablation study, sensitive analysis and clustering visualization.. Due to space limits, the experimental configurations, supplementary experiments are differed to Appendix C, E, F, G and H.

4.1 Experimental Settings

Benchmark Datasets: We construct comprehensive experiments on 6 benchmark multi-view datasets namely ALOI_100, Animal, Digit4k, 100Leaves, Reuters_21578, and VGGFace2_50. The detailed statistics of these datasets are shown in Appendix D, TABLE 3.

Table 1: Performance Comparison on Six Datasets Under Different Missing rates. The Optimal Results are In **Bold** and The Second-best Results are Underlined. O/M Represents Out of Memory.

	Missing Rate	0.1			0.3			0.5			0.7		
	Metric	ACC	ARI	F-score	ACC	ARI	F-score	ACC	ARI	F-score	ACC	ARI	F-score
Animal	CDIMC-net(IJCAI'21) Wen et al. (2021)	5.94	0.66	0.60	3.76	-0.01	0.72	3.90	-0.01	0.77	3.90	0.00	0.77
	COMPLETER (CVPR'21) Lin et al. (2021)	21.29	15.40	15.84	19.67	13.42	13.12	20.37	15.59	14.22	20.93	11.52	14.84
	SURE(TPAMI'22) Yang et al. (2022)	35.67	23.00	32.72	31.42	20.66	27.69	27.46	15.90	24.92	25.18	13.47	21.98
	ProImp (IJCAI'23) Li et al. (2023a)	41.40	30.05	38.02	38.51	27.47	34.63	38.14	26.90	34.52	37.65	26.06	34.44
	APADC (TIP'23) Xu et al. (2023a)	22.73	5.39	20.78	34.61	21.69	31.91	32.50	21.23	27.25	23.60	12.70	16.41
	IMVC-IE (ICASSP'24) Huang et al. (2024)	17.55	2.96	18.82	11.36	0.37	12.23	7.95	0.05	7.60	8.26	1.44	6.92
	RPCIC (ACM MM'24) Yuan et al. (2024)	43.84	31.05	45.41	46.05	34.39	47.42	44.93	32.40	46.42	35.64	23.27	36.44
	BURG (ICCV'25) Jin et al. (2025)	60.82	50.40	63.81	50.57	38.36	54.34	49.67	36.18	54.03	39.46	25.97	42.97
	BRIDGE (ICCV'25) Jiang et al. (2025)	20.12	14.60	17.46	18.80	13.16	15.92	17.54	11.94	15.75	27.58	17.76	25.74
	HSACC (NeurIPS'25) Ding et al. (2025)	20.29	12.98	16.82	24.03	17.63	17.55	24.77	18.89	20.10	23.02	16.33	17.60
	IMC-MCL (TKDE'25) Yin et al. (2025)	25.98	17.17	15.39	24.63	16.82	0.41	25.35	16.17	0.49	26.45	18.16	16.36
	MGCCFF(AAAI'25) Zhao et al. (2025)	11.69	4.45	9.70	11.53	4.43	9.40	10.92	3.86	9.13	10.81	3.87	8.46
	DGIMVCM (AAAI'26) Zhang et al. (2026)	62.45	52.07	59.09	49.10	40.75	44.72	45.27	34.44	41.50	36.10	23.76	35.53
	GMAE (TPAMI'26) Zou et al. (2026)	56.22	43.41	47.78	29.53	15.92	16.20	14.60	7.40	6.32	20.55	12.56	10.97
	SPORT (Ours)	**64.71**	**54.80**	**64.39**	**55.98**	**46.78**	**54.90**	**58.48**	**45.88**	**57.57**	**51.66**	**39.47**	**51.20**
ALOI_100	CDIMC-net(IJCAI'21) Wen et al. (2021)	62.02	48.70	57.36	37.92	18.34	38.04	21.52	7.85	21.85	13.48	3.77	14.05
	COMPLETER (CVPR'21) Lin et al. (2021)	20.83	11.77	20.12	20.28	12.02	19.29	16.96	9.83	14.39	18.94	10.68	18.28
	SURE(TPAMI'22) Yang et al. (2022)	61.75	52.36	60.64	63.24	53.62	62.20	56.37	46.45	54.52	57.09	46.53	55.45
	ProImp (IJCAI'23) Li et al. (2023a)	45.67	33.92	44.09	48.06	35.24	46.86	43.44	31.32	42.42	42.16	29.24	41.12
	APADC (TIP'23) Xu et al. (2023a)	26.55	10.12	20.59	16.41	8.86	10.50	16.41	7.32	11.70	22.35	10.14	19.31
	IMVC-IE (ICASSP'24) Huang et al. (2024)	35.92	23.28	33.07	30.04	18.13	27.47	26.64	16.48	25.08	23.23	13.27	21.61
	RPCIC (ACM MM'24) Yuan et al. (2024)	74.99	64.99	73.20	71.35	**63.07**	68.91	55.36	45.21	56.03	63.23	54.50	61.43
	BURG (ICCV'25) Jin et al. (2025)	61.80	48.14	60.02	50.19	29.84	51.06	41.10	19.06	43.49	29.94	14.34	32.29
	BRIDGE (ICCV'25) Jiang et al. (2025)	**78.63**	**72.04**	**75.47**	70.30	60.89	66.75	56.72	43.21	53.49	32.61	18.73	31.89
	HSACC (NeurIPS'25) Ding et al. (2025)	20.19	12.56	18.42	20.30	11.74	19.50	19.94	11.20	19.21	19.10	11.63	17.76
	IMC-MCL (TKDE'25) Yin et al. (2025)	17.47	10.36	0.90	17.76	11.38	0.91	15.44	9.46	0.65	16.20	9.89	0.44
	MGCCFF(AAAI'25) Zhao et al. (2025)	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M
	DGIMVCM (AAAI'26) Zhang et al. (2026)	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M
	GMAE (TPAMI'26) Zou et al. (2026)	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M
	SPORT (Ours)	77.15	69.05	74.18	**72.93**	62.87	**70.27**	**74.56**	**64.04**	**71.32**	**70.51**	**57.88**	**67.27**
Digit4k	CDIMC-net(IJCAI'21) Wen et al. (2021)	34.10	2.91	21.57	46.02	24.47	27.31	33.25	1.84	25.46	57.65	34.42	51.13
	COMPLETER (CVPR'21) Lin et al. (2021)	30.10	0.86	19.85	29.20	0.41	18.30	30.12	0.73	19.90	32.28	2.16	23.35
	SURE(TPAMI'22) Yang et al. (2022)	64.30	33.29	64.61	46.78	24.15	46.36	48.38	18.59	47.36	42.55	13.69	41.37
	ProImp (IJCAI'23) Li et al. (2023a)	61.96	35.17	62.22	53.10	21.06	53.30	78.57	52.51	78.51	62.32	29.83	62.86
	APADC (TIP'23) Xu et al. (2023a)	38.48	5.86	30.67	30.70	5.89	20.43	36.78	9.39	27.55	34.82	7.26	25.47
	IMVC-IE (ICASSP'24) Huang et al. (2024)	82.05	60.85	60.80	62.72	31.13	46.10	56.50	20.68	44.36	57.45	22.63	44.70
	RPCIC (ACM MM'24) Yuan et al. (2024)	80.60	57.70	**80.34**	55.60	26.40	56.39	39.92	18.77	39.90	42.58	14.25	43.89
	BURG (ICCV'25) Jin et al. (2025)	69.45	38.68	69.89	68.85	27.49	**69.35**	79.35	56.68	**78.61**	53.72	35.71	50.60
	BRIDGE (ICCV'25) Jiang et al. (2025)	47.17	16.83	36.89	71.48	45.76	53.03	68.14	46.00	49.26	68.04	44.02	49.74
	HSACC (NeurIPS'25) Ding et al. (2025)	35.45	4.42	28.63	35.22	3.95	28.26	35.20	6.71	28.17	37.08	6.46	30.73
	IMC-MCL (TKDE'25) Yin et al. (2025)	73.85	40.04	54.50	**81.58**	59.56	59.18	64.62	32.49	48.02	61.88	30.17	48.93
