Robust Multi-view Clustering against Imperfect Information

Zhichao Huang[1] **Haochen Zhou**[1] **Hao Wang**[1] **Mouxing Yang**[1*] **Xi Peng**[2,3]

[1]College of Computer Science, Sichuan University, China.
[2]School of Artificial Intelligence, Sichuan University, China.
[3]Tianfu Jincheng Laboratory, Chengdu, China.

{zhichaohuang.gm, haochenzhou.xl, cshaowang, yangmouxing, pengx.gm}@gmail.com

Abstract

Real-world multi-view data always suffer from imperfect information problem, where the view-specific observations are absent (*i.e.*, Incomplete Views, IV) and cross-view correspondences are mismatched (*i.e.*, Noisy Correspondences, NC) for certain instances. As a remedy, numerous IV- and NC-oriented multi-view clustering (MvC) methods have been proposed, which however require either reliable correspondences or sufficiently complete instances, thus stopping short of addressing the imperfect information problem. In contrast, we observe that both IV and NC challenges originate from the same issue of imperfect cross-view counterpart information, where the counterpart of an anchor instance in another view might be either unavailable or unreliable. Based on the observation, we propose a novel robust MvC framework, termed Posterior-guided Latent Counterpart Inference (PLCI), which could handle both IV and NC in a unified manner. Specifically, PLCI formulates the desired cross-view counterpart of each anchor instance as a latent variable, and integrates both instance-level reliability and prototype-level semantic transport to infer the posterior distribution of the latent counterpart. Extensive experiments on six widely-used multi-view datasets against 10 state-of-the-art MvC methods demonstrate the effectiveness of PLCI for tackling the imperfect information problem. Code is available at https://github.com/XLearning-SCU/2026-NeurIPS-PLCI.

1 Introduction

Real-world instances are always characterized by multiple views or modalities, including but not limited to RGB images, depth maps, audio signals, and text descriptions. As one of the most fundamental tools for analyzing such heterogeneous data, multi-view clustering (MvC) aims to partition unlabeled instances into semantic groups by exploiting complementary information within individual views and consistency information across cross-view correspondence [7, 9, 30, 31, 34, 42, 43, 51]. In practice, however, the success of most existing MvC methods [36, 37] is often hindered by the imperfect information problem, which is mainly manifested as incomplete views (IV) and noisy correspondences (NC). Specifically, as illustrated in Fig. 1(a), IV and NC refer to the absence of view-specific observations and the mismatched inter-view correspondences for certain instances, respectively. As a result, these two challenges would impede the exploitation of both view-specific and cross-view information, thereby significantly degrading MvC performance.

To conquer these challenges, numerous robust MvC methods [24, 27–29] have been proposed. Specifically, as shown in Fig. 1(b)-(c), for the IV challenge, IV-oriented MvC methods typically learn the cross-view generative mapping from the complete instances, and accordingly recover the missing

*Corresponding authors.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

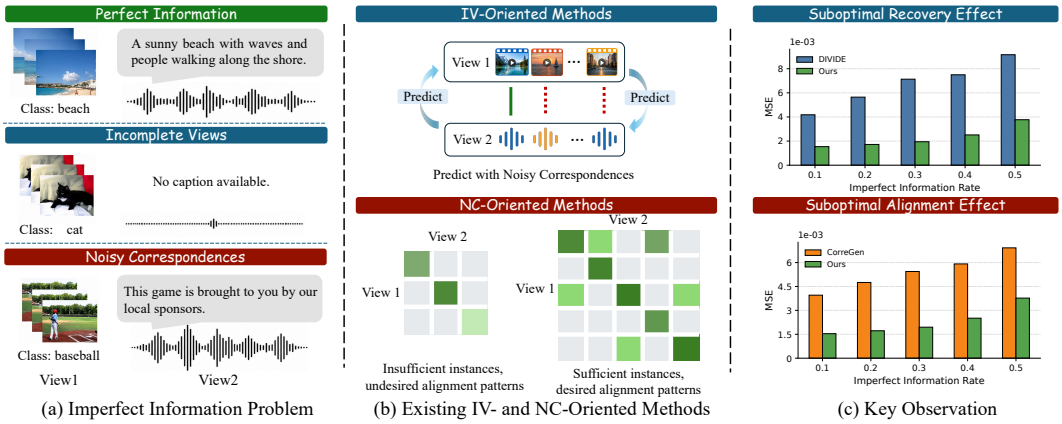


Figure 1: Our motivations and observations. (a) The studied problem. Ideally, the video and audio views should provide consistent semantics. However, in the noisy correspondences (NC) case, the audio view is irrelevant to the visual content. In the incomplete views (IV) case, the audio view is missing. (b) Limitations of existing IV- and NC-oriented methods. Under imperfect information, IV-oriented methods are susceptible to unreliable correspondences during training, which could result in erroneous view recovery. Meanwhile, NC-oriented methods would establish inferior alignment effects across views as the number of complete instances for correspondence inference is insufficient. (c) Analytic studies. With the increasing imperfect information level, the recovery/alignment effect of both state-of-the-art IV- and NC-oriented methods significantly degrades, demonstrating that neither IV- nor NC-oriented methods can handle the imperfect information problem. The analyses are conducted on the Scene15 dataset, and the recovery/alignment effect is measured by the mean squared error (MSE) between the recovered/aligned view and the ground-truth view.

views of incomplete instances with the help of the observed ones. However, mismatched cross-view correspondences might steer the optimization of generators toward incorrect cross-view mappings, resulting in suboptimal or even erroneous recovery. Different from the widely-studied IV challenge, the NC challenge has attracted increasing attention in recent years. In general, NC-oriented methods aim to re-establish reliable inter-view correspondence by leveraging the alignment pattern learned from a number of complete instances. Despite their robustness against NC, the lack of sufficient complete instances would undermine the learning of reliable cross-view alignment patterns. Clearly, existing studies [44, 45, 49, 50] are mainly designed to handle either the IV or NC challenge. In other words, how to address both challenges in a unified manner and thus tackle the imperfect information problem remains under-explored.

