COMPLETE-SAMPLE-DRIVEN SHARED SEMANTIC PROTOTYPE LEARNING FOR HIGHLY INCOMPLETE MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Incomplete multi-view clustering (IMVC) aims to discover latent cluster structures from multi-source data with missing views. However, under high missing rates, the limited cross-view correspondences often lead to unstable representations, semantic misalignment across views, and biased cluster prototypes. To address these issues, we propose a complete-sample-driven framework with shared semantic prototypes, denoted as CSSP, for highly incomplete multi-view clustering. Specifically, a small set of complete samples is first used to pretrain view-specific autoencoders for stable representation initialization. During prototype training, artificial missing-view tasks are constructed by randomly masking complete samples, and batch-wise missing-view consistency regularization is imposed on latent representations, reconstructions, and prototype assignments, enhancing the model's adaptability to diverse missing patterns. Furthermore, shared semantic prototypes are constructed from complete samples to provide a reliable global semantic reference across views. Shared assignment consistency and prototype-level soft bridging are then introduced to align view-specific prototypes with the shared semantic space. Finally, the learned view prototypes are used to recover missing latent representations for clustering. The proposed method effectively exploits reliable cross-view semantics from a limited number of complete samples without relying on complex generative view recovery. Extensive experiments on multiple multi-view datasets under various high-missing-rate settings demonstrate competitive clustering performance and strong robustness, particularly in highly incomplete scenarios.

1 INTRODUCTION

Missing views are common in multi-view clustering (Wen et al., 2022; Lin et al., 2021; Ma et al., 2026), particularly in real-world medical applications, where different modalities are acquired through heterogeneous examination devices, clinical workflows, and information systems (Hou et al., 2023). Differences in acquisition cost, examination indications, and data quality further exacerbate the missing-view problem. For example, lung cancer (Captier et al., 2025) diagnosis and treatment may involve CT, PET-CT, pathological data, genomic testing, tumor markers, treatment records, and follow-up data, while individual patients rarely have all modalities. Patients undergoing early screening may have only CT scans, pathological data may be unavailable without biopsy or surgery, and genomic testing may be absent because of cost or sample-quality limitations. Consequently, multimodal lung cancer data often exhibit high missing rates and complex missing patterns, leaving only a limited number of samples observed across all views (see Figure A.1). This scarcity weakens reliable cross-view semantic correspondence and destabilizes shared representation learning and cluster prototype estimation (Dai et al., 2025; Li et al., 2023).

In high-missing-rate incomplete multi-view scenarios (Zhao et al., 2025), the scarcity of complete cross-view information is conceptually related to few-shot and semi-supervised learning. Few-shot classification (Zhang et al., 2025a) typically learns class decision boundaries from a small number of labeled support samples, whereas semi-supervised classification combines limited labeled samples with abundant unlabeled data (Li et al., 2025). However, both paradigms primarily target discrimination among known classes and are therefore not directly suited to discovering latent cluster structures

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.

without labels. Semi-supervised clustering (Chen et al., 2024) is more closely related to our setting, as it guides clustering using limited labels, pairwise constraints, or class priors. Nevertheless, under severe missingness, samples may differ substantially in missing patterns and numbers of observed views, such that sparse supervision may cover only local regions of the data distribution and bias the resulting cluster structures. Therefore, rather than relying on manual labels or pairwise supervision, we exploit a small number of complete samples as reliable cross-view semantic references while retaining a fully unsupervised clustering setting (Cui et al., 2025; Ma et al., 2025).

Existing studies have also investigated high-missing-rate scenarios. SCVT (Dong et al., 2025) constructs cross-view topology graphs using optimal transport distances and selects correlated views for missing-view imputation and contrastive alignment. DCG (Zhang et al., 2025b) employs a diffusion model to generate missing views in the latent space and combines contrastive learning with instance- and cluster-level interaction. PMIMC (Yuan et al., 2025) alleviates multi-view heterogeneity through relational consistency learning, reduces noise from pseudo-unpaired samples via robust prototype contrastive learning, and improves clustering stability through prototype completion. FreeCSL (Dai et al., 2025) adopts an imputation- and alignment-free strategy by learning consensus prototypes from available observations to construct a shared semantic space, together with view-specific cluster structure enhancement. Overall, existing methods mainly rely on missing-view imputation, generative recovery, or cross-view prototype matching. SCVT exploits cross-view topology to guide imputation, but its performance depends on the reliability of observable neighborhood structures under severe missingness. DCG improves representation learning through diffusion-based generation, at the cost of considerable computation induced by the multi-step diffusion process. PMIMC enhances robustness through prototype matching and completion, yet its prototype correspondences and pseudo-sample construction may remain susceptible to noise when view distributions differ substantially. FreeCSL avoids explicit view recovery, but its consensus semantic space is learned from available incomplete observations rather than being anchored by a small set of fully observed cross-view samples.

To address these limitations, our method neither relies on complex generative models for global missing-view recovery nor constructs shared prototypes from large numbers of incomplete samples. Instead, it exploits a small set of complete samples that retain reliable cross-view semantic correspondences. These samples provide shared semantic information and are randomly masked to construct artificial missing-view tasks, while their complete observations serve as stable semantic references for learning representations robust to diverse missing patterns.

Specifically, complete samples are first used to pretrain the autoencoders, providing stable initialization for subsequent learning. During prototype training, artificial incomplete inputs are constructed by randomly masking complete samples within each batch, and batch-wise missing-view consistency regularization jointly imposes consistency on latent representations, reconstructions, and prototype assignments, thereby improving adaptability to diverse missing-view patterns. Furthermore, shared semantic prototypes are constructed from complete-sample representations to provide a reliable global semantic reference. Based on this reference, shared assignment consistency and prototype-level soft bridging are jointly introduced to align view-specific prototype structures with the shared semantic space. The main contributions are summarized as follows:

- We propose a complete-sample-driven batch-wise missing-view consistency regularization strategy. During prototype training, artificial missing-view tasks are constructed by randomly masking complete samples, while consistency constraints on latent representations, reconstructions, and prototype assignments are jointly imposed to improve representation stability and semantic consistency under diverse missing patterns.
- We develop a complete-sample-supported shared semantic prototype mechanism. Shared semantic prototypes are constructed exclusively from samples observed across all views and serve as a reliable global semantic reference for constraining the semantic assignments and positions of view-specific prototypes, thereby reducing the influence of highly incomplete observations on prototype estimation.
- We introduce a shared-semantic-guided prototype-level soft alignment mechanism. Using the shared semantic prototypes as a global reference, shared assignment consistency and prototype-level soft bridging are jointly enforced to align the semantic assignments and positions of view-specific prototypes, thereby improving cross-view semantic consistency under high missing rates.

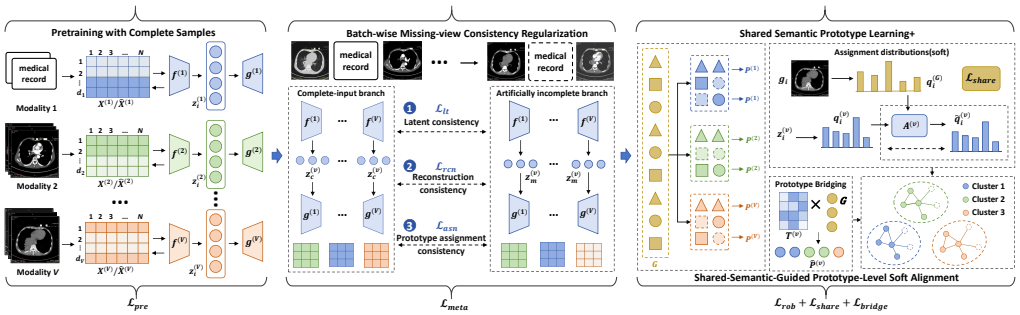

Figure 1: The figure illustrates the overall pipeline of the proposed method, which mainly consists of three stages. First, view-specific autoencoders are pretrained using complete samples to obtain stable basic latent representations. Second, during the prototype training stage, batch-wise missing-view consistency regularization is introduced by constructing artificial missing-view tasks, and cross-view semantic alignment is achieved by combining shared semantic prototype learning, shared assignment consistency constraint, and prototype-level soft bridging mechanism. Finally, the learned view-specific prototypes are utilized to recover missing latent representations, and the completed multi-view representations are clustered.