	MGCCFF(AAAI'25) Zhao et al. (2025)	64.84	40.80	64.31	61.49	39.43	61.01	65.05	38.94	64.27	62.37	35.82	61.28
	DGIMVCM (AAAI'26) Zhang et al. (2026)	64.18	56.35	59.66	70.18	**60.80**	60.02	63.58	43.07	63.56	63.80	41.66	63.68
	GMAE (TPAMI'26) Zou et al. (2026)	65.45	57.50	61.74	66.15	28.99	61.28	59.62	46.66	56.18	55.70	37.42	52.43
	SPORT (Ours)	**85.70**	**67.06**	65.14	81.22	58.41	62.58	**83.95**	**63.52**	63.90	**84.12**	**64.40**	**63.92**
100Leaves	CDIMC-net(IJCAI'21) Wen et al. (2021)	70.75	61.18	64.96	47.56	38.38	43.82	31.37	20.27	29.35	24.88	11.27	23.95
	COMPLETER (CVPR'21) Lin et al. (2021)	29.38	16.97	27.13	26.44	14.10	24.21	37.44	21.27	37.81	29.06	15.54	28.07
	SURE(TPAMI'22) Yang et al. (2022)	56.62	41.57	53.66	51.38	37.40	48.84	48.00	35.09	45.19	37.62	24.15	36.05
	ProImp (IJCAI'23) Li et al. (2023a)	63.44	56.51	58.70	48.62	37.08	45.33	45.94	34.29	43.37	41.94	27.43	41.07
	APADC (TIP'23) Xu et al. (2023a)	24.06	12.29	19.01	18.75	8.12	15.32	20.81	8.00	18.06	19.31	8.18	16.37
	IMVC-IE (ICASSP'24) Huang et al. (2024)	39.38	24.22	38.86	29.19	13.76	28.67	22.75	7.92	23.10	15.44	3.11	15.28
	RPCIC (ACM MM'24) Yuan et al. (2024)	74.50	62.36	74.70	59.25	38.22	61.77	47.06	30.51	48.42	45.56	31.63	44.40
	BURG (ICCV'25) Jin et al. (2025)	71.19	58.87	69.12	57.06	37.48	56.85	45.38	21.17	47.49	34.62	14.42	36.58
	BRIDGE (ICCV'25) Jiang et al. (2025)	49.30	35.14	47.68	29.07	13.63	28.67	26.29	9.57	26.53	28.66	11.95	28.66
	HSACC (NeurIPS'25) Ding et al. (2025)	27.31	15.61	24.49	29.38	18.24	25.82	29.38	18.14	26.49	31.00	16.15	28.90
	IMC-MCL (TKDE'25) Yin et al. (2025)	9.00	7.87	0.07	7.12	5.54	2.07	6.62	4.37	0.06	6.75	3.40	0.18
	MGCCFF(AAAI'25) Zhao et al. (2025)	6.56	3.88	0.21	9.80	5.88	0.41	10.66	5.09	0.24	10.84	6.72	4.72
	DGIMVCM (AAAI'26) Zhang et al. (2026)	**85.21**	**81.28**	**83.76**	**72.56**	**64.68**	70.14	63.50	53.01	61.11	48.88	36.25	46.71
	GMAE (TPAMI'26) Zou et al. (2026)	67.81	56.83	64.96	42.44	23.94	42.06	37.56	26.55	35.71	35.19	20.11	34.82
	SPORT (Ours)	80.62	72.95	78.71	74.06	62.31	**71.18**	**69.56**	**54.35**	**67.47**	**56.06**	**36.68**	**54.24**
Reuters_21578	CDIMC-net(IJCAI'21) Wen et al. (2021)	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M
	COMPLETER (CVPR'21) Lin et al. (2021)	33.87	1.03	18.64	30.93	1.99	17.31	34.40	1.27	18.92	32.40	1.91	19.71
	SURE(TPAMI'22) Yang et al. (2022)	32.27	9.36	28.75	31.40	7.68	29.01	29.80	4.00	23.90	29.33	4.64	25.29
	ProImp (IJCAI'23) Li et al. (2023a)	40.87	16.06	38.46	41.13	17.09	28.34	31.00	9.49	29.28	35.33	10.58	31.42
	APADC (TIP'23) Xu et al. (2023a)	28.07	18.30	41.09	27.40	7.29	29.05	28.53	18.83	47.69	29.93	9.84	38.89
	IMVC-IE (ICASSP'24) Huang et al. (2024)	3.93	18.95	47.43	7.33	12.10	39.40	6.80	11.21	39.11	3.93	8.29	32.06
	RPCIC (ACM MM'24) Yuan et al. (2024)	46.00	18.38	46.40	46.93	17.54	46.54	46.13	17.77	44.78	**48.07**	20.32	**46.16**
	BURG (ICCV'25) Jin et al. (2025)	45.93	11.81	35.53	44.20	10.08	35.28	27.67	−0.05	12.28	47.07	14.33	42.47
	BRIDGE (ICCV'25) Jiang et al. (2025)	5.40	**25.94**	44.69	6.39	**23.57**	46.43	7.58	23.39	44.71	2.77	8.08	35.69
	HSACC (NeurIPS'25) Ding et al. (2025)	38.20	5.14	31.71	34.93	1.59	30.04	37.20	4.69	31.83	32.93	2.94	24.51
	IMC-MCL (TKDE'25) Yin et al. (2025)	33.67	11.72	37.56	38.27	11.85	41.65	29.67	7.27	31.63	35.33	7.07	32.39
	MGCCFF(AAAI'25) Zhao et al. (2025)	44.73	16.87	39.15	46.79	22.50	46.09	45.81	17.21	40.26	46.53	**21.86**	**46.16**
	DGIMVCM (AAAI'26) Zhang et al. (2026)	40.60	9.03	20.62	38.20	6.14	21.39	32.87	5.37	22.39	29.00	3.29	20.31
	GMAE (TPAMI'26) Zou et al. (2026)	34.00	2.06	20.60	31.80	0.07	14.43	25.87	2.40	19.89	27.13	1.57	23.98
	SPORT (Ours)	**46.33**	21.73	48.27	**51.00**	23.45	**51.73**	**53.47**	**26.44**	**51.39**	38.13	15.87	39.20
VGGFace2_50	CDIMC-net(IJCAI'21) Wen et al. (2021)	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M
	COMPLETER (CVPR'21) Lin et al. (2021)	4.72	0.73	4.10	5.27	0.94	3.74	4.70	0.74	3.91	4.76	0.82	4.05
	SURE(TPAMI'22) Yang et al. (2022)	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M
	ProImp (IJCAI'23) Li et al. (2023a)	7.62	1.48	7.57	7.15	1.37	7.05	8.30	1.66	8.19	7.18	1.41	7.13
	APADC (TIP'23) Xu et al. (2023a)	5.42	0.94	4.02	5.58	0.92	3.84	5.24	0.84	3.21	5.62	1.12	2.87
	IMVC-IE (ICASSP'24) Huang et al. (2024)	6.82	1.23	6.43	5.44	0.63	5.32	4.79	0.38	4.76	4.62	0.38	4.20
	RPCIC (ACM MM'24) Yuan et al. (2024)	8.21	1.92	8.19	9.62	2.53	9.62	9.96	2.58	10.03	**10.31**	2.82	**10.46**
	BURG (ICCV'25) Jin et al. (2025)	10.04	2.90	9.96	9.40	1.82	9.52	8.11	1.44	8.26	5.79	0.78	5.81
	BRIDGE (ICCV'25) Jiang et al. (2025)	5.03	0.78	4.82	5.49	1.25	4.78	5.60	0.79	5.61	5.53	0.88	5.43
	HSACC (NeurIPS'25) Ding et al. (2025)	5.39	1.06	4.21	5.47	0.97	4.04	5.33	0.94	4.08	5.51	0.92	4.10
	IMC-MCL (TKDE'25) Yin et al. (2025)	5.05	0.73	3.56	4.93	0.63	3.89	4.64	0.59	2.66	5.00	0.75	3.54
	MGCCFF(AAAI'25) Zhao et al. (2025)	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M
	DGIMVCM (AAAI'26) Zhang et al. (2026)	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M
	GMAE (TPAMI'26) Zou et al. (2026)	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M	O/M
	SPORT (Ours)	**10.42**	**3.05**	**10.23**	**10.41**	**3.01**	**10.33**	**11.29**	**3.17**	**11.23**	10.24	**2.92**	9.82