In this paper, we observe that both IV and NC originate from imperfect cross-view counterpart information. Specifically, IV makes the counterpart unavailable, whereas NC renders the observed counterpart unreliable. One straightforward way to address these challenges is to combine the existing solutions of IV and NC. To be more specific, one could first recover the unavailable counterparts and then identify the reliable ones in other views, or vice versa. However, such a two-stage strategy is inherently suboptimal, as errors introduced in either recovery or re-alignment would inevitably propagate to the other stage. More importantly, as aforementioned, the effectiveness of view recovery and correspondence re-alignment fundamentally depends on reliable correspondences and sufficiently complete instances, both of which are daunting to satisfy in the context of imperfect information. To support the above claims, we present analysis experiments in Table 12.

To solve the imperfect information problem, we propose a novel MvC method termed Posterior-guided Latent Counterpart Inference (PLCI), which could address both IV and NC challenges in a unified framework. The key idea behind PLCI is reformulating the counterpart of instances as a latent variable and inducing the posterior distribution for the latent counterpart by integrating both instance-level reliability and prototype-level semantic transport. Specifically, PLCI first estimates the correspondence reliability of complete instances by resorting to the memory effect of deep neural networks [2]. After that, PLCI performs reliability-aware optimal transport upon view-specific prototypes constructed from all observed samples, thus propagating semantics across views. Finally, by combining the transported semantic structure with the observed information, PLCI could infer

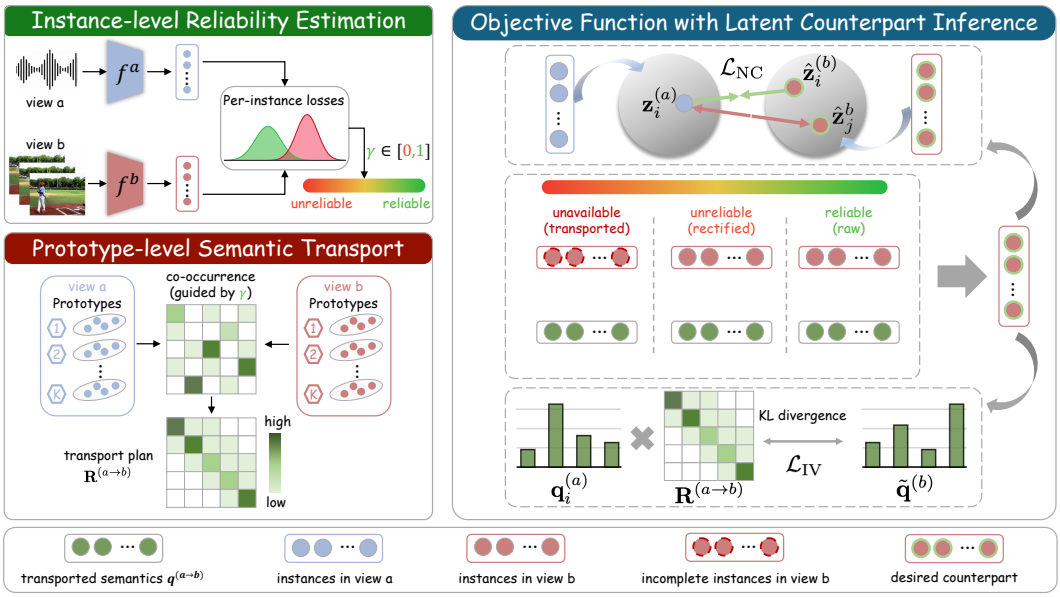


Figure 2: **Overview of PLCI.** To begin with, PLCI estimates the instance-level reliability γ_i from the pair-loss distribution of complete instances. Then, PLCI constructs view-specific prototypes and computes a reliability-aware co-occurrence matrix, which serves as a robust proxy for cross-view prototype correspondence and thus facilitates prototype-level semantic transport. Finally, PLCI integrates the observed counterpart information and the transported semantic counterpart to infer the latent counterpart for achieving robust multi-view clustering against imperfect information.

the posterior distribution for each instance and derive the reliable pseudo counterpart for achieving robust multi-view clustering. In brief, the main contributions of this work are outlined as follows:

- We study a practical yet under-explored problem in multi-view clustering, termed imperfect information, where incomplete views and noisy correspondences simultaneously exist. Moreover, we investigate the negative impact of the imperfect information problem and reveal that existing IV- and NC-oriented methods are incapable of handling the problem.
- To solve the imperfect information problem, we propose a novel MvC method, termed Posterior-guided Latent Counterpart Inference (PLCI). Instead of handling unavailable or unreliable counterparts separately, PLCI formulates the counterpart of each instance as a latent variable, thus addressing the imperfect information problem in a unified framework.
- Extensive comparative experiments against 10 state-of-the-art robust MvC methods on six widely-used multi-view clustering datasets covering diverse modalities (*e.g.*, video-audio, RGB-depth), verify the effectiveness of PLCI. Furthermore, we demonstrate that PLCI is a generalizable framework that could be seamlessly incorporated into a broad spectrum of existing MvC methods to enhance their robustness against imperfect information.

2 Method

In this section, we present Posterior-guided Latent Counterpart Inference (PLCI) for robust multi-view clustering under imperfect information. As shown in Fig. 2, PLCI formulates the desirable target-view counterpart as a latent variable, and infers it by combining instance-level correspondence reliability with prototype-level semantic transport. Specifically, Sec. 2.1 defines the imperfect information problem and the latent counterpart formulation. Sec. 2.2 estimates reliable correspondences from complete instances, while Sec. 2.3 propagates semantic structures across views via prototype-level transport. Finally, Sec. 2.4 presents the latent counterpart inference criterion and the joint objective for learning with both noisy correspondences and incomplete views.

2.1 Problem Formulation

Given a multi-view dataset $\mathcal{D} = \{\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_N\}$, the objective of MvC is to construct a common space where N instances could be partitioned into C groups with different semantics. For clarity, let $\mathcal{X}_i = \{(\mathbf{x}_i^{(v)} \mid m_i^{(v)} \in \{0,1\})\}_{v=1}^{V}$ denote the i-th instance, where v is the view index and $m_i^{(v)} = 0/1$ indicates that the view v is unavailable/available. Besides, in the following, we use a and b ($b \neq a$) to denote the view indices of the anchor and target views, respectively. To begin with, we define the imperfect information problem as follows.