2 METHOD

Figure 1 outlines CSSP. Let the incomplete multi-view dataset be denoted as $\mathcal{X} = \{\mathbf{X}^{(v)}\}_{v=1}^{V}$, where V is the number of views, $\mathbf{X}^{(v)} \in \mathbb{R}^{N \times d_v}$ denotes the feature matrix of the v-th view, N is the number of samples, and d_v is the original feature dimension of the v-th view. To characterize view missingness, we introduce a missing-indicator matrix $\mathbf{M} \in \{0,1\}^{N \times V}$, where $M_i^{(v)} = 1$ indicates that the i-th sample is observed in the v-th view, while $M_i^{(v)} = 0$ indicates that the corresponding view is missing. Based on the missing-indicator matrix, the set of complete samples is defined as

$$\mathcal{C} = \left\{ i \mid M_i^{(v)} = 1,\ \forall v = 1, \ldots, V \right\}, \tag{1}$$

where $N_c = |\mathcal{C}|$ denotes the number of complete samples. During training, let $\mathcal{B} \subseteq \{1, \ldots, N\}$ denote the index set of samples in the current mini-batch, and let $N_B = |\mathcal{B}|$ denote its batch size. The complete samples contained in the current mini-batch are therefore indexed by $\mathcal{C} \cap \mathcal{B}$, whose number is denoted by

$$N_c^B = |\mathcal{C} \cap \mathcal{B}|. \tag{2}$$

2.1 BATCH-WISE MISSING-VIEW CONSISTENCY REGULARIZATION

For each prototype-training batch, complete samples are randomly masked to construct artificial missing-view tasks while retaining at least one observed view. The complete-sample branch serves as a stop-gradient reference, and the artificially incomplete branch is constrained by three consistency losses. Quantities from the artificially incomplete branch are denoted with a hat. First, latent representation and reconstruction consistencies are enforced as:

$$\mathcal{L}_{\text{lt}} = \frac{1}{VN_c^B} \sum_{i \in C \cap B} \sum_{v=1}^{V} \left\| \hat{\mathbf{z}}_i^{(v)} - \mathbf{z}_i^{(v)} \right\|_2^2, \qquad \mathcal{L}_{\text{rcn}} = \frac{1}{VN_c^B} \sum_{i \in C \cap B} \sum_{v=1}^{V} \left\| \hat{\mathbf{x}}_{i,\text{rec}}^{(v)} - \mathbf{x}_i^{(v)} \right\|_2^2, \tag{3}$$

where $z_i^{(v)}$ and $\hat{z}_i^{(v)}$ are latent representations from the complete and incomplete branches, and $\hat{x}_{i,\text{rec}}^{(v)}$ is the reconstructed view.

Second, prototype-assignment consistency is imposed to preserve semantic assignments under view missingness. Using the same current view-specific prototypes, soft prototype-assignment distributions are computed for both the complete and artificially incomplete representations. We denote these distributions by $\mathbf{q}_i^{(v)}$ and $\hat{\mathbf{q}}_i^{(v)}$, respectively, where $\mathbf{q}_i^{(v)}$ is defined in Eq. (11). Their consistency is enforced by:

$$\mathcal{L}_{\text{asn}} = \frac{1}{VN_c^B} \sum_{i \in \mathcal{C} \cap \mathcal{B}} \sum_{v=1}^{V} D_{\text{KL}} \left(\mathbf{q}_i^{(v)} \middle\| \hat{\mathbf{q}}_i^{(v)} \right), \tag{4}$$

where $\hat{q}_i^{(v)}$ denotes the prototype-assignment distribution from the artificially incomplete branch. The overall batch-wise missing-view consistency regularization is defined as

$$\mathcal{L}_{\text{mvc}} = \mathcal{L}_{\text{lt}} + \mathcal{L}_{\text{rcn}} + \mathcal{L}_{\text{asn}}. \tag{5}$$

Joint optimization over these artificial missing-view tasks promotes representation stability across diverse missing patterns and strengthens robustness under severe missingness.

2.2 Shared Semantic Prototype Learning

During prototype training, two complementary prototype structures are constructed from the latent representations. Shared semantic prototypes are built from complete samples to provide a global cross-view semantic reference, while view-specific prototypes capture the local latent structure of each view. For each complete sample $i \in \mathcal{C}$, its latent representations from all views are concatenated as

$$\mathbf{g}_i = \left[(\mathbf{z}_i^{(1)})^T, (\mathbf{z}_i^{(2)})^T, \ldots, (\mathbf{z}_i^{(V)})^T \right]^{\mathbf{T}}. \tag{6}$$

The resulting shared semantic prototype bank is written as

$$\mathbf{G} = \left[\mathbf{g}_1^T; \mathbf{g}_2^T; \ldots; \mathbf{g}_{N_c}^T \right] \in \mathbb{R}^{N_c \times d_s}, \qquad d_s = \sum_{v=1}^{V} d_z^{(v)}, \tag{7}$$

where $d_z^{(v)}$ denotes the latent dimension of the v-th view. Thus, $\mathbf{G}$ serves as a global semantic reference rather than batch-wise cluster centers.

View-specific sample-anchor prototypes $\mathbf{P}^{(v)}$ are constructed from visible latent representations in each mini-batch, with the prototype number adapted to the available observations. The prototypes are subsequently refined using assignment-based updates. Detailed prototype construction is provided in Appendix A.2. The view-specific prototype matrix is defined as

$$\mathbf{P}^{(v)} = \left[\mathbf{p}_1^{(v)}, \ldots, \mathbf{p}_{K_p^{(v)}}^{(v)} \right]^{\mathbf{T}}, \tag{8}$$

where $K_p^{(v)}$ denotes the number of prototypes. Based on these prototypes, we further introduce a noise-robust loss to enhance cross-view consistency. Since the view-specific sample anchors are constructed from aligned samples, anchors sharing the same sample index naturally preserve cross-view correspondence. For views u and v, let $K_p^{(u,v)} = \min\left(K_p^{(u)}, K_p^{(v)}\right)$. For $1 \le k \le K_p^{(u,v)}$, the prototype correspondence probability is computed as

$$Q_k^{(u,v)} = \frac{\exp\left(\text{sim}\left(\mathbf{p}_k^{(u)}, \mathbf{p}_k^{(v)}\right)/\tau\right)}{\sum_{j=1}^{K_p^{(v)}} \exp\left(\text{sim}\left(\mathbf{p}_k^{(u)}, \mathbf{p}_j^{(v)}\right)/\tau\right)}, \tag{9}$$

where τ is the temperature parameter. The noise-robust prototype loss is defined as

$$\mathcal{L}_{\text{rob}} = \sum_{u<v} \sum_{r=1}^{t-1} \frac{1}{r} \mathbb{E}_k \left[\left(1 - Q_k^{(u,v)}\right)^r \right], \tag{10}$$

where t is a positive integer controlling the truncation order of the robust loss, and $\mathbb{E}_k[\cdot]$ denotes the average over all valid prototype correspondences, i.e., $k = 1, \ldots, K_p^{(u,v)}$.

This loss constrains the local prototype structures across different views through prototype-level correspondences while reducing the influence of erroneous prototype matching during training.

2.3 SHARED-SEMANTIC-GUIDED PROTOTYPE-LEVEL SOFT ALIGNMENT

To reduce cross-view semantic inconsistency, shared semantic prototypes constructed from complete samples are used as a global reference to guide view-specific prototypes, as illustrated in Figure A.2. For each complete sample in the current batch, its shared representation is denoted by $\mathbf{z}_i^s = \mathbf{g}_i$.

The soft assignments of sample i to the shared semantic prototypes $\mathbf{G}$ and the view-specific prototypes $\mathbf{P}^{(v)}$ are defined as

$$\mathbf{q}_i^{(G)} = \mathrm{softmax}\left(\frac{\mathrm{sim}(\mathbf{z}_i^s, \mathbf{G})}{\tau}\right), \qquad \mathbf{q}_i^{(v)} = \mathrm{softmax}\left(\frac{\mathrm{sim}(\mathbf{z}_i^{(v)}, \mathbf{P}^{(v)})}{\tau}\right), \tag{11}$$

where $\mathbf{q}_i^{(G)} \in \mathbb{R}^{1\times N_c}$, $\mathbf{q}_i^{(v)} \in \mathbb{R}^{1\times K_p^{(v)}}$.