Evaluating Metrics: We used three widely-used evaluation criteria to demonstrate the validity of the experiment. These metrics are accuracy(ACC), Adjusted Rand Index(ARI) and F-Score(F-Sco).

Table 2: Ablation Study Results on Six Benchmark Datasets With 0.5 Missing Rate. The Best Results are in **Bold**, and the Second-best Results are Underlined.

Components					Reuters_21578			100Leaves			Digit4k			Animal			ALoi100			VGGFace2_50		
$\mathcal{L}_{REC}$	$\mathcal{L}_{CON}$	$\mathcal{L}_{PRO}$	$\mathcal{L}_{SHARED}$	$\mathcal{L}_{DE}$	ACC	ARI	F-Sco	ACC	ARI	F-Sco	ACC	ARI	F-Sco	ACC	ARI	F-Sco	ACC	ARI	F-Sco	ACC	ARI	F-Sco
✓					44.08	18.50	44.57	66.50	50.96	63.84	83.90	63.44	63.86	53.72	42.08	53.80	70.59	58.71	67.73	10.50	3.01	10.20
	✓				43.40	17.45	43.21	69.06	53.83	67.25	81.52	58.85	62.47	39.88	27.11	40.72	66.31	50.02	63.47	4.75	0.71	4.66
		✓	✓	✓	51.87	25.88	**52.84**	67.62	51.59	64.71	83.42	62.44	63.63	52.64	40.35	52.93	71.30	60.35	67.58	10.02	2.88	9.74
✓	✓				45.27	20.37	47.12	67.38	51.78	65.63	83.92	63.48	63.87	52.81	42.04	52.69	73.06	61.20	69.90	10.44	3.02	10.03
✓		✓	✓	✓	44.80	18.53	44.58	66.44	50.26	63.10	83.38	62.46	63.44	56.65	44.34	56.24	70.80	59.98	67.31	9.90	2.75	9.95
	✓	✓	✓	✓	46.67	19.57	46.85	68.25	53.01	66.12	81.50	58.80	62.46	41.77	28.46	41.86	67.15	55.43	64.53	4.87	0.73	4.73
✓	✓	✓	✓		45.07	20.29	45.92	69.38	53.46	66.70	81.50	58.80	62.46	56.91	45.21	56.89	73.36	61.52	70.33	11.04	3.11	10.91
✓	✓	✓		✓	45.13	20.28	47.00	69.38	53.73	**67.47**	81.52	58.85	62.47	52.96	42.24	52.77	72.22	60.97	69.12	11.13	**3.18**	10.99
✓	✓	✓	✓	✓	**53.47**	**26.44**	51.39	**69.56**	**54.35**	**67.47**	**83.95**	**63.52**	**63.90**	**58.48**	**45.88**	**57.57**	**74.56**	**64.04**	**71.32**	**11.29**	3.17	**11.23**

Comparing Methods: We choose 14 state-of-the-art IMVC methods as baselines and compare them with our SPORT framework to investigate SPORT's effectiveness. These baselines are CDIMC-net (IJCAI'21) Wen et al. (2021), COMPLETER (CVPR'21) Lin et al. (2021), SURE (TPAMI'22) Yang et al. (2022), ProImp (IJCAI'23) Li et al. (2023a), APADC (TIP'23) Xu et al. (2023a), IMVC-IE (ICASSP'24) Huang et al. (2024), RPCIC (ACM MM'24) Yuan et al. (2024), BURG (ICCV'25) Jin et al. (2025), BRIDGE (ICCV'25) Jiang et al. (2025), HSACC (NeurIPS'25) Ding et al. (2025), IMC-MCL (TKDE'25) Yin et al. (2025), MGCCFF (AAAI'25) Zhao et al. (2025), DGIMVCM (AAAI'26) Zhang et al. (2026) and GMAE (TPAMI'26) Zou et al. (2026).

4.2 Performance Comparisons

The clustering performance of SPORT and 14 competing methods on six multi-view datasets under different missing rates is reported in Table 1. We make the following observations.