Definition 1 (Imperfect Information)**.** *The dataset $\mathcal{D}$ is contaminated with imperfect information when there exists a subset $\mathcal{D}_{\text{imp}}$ that satisfies the following condition:*

$$\mathcal{D}_{\text{imp}} = \left\{ \mathcal{X}_i \;\middle|\; \underbrace{0 < \sum_{v=1}^{V} m_i^{(v)} < V}_{\text{incomplete views}} \vee \underbrace{\sum_{b=1, b\neq a}^{V} \mathbb{I}\left(\mathbf{x}_i^{(b)}, \mathbf{x}_i^{(a)}\right) < (V-1)}_{\text{noisy correspondences}} \right\}, \tag{1}$$

where $\mathbb{I}(\cdot,\cdot)$ is the indicator function evaluating to 1 i.f.f. $\mathbf{x}_i^{(b)}$ and $\mathbf{x}_i^{(a)}$ belong to the same instance.

Definition 1 demonstrates that imperfect information is mainly manifested as Incomplete Views (IV) and Noisy Correspondences (NC). To be specific, for an observed anchor $\mathbf{x}_i^{(a)}$, IV contributes to the absence of its counterparts in other views, while NC results in the presence of unreliable counterparts, both of which would heavily degrade the MvC performance. However, most existing IV-oriented and NC-oriented MvC methods are capable of handling either one of these two challenges, stopping short of addressing the imperfect information problem, as verified in Table 1 and Figs. 4-5.

To tackle the imperfect information problem, one can formulate the desirable counterpart representation in the target view as a latent variable $\hat{\mathbf{z}}_i^{(b)}$ and model its posterior distribution as

$$p_\phi\left(\hat{\mathbf{z}}_i^{(b)} \mid \mathcal{O}_i\right), \tag{2}$$

where $\mathcal{O}_i$ denotes the combination of i-th instance observations and the whole dataset $\mathcal{D}$ acting as prior for the latent counterpart $\hat{\mathbf{z}}_i^{(b)}$. Ideally, Eq. 2 should shed light on the information inherent in $\mathcal{O}_i$, so that one could obtain the desirable counterpart for each anchor instance. However, the potential noisy correspondences in the complete instances and the absence of counterparts in incomplete instances would prevent the effective inference of the latent counterparts.

As a remedy, we propose PLCI to enable effective latent counterpart inference for robust multi-view clustering under imperfect information. PLCI consists of two modules (instance-level reliability estimation and prototype-level semantic transport) to explore and exploit information in complete instances and incomplete instances for counterpart inference, and a joint objective function to achieve robust multi-view clustering against imperfect information.

2.2 Instance-level Reliability Estimation

The obstacle behind the information mining for complete instances is the potential noisy correspondences, which would mislead the cross-view common space construction. As a remedy, the proposed method aims to estimate the correspondence reliability of complete instances, so that the NC-contaminated complete instances could be uncovered. To this end, the proposed method resorts to the widely-acknowledged memorization effect of deep neural networks (DNNs) [2], which indicates that DNNs tend to fit simple patterns and gradually learn harder ones, thus leading to relatively lower losses for clean samples. Motivated by this, one can regard the loss distribution of complete instances as a proxy for estimating the reliability of their correspondence. To be specific, for each complete instance, the per-instance losses are computed as follows,

$$\left\{ \ell_{\text{con}}\left(f^{(a)}(\mathbf{x}_i^{(a)}), f^{(b)}(\mathbf{x}_i^{(b)})\right) \;\middle|\; m_i^{(a)} m_i^{(b)} = 1 \right\}, \tag{3}$$

where $f^{(v)}(\cdot)$ denotes the view-specific encoder for view v and $\ell_{\text{con}}(\cdot,\cdot)$ is the loss function for measuring the consistency between two views. In the implementation, one can adopt the widely-used contrastive loss [5, 15, 32] as $\ell_{\text{con}}(\cdot,\cdot)$.

After that, following the typical paradigm in noisy label learning [22], one could fit the loss value set in Eq. 3 with a two-component Gaussian Mixture Model (GMM). Formally,

$$p(\ell_{\text{con}} \mid \theta) = \sum_{k=1}^{2} \pi_k \Psi(\ell_{\text{con}} \mid k), \tag{4}$$

where π_k and $\Psi(\ell_{\text{con}} \mid k)$ denote the mixture coefficient and the Gaussian distribution of the k-th component, respectively. According to the memory effect of DNNs, the component with a smaller mean is expected to represent instances with reliable correspondence. Therefore, one could estimate the correspondence reliability γ_i for complete instances by the posterior probability of their losses belonging to the low-loss component κ, namely,

$$\gamma_i = p(\kappa \mid \ell_{\text{con}}). \tag{5}$$

2.3 Prototype-level Semantic Transport

To mine the information from incomplete instances, it is expected to complement the missing views with the help of observed ones. Unlike the existing IV-oriented methods that directly recover the missing views at the instance level, PLCI transports the semantic structure across views. The motivation behind this design is two-fold. First, the recovery effect of existing methods heavily relies on sufficient complete instances with reliable correspondences, which is hard to satisfy in the context of imperfect information. In contrast, the semantic structure of each view can be estimated from all available observations, thus being more robust against imperfect information. Second, the propagated semantics across views encourage cross-view consistency, which could further facilitate the cross-view common space construction.

Specifically, PLCI first reveals the semantic structure of each view by maintaining a set of view-specific prototypes. Formally speaking, for each view v, PLCI maintains C prototypes $\{\boldsymbol{\mu}_c^{(v)}\}_{c=1}^{C}$. After that, PLCI estimates the semantic structure of each view by assigning each observed instance to the prototypes. Formally, for an observed sample $\mathbf{x}_i^{(v)}$, the semantic structure is defined as

$$q_{ic}^{(v)} = \frac{\exp\left(\text{sim}(\mathbf{z}_i^{(v)}, \boldsymbol{\mu}_c^{(v)})/\tau_p\right)}{\sum_{r=1}^{C} \exp\left(\text{sim}(\mathbf{z}_i^{(v)}, \boldsymbol{\mu}_r^{(v)})/\tau_p\right)}, \tag{6}$$

where $\mathbf{z}_i^{(v)} = f^{(v)}(\mathbf{x}_i^{(v)})$, τ_p is the temperature parameter and $\text{sim}(\cdot,\cdot)$ denotes the similarity function. Therefore, the semantic structure of $\mathbf{x}_i^{(v)}$ is characterized by the assignment vector $\mathbf{q}_i^{(v)} = [q_{i1}^{(v)}, \cdots, q_{iC}^{(v)}]$.