Since $\mathbf{q}_i^{(G)}$ and $\mathbf{q}_i^{(v)}$ are defined over different prototype spaces and generally have different dimensions, they cannot be directly compared. We therefore extract the component of the shared semantic prototype bank corresponding to the v-th view:

$$\mathbf{G}^{(v)} = \left[(\mathbf{z}_{i_1}^{(v)})^T; \ldots; (\mathbf{z}_{i_{N_c}}^{(v)})^T\right] \in \mathbb{R}^{N_c\times d_z^{(v)}}. \tag{12}$$

A soft mapping from the shared semantic prototype space to the v-th view-specific prototype space is then constructed as

$$\mathbf{A}^{(v)} = \mathrm{softmax}_{\mathrm{row}}\left(\frac{\mathrm{sim}(\mathbf{G}^{(v)}, \mathbf{P}^{(v)})}{\tau}\right) \in \mathbb{R}^{N_c\times K_p^{(v)}}. \tag{13}$$

The shared semantic assignment is projected into the v-th view-specific prototype space:

$$\tilde{\mathbf{q}}_i^{(v)} = \mathbf{q}_i^{(G)}\mathbf{A}^{(v)}, \tag{14}$$

where $\tilde{\mathbf{q}}_i^{(v)} \in \mathbb{R}^{1\times K_p^{(v)}}$. Using the projected shared semantic assignment as the semantic reference, assignment consistency is enforced by

$$\mathcal{L}_{\mathrm{share}} = \frac{1}{VN_c^B}\sum_{i\in\mathcal{C}\cap\mathcal{B}}\sum_{v=1}^{V} D_{\mathrm{KL}}\left(\tilde{\mathbf{q}}_i^{(v)} \middle\| \mathbf{q}_i^{(v)}\right). \tag{15}$$

This constraint promotes cross-view semantic consistency through global semantic guidance. To further reduce the structural discrepancy between view-specific and shared semantic prototypes, we introduce a prototype-level soft bridging mechanism. For the complete samples in the current batch, let $\mathbf{Q}^{(v)} \in \mathbb{R}^{N_c^B\times K_p^{(v)}}$ and $\mathbf{Q}^{(G)} \in \mathbb{R}^{N_c^B\times N_c}$ denote their assignment matrices to the view-specific and shared prototypes, respectively. Their semantic co-occurrence is then used to construct a soft bridging matrix from the view-specific prototypes to the shared semantic prototypes:

$$\mathbf{T}^{(v)} = \mathrm{RowNorm}\left((\mathbf{Q}^{(v)})^T\mathbf{Q}^{(G)}\right) \in \mathbb{R}^{K_p^{(v)}\times N_c}, \tag{16}$$

where $\mathrm{RowNorm}(\cdot)$ denotes row-wise normalization. The matrix $\mathbf{T}^{(v)}$ characterizes soft correspondences between view-specific and shared semantic prototypes based on their assignment co-occurrence, avoiding restrictive one-to-one hard matching. The prototype bridging loss is then defined as

$$\mathcal{L}_{\mathrm{bridge}} = \frac{1}{V}\sum_{v=1}^{V}\left\|\mathbf{P}^{(v)} - \mathbf{T}^{(v)}\mathbf{G}^{(v)}\right\|_F^2. \tag{17}$$

2.4 OVERALL OPTIMIZATION OBJECTIVE

The model is trained in two stages: pretraining and prototype training. During pretraining, only complete samples are used to optimize the view-specific autoencoders with the reconstruction loss $\mathcal{L}_{\mathrm{pre}}$. During prototype training, prototype semantic learning and batch-wise missing-view consistency regularization are jointly optimized:

$$\mathcal{L} = \mathcal{L}_{\mathrm{pro}} + \lambda_{\mathrm{mvc}}\mathcal{L}_{\mathrm{mvc}}. \tag{18}$$

The prototype semantic learning loss is defined as

$$\mathcal{L}_{\text{pro}} = \lambda_{\text{rob}}\mathcal{L}_{\text{rob}} + \lambda_{\text{share}}\mathcal{L}_{\text{share}} + \beta\mathcal{L}_{\text{bridge}}. \tag{19}$$

Here, λ_{rob}, λ_{share}, β, and λ_{mvc} control the contributions of the corresponding loss terms. Joint optimization strengthens cross-view semantic consistency while enhancing adaptability to diverse missing patterns.

After training, missing-view representations are recovered by identifying the nearest prototype from an observed view and using the corresponding prototype for the missing view. The completed latent representations from all views are then fused for clustering.

3 EXPERIMENT

In this section, we evaluate CSSP on six widely used incomplete multi-view clustering benchmark datasets and compare it with representative state-of-the-art methods, including DSIMVC (Tang & Liu, 2022), DIMVC (Xu et al., 2022), SURE (Yang et al., 2023), DCP (Lin et al., 2023), APADC (Xu et al., 2023), IMVC-IE (Huang et al., 2024), ICMVC (Chao et al., 2024), PMIMC (Yuan et al., 2025), DCG (Zhang et al., 2025b), BURG (Jin et al., 2025), and GMAE (Zou et al., 2026). All methods are evaluated under the same missing-rate settings, and clustering performance is measured by Accuracy (ACC), Normalized Mutual Information (NMI), and F-measure (F1).

Table 1: Clustering performance on ALOI, Caltech101-7, and HandWritten under four missing rates.

Dataset	Method	0.7			0.8			0.9			0.95		
		ACC	NMI	F1	ACC	NMI	F1	ACC	NMI	F1	ACC	NMI	F1
ALOI	DSIMVC(ICML'22)	57.02	66.60	49.68	56.95	64.28	46.46	55.48	63.13	45.57	53.82	60.55	43.21
	DIMVC(AAAI'22)	54.32	67.78	51.71	30.84	46.93	27.59	34.36	49.66	35.72	24.43	38.78	15.50
	SURE(TPAMI'23)	36.19	64.92	22.06	38.57	63.49	28.75	35.89	57.93	31.20	34.16	56.55	33.46
	DCP(PAMI'23)	39.24	68.32	36.84	35.28	62.98	34.34	29.20	66.01	37.27	27.43	53.55	27.87
	APADC(TIP'23)	23.39	50.79	23.62	24.98	53.13	20.18	20.33	48.39	21.09	16.69	37.63	17.55
	IMVC-IE(ICASSP'24)	27.12	53.83	28.95	26.46	48.08	26.41	25.75	46.32	26.99	23.82	43.91	24.86
	ICMVC(AAAI'24)	58.46	74.26	57.90	38.20	58.06	36.70	22.50	50.42	22.72	4.45	12.65	1.23
	PMIMC(TIP'25)	66.28	79.99	65.09	**70.79**	**83.43**	**69.29**	**70.31**	**83.35**	**72.77**	**70.14**	**83.26**	**68.88**
	DCG(AAAI'25)	22.26	54.76	18.54	17.16	49.78	13.31	14.16	37.66	10.83	4.66	25.81	1.58
	BURG(ICCV'25)	41.70	65.52	42.84	43.60	66.20	44.18	37.10	62.06	38.29	33.70	59.39	35.71
	GMAE(TPAMI'26)	20.65	50.39	1.32	19.85	49.53	0.65	22.90	52.32	2.48	24.25	53.59	1.95
	CSSP	**67.35**	**81.11**	**66.26**	69.24	81.80	68.28	67.16	81.21	65.84	67.89	81.38	66.73
Caltech101-7	DSIMVC(ICML'22)	54.79	42.62	54.71	52.43	43.32	52.16	44.64	33.52	44.09	39.81	28.10	37.88
	DIMVC(AAAI'22)	60.43	41.47	46.34	61.05	44.16	53.44	60.64	56.20	53.18	52.76	36.66	48.14
	SURE(TPAMI'23)	75.64	65.67	75.11	54.52	40.15	55.03	48.29	38.11	48.07	37.81	22.90	37.73
	DCP(PAMI'23)	74.00	74.45	71.62	72.43	56.05	73.64	42.50	37.06	39.52	36.93	33.32	32.94
	APADC(TIP'23)	59.14	56.00	53.44	50.83	37.67	48.89	43.66	40.40	37.38	44.50	29.09	37.27
	IMVC-IE(ICASSP'24)	51.42	34.16	51.92	47.14	31.52	47.75	43.92	31.06	44.53	43.81	30.85	44.20
	ICMVC(AAAI'24)	72.07	65.64	71.95	66.14	60.04	66.13	62.07	58.12	62.16	56.97	47.15	57.18
	PMIMC(TIP'25)	83.30	**76.52**	81.45	84.07	**76.72**	85.00	84.47	**75.40**	83.18	80.70	73.69	78.75
	DCG(AAAI'25)	66.29	59.38	66.25	60.79	53.08	59.41	68.36	53.34	68.68	60.86	44.97	60.32
	BURG(ICCV'25)	64.43	48.40	64.52	63.00	47.01	61.63	62.35	53.91	60.87	58.57	39.25	57.88
	GMAE(TPAMI'26)	73.00	63.08	43.26	76.57	65.99	43.42	75.14	65.91	1.74	74.86	63.20	8.41
	CSSP	**85.55**	75.06	**85.20**	**86.78**	76.47	**86.43**	**85.97**	75.13	**85.64**	**84.12**	**75.67**	**82.78**
HandWritten	DSIMVC(ICML'22)	69.06	64.57	68.95	66.48	61.12	69.37	45.46	42.85	44.38	45.40	45.25	43.11
	DIMVC(AAAI'22)	**84.50**	**77.16**	38.55	68.28	61.24	66.44	30.65	22.07	23.10	58.45	48.91	55.30
	SURE(TPAMI'23)	69.85	59.95	68.36	61.23	58.30	60.38	51.80	44.00	50.62	47.37	43.39	46.70
	DCP(PAMI'23)	58.43	63.44	32.10	31.75	31.00	28.85	33.14	31.08	8.25	26.05	17.84	23.22
	APADC(TIP'23)	67.03	65.26	61.90	64.60	64.04	62.49	65.61	64.05	64.58	66.90	62.72	66.57
	IMVC-IE(ICASSP'24)	69.45	53.00	69.34	64.13	52.23	63.35	51.45	42.32	52.45	45.02	33.68	41.72
	ICMVC(AAAI'24)	74.05	70.37	75.15	56.77	59.97	56.88	47.65	45.21	48.10	43.22	48.11	43.20
	PMIMC(TIP'25)	75.76	72.87	75.19	74.44	73.29	73.67	74.86	72.26	**76.43**	76.68	**73.57**	75.70
	DCG(AAAI'25)	63.90	67.31	62.27	73.90	71.99	72.54	64.70	57.79	64.81	55.40	48.42	54.88
	BURG(ICCV'25)	58.65	63.00	55.35	70.80	67.06	70.12	65.45	61.55	63.43	63.65	63.66	60.08
	GMAE(TPAMI'26)	83.10	76.48	1.71	**88.10**	**81.07**	1.30	**77.25**	**75.47**	0.97	70.25	69.85	19.04
	CSSP	77.33	73.18	**76.69**	76.76	72.33	**75.78**	76.55	74.38	75.64	**77.15**	68.43	**77.03**