(1) **SPORT consistently achieves superior performance across most datasets.** On Animal, for example, SPORT achieves the best ACC, ARI, and F-score under all missing rates. At a missing rate of 0.5, SPORT outperforms the second-best method, BURG, by 17.7%, 26.8%, and 6.5% in ACC, ARI, and F-score, respectively. This advantage can be attributed to the structure-aware relational contrastive learning and prototype partial alignment modules, which jointly capture structural and cross-view consistency while preserving view-specific information.

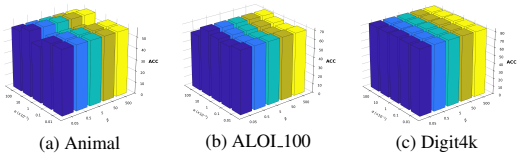
(2) **SPORT exhibits strong robustness to increasing missing rates.** On 100Leaves, its ACC decreases by only 8.1% and 6.7% as the missing rate increases from 0.1 to 0.3 and from 0.3 to 0.5, respectively, whereas GMAE, despite performing well under low missing rates, drops by 14.8% and 12.5%. On Digit4k, SPORT maintains an ACC above 80% across all missing rates, with only a 1.8% decline as the missing rate increases from 0.1 to 0.7. These results demonstrate the effectiveness of the hybrid imputation strategy in jointly exploiting global prototype semantics and local neighborhood context, enabling reliable missing-view recovery and robust clustering under severe missingness.

4.3 Ablation Study

As defined in Eq.(13), our SPORT model consists of three major modules namely the feature representation learning module, the structure-aware relational contrastive learning module and the prototype partial alignment module. To further validate the contributions of each individual module, in this subsection we conduct ablation study on all six datasets with a missing rate of 0.5. The components are systematically removed while the model is retrained under the same configuration. The results are reported in TABLE 2. Overall, removing any major module generally degrades clustering performance, demonstrating their complementary roles. For example, without prototype partial alignment, the ACC on Reuters_21578 decreases from 53.47% to 45.27%. Removing $\mathcal{L}_{REC}$ causes the ARI on Animal to drop from 45.88% to 28.46%, while removing $\mathcal{L}_{CON}$ reduces the ACC on Reuters_21578 by 8.67 percentage points. Moreover, the components within prototype partial alignment are also important; for instance, removing $\mathcal{L}_{SHARED}$ decreases the ACC on Animal by 5.52%. These results verify that the joint optimization of the three modules enables SPORT to learn more discriminative and semantically consistent representations, leading to superior overall clustering performance.

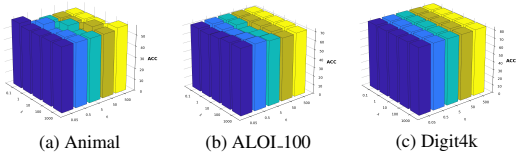
4.4 SENSITIVE ANALYSIS

In this paper, there are four major hyperparameters, α, β, λ, and η. Specifically, α and β are used to balance the structure-aware relational contrastive learning loss and the prototype partial alignment loss, while λ and η are used to balance the shared and view-specific losses. All six datasets are used to evaluate the stability of the model with respect to α, β, λ, and η, where $\alpha \in \{10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}\}$, $\beta \in \{5 \times 10^{-2}, 5 \times 10^{-1}, 5 \times 10^{0}, 5 \times 10^{1}, 5 \times 10^{2}\}$, $\lambda \in \{0.1, 1, 10, 100, 1000\}$, and $\eta \in \{0.05, 0.5, 5, 50, 500\}$. The missing rate for all datasets is set to 0.5. The detailed results are visualized in Fig. 2, Fig. 3, and Appendix F. Overall, despite some minor fluctuations under varying hyperparameters, SPORT exhibits strong robustness to hyperparameter variations.



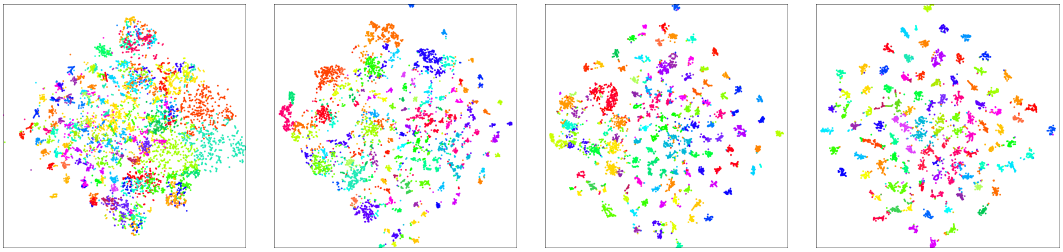
(a) Animal (b) ALOI_100 (c) Digit4k

Figure 2: Sensitivity analysis of α and β at a missing rate of 0.5.



(a) Animal (b) ALOI_100 (c) Digit4k

Figure 3: Sensitivity analysis of λ and η at a missing rate of 0.5.



(a) Before pretraining (ACC:0.4142 ARI:0.2718) (b) After pretraining (ACC:0.4650 ARI:0.3115) (c) After fine-tuning (ACC:0.5045 ARI:0.3303) (d) After imputation (ACC:0.7293 ARI:0.6287)

Figure 4: Visualization of feature representations of ALOI_100 at four stages.

4.5 REPRESENTATION VISUALIZATION

Fig.4 presents the t-SNE visualizations of the learned multi-view feature representations at different training stages, including: (a) before pretraining with raw features, (b) after pretraining, (c) after fine-tuning, and (d) after imputation. The experiments are conducted on the ALOI_100 dataset under a missing rate of 0.3. Across these four stages, the learned representations become progressively more discriminative, evolving from highly mixed and scattered distributions to compact and well-separated cluster structures. This progression aligns with the staged training strategy, confirming that pretraining, fine-tuning, and prototype imputation each contribute distinctly to improving feature separability under incomplete multi-view conditions.

5 CONCLUSION

In this paper, we propose SPORT, an incomplete multi-view clustering framework that jointly preserves cross-view consistency, structural relationships, and view-specific semantics. SPORT learns shared semantic representations and captures neighborhood-level relations through structure-aware relational contrastive learning, while partially aligning prototypes to preserve complementary view-specific information. To recover missing views, we further introduce a hybrid prototype imputation strategy that combines global prototype semantics with local neighborhood context for more reliable cross-view reconstruction. Extensive experiments on multiple benchmark datasets, together with comprehensive analyses and visualizations, demonstrate the effectiveness of SPORT in learning discriminative representations and achieving superior clustering performance under incomplete multi-view settings.

AI USE STATEMENT

In this work, generative AI tools were used solely to improve the readability, grammar, clarity of the manuscript and organize the code. They were not used to develop the methodology, formulate theoretical claims, interpret results, or generate scientific conclusions. All AI-assisted edits were reviewed and verified by the authors, who take full responsibility for the final content of the pap

REFERENCES

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*, volume 38, pp. 11221–11229, 2024.