To propagate semantic structure across views, PLCI establishes cross-view prototype correspondence by employing reliability-aware optimal transport. To be specific, PLCI estimates the co-occurrence matrix of prototypes across views, which could reflect the likelihood of different cross-view prototypes describing the same underlying semantics. However, the estimation of co-occurrence matrix is susceptible to noisy correspondences. To mitigate the adverse effect of noisy correspondences, the estimated reliability of complete instances is further incorporated into the co-occurrence estimation. Mathematically, the co-occurrence matrix is defined as

$$\mathbf{U}^{(a,b)} = \sum \gamma_i \left(\mathbf{q}_i^{(a)}\right)^{\top} \mathbf{q}_i^{(b)}, \tag{7}$$

where γ_i denotes the estimated reliability of the correspondence between $\mathbf{x}_i^{(a)}$ and $\mathbf{x}_i^{(b)}$. Clearly, a larger value of $\mathbf{U}_{tr}^{(a,b)}$ indicates that the t-th prototype in view a and the r-th prototype in view b are more likely to describe the same underlying semantics.

After that, one can employ the optimal transport to derive the cross-view prototype correspondence, which indicates how the semantic structure of one view should be transported to the other view. To be specific, the cost matrix for the optimal transport is defined as

$$\mathbf{D}^{(a,b)} = -\frac{\mathbf{U}^{(a,b)}}{\|\mathbf{U}^{(a,b)}\|_{\infty}}, \tag{8}$$

where $\|\cdot\|_\infty$ denotes the infinity norm. After solving the optimal transport problem with the cost matrix $\mathbf{D}^{(a,b)}$, one could obtain the transport plan $\mathbf{R}^{(a\to b)}$. Finally, the prototype-level semantic transport from view a to view b could be achieved as follows,

$$\mathbf{q}_i^{(a\to b)} = \mathbf{q}_i^{(a)}\mathbf{R}^{(a\to b)}. \tag{9}$$

2.4 Objective Function with Latent Counterpart Inference

After obtaining the instance-level correspondence reliability and the prototype-level semantic transport, PLCI could infer the latent counterpart by combining the observed counterpart information with the transported semantic prior. Specifically, one could use the posterior mean as the inferred counterpart, which is approximated as

$$\hat{\mathbf{z}}_i^{(b)} \triangleq \mathbb{E}\left[\hat{\mathbf{z}}_i^{(b)} \mid \lambda_i, \mathcal{O}_i\right] \approx \lambda_i \tilde{\mathbf{z}}_i^{(b)} + (1-\lambda_i)\,\mathbf{s}_i^{(a\to b)}, \tag{10}$$

where $\tilde{\mathbf{z}}_i^{(b)}$ denotes the counterpart representation in the target view, $\mathbf{s}_i^{(a\to b)} = \sum_{c=1}^{C} q_{ic}^{(a\to b)}\boldsymbol{\mu}_c^{(b)}$ denotes the transported semantic counterpart obtained from prototype-level semantic transport, and λ_i controls their relative contributions. Under the imperfect information problem, the counterpart may be unreliable or unavailable, resulting in the following two situations:

- **For complete but reliability-agnostic instances.** The target-view observation is available, and $\tilde{\mathbf{z}}_i^{(b)}$ is given by the observed target-view representation $\mathbf{z}_i^{(b)}$. However, this observed counterpart may be mismatched with the anchor view due to noisy correspondences. To alleviate this issue, $\lambda_i = \gamma_i \left\langle \mathbf{q}_i^{(a\to b)}, \mathbf{q}_i^{(b)} \right\rangle$ serves as an adaptive weight in Eq. 10, controlling the relative contribution of the observed counterpart and the transported semantic counterpart. Specifically, it combines the estimated correspondence reliability γ_i with the semantic consistency $\left\langle \mathbf{q}_i^{(a\to b)}, \mathbf{q}_i^{(b)} \right\rangle$ between the transported prototype assignment and the observed target-view assignment. Consequently, reliable complete instances assign larger weights to the observed counterpart, whereas unreliable ones are calibrated by the transported semantic counterpart.
- **For incomplete and availability-deficient instances.** The target-view observation is unavailable, and thus $\mathbf{z}_i^{(b)}$ cannot be directly observed. Accordingly, $\tilde{\mathbf{z}}_i^{(b)}$ has to be constructed from the anchor-view representation through the cross-view generator $g^{(a\to b)}(\mathbf{z}_i^{(a)})$, following the common practice of IV-oriented MvC methods [25, 32]. However, this generated counterpart may be affected by inaccurate cross-view mappings, especially when complete instances are contaminated by noisy correspondences, as depicted in Fig. 1. To reduce the dependence on the generated counterpart alone, λ_i is set to $1/2$ in Eq. 10, allowing the generated counterpart and the transported semantic counterpart to contribute equally to the inferred latent counterpart. In this way, the inferred latent counterpart is constrained by the target-view semantic structure rather than relying solely on the generated counterpart.

With the inferred counterparts, one could optimize PLCI by the following joint objective:

$$\mathcal{L}_{\mathrm{PLCI}} = \mathcal{L}_{\mathrm{NC}} + \alpha\mathcal{L}_{\mathrm{IV}}, \tag{11}$$

where α is a trade-off hyperparameter fixed as 0.2 in our experiments. $\mathcal{L}_{\mathrm{NC}}$ and $\mathcal{L}_{\mathrm{IV}}$ are designed for learning with noisy correspondences and incomplete views, respectively.

Specifically, $\mathcal{L}_{\mathrm{NC}}$ is imposed on complete instances and thus given by:

$$\mathcal{L}_{\mathrm{NC}} = \sum \ell_{\mathrm{con}}\left(f^{(a)}(\mathbf{x}_i^{(a)}), \hat{\mathbf{z}}_i^{(b)}\right), \tag{12}$$

where $f^{(a)}(\cdot)$ is the view-specific encoder, $\ell_{\mathrm{con}}(\cdot,\cdot)$ denotes the cross-view contrastive loss [41]. By utilizing the inferred latent counterpart, $\mathcal{L}_{\mathrm{NC}}$ encourages robust cross-view alignment under noisy correspondences.