3.1 DATASETS AND EXPERIMENTAL SETTINGS

To evaluate the effectiveness of the proposed method for incomplete multi-view clustering, experiments are conducted on six benchmark datasets, including ALOI, Caltech101-7, HandWritten, Scene15, 100leaves, and MSRCV1. ALOI contains 10,800 object images from 100 categories represented by two visual feature views. Caltech101-7 consists of 1,400 images from seven categories with five views. HandWritten contains 2,000 handwritten digit samples from 10 classes

represented by six feature views. Scene15 includes 4,485 scene images from 15 categories with three views, while 100leaves contains 1,600 leaf images from 100 categories represented by three views. MSRCV1 consists of 210 image samples from seven categories with five visual feature views.

Table 2: Clustering performance on Scene15, 100leaves, and MSRCV1 under four missing rates.

Dataset	Method	0.7			0.8			0.9			0.95		
		ACC	NMI	F1	ACC	NMI	F1	ACC	NMI	F1	ACC	NMI	F1
Scene15	DSIMVC(ICML'22)	29.11	34.53	29.91	24.40	22.00	21.46	27.97	32.99	29.14	16.90	12.00	10.24
	DIMVC(AAAI'22)	29.48	26.52	21.82	29.64	24.23	28.37	24.91	21.09	25.95	27.42	19.19	25.60
	SURE(TPAMI'23)	38.64	41.11	36.38	37.43	40.06	34.51	32.02	32.63	29.65	26.18	26.02	25.62
	DCP(PAMI'23)	37.26	40.35	34.78	35.14	38.64	33.65	32.54	37.14	28.92	28.15	29.75	28.07
	APADC(TIP'23)	38.92	41.84	38.50	34.19	36.86	30.78	34.73	37.62	30.38	32.55	36.42	29.84
	IMVC-IE(ICASSP'24)	30.24	26.89	33.37	26.18	29.45	24.88	28.82	27.98	29.14	18.43	15.73	13.26
	ICMVC(AAAI'24)	29.39	24.18	28.87	28.16	31.85	27.27	22.16	19.74	21.78	20.42	21.74	17.73
	PMIMC(TIP'25)	32.77	33.46	33.24	35.35	36.32	35.87	36.21	37.53	**40.81**	37.61	39.96	37.58
	DCG(AAAI'25)	37.24	40.38	36.73	38.48	38.50	38.11	34.72	32.19	34.08	29.52	30.18	26.28
	BURG(ICCV'25)	30.60	28.26	29.88	32.35	27.31	31.51	30.95	28.97	31.64	30.35	27.34	29.82
	GMAE(TPAMI'26)	25.85	26.63	12.62	22.35	21.01	12.86	20.45	18.28	8.92	21.80	18.55	11.06
	CSSP	**42.46**	**43.67**	**42.33**	**42.59**	**40.83**	**42.83**	**41.17**	**42.74**	40.54	**41.13**	**41.53**	**40.44**
100leaves	DSIMVC(ICML'22)	12.35	41.05	3.82	14.37	44.57	3.85	8.25	31.09	0.21	9.81	35.38	0.31
	DIMVC(AAAI'22)	47.31	67.34	45.31	40.33	60.65	38.19	34.73	57.06	32.15	0.00	0.00	0.00
	SURE(TPAMI'23)	26.19	60.28	24.77	23.92	56.86	22.25	22.06	53.49	21.54	21.65	52.82	20.26
	DCP(PAMI'23)	40.94	65.30	40.36	33.25	60.84	32.72	26.35	55.74	26.45	19.79	50.96	18.76
	APADC(TIP'23)	54.96	75.33	53.88	48.48	71.80	47.05	37.73	66.52	36.34	31.42	62.02	29.26
	IMVC-IE(ICASSP'24)	9.71	34.71	2.93	9.81	34.62	5.98	10.50	35.78	3.78	8.40	30.00	2.99
	ICMVC(AAAI'24)	33.94	65.06	30.10	26.58	58.93	23.05	9.77	40.16	1.32	5.19	25.35	0.80
	PMIMC(TIP'25)	**81.27**	91.59	**79.44**	**78.38**	90.31	**76.66**	75.91	89.00	**78.69**	76.10	89.38	74.01
	DCG(AAAI'25)	16.06	49.14	10.43	9.19	38.93	5.05	7.38	33.45	3.18	9.12	35.08	6.50
	BURG(ICCV'25)	68.00	82.01	68.08	65.06	80.58	64.63	63.38	79.39	63.78	60.81	77.63	61.59
	GMAE(TPAMI'26)	28.38	59.58	2.23	22.94	56.48	2.45	22.88	55.00	2.44	22.62	54.86	3.16
	CSSP	80.04	**91.78**	78.14	77.59	**90.74**	75.63	**77.69**	**90.32**	75.63	**76.24**	**90.00**	**74.15**
MSRCV1	DSIMVC(ICML'22)	29.05	15.31	23.12	29.21	15.16	18.54	28.10	15.35	25.24	28.89	16.81	23.54
	DIMVC(AAAI'22)	40.00	27.93	36.02	43.81	30.89	40.81	44.92	32.43	44.11	35.08	24.07	33.42
	SURE(TPAMI'23)	40.79	26.89	41.31	40.00	27.50	40.09	39.68	34.11	39.97	42.06	37.50	41.98
	DCP(PAMI'23)	18.73	8.89	11.61	20.95	11.51	14.18	25.87	12.43	21.17	27.94	16.87	25.48
	APADC(TIP'23)	32.54	27.57	30.64	28.73	24.07	26.58	39.37	36.10	38.58	41.59	33.13	39.53
	IMVC-IE(ICASSP'24)	32.54	22.35	29.30	31.11	19.86	27.80	30.32	18.60	26.07	21.43	5.07	18.93
	ICMVC(AAAI'24)	54.60	45.12	54.30	54.76	46.39	54.39	43.02	37.43	43.31	45.87	38.28	45.76
	PMIMC(TIP'25)	58.00	49.30	56.87	61.31	54.07	59.86	59.43	50.15	61.90	60.76	50.45	60.37
	DCG(AAAI'25)	54.29	41.20	53.56	48.57	42.69	47.23	45.71	32.66	44.88	48.57	35.77	48.31
	BURG(ICCV'25)	64.29	57.04	**64.83**	63.81	54.04	**63.68**	57.14	44.51	57.39	55.71	44.45	55.32
	GMAE(TPAMI'26)	**70.00**	**57.57**	49.61	**65.71**	**57.58**	38.06	56.19	47.78	20.63	60.48	50.47	22.57
	CSSP	62.86	57.34	60.55	62.52	54.02	60.06	**63.81**	**53.37**	**62.33**	**64.05**	**54.84**	**62.54**

To simulate high-missing-rate incomplete multi-view scenarios, we randomly remove observations from each view and set the missing rates to 0.7, 0.8, 0.9, and 0.95. The first three settings characterize clustering performance as missingness increases, while the 0.95 setting provides an extreme case with only limited cross-view observations remaining. Hyperparameters λ_{share}, λ_{rob}, and β are selected by grid search, and all experiments are repeated 10 times with different random seeds, with the average results reported. Detailed hyperparameter settings are provided in Appendix A.1, and the prototype initialization analysis is presented in Appendix A.3 (see Figure A.5).

3.2 Clustering Results and Analysis

Tables 1 and 2 compare CSSP with state-of-the-art incomplete multi-view clustering methods at different missing rates on six benchmark datasets. The best results are shown in bold, and the second-best results are underlined. ACC, NMI, and F1 are reported for all settings.

1) As shown in Tables 1 and 2, CSSP achieves the best or competitive results under most settings across ALOI, Caltech101-7, HandWritten, Scene15, 100leaves, and MSRCV1. While the performance of many existing methods decreases noticeably as the missing rate increases, CSSP remains comparatively stable, as shown in Figure A.3. This behavior benefits from the complete-sample-driven shared semantic modeling strategy, where the shared semantic prototype bank exploits reliable cross-view correspondences retained in complete samples and prototype-level soft bridging aligns view-specific prototypes with the shared semantic space, improving the use of limited cross-view information under high missing rates.