Gabriele Corso, Hannes Stark, Stefanie Jegelka, Tommi Jaakkola, and Regina Barzilay. Graph neural networks. *Nature Reviews Methods Primers*, 4(1):17, 2024.

Xiaojian Ding, Lin Zhao, Xian Li, and Xiaoying Zhu. Incomplete multi-view clustering via hierarchical semantic alignment and cooperative completion. *Advances in Neural Information Processing Systems*, 38:52175–52195, 2025.

Yiming Du, Yao Wang, Ziyu Wang, Rui Ning, and Lusi Li. Pgformer: A prototype-graph transformer for incomplete multiview clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 37(3):1163–1175, 2026.

Sifan Fang, Zuyuan Yang, and Junhang Chen. Incomplete multi-view clustering via diffusion completion. *Multimedia Tools and Applications*, 83(18):55889–55902, 2024.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

Zhibin Gu and Weili Wang. Hypergraph-enhanced contrastive learning for multi-view clustering with hyper-laplacian regularization. *Advances in Neural Information Processing Systems*, 38: 54600–54620, 2025.

Jun Guo and Jiahui Ye. Anchors bring ease: An embarrassingly simple approach to partial multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 118–125, 2019.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Binqiang Huang, Zhijie Huang, Shoujie Lan, Qinghai Zheng, and Yuanlong Yu. Incomplete multi-view clustering via inference and evaluation. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8180–8184, 2024.

Michael Jauch, Peter D. Hoff, and David B. Dunson. Random orthogonal matrices and the cayley transform, 2018. URL `https://arxiv.org/abs/1810.02881`.

Xiaorui Jiang, Buyun He, Peng Yuan Zhou, Xinyue Chen, Jingcai Guo, Jie Xu, and Yong Liao. A unified framework to bridge complete and incomplete deep multi-view clustering under non-iid missing patterns. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 594–603, 2025.

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

Jiaqi Jin, Siwei Wang, Zhibin Dong, Xihong Yang, Xinwang Liu, En Zhu, and Kunlun He. Deep incomplete multi-view clustering with distribution dual-consistency recovery guidance. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1016–1026, 2025.

Ao Li, Sanlin Mei, Fengwei Gu, Dehua Miao, and Tianyu Gao. Attention-based deep incomplete multi-view clustering via bi-alignment guidance. *Complex & Intelligent Systems*, 11(8):349, 2025.

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*, 2023a.

Pengzhi Li, Yan Pei, and Jianqiang Li. A comprehensive survey on design and application of autoencoder in deep learning. *Applied Soft Computing*, 138:110176, 2023b.

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*, pp. 11174–11183, 2021.

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.

Chengliang Liu, Jie Wen, Yong Xu, Bob Zhang, Liqiang Nie, and Min Zhang. Reliable representation learning for incomplete multi-view missing multi-label classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6):4940–4956, 2025.

Xinwang Liu, Ke Liang, Jun Wang, Suyuan Liu, Xiangke Wang, and Huaimin Wang. Two decades of multi-view clustering: Taxonomy, application, and challenge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(3):3744–3764, 2026.

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*, volume 38, pp. 14193–14201, 2024.

Leif E Peterson. K-nearest neighbor. *Scholarpedia*, 4(2):1883, 2009.

Qianyao Qiang, Bin Zhang, Yunjia Hua, and Feiping Nie. Multi-view clustering via bilaterally constrained anchor graph. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–14, 2026.

Yang Qin, Yanglin Feng, Yuan Sun, Dezhong Peng, Xi Peng, and Peng Hu. Deep information-balanced multimodal learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–13, 2026.

Chao Shang, Aaron Palmer, Jiangwen Sun, Ko-Shin Chen, Jin Lu, and Jinbo Bi. Vigan: Missing view imputation with generative adversarial networks. In *2017 IEEE International conference on big data (Big Data)*, pp. 766–775, 2017.

Huayi Tang and Yong Liu. Deep safe incomplete multi-view clustering: Theorem and algorithm. In *International conference on machine learning*, pp. 21090–21110, 2022.

Ge Teng, Ting Mao, Chen Shen, Xiang Tian, Xuesong Liu, Yaowu Chen, and Jieping Ye. Urrl-imvc: Unified and robust representation learning for incomplete multi-view clustering. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2888–2899, 2024.

Daniel J Trosten, Sigurd Løkse, Robert Jenssen, and Michael C Kampffmeyer. On the effects of self-supervision and contrastive alignment in deep multi-view clustering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 23976–23985, 2023.

Jiatai Wang, Zhiwei Xu, Xuewen Yang, Dongjin Guo, and Limin Liu. Self-supervised image clustering from multiple incomplete views via constrastive complementary generation. *IET Computer Vision*, 17(2):189–202, 2023.

Qianqian Wang. Llm-damvc: A large language model assisted dynamic agent for multi-view clustering. *Advances in Neural Information Processing Systems*, 38:119515–119534, 2025.

Qianqian Wang, Zhengming Ding, Zhiqiang Tao, Quanxue Gao, and Yun Fu. Partial multi-view clustering via consistent gan. In *2018 IEEE International Conference on Data Mining (ICDM)*, pp. 1290–1295, 2018.

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.

Qianqian Wang, Haiming Xu, Wei Feng, and Quanxue Gao. Adversarial fair incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 10011–10019, 2026.

Yiming Wang, Qun Li, Dongxia Chang, Jie Wen, Hua Dai, Fu Xiao, and Yao Zhao. Learning from disjoint views: A contrastive prototype matching network for fully incomplete multi-view clustering. *Advances in Neural Information Processing Systems*, 38:18265–18284, 2025.

Jie Wen, Yong Xu, and Hong Liu. Incomplete multiview spectral clustering with adaptive graph learning. *IEEE Transactions on Cybernetics*, 50(4):1418–1429, 2020.

Jie Wen, Zhihao Wu, Zheng Zhang, Lunke Fei, Bob Zhang, and Yong Xu. Structural deep incomplete multi-view clustering network. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pp. 3538–3542, 2021.

Jie Wen, Shijie Deng, Waikeung Wong, Guoqing Chao, Chao Huang, Lunke Fei, and Yong Xu. Diffusion-based missing-view generation with the application on incomplete multi-view clustering. In *Forty-first international conference on machine learning*, volume 235, pp. 52762–52778, 2024.

Benyu Wu, Wei Du, Jun Wang, and Guoxian Yu. Imputation-free incomplete multi-view clustering via knowledge distillation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 5071–5081, 2025a.

Tingting Wu, Zhendong Li, Zhibin Gu, Jiazheng Yuan, and Songhe Feng. Koala: Kernel coupling and element imputation induced multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 21581–21589, 2025b.