As for incomplete instances, instead of directly applying to the inferred counterparts, the proposed method manipulates the transported semantic structure to regularize the inferred counterparts. Formally,

$$\mathcal{L}_{\mathrm{IV}} = \sum D_{\mathrm{KL}}\left(\mathbf{q}_i^{(a\to b)} \| \tilde{\mathbf{q}}_i^{(b)}\right), \tag{13}$$

where $D_{\mathrm{KL}}(\cdot\|\cdot)$ denotes the KL divergence, $\mathbf{q}_i^{(a\to b)}$ is the transported prototype assignment defined in Eq. 9, and $\tilde{\mathbf{q}}_i^{(b)}$ is the prototype assignment of the inferred latent counterpart $\hat{\mathbf{z}}_i^{(b)}$. By minimizing $\mathcal{L}_{\mathrm{IV}}$, the inferred counterpart is encouraged to preserve the transported target-view semantic structure, thereby facilitating robust clustering when target-view observations are missing.

3 Experiments

In this section, we carry out extensive experiments to evaluate the effectiveness of the proposed PLCI in addressing the imperfect information problem. Due to space limitations, we present more details and results in the supplementary material.

3.1 Experimental Setup

Datasets. We evaluate our method on six widely used multi-view benchmarks including Scene15 [10], LandUse21 [23], Reuters [1], CCV20 [19], HandWritten [8], and SUN RGB-D [14]. These datasets cover a wide range of modalities, including video-audio, RGB-depth, image-text, and multilingual text, thus providing a comprehensive testbed for evaluating the robustness of MvC methods against imperfect information.

Table 1: Clustering performance comparisons on five widely-used multi-view datasets under different IIR. The results are the mean of five runs. The best and second-best results are marked as **bold** and underline, respectively. The background colors , , and indicate methods designed for handling incomplete information, incomplete views, and noisy correspondences, respectively.