2) The robustness of CSSP becomes more evident as the missing rate approaches 0.9–0.95, where observable cross-view information is substantially reduced. Methods that rely strongly on direct

inter-view relationships are more susceptible to representation shifts when only limited paired information remains. In comparison, CSSP combines semantic references derived from complete samples with batch-wise missing-view consistency regularization over simulated missing-view patterns, which helps maintain stable representations under highly incomplete settings. For example, on Caltech101-7, the ACC of CSSP decreases only from 85.55% at a missing rate of 0.7 to 84.12% at 0.95, showing relatively small performance variation despite the substantial increase in missingness.

3) The results also reflect the role of shared semantic alignment in reducing cross-view inconsistency. Instead of relying on large-scale missing-view recovery or direct hard prototype matching, CSSP uses complete samples to establish a reliable global semantic reference and constructs soft associations between view-specific prototypes and the shared semantic space. Such a design reduces the dependence on potentially unstable correspondences inferred from highly incomplete observations. The stable performance observed at a 95% missing rate is consistent with the effectiveness of the proposed shared-semantic alignment strategy.

4) On ALOI and 100leaves, which contain relatively large numbers of categories, the advantage of CSSP becomes narrower in several settings. One reason is that when the effective prototype number is limited relative to the underlying semantic complexity, different latent structures may not be represented by sufficiently distinct prototypes, particularly under extreme missingness. This result suggests that the relationship between prototype capacity and latent structural complexity remains an important factor affecting clustering performance.

5) A similar trend can be observed on MSRCV1. At missing rates of 0.7 and 0.8, several generation- or graph-based methods remain competitive because sufficient observable cross-view information is still available for learning effective representations. When the missing rate increases to 0.9 and 0.95, however, reduced cross-view evidence introduces greater matching noise and representation shifts. Under these conditions, the complete-sample semantic references and simulated missing-view tasks used by CSSP help preserve representation stability, making its relative advantage more evident in highly incomplete scenarios.

3.3 Ablation Experiments

Table 3: Ablation study under different missing rates on Caltech101-7, Scene15, and 100leaves. The best results are shown in bold and the second-best results are underlined.

Dataset	Variant	0.70			0.80			0.90			0.95		
		ACC	NMI	F1	ACC	NMI	F1	ACC	NMI	F1	ACC	NMI	F1
Caltech101-7	full	**85.55**	75.06	**85.20**	**86.78**	**76.47**	**86.43**	**85.97**	75.13	**85.64**	**84.12**	**75.67**	**82.78**
	w/o $\mathcal{L}_{mvc}$	80.99	73.87	79.42	84.43	75.12	83.58	82.51	74.80	81.18	80.31	70.32	79.74
	w/o $\mathcal{L}_{share}$	83.68	**76.00**	83.14	81.52	75.36	79.31	85.20	**76.05**	84.25	82.54	74.00	81.55
	w/o $\mathcal{L}_{bridge}$	81.94	72.90	81.21	79.20	71.88	77.55	85.93	74.92	85.42	83.16	73.78	81.98
Scene15	full	**42.46**	43.67	**42.33**	42.59	40.83	**42.83**	**41.17**	**42.74**	**40.54**	41.13	41.53	40.44
	w/o $\mathcal{L}_{mvc}$	41.57	42.81	40.10	41.78	**44.03**	41.14	40.08	42.22	39.87	**42.55**	42.82	**41.90**
	w/o $\mathcal{L}_{share}$	42.05	**44.04**	41.42	**42.98**	43.39	42.73	40.78	42.55	40.00	41.89	**43.24**	41.29
	w/o $\mathcal{L}_{bridge}$	34.72	38.49	34.65	37.64	38.67	37.70	37.42	39.23	37.88	36.85	39.59	36.76
100leaves	full	**80.04**	**91.78**	**78.14**	77.59	90.74	75.63	77.69	90.32	75.63	76.24	90.00	74.15
	w/o $\mathcal{L}_{mvc}$	79.36	91.14	77.43	78.42	90.69	75.63	76.11	89.98	73.69	76.10	90.05	74.01
	w/o $\mathcal{L}_{share}$	78.91	91.41	76.85	**78.50**	**91.95**	**76.57**	**79.98**	**91.10**	**78.31**	**78.13**	**90.76**	**76.17**
	w/o $\mathcal{L}_{bridge}$	75.09	88.90	73.03	74.69	88.26	72.67	72.81	87.76	70.81	72.19	86.98	70.01

Table 3 presents the ablation results of different loss components in CSSP. The Full model denotes the complete framework, while w/o $\mathcal{L}_{mvc}$, w/o $\mathcal{L}_{share}$, and w/o $\mathcal{L}_{bridge}$ represent the variants without batch-wise missing-view consistency regularization, shared distribution consistency constraint, and prototype bridging loss, respectively. It can be observed that the complete model achieves the best or near-best clustering performance across three datasets and different missing rate settings, validating the effectiveness of the proposed components.

First, removing $\mathcal{L}_{mvc}$ leads to different degrees of performance degradation in most experimental settings. For example, on the Caltech101-7 dataset with a missing rate of 0.7, the ACC, NMI, and F1 of w/o $\mathcal{L}_{mvc}$ decrease to 80.99, 73.87, and 79.42, respectively. This result indicates that the batch-wise missing-view consistency regularization enables the model to learn stable representations

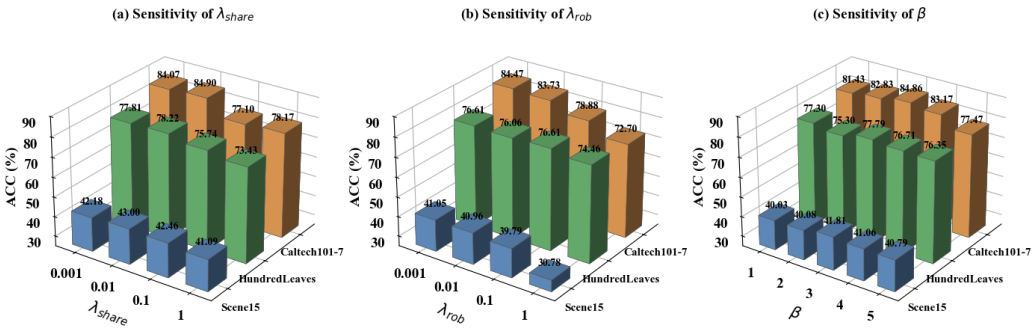

Figure 2: Sensitivity analysis of λ_{share}, λ_{rob}, and β on three benchmark datasets.

under different missing patterns by constructing artificial missing-view tasks, thereby improving its adaptability under high-missing-rate scenarios.

Second, removing $\mathcal{L}_{share}$ leads to performance degradation in a majority of the evaluated settings, indicating that the shared distribution consistency constraint contributes to maintaining cross-view semantic consistency. Since different views exhibit different data distributions, learning view-specific prototypes alone is insufficient to guarantee semantic correspondence across views. In contrast, $\mathcal{L}_{share}$ utilizes the global reference provided by shared semantic prototypes to further constrain the semantic assignments of different views.

Finally, removing $\mathcal{L}_{bridge}$ results in the most significant performance degradation, as shown in Figure A.4. For example, on the Scene15 dataset with a missing rate of 0.7, the ACC and F1 of w/o $\mathcal{L}_{bridge}$ decrease from 42.46 to 34.72 and from 42.33 to 34.65, respectively. Similar degradation is also observed on the 100leaves dataset under high missing rate conditions. This demonstrates that the prototype bridging mechanism is an important component for connecting view-specific prototypes with the shared semantic space. Without this constraint, the interaction among view-specific prototypes is weakened, resulting in reduced cross-view semantic consistency.

3.4 Parameter Sensitivity Analysis

Figure 2 illustrates the effects of λ_{share}, λ_{rob}, and β on the clustering performance. Overall, the three parameters exhibit relatively consistent trends across different datasets. For λ_{share}, the model achieves the best or near-best performance around 0.01, while excessively large values lead to noticeable degradation, indicating that overly strong shared semantic constraints may disturb the balance among different learning objectives. For λ_{rob}, smaller values generally yield more stable performance, with 0.001 providing favorable results. When λ_{rob} increases to 1, the performance drops considerably, suggesting that an overly strong prototype robustness constraint may limit the representation capability of the prototypes.

In comparison, the model is less sensitive to β and achieves the best overall performance at $\beta = 3$. A small β may provide insufficient prototype bridging, whereas an excessively large value may weaken view-specific structures. Based on these observations, we set

$$\lambda_{share} = 0.01, \quad \lambda_{rob} = 0.001, \quad \beta = 3.$$

4 Conclusion

In this paper, we proposed CSSP for highly incomplete multi-view clustering. By exploiting a small number of complete samples, CSSP introduces batch-wise missing-view consistency regularization to improve adaptability to diverse missing patterns and constructs shared semantic prototypes to provide a reliable global semantic reference across views. Prototype-level soft alignment further strengthens the consistency between view-specific prototypes and the shared semantic space. Extensive experiments on six benchmark datasets demonstrate the effectiveness and robustness of CSSP, particularly under severe missingness.