Cai Xu, Ziyu Guan, Wei Zhao, Hongchang Wu, Yunfei Niu, and Beilei Ling. Adversarial incomplete multi-view clustering. In *Ijcai*, volume 7, pp. 3933–3939, 2019.

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, 2023a.

Jie Xu, Yazhou Ren, Xiaoshuang Shi, Heng Tao Shen, and Xiaofeng Zhu. Untie: Clustering analysis with disentanglement in multi-view information fusion. *Information Fusion*, 100:101937, 2023b.

Jinglin Xu, Junwei Han, and Feiping Nie. Discriminatively embedded k-means for multi-view clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5356–5364, 2016.

Wenbiao Yan, Jihua Zhu, Yiyang Zhou, Jinqian Chen, Haozhe Cheng, Kun Yue, and Qinghai Zheng. Neighbor-based completion for addressing incomplete multiview clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

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

Jun Yin, Pei Wang, Shiliang Sun, and Zhonglong Zheng. Incomplete multi-view clustering via multi-level contrastive learning. *IEEE Transactions on Knowledge and Data Engineering*, 2025.

Honglin Yuan, Shiyun Lai, Xingfeng Li, Jian Dai, Yuan Sun, and Zhenwen Ren. Robust prototype completion for incomplete multi-view clustering. In *Proceedings of the 32nd ACM international conference on multimedia*, pp. 10402–10411, 2024.

Honglin Yuan, Yuan Sun, Fei Zhou, Jing Wen, Shihua Yuan, Xiaojian You, and Zhenwen Ren. Prototype matching learning for incomplete multi-view clustering. *IEEE Transactions on Image Processing*, 34:828–841, 2025.

Yong Zhang, Li Jiang, Da Liu, and Wenzhe Liu. Incomplete multi-view clustering via self-attention networks and feature reconstruction: Y. zhang et al. *Applied Intelligence*, 54(4):2998–3016, 2024.

Yuanyang Zhang, Yijie Lin, Weiqing Yan, Li Yao, Xinhang Wan, Guangyuan Li, Chao Zhang, Guanzhou Ke, and Jie Xu. Incomplete multi-view clustering via diffusion contrastive generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 22650–22658, 2025.

Zhenghao Zhang, Jun Xie, Xingchen Chen, Tao Yu, Hongzhu Yi, Kaixin Xu, Yuanxiang Wang, Tianyu Zong, Xinming Wang, Jiahuan Chen, et al. Dynamic deep graph learning for incomplete multi-view clustering with masked graph reconstruction loss. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 28600–28608, 2026.

Handong Zhao, Zhengming Ding, and Yun Fu. Multi-view clustering via deep matrix factorization. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

Liang Zhao, Ziyue Wang, Xiao Wang, Zhikui Chen, and Bo Xu. Incomplete and unpaired multi-view graph clustering with cross-view feature fusion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 22786–22794, 2025.

Wei Zhou, Hao Wang, and Yan Yang. Consensus graph learning for incomplete multi-view clustering. In *Lecture Notes in Computer Science*, pp. 529–540, 2019.

Zhengzhong Zhu, Pei Zhou, Lanxi Bai, Jia Nie, Li Cheng, Shiquan Min, and Jiangping Zhu. Prototype and sample level semantic alignment for incomplete multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5818–5827, 2026.

Xin Zou, Ruimeng Liu, Chang Tang, Zhenglai Li, Xinwang Liu, Kunlun He, and Wanqing Li. Learning disentangled representations for generalized multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–16, 2026.

A SUPPLEMENTARY RELATED WORKS

A.1 GENERATOR-BASED METHODS

These methods recover missing views with the help pf generative models.Shang et al. (2017); Wang et al. (2023); Xu et al. (2019); Wang et al. (2018); Wen et al. (2024); Fang et al. (2024); Zhang et al. (2025); Goodfellow et al. (2020); Wang et al. (2021). Early works predominantly adopt GANs due to their ability to learn high-fidelity representations via adversarial training. For example, Xu et al. (2019) proposed an adversarial feature recovery framework based on GANs, while Wang et al. (2021) incorporated cycle-consistency and adaptive fusion to handle incomplete views. More recently, diffusion models have attracted increasing attention due to their training stability and strong mode coverage. Representative works include Fang et al. (2024), which initialize missing features with Gaussian noise and progressively recover them via a diffusion-based completion process, and Zhang et al. (2025), which further integrates diffusion modeling with contrastive learning, yielding more discriminative and compact clustering representations.

A.2 PREDICTOR-BASED METHODS

These methods infer missing features from observed ones through an information-theoretic framework. Lin et al. (2021; 2022); Xu et al. (2023a); Lu et al. (2024); Ding et al. (2025). For example, Lin et al. (2021) first established a unified information-theoretic framework that jointly learned consistent representations and recovers missing views by maximizing cross-view mutual information. Building upon this work, Lin et al. (2022) introduced a dual contrastive loss and a dual prediction loss to simultaneously enforce cross-view consistency and data recoverability. To further mitigate the distribution discrepancy between complete and incomplete data, Xu et al. (2023a) designed an adaptive feature projection mechanism that aligned the distributions between complete and incomplete data by optimizing mutual information, yielding more faithful feature recoverability. Beyond single-level prediction, Yin et al. (2025) formulated a multi-level dual prediction strategy in which both high-level and semantic features are extracted via an attention layer and optimized for missing data prediction.

A.3 NEIGHBORHOOD-BASED METHODS

These methods address data incompleteness by exploiting local structural relationships within and across views. Yang et al. (2022); Tang & Liu (2022); Wen et al. (2021); Teng et al. (2024); Yan et al. (2025). K-nearest neighbor (kNN) Peterson (2009) is a classic machine learning method that is used to integrate features from the top-K spatial neighbors. For example, Teng et al. (2024) integrated multi-view and neighborhood information using complementary encoders with kNN-based augmentation and reconstruction. Yan et al. (2025) explored the latent common clustering information embedded in multi-view data and performed neighbor-based completion via kNN. Beyond instance-level aggregation, graph neural networks (GNNs) Corso et al. (2024) provide an alternative strategy for recovering missing views from adjacent samples. Typical works include Wen et al. (2021) used a GCN to aggregate adjacent information and fused a consensus graph to repair missing view features and Chao et al. (2024) which transferred multi-view consistency relations via GCN, while applying high-confidence guiding with instance-level contrastive learning for missing view reconstruction.

A.4 PROTOTYPE-BASED METHODS

These methods learn representative prototypes for individual views to encode global semantic information and use them to reconstruct missing data Yuan et al. (2025); Jin et al. (2023); Li et al. (2025); Wang et al. (2025); Li et al. (2023a); Du et al. (2026); Zhu et al. (2026). These prototypes are often misaligned across views, which has motivated extensive research on cross-view prototype alignment. For example, Yuan et al. (2025) addressed the prototype unaligned problem (PUP) by aligning the generated prototypes across views via contrastive learning and proposed a novel prototype matching framework to impute the missing data. Jin et al. (2023) considered the IMVC problem from a partially-aligned setting and generated view-specific prototypes via l_2 loss minimization while reconstructing missing views through contrastive learning-free aligned prototypes.