Setting	Method	Scene15			LandUse21			Reuters			CCV20			HandWritten			Average		
		ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
IIR=0	SURE (TPAMI'23)	40.60	42.46	24.23	26.59	30.78	12.90	49.12	30.74	24.70	19.94	16.82	6.31	66.73	60.15	49.01	40.60	36.19	23.43
	CAMERA (TPAMI'26)	45.98	46.66	28.88	29.32	35.97	15.73	52.80	32.16	26.48	23.30	19.47	8.09	92.17	87.96	85.35	48.71	44.44	32.91
	GCFAgg (CVPR'23)	42.34	41.96	24.46	26.59	30.31	12.79	32.19	15.35	7.79	19.16	15.89	5.75	92.41	86.10	84.14	42.54	37.92	26.99
	DIVIDE (AAAI'24)	45.17	45.74	28.62	32.17	**39.71**	17.92	53.40	**36.64**	27.96	22.89	17.51	7.24	**94.48**	**88.55**	**88.17**	49.62	45.63	33.98
	FreeCSL (CVPR'25)	41.54	44.18	25.02	28.51	34.03	15.29	48.92	32.59	28.14	21.39	17.60	6.86	81.77	83.83	76.78	44.43	42.45	30.42
	CANDY (NeurIPS'24)	41.70	41.10	24.84	30.63	35.95	16.11	48.43	32.57	24.67	20.88	16.54	6.85	90.21	82.28	79.88	46.37	41.69	30.47
	ROLL (CVPR'25)	**47.10**	**48.96**	**30.13**	29.15	34.55	15.16	47.06	23.12	22.47	23.10	19.09	7.84	73.85	56.12	51.93	44.05	36.37	25.51
	CorreGen (ICLR'26)	45.62	45.83	29.09	31.74	38.72	18.05	46.41	29.04	20.74	22.45	19.26	7.95	92.33	86.14	84.05	47.71	43.80	31.98
	PLCI (Ours)	46.55	46.73	29.86	**32.35**	38.40	**18.52**	**56.37**	35.12	**29.22**	**25.25**	**20.85**	**9.09**	93.97	88.40	87.12	**50.90**	**45.90**	**34.76**
IIR=0.2	SURE (TPAMI'23)	38.95	36.07	20.45	24.31	25.61	10.02	40.66	17.44	12.97	18.39	13.43	4.93	67.15	57.83	48.49	37.89	30.08	19.37
	SURE* (TPAMI'23)	39.11	40.82	23.18	26.82	30.49	13.30	44.69	27.49	21.53	20.07	16.73	6.39	73.22	64.91	57.82	40.78	36.09	24.45
	CAMERA (TPAMI'26)	41.83	40.28	23.72	25.95	29.14	11.71	51.82	28.58	25.22	21.28	16.46	6.64	77.37	64.76	61.67	43.65	35.84	25.79
	CAMERA* (TPAMI'26)	42.57	40.99	24.23	25.63	28.40	11.45	50.75	28.63	25.1	21.22	16.52	6.69	79.72	64.61	61.98	43.98	35.83	25.89
	GCFAgg (CVPR'23)	38.02	34.76	19.97	25.00	26.93	11.44	29.79	10.98	4.76	17.76	13.39	4.82	79.78	70.81	66.80	38.07	31.37	21.56
	DIVIDE (AAAI'24)	43.01	40.82	25.90	29.94	34.75	15.84	51.51	30.21	25.31	20.59	13.85	5.60	88.00	75.70	75.36	46.61	39.07	29.60
	FreeCSL (CVPR'25)	38.18	37.43	22.07	27.94	30.58	13.74	44.10	23.39	20.74	19.69	14.86	5.83	73.61	66.65	61.60	40.70	34.58	24.80
	CANDY (NeurIPS'24)	40.11	39.68	24.04	27.74	29.09	13.44	52.47	**33.59**	26.77	18.52	14.33	5.57	88.13	78.28	75.98	45.39	38.99	29.16
	ROLL (CVPR'25)	45.52	44.47	27.51	28.75	32.04	14.25	45.94	20.61	20.00	20.86	15.63	6.23	79.90	64.14	61.12	44.19	35.38	25.82
	CorreGen (ICLR'26)	41.63	43.05	26.08	30.25	35.92	16.70	38.53	21.51	13.24	21.20	17.52	6.93	89.56	80.65	78.66	44.23	39.73	28.32
	PLCI (Ours)	**46.17**	**45.20**	**28.62**	**32.04**	**36.27**	**17.58**	**54.58**	32.22	**28.51**	**24.29**	**19.94**	**8.54**	**90.97**	**81.84**	**81.24**	**49.61**	**43.09**	**32.90**
IIR=0.5	SURE (TPAMI'23)	32.75	31.13	16.53	19.20	18.65	5.98	39.64	18.14	14.80	13.64	9.23	2.86	53.28	41.46	32.05	31.70	23.72	14.44
	SURE* (TPAMI'23)	38.93	39.28	22.97	23.58	26.32	10.52	45.71	27.61	22.09	18.27	15.17	5.52	66.91	58.45	49.94	38.68	33.37	22.21
	CAMERA (TPAMI'26)	33.57	33.02	17.30	20.66	21.10	7.10	45.63	21.39	19.01	17.28	12.50	4.70	68.63	48.16	44.27	37.15	27.23	18.48
	CAMERA* (TPAMI'26)	36.96	36.04	19.46	21.42	22.71	7.84	48.35	24.68	20.85	19.04	13.74	5.41	71.12	51.79	47.91	39.38	29.79	20.29
	GCFAgg (CVPR'23)	32.82	30.17	16.35	21.72	21.74	8.14	32.55	10.65	6.88	16.21	11.04	3.79	58.55	45.34	37.83	32.37	23.79	14.60
	DIVIDE (AAAI'24)	37.51	32.52	19.83	27.14	28.29	12.68	**51.45**	26.37	22.72	15.25	8.98	2.89	77.89	61.58	59.23	41.85	31.55	23.47
	FreeCSL (CVPR'25)	32.03	28.24	15.76	22.73	22.11	8.83	34.42	10.45	9.31	15.54	10.18	3.33	66.99	52.88	48.16	34.34	24.77	17.08
	CANDY (NeurIPS'24)	36.40	36.31	20.94	22.46	22.44	8.85	42.46	22.23	16.53	15.23	10.38	3.49	79.67	64.75	61.02	39.24	31.22	22.17
	ROLL (CVPR'25)	39.85	36.68	21.89	25.45	27.25	11.60	40.39	15.34	14.90	16.61	10.43	3.82	72.30	57.21	50.82	38.92	29.38	20.61
	CorreGen (ICLR'26)	38.31	38.13	21.89	27.27	30.02	12.80	45.04	24.79	18.83	19.11	14.77	5.51	80.50	68.49	64.36	42.05	35.24	24.68
	PLCI (Ours)	**42.17**	**41.39**	**25.00**	**30.09**	**34.58**	**16.13**	49.80	**26.93**	**23.56**	**22.97**	**18.78**	**8.06**	**86.94**	**76.08**	**74.05**	**46.39**	**39.55**	**29.36**
IIR=0.8	SURE (TPAMI'23)	21.81	18.34	8.07	14.72	12.53	3.23	31.41	10.49	8.17	11.83	6.26	1.8	37.54	26.93	17.43	23.47	14.91	7.74
	SURE* (TPAMI'23)	23.92	20.60	9.30	15.46	13.87	3.65	43.02	**23.25**	18.16	14.17	9.81	3.15	41.63	34.42	22.05	27.64	20.39	11.26
	CAMERA (TPAMI'26)	27.44	28.06	12.94	18.27	18.95	5.37	38.13	13.54	12.03	13.70	8.45	2.61	46.99	34.25	24.28	28.91	20.65	11.45
	CAMERA* (TPAMI'26)	22.94	23.42	9.08	15.04	18.93	4.26	**45.15**	22.06	**19.06**	15.93	11.02	3.82	30.17	27.19	12.35	25.85	20.52	9.71
	GCFAgg (CVPR'23)	26.49	23.77	11.39	16.36	15.75	4.48	28.26	6.71	4.45	12.54	8.07	2.31	42.81	36.82	23.52	25.29	18.22	9.23
	DIVIDE (AAAI'24)	29.24	21.83	11.93	20.29	20.73	7.14	28.92	8.18	6.11	11.28	5.16	1.35	51.61	36.64	28.27	28.27	18.51	10.96
	FreeCSL (CVPR'25)	23.04	18.31	8.62	17.73	15.23	4.79	29.46	10.02	6.40	13.48	8.23	2.22	38.86	22.22	15.47	24.51	14.80	7.50
	CANDY (NeurIPS'24)	28.46	28.36	13.82	17.46	18.11	5.64	31.89	11.86	7.57	13.60	8.57	2.51	52.98	37.55	29.48	28.88	20.89	11.80
	ROLL (CVPR'25)	28.65	24.45	12.68	19.38	16.78	6.35	34.33	10.71	9.94	11.14	4.43	1.14	46.21	34.95	21.03	27.94	18.26	10.23
	CorreGen (ICLR'26)	28.12	31.41	14.99	21.15	24.67	8.58	34.73	15.44	10.53	12.46	10.89	3.05	43.32	39.60	25.70	27.96	24.40	12.57
	PLCI (Ours)	**37.78**	**36.71**	**22.21**	**26.66**	**30.68**	**11.68**	40.72	18.56	15.00	**16.47**	**12.39**	**4.35**	**72.54**	**63.07**	**57.13**	**38.83**	**32.28**	**22.07**

Implementation Details. PLCI is a general framework that can enhance existing MvC methods with robustness against imperfect information. In this work, we verify the effectiveness of PLCI on top of the recently proposed DIVIDE [32]. Specifically, to endow the backbone with robustness, we retain the backbone and training pipeline of DIVIDE, and incorporate the proposed instance-level reliability estimation, prototype-level semantic transport, and latent-counterpart-based joint objective. To evaluate robustness against imperfect information, we simulate different imperfect information rate (IIR) settings by randomly introducing incomplete views and noisy correspondences into the original datasets. Specifically, for each instance, we randomly mask its views with a certain probability to simulate incomplete views, and randomly shuffle the correspondences of a certain proportion of complete instances to simulate noisy correspondences.

3.2 Comparison with State of the Arts

We compare our method with 10 state-of-the-art MvC methods, including the IV-oriented methods (GCFAgg [47], DIVIDE [32], and FreeCSL [6]), the NC-oriented methods (CANDY [12], ROLL [40], and CorreGen [52]), and the methods designed for handling incomplete information (SURE [48] and CAMERA [21]). In addition, we report the results of SURE and CAMERA using instances with fully clean correspondences, which serve as quite strong baselines, as the used data does not contain any noise. Following widely used evaluation protocols, we adopt clustering accuracy (ACC), normalized mutual information (NMI), and adjusted Rand index (ARI) as evaluation metrics.