AI USE STATEMENT

In this work, we used generative AI tools for language polishing and for assisting with the drafting of parts of the implementation code. All AI-assisted text was manually reviewed and revised by the authors, and all AI-assisted code was checked, tested, and verified for correctness. Generative AI tools were not used to generate synthetic datasets or prove mathematical claims. We take responsibility for the final content of this work, including text, claims, experimental results, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The model architecture, training procedure, and experimental settings are described in the main paper, with additional implementation details provided in Appendix A.2 and hyperparameter settings in Appendix A.1. Anonymous source code will be provided as supplementary material.

REFERENCES

Nicolas Captier, Marvin Lerousseau, Fanny Orlhac, Narinée Hovhannisyan-Baghdasarian, Marie Luporsi, Erwin Woff, Sarah Lagha, Paulette Salamoun Feghali, Christine Lonjou, Clément Beaulaton, et al. Integration of clinical, pathological, radiological, and transcriptomic data improves prediction for first-line immunotherapy outcome in metastatic non-small cell lung cancer. *Nature communications*, 16(1):614, 2025.

Guoqing Chao, Yi Jiang, and Dianhui Chu. Incomplete contrastive multi-view clustering with high-confidence guiding. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(10):11221–11229, 2024.

Jingwei Chen, Shiyu Xie, Hui Yang, and Feiping Nie. Effective semi-supervised graph clustering with pairwise constraints. *Information Sciences*, 681:121249, 2024. ISSN 0020-0255.

Jinrong Cui, Bang Liufu, Chongjie Dong, Jingcheng Ke, and Jie Wen. Deep multi-view clustering with meta information compression. *IEEE Transactions on Image Processing*, 34:7514–7527, 2025.

Yuzhuo Dai, Jiaqi Jin, Zhibin Dong, Siwei Wang, Xinwang Liu, En Zhu, Xihong Yang, Xinbiao Gan, and Yu Feng. Imputation-free and alignment-free: Incomplete multi-view clustering driven by consensus semantic learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5071–5081, June 2025.

Zhibin Dong, Dayu Hu, Jiaqi Jin, Siwei Wang, Xinwang Liu, and En Zhu. Selective cross-view topology for deep incomplete multi-view clustering. *IEEE Transactions on Image Processing*, 34:4792–4805, 2025.

Wentai Hou, Chengxuan Lin, Lequan Yu, Jing Qin, Rongshan Yu, and Liansheng Wang. Hybrid graph convolutional network with online masked autoencoder for robust multimodal cancer survival prediction. *IEEE Transactions on Medical Imaging*, 42(8):2462–2473, 2023.

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.

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 *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1016–1026. IEEE, 2025.

Guojie Li, Zhiwen Yu, Kaixiang Yang, C. L. Philip Chen, and Xuelong Li. Ensemble-enhanced semi-supervised learning with optimized graph construction for high-dimensional data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(2):1103–1119, 2025.

Haobin Li, Yunfan Li, Mouxing Yang, Peng Hu, Dezhong Peng, and Xi Peng. Incomplete multi-view clustering via prototype-based imputation. In Edith Elkind (ed.), *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, pp. 3911–3919. International Joint Conferences on Artificial Intelligence Organization, 8 2023. Main Track.

Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. Completer: Incomplete multi-view clustering via contrastive prediction. In *2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 11169–11178. IEEE, 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, 2023.

Shubin Ma, Liang Zhao, Mingdong Lu, Yifan Guo, and Bo Xu. Consistency-aware padding for incomplete multi-modal alignment clustering based on self-repellent greedy anchor search. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 5923–5931, 2025.

Shubin Ma, Zhikui Chen, Lin Yuanbo Wu, and Liang Zhao. Multi-view aligned clustering via sample-bundled optimization: Anchor graph enhancement and contrastive propagation. *IEEE Transactions on Multimedia*, pp. 1–12, 2026.

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

Jie Wen, Zheng Zhang, Lunke Fei, Bob Zhang, Yong Xu, Zhao Zhang, and Jinxing Li. A survey on incomplete multiview clustering. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(2):1136–1149, 2022.

Jie Xu, Chao Li, Yazhou Ren, Liang Peng, Yujie Mo, Xiaoshuang Shi, and Xiaofeng Zhu. Deep incomplete multi-view clustering via mining cluster complementarity. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(8):8761–8769, 2022.

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.

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, 2023.

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.

Ji Zhang, Jingkuan Song, Lianli Gao, Nicu Sebe, and Heng Tao Shen. Reliable few-shot learning under dual noises. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(10): 9005–9022, 2025a.

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. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(21):22650–22658, 2025b.

Liang Zhao, Shubin Ma, Bo Xu, and Qingchen Zhang. Dual-learning based penalized multi-align clustering for multi-view incomplete and disorderly data. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 1318–1326, 2025.

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*, 48(9):10489–10503, 2026.

A APPENDIX

A.1 EXPERIMENTAL SETTINGS

Missing-rate definition. The missing rate in our experiments denotes the proportion of incomplete samples, rather than the proportion of missing entries in the sample–view indicator matrix or an independent missing probability for each view. Every sample retains at least one observed view, i.e., $\sum_{v=1}^{V} M_i^{(v)} \geq 1$ for all i. An incomplete sample has at least one observed view but fewer than V observed views. Accordingly, the realized sample-level missing rate is

$$\rho = \frac{1}{N}\left|\left\{i \mid 1 \leq \sum_{v=1}^{V} M_i^{(v)} < V\right\}\right| = 1 - \frac{N_c}{N}. \tag{20}$$

Thus, the nominal missing rates of 0.7, 0.8, 0.9, and 0.95 correspond to approximately 30%, 20%, 10%, and 5% complete samples, respectively, with integer sample counts. In particular, the 95% setting means that approximately 95% of samples retain a nonempty, proper subset of views, while the remaining samples retain all views. It does not mean that each view is independently removed with probability 0.95.

Conditions for batch-wise loss evaluation. Sample-level observability and batch-level completeness are distinct conditions. The guarantee that every sample retains at least one view ensures that a nonempty batch contains observed data, but does not by itself imply that $N_c^B = |\mathcal{C} \cap \mathcal{B}|$ is positive. Likewise, it does not guarantee that every individual view has visible samples in that batch. The complete-sample consistency losses and shared assignment consistency loss, whose normalization contains N_c^B, require $N_c^B > 0$ for evaluation as written. The co-occurrence-based bridging matrix also requires complete samples in the batch for its row normalization to be defined. If $N_c^B > 0$, then every view contains at least those N_c^B complete samples, so $n_v \geq N_c^B > 0$ for all views. Therefore, the per-sample observation constraint alone is not sufficient to establish that all batch-wise loss expressions are well-defined. In implementation, each prototype-training mini-batch is sampled to include at least one complete sample, ensuring $N_c^B > 0$ and making the above batch-wise objectives well-defined.

For hyperparameter selection, a grid search strategy is adopted to determine three key parameters, including the shared semantic consistency weight λ_{share}, the robust prototype loss weight λ_{rob}, and the prototype bridging weight β. Specifically, the search ranges of λ_{share}, λ_{rob}, and β are set to $\{0.001, 0.01, 0.1, 1\}$, $\{0.001, 0.01, 0.1, 1\}$, and $\{1.0, 2.0, 3.0, 4.0, 5.0\}$, respectively. The final parameter configurations are selected according to the clustering performance. The influence of these hyperparameters is further investigated in the parameter sensitivity analysis. In addition, prototype initialization analysis and ablation studies are conducted to validate the effectiveness of each component.

To reduce the influence of random initialization, each experiment is repeated 10 times with different random seeds, and the average results are reported. All experiments are implemented in Python and conducted on a workstation equipped with an NVIDIA GeForce RTX 4090 GPU.

A.2 View-Specific Prototype Construction and Update

To characterize the local latent structure of each view, view-specific prototypes are further constructed from visible samples within the current mini-batch $\mathcal{B}$. For the v-th view, the set of visible latent representations is

$$\mathcal{Z}_{\mathcal{B}}^{(v)} = \left\{ \mathbf{z}_i^{(v)} \mid i \in \mathcal{B}, M_i^{(v)} = 1 \right\}. \tag{21}$$

The base prototype number is determined from the visible-sample scale as

$$K_p = \left\lceil \frac{N_{\text{vis}}^{\max} r_p}{N / B_{\text{rob}}} \right\rceil, \qquad N_{\text{vis}}^{\max} = \max_v \sum_{i=1}^{N} M_i^{(v)}. \tag{22}$$

r_p denotes the prototype rate, and B_{rob} is the batch size used during prototype training. Since the number of visible samples may vary substantially across views under high missing rates, the effective prototype number for the v-th view is adaptively determined as

$$K_p^{(v)} = \min(K_p, n_v), \qquad n_v = \left| \mathcal{Z}_{\mathcal{B}}^{(v)} \right|. \tag{23}$$

Thus, the number of prototypes never exceeds the number of visible samples in the current view.