Other works utilized graphs to capture semantic dependencies among prototypes to enhance prototype alignment. For instance, Du et al. (2026) formulated a prototype-guided graph transformer framework that integrated dual attention mechanisms with cross-view alignment which mitigated the influence of noise and outliers. Zhu et al. (2026) introduced a similarity graph-based framework that captures semantic-invariant structural information across prototypes for prototype alignment, further improving quality of the aligned prototypes.

B THE TRAINING PROCEDURE OF SPORT

Algorithm 1 The training procedure of SPORT

Input: Multi-view raw data $\{\mathbf{X}^v\}_{v=1}^m$, number of clusters k, pretraining epoch E_{PRE}, fine-tuning epoch E_{FIN},
Multi-view raw complete data $\{\tilde{\mathbf{X}}^v\}_{v=1}^m$ with size s, Batch size B
Output: Clustering results via k-means

Pretraining:
for $t = 1, \ldots, E_{PRE}$ **do**
  for $i = 1, \ldots, s/B$ **do**
    Pretrain the autoencoder by optimizing Eq. (2)
  end for
end for
Feature construction:
Obtain features $\{\mathbf{H}^v\}_{v=1}^V$ from the encoders
Fine-Tuning:
for $t = 1, \ldots, E_{FIN}$ **do**
  for $i = 1, \ldots, n/B$ **do**
    Features are aligned by Eq. (5)
    Initial prototypes $\{\mathbf{P}_v\}_{v=1}^m$ via k-means
    Prototypes are decomposed by Eq. (7)
    Prototypes are optimized by Eq. (12)
    Model optimized by overall loss in Eq. (13)
  end for
end for
Prototype imputation:
Impute missing data via hybrid prototype imputation
Clustering:
Apply k-means to obtain final clustering results

In the pretraining phase, only the reconstruction loss defined in Eq.(2) is used to train the autoencoder. After the pretraining phase, we enter a fine-tuning phase where the model uses the overall loss defined in Eq.(13) to better align the features and prototypes across different views. The missing data is then padded via a hybrid prototype imputation strategy discussed in the next subsection. Finally, k-means is performed on the imputed complete data to obtain clustering results.

C EXPERIMENTAL CONFIGURATIONS

All experiments are conducted on an NVIDIA RTX L20 GPU with 48 GB memory using PyTorch 2.11.0+CUDA 13.0. For each dataset we set the missing rate as $[0.1, 0.3, 0.5, 0.7]$. The implementation of SPORT consists of two phases, namely the pretraining phase and fine-tuning phase with the learning rate set to 0.0005 and 0.00001 respectively. We set the pretraining epoch E_{PRE} to 200 and fine-tuning epoch E_{FIN} to 100. Our method adopts an autoencoder with the encoder dimensionality structured as Input dimension-512-512-1024-d, and the decoder dimensionality structured as d-1024-512-512-Input dimension where d is set to 256. Additionally, we set the the batch size to 512, and d_c to 128. For the loss function, α and β, the trade-off parameters of the overall loss are set to 1.6×10^{-5} and 5 respectively, while λ and η which balance the shared alignment loss and view-specific de-correlation loss are set to 10 and 5. Moreover, the temperature parameters $[\tau, \tau_w]$ are set to $[0.75, 0.5]$ and t the exponent of the shared-alignment loss is set to 3. Furthermore, γ is set to 0.3

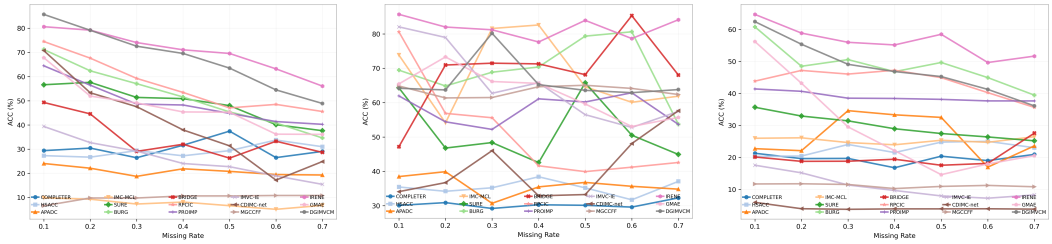
for our hybrid imputation strategy. During our experiment, we selectively fine-tune a subset of parameters certain datasets to accommodate varying dataset characteristics. Detailed implementation configurations are provided in the supplementary materials.

D STATISTICS OF 6 BENCHMARK DATASETS

Table 3: Statistics of 6 Benchmark Datasets.

Dataset	**Samples** (n)	**Views** (v)	**Clusters** (k)	**Dimensionality** (d_v)
Reuters_21578	1500	5	6	21531/24892/34251/15506/11547
100Leaves	1600	3	100	64/64/64
Digit4k	4000	2	4	30/30
Animal	10158	2	50	4096/4096
ALOI_100	10800	4	100	77/13/64/125
VGGFace2_50	34027	4	50	944/576/512/640

E VIZUALIZATION OF CLUSTERING PERFORMANCE COMPARISON

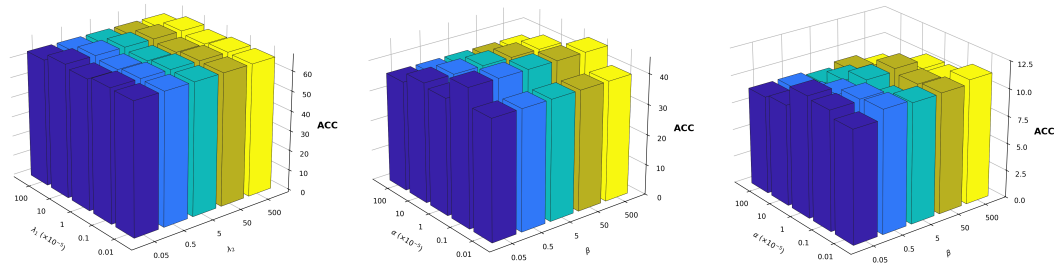


(a) 100Leaves (b) Digit4k (c) Animal

Figure 5: The clustering performance results with different missing rates on 100Leaves, Digit4k and Animal.