Table 1 summarizes the clustering performance under four IIR settings, where one could observe that the proposed PLCI consistently achieves the best average performance across all settings, demonstrating its superior robustness against imperfect information.

3.3 Experiments on Raw Data Inputs

To further assess whether PLCI can be applied beyond pre-extracted multi-view features, we conduct additional experiments on SUN RGB-D with raw visual inputs. The two views are RGB images and depth maps. We replace the original feature encoder with a ResNet-18 backbone for each view, while keeping the remaining structure of PLCI unchanged. We compare PLCI with DIVIDE and CorreGen, which are the most competitive IV- and NC-oriented baselines in the main experiments.

As shown in Table 2, PLCI remains competitive under the clean setting and achieves the best NMI and ARI. When NC and IV coexist, PLCI consistently outperforms both baselines across all metrics. The advantage becomes more pronounced as the degradation increases. Under the most challenging setting, *i.e.*, IIR=0.8, PLCI improves over the second-best method by 4.19, 3.99, and 2.07 in ACC, NMI, and ARI, respectively. These results indicate that PLCI is not restricted to pre-extracted feature inputs and can also work effectively with raw data inputs.

Table 2: Clustering performance comparisons on SUN RGB-D dataset under different NC and IV settings. The best and second-best results are shown in **bold** and underline, respectively.

Dataset	Metric	IIR=0			IIR=0.2			IIR=0.5			IIR=0.8		
		DIVIDE	CorreGen	Ours	DIVIDE	CorreGen	Ours	DIVIDE	CorreGen	Ours	DIVIDE	CorreGen	Ours
SUN RGB-D	ACC	**35.95**	34.90	34.98	32.00	33.46	**34.06**	26.80	25.13	**28.27**	22.21	23.93	**28.12**
	NMI	34.30	34.53	**35.04**	27.40	28.77	**31.93**	19.91	19.81	**22.68**	18.21	20.56	**24.55**
	ARI	22.14	20.70	**22.21**	17.15	17.93	**20.17**	12.51	10.65	**13.08**	8.92	10.39	**12.46**

3.4 Ablation Studies

To investigate the effectiveness of each component in PLCI, we conduct ablation studies on three datasets, *i.e.*, Scene15, LandUse21, and CCV20, under two IIR settings, *i.e.*, IIR$= 0.5$ and IIR$= 0.8$. More specifically, we compare the following variants of PLCI: i) w/o $\mathcal{L}_{\mathrm{NC}}$: PLCI without the prototype-guided counterpart correction for unreliable correspondences, which directly applies the cross-view contrastive loss on the observed counterparts; ii) w/o $\mathcal{L}_{\mathrm{IV}}$: PLCI without the prototype-guided counterpart substitution for missing views, which does not utilize the inferred latent counterparts for regularization; iii) Full PLCI: the complete version of PLCI with all components used. The results are shown in Table 3, where one could observe that each component serves as an inseparable part of PLCI for achieving robust MvC against imperfect information.

3.5 Study on the Generalizability

To verify the generalizability of the proposed method, we further apply PLCI to another representative MvC method, Completer [25]. For a comprehensive evaluation, we fix the incomplete view rate (IVR) at 0.5 and vary the noisy correspondence rate (NCR) from 0.0 to 0.9 with an interval of 0.1 on Scene15. As shown in Fig. 3, Completer+Ours consistently outperforms Completer across all NCR settings, demonstrating the effectiveness of PLCI in enhancing the robustness of existing MvC methods against imperfect information.

Table 3: Ablation studies of PLCI on datasets Scene15, LandUse21 and CCV20. ✓ denotes that the corresponding module is used.

Setting	$\mathcal{L}_{\mathrm{NC}}$	$\mathcal{L}_{\mathrm{IV}}$	Scene15			LandUse21			CCV20		
			ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
IIR=0.5	✓		32.99	34.71	17.99	24.17	28.11	10.96	19.11	15.27	6.00
		✓	31.57	30.36	15.66	23.75	26.32	10.17	15.47	10.20	3.53
	✓	✓	**42.17**	**41.39**	**25.00**	**30.09**	**34.58**	**16.13**	**22.97**	**18.78**	**8.06**
IIR=0.8	✓		28.80	31.24	15.09	20.38	23.18	7.71	14.31	9.76	3.12
		✓	24.40	21.90	9.97	20.34	23.12	7.80	12.45	6.69	1.93
	✓	✓	**37.78**	**36.71**	**22.21**	**26.66**	**30.68**	**11.68**	**16.47**	**12.39**	**4.35**

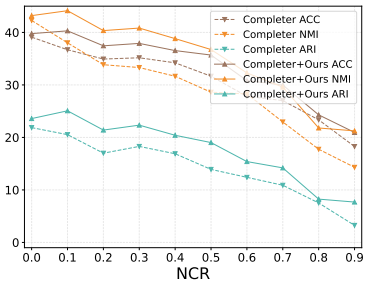


Figure 3: Generalizability study on Scene15.

3.6 Real-World and Structured Imperfect Information

Naturally occurring IV and NC patterns. To assess the effectiveness of PLCI under real-world imperfect information, we evaluate PLCI on BBCSport with a natural IVR of 26.19% and UPMC Food-101 with a natural NCR of 5%. Table 4 compares PLCI with the representative IV-oriented method DIVIDE and NC-oriented method CorreGen, respectively. PLCI improves ACC/NMI/ARI by 9.09/13.34/17.53 percentage points on BBCSport and 2.91/2.15/5.26 points on UPMC Food-101. These improvements support the effectiveness of PLCI under naturally occurring incomplete views and noisy correspondences.

Table 4: Clustering performance with naturally occurring IV and NC on BBCSport and UPMC Food-101, respectively.

Dataset	Method	ACC	NMI	ARI
BBCSport	DIVIDE	75.58	49.11	46.51
	PLCI	**84.67**	**62.45**	**64.04**
UPMC Food-101	CorreGen	67.89	53.75	46.15
	PLCI	**70.80**	**55.90**	**51.41**

Table 5: Clustering performance (%) with two and three views on Scene15 at IIR= 0.5.