In implementation, visible samples are randomly shuffled according to their aligned sample indices. The visible latent matrix of the v-th view is written as

$$\mathbf{Z}_{\mathcal{B}}^{(v)} = \left[(\mathbf{z}_{b_1}^{(v)})^T; (\mathbf{z}_{b_2}^{(v)})^T; \ldots; (\mathbf{z}_{b_{n_v}}^{(v)})^T \right]. \tag{24}$$

The first $K_p^{(v)}$ visible representations are selected as sample-anchor prototypes:

$$\mathbf{P}^{(v)} = \left[(\mathbf{p}_1^{(v)})^T; \ldots; (\mathbf{p}_{K_p^{(v)}}^{(v)})^T \right] \in \mathbb{R}^{K_p^{(v)} \times d_z^{(v)}}, \qquad \mathbf{p}_k^{(v)} = \mathbf{z}_{b_k}^{(v)}. \tag{25}$$

These prototypes act as local sample anchors rather than strict cluster centroids and are accumulated as a view-specific prototype memory for missing-representation recovery.

To further refine the sample-anchor prototypes, we additionally consider a centralized prototype update initialized from these anchors. For a visible sample $\mathbf{z}_i^{(v)}$, its hard prototype assignment is defined as

$$a_i^{(v)} = \underset{1 \le k \le K_p^{(v)}}{\arg\min} \left(1 - \text{sim}\left(\mathbf{z}_i^{(v)}, \mathbf{p}_k^{(v)} \right) \right), \tag{26}$$

where $\text{sim}(\cdot, \cdot)$ denotes cosine similarity. Let

$$\mathcal{I}_k^{(v)} = \left\{ i \mid a_i^{(v)} = k \right\} \tag{27}$$

denote the set of samples assigned to the k-th prototype. The prototype is then updated as

$$\mathbf{p}_k^{(v)} \leftarrow \frac{1}{|\mathcal{I}_k^{(v)}|} \sum_{i \in \mathcal{I}_k^{(v)}} \mathbf{z}_i^{(v)}. \tag{28}$$

When $\mathcal{I}_k^{(v)}$ is empty, the corresponding prototype remains unchanged to avoid numerical instability caused by empty clusters.

A.3 Prototype Initialization Strategy Analysis

Table 4 compares the impact of different prototype initialization strategies on the clustering performance of CSSP. Here, unba denotes the strategy that randomly selects visible samples for prototype initialization, while ba represents the strategy that selects complete samples based on a balanced coverage strategy to initialize prototypes. The results show that the balanced prototype initialization strategy achieves better clustering performance on most datasets and missing rate settings. For example, on the Caltech101-7 dataset, when the missing rate is 0.7, ba achieves ACC/NMI/F1 of 85.55/75.06/85.20, outperforming unba (84.82/75.12/83.80) in terms of ACC and F1. On the Scene15 dataset, ba consistently maintains advantages under all missing rate settings. Moreover, on the 100leaves dataset, ba also achieves better performance under relatively low missing rates. For example, at a missing rate of 0.7, ba obtains ACC/NMI/F1 of 80.04/91.78/78.14.

However, the advantage of the balanced prototype initialization strategy becomes less significant under some extremely high missing rate scenarios. For example, under the missing rate of 0.95 on Caltech101-7 and the missing rates of 0.9 and 0.95 on 100leaves, the unba strategy achieves better or comparable performance. This is mainly because these datasets contain more complex latent cluster structures, while the number of prototypes adopted in this study is relatively limited. In such cases, the balanced coverage strategy can improve the coverage of different semantic regions, but a limited number of prototypes may still be insufficient to represent all latent structures. In contrast, the unba strategy tends to select more representative samples according to the data distribution, resulting in more compact prototype representations. Therefore, when the number of prototypes is insufficient to cover complex data structures, prototype initialization strategies need to balance structural coverage capability and sample representativeness.

Furthermore, Table 4 shows that balanced prototype initialization yields more stable training in several settings. For example, on Caltech101-7 at a missing rate of 0.7, the standard deviation of ba (2.11) is lower than that of unba (3.45), indicating reduced fluctuation caused by random sample selection. Overall, balanced prototype initialization improves the coverage of latent semantic structures and generally provides more stable clustering performance under severe missingness.

Table 4: Clustering performance under different missing rates on Caltech101-7, Scene15, and 100leaves. The best results are shown in bold and the second-best results are underlined.

Dataset	Method	0.70			0.80			0.90			0.95		
		ACC	NMI	F1	ACC	NMI	F1	ACC	NMI	F1	ACC	NMI	F1
Caltech101-7	unba–mean	84.82	75.12	83.80	84.76	75.05	84.76	84.24	73.51	80.95	84.44	75.35	83.40
	unba–std	3.45	2.09	4.74	4.38	1.71	5.61	6.53	3.61	7.65	3.86	2.13	5.25
	ba–mean	**85.55**	75.06	**85.20**	**86.78**	**76.47**	**86.43**	**85.97**	**75.13**	**85.64**	84.12	**75.67**	82.78
	ba–std	**2.11**	**2.02**	**2.27**	**3.74**	2.49	**4.88**	**4.16**	**3.30**	**4.50**	4.45	3.03	**4.96**
Scene15	unba–mean	40.94	42.98	40.64	40.06	42.04	39.02	41.12	41.98	40.37	39.69	40.91	39.40
	unba–std	2.57	1.33	2.60	2.15	1.27	2.27	1.70	1.20	2.14	1.62	1.26	1.51
	ba–mean	**42.46**	**43.67**	**42.33**	**42.59**	**40.83**	**42.83**	**41.17**	**42.74**	**40.54**	**41.13**	**41.53**	**40.44**
	ba–std	**2.27**	**0.86**	**2.49**	**1.59**	**0.92**	**1.59**	**1.59**	1.33	**1.73**	2.01	**1.13**	2.25
100leaves	unba–mean	78.76	91.44	76.70	78.12	90.68	76.23	77.96	90.22	76.15	76.51	89.67	74.46
	unba–std	1.79	0.82	1.99	2.36	0.98	2.54	3.02	1.29	3.31	1.59	0.66	1.94
	ba–mean	**80.04**	**91.78**	**78.14**	77.59	**90.74**	75.63	77.69	**90.32**	75.63	76.24	**90.00**	74.15
	ba–std	2.07	**0.68**	2.42	**2.06**	**0.59**	**2.28**	**2.23**	**0.86**	**2.60**	2.91	1.22	3.07

A.4 ALGORITHM

For clarity, the overall training and inference procedure of the proposed method is summarized in Algorithm 1.

Algorithm 1 Complete-Sample-Driven Missing-View Consistency Learning with Shared Semantic Prototypes

Require: Incomplete multi-view data $\{\mathbf{X}^{(v)}\}_{v=1}^{V}$, missing indicator matrix $\mathbf{M}$, prototype rate r_p, and prototype-training batch size B_{rob}
Ensure: Cluster assignments $\mathbf{Y}$

Pretraining Stage:
Select complete samples $\mathcal{C} = \{i | M_i^{(v)} = 1, \forall v\}$
Optimize view-specific autoencoders using reconstruction loss $\mathcal{L}_{pre}$
Obtain initial latent representations $\{\mathbf{Z}^{(v)}\}_{v=1}^{V}$
Prototype Training Stage:
for each training iteration **do**
  Sample a mini-batch $\mathcal{B}$ and construct artificial missing-view tasks from complete samples
  Compute K_p by Eq. (22) and $K_p^{(v)} = \min(K_p, n_v)$ by Eq. (23)
  Compute batch-wise missing-view consistency regularization:

$$\mathcal{L}_{\text{mvc}} = \mathcal{L}_{\text{lt}} + \mathcal{L}_{\text{rcn}} + \mathcal{L}_{\text{asn}}$$

  Construct the shared semantic prototype bank $\mathbf{G}$ and view-specific prototype matrices $\{\mathbf{P}^{(v)}\}_{v=1}^{V}$
  Compute prototype semantic learning loss:

$$\mathcal{L}_{\text{pro}} = \lambda_{\text{rob}}\mathcal{L}_{\text{rob}} + \lambda_{\text{share}}\mathcal{L}_{\text{share}} + \beta\mathcal{L}_{\text{bridge}}$$

  Update model parameters by minimizing:

$$\mathcal{L} = \mathcal{L}_{\text{pro}} + \lambda_{\text{mvc}}\mathcal{L}_{\text{mvc}}$$

end for
Inference Stage:
Recover missing latent representations using learned prototypes, concatenate completed multi-view representations, and apply K-means to obtain final clustering assignments
return $\mathbf{Y}$

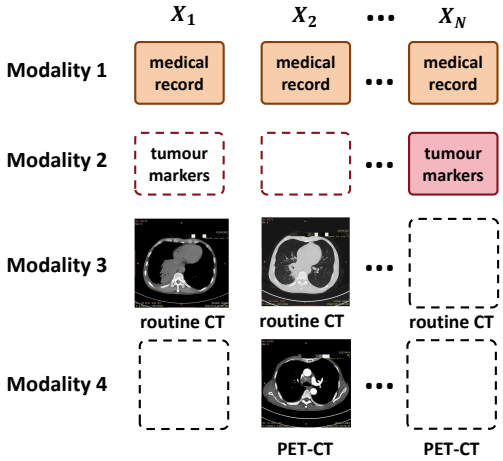

Figure A.1: Example of incomplete multimodal data.