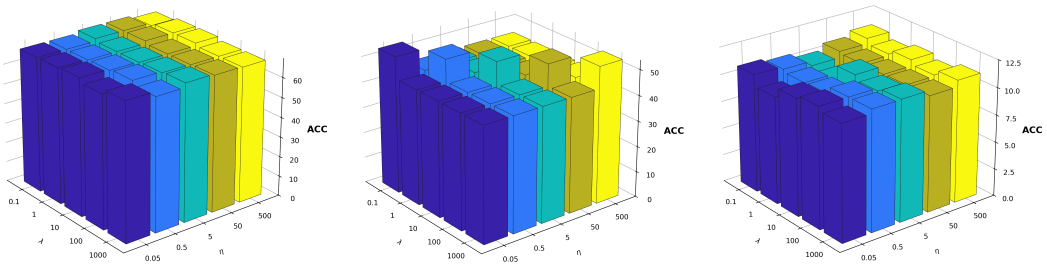
To further validate the superiority of SPORT, Figure.5 provides an intuitive visualization of the performance degradation trends for all models with increasing missing rates. As shown in Figure.5 (a), (b) and (c), when the missing rate increases, almost all methods suffer a performance decline, indicating that missing information impairs model performance. However, comparing to other methods, in most scenarios the ACC curves of SPORT are positioned at the top while remaining relative smooth and flat with increasing missing rates. This observation further demonstrates the effectiveness and robustness of the proposed method under different levels of missingness.

F SUPPLEMENTARY SENSITIVE ANALYSIS



(a) 100Leaves (b) Reuters_21578 (c) VGGFace_50

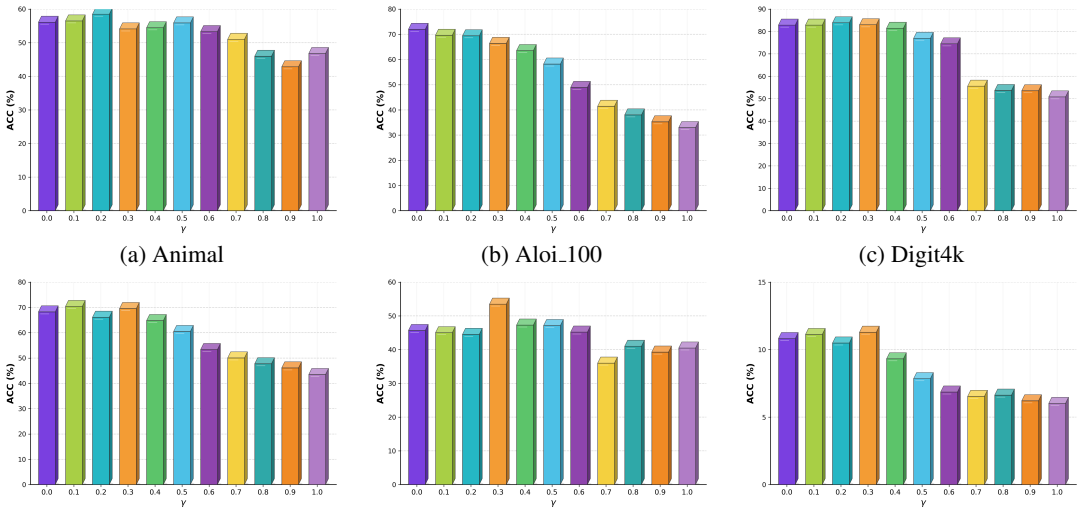
Figure 6: Sensitivity analysis of α and β at a missing rate of 0.5 on 100Leaves, Reuters_21578 and VGGFace_50.



(a) 100Leaves (b) Reuters_21578 (c) VGGFace_50

Figure 7: Sensitivity analysis of λ and η at a missing rate of 0.5 on 100Leaves, Reuters_21578 and VGGFace_50.

G Analysis of Hybrid Imputation



(a) Animal (b) Aloi_100 (c) Digit4k

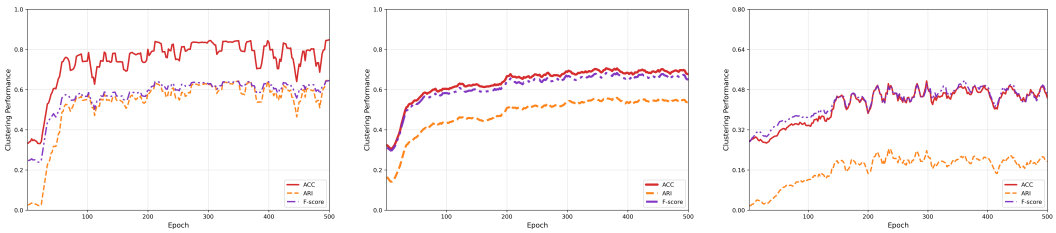
(d) 100Leaves (e) Reuters_21578 (f) VGGFace_50

Figure 8: Analysis of hybrid imputation under a missing rate of 0.5.

A key component of the proposed framework lies in its hybrid imputation strategy, which reconstructs missing data by jointly leveraging informative prototypes and neighboring samples, where γ controls their relative contributions. To better understand the behavior of this design, Fig. 8 reports the ACC of SPORT on six datasets with a missing rate of 0.5, where γ varies from 0 to 1 with a step size of 0.1. Here, $\gamma = 0$ and $\gamma = 1$ correspond to the neighbor-only and prototype-only imputation strategies, respectively. As shown, the hybrid strategy outperforms both degenerate variants across all datasets. For example, in Fig. 8 (e), on the Reuters_21578 dataset, SPORT achieves an ACC of 53.47% when $\gamma = 0.3$, while the neighbor-only and prototype-only variants obtain 45.67% and 40.47%, respectively. Similarly, in Fig. 8 (f), setting $\gamma = 0.3$ yields relative ACC gains of 4.33% over the neighbor-only strategy and 87.27% over the prototype-only strategy. It can therefore be concluded that the hybrid imputation strategy effectively balances prototype-level global structure and neighborhood-level local consistency. By integrating these two complementary sources of information, the model achieves more reliable reconstruction of missing views, which in turn leads to improved feature quality and more robust clustering performance across different datasets.

H Convergence Analysis

To further validate the convergence behavior of SPORT, we extend the training to 500 epochs on (a) Digit4k, (b) 100Leaves, and (c) Reuters_21578 under a missing rate of 0.5, with detailed results



(a) Digit4k (b) Leaves_100 (c) Reuters_21578

Figure 9: Convergence analysis of Digit4k, Leaves_100 and Reuters_21578.

presented in Fig. 9. Taking the 100Leaves dataset as an example, during the first 200 epochs, SPORT is pretrained to learn feature representations with high reconstruction fidelity. In this stage, the ACC smoothly increases from 33.06% to 65.31%, the ARI rises from 6.23% to 52.45%, and the F-Score improves from 28.45% to 64.37%. In the subsequent 100 epochs, where structure-aware relational contrastive learning and prototype partial alignment are further applied, SPORT continues to refine the learned representations and prototypes, leading to a improvement by 6.51%, 3.7%, and 4.75% in ACC, ARI, and F-Score, respectively. However, extending training to 300–500 epochs results in only minor fluctuations across all three metrics without consistent improvement. This behavior suggests that the proposed framework reaches stable convergence within a moderate number of epochs, and further training yields negligible improvements.