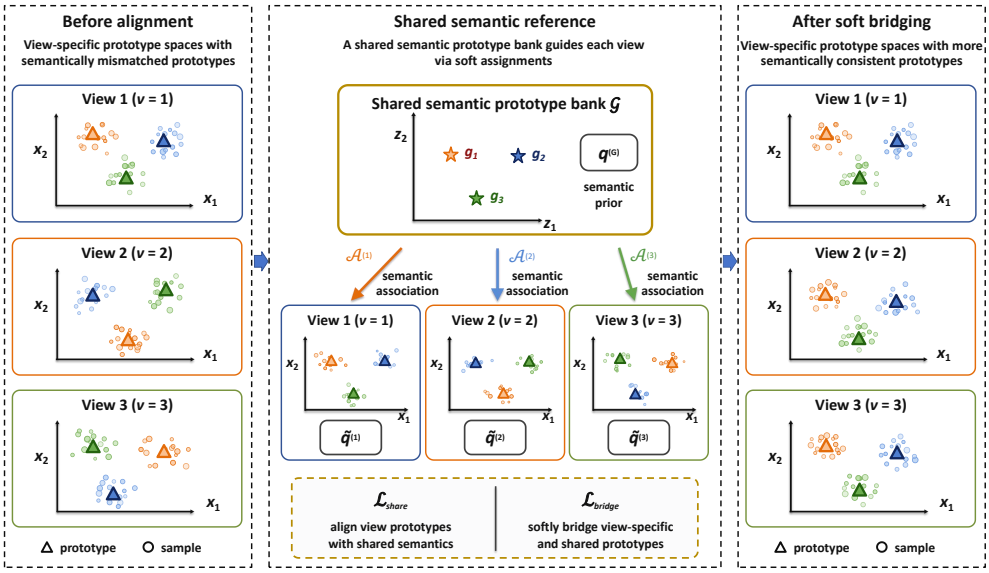

Figure A.2: Example Illustration of Shared-Semantic-Guided Prototype Alignment.

Figure A.2 presents a conceptual illustration of shared-semantic-guided prototype alignment. Shared semantic prototypes provide a global semantic reference, while $\mathcal{L}_{\text{share}}$ and $\mathcal{L}_{\text{bridge}}$ jointly guide view-specific prototype structures toward greater cross-view semantic consistency.

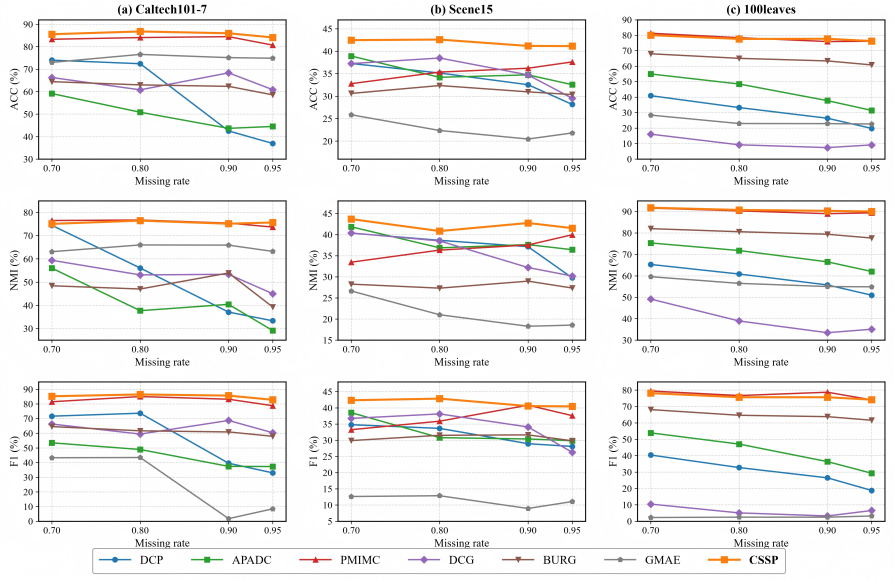

Figure A.3: Clustering Performance of Representative Methods Across Different Missing Rates.

Figure A.3 compares CSSP with representative incomplete multi-view clustering methods on Caltech101-7, Scene15, and 100leaves at different missing rates. CSSP generally achieves superior or competitive performance in ACC, NMI, and F1, while maintaining relatively stable results as the missing rate increases.

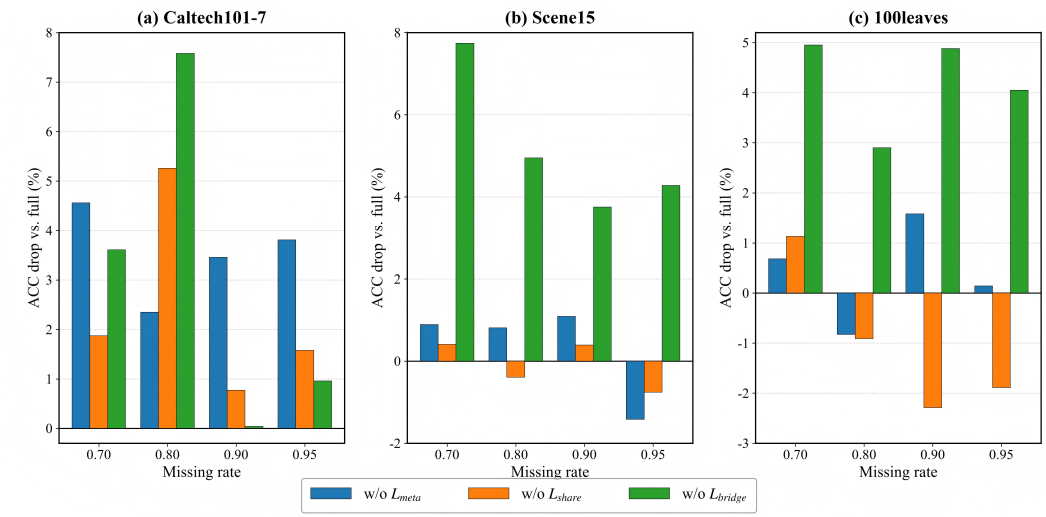

Figure A.4: Ablation Impact on ACC.

Figure A.4 shows the ACC drop after removing $\mathcal{L}_{\text{mvc}}$, $\mathcal{L}_{\text{share}}$, or $\mathcal{L}_{\text{bridge}}$ from the full CSSP model. Removing $\mathcal{L}_{\text{bridge}}$ generally causes the largest performance degradation among the three ablations, highlighting the contribution of prototype-level soft bridging.

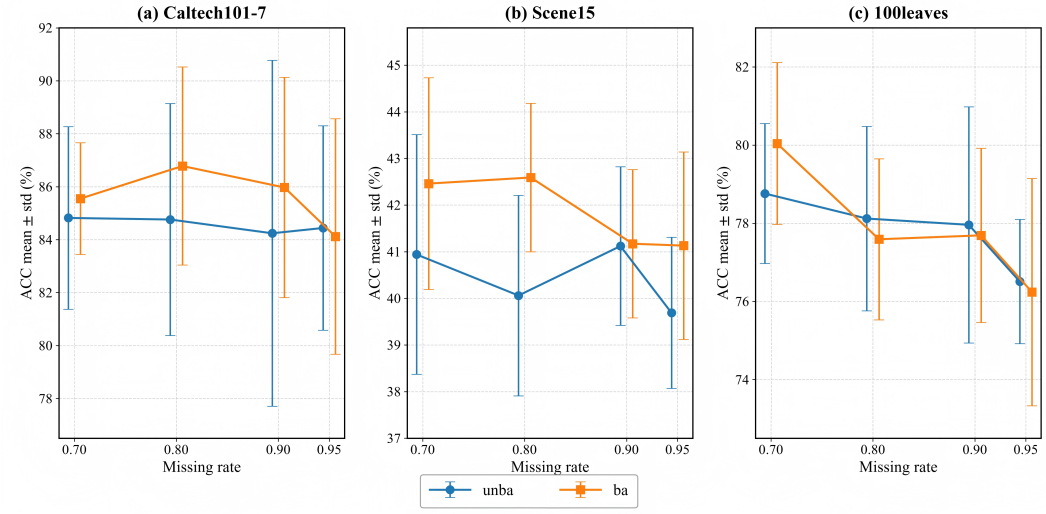

Figure A.5: Prototype Initialization Strategy Analysis.

Figure A.5 compares unbalanced random initialization (unba) and balanced initialization (ba) in terms of mean ACC and standard deviation. Balanced initialization generally yields better or more stable performance, although its advantage becomes less pronounced in some highly incomplete settings.