Method	Views	ACC	NMI	ARI
CorreGen	2	38.31	38.13	21.89
CorreGen	3	44.42	41.21	25.19
PLCI	2	42.17	41.39	25.00
PLCI	3	**45.23**	**42.01**	**27.41**

Heterogeneous IV and class-conditional NC patterns. We further evaluate PLCI beyond random corruption on Scene15 and LandUse21 under IIR= 0.5. Specifically, *class-conditional* NC introduces mismatches between semantically similar but different classes, such as baseball diamond and tennis court on LandUse21. *Heterogeneous* IV assigns different IVRs to different categories, yielding category-dependent view availability. To examine the individual and combined effects of these two patterns, we vary them separately and jointly and report the average results over five runs. As shown in Table 6, PLCI achieves the best NMI and ARI across all eight settings and the best ACC in seven of them. Moreover, when class-conditional NC and heterogeneous IV coexist, PLCI maintains its advantage across all three metrics, with ACC gains of 1.73 and 1.88 percentage points over the strongest competing baseline on Scene15 and LandUse21, respectively. These results support the robustness of the proposed PLCI beyond random masking and shuffling.

3.7 Effect of Additional Views

At IIR= 0.5, we extend PLCI to three views on Scene15 by applying it to each pair of views and summing the pairwise losses. Table 5 shows ACC/NMI/ARI gains of 3.06/0.62/2.41 points over two-view PLCI and 0.81/0.80/2.22 points over three-view CorreGen, indicating that PLCI can exploit additional complementary information with more views.

Table 6: Clustering performance (%) under random and structured corruption patterns on Scene15 and LandUse21 at IIR$= 0.5$. Class-cond. denotes class-conditional NC, and Heterog. denotes view missingness with class-dependent rates. Results are averaged over five runs. Best results for each setting are in **bold**.

Corruption pattern		DIVIDE			CorreGen			PLCI		
NC	IV	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
Scene15										
Random	Random	37.51	32.52	19.83	38.31	38.13	21.89	**42.17**	**41.39**	**25.00**
Class-cond.	Random	40.56	43.06	25.29	38.57	38.05	21.71	**40.85**	**45.98**	**27.86**
Random	Heterog.	**40.87**	36.80	23.24	40.20	40.22	23.76	40.10	**42.05**	**25.35**
Class-cond.	Heterog.	39.66	42.58	25.40	39.03	40.13	23.00	**41.39**	**45.18**	**27.28**
LandUse21										
Random	Random	27.14	28.29	12.68	27.27	30.02	12.80	**30.09**	**34.58**	**16.13**
Class-cond.	Random	28.37	32.14	14.26	26.30	30.09	12.52	**30.33**	**37.03**	**16.77**
Random	Heterog.	27.70	28.51	12.89	28.51	31.95	13.99	**29.09**	**33.22**	**14.86**
Class-cond.	Heterog.	28.55	32.53	14.52	28.66	31.87	13.75	**30.54**	**37.43**	**16.94**

3.8 Experiments on Downstream Image–Text Retrieval

To assess the utility of the learned representations beyond clustering, we conduct image–text retrieval experiments on Wikipedia [35] under the same imperfect-information setting used in our clustering experiments, namely IIR$= 0.5$. We compare PLCI with DIVIDE, CANDY, and CorreGen, as these methods are the most competitive baselines in our main experiments under the same setting. Following the retrieval evaluation protocol in [17], we rank all gallery samples from the other modality for each query according to their cosine similarity in the learned representation space. Samples sharing the query's category label are treated as relevant. Average precision is computed over the full ranking for each query and then averaged across queries to obtain mean average precision (mAP). Category labels are used only for retrieval evaluation and are not involved in representation learning.

Table 7: Image–text retrieval on Wikipedia at IIR$= 0.5$. Results are mean $\pm$ standard deviation of mAP (%) over five runs. Average denotes the mean of the two retrieval directions. Best results are in **bold**.

Method	Image $\rightarrow$ Text	Text $\rightarrow$ Image	Average
DIVIDE	23.82 ± 0.80	25.86 ± 0.84	24.84 ± 0.76
CANDY	27.77 ± 0.79	26.06 ± 0.76	26.91 ± 0.58
CorreGen	25.82 ± 0.56	25.12 ± 0.86	25.47 ± 0.69
PLCI	$\mathbf{31.70 \pm 1.12}$	$\mathbf{28.54 \pm 1.84}$	$\mathbf{30.12 \pm 1.43}$

As shown in Table 7, PLCI achieves the highest mAP in both retrieval directions. Averaged over five runs, its bidirectional average mAP reaches 30.12%, outperforming the strongest baseline CANDY by 3.21 percentage points. These results indicate that the cross-view representations learned by PLCI also benefit downstream image–text retrieval under imperfect information.

4 Conclusion

In this paper, we study the imperfect information problem in multi-view clustering (MvC), where incomplete views (IV) and noisy correspondences (NC) arise simultaneously in real-world multi-view data. We demonstrate that these two challenges could be unified from the perspective of imperfect cross-view counterpart information, in which the desired counterpart of an anchor instance may be either unavailable or unreliable. Based on this observation, we propose Posterior-guided Latent Counterpart Inference (PLCI), a robust MvC framework that formulates the desired counterpart as a latent variable and infers it by jointly exploiting instance-level reliability and prototype-level semantic transport. Extensive experiments on widely used multi-view benchmarks demonstrate that PLCI consistently outperforms state-of-the-art methods under diverse imperfect information settings. As the multi-modal/view data has an incentive to be contaminated by imperfect information, in the future, we plan to extend the imperfect information problem to more complex scenarios and applications, including but not limited to cross-modal retrieval, multi-modal person re-identification, and multi-view representation learning.

Author Contributions

All authors contributed significantly to this work. Xi Peng and Mouxing Yang conceived the study, designed the PLCI algorithm, revised the manuscript, and supervised the project. Zhichao Huang co-designed and implemented the PLCI algorithm, conducted the baseline evaluations, and drafted the manuscript. Haochen Zhou and Hao Wang analyzed the experimental results and contributed to the formulation of the manuscript. All authors reviewed and approved the final version.

Acknowledgments and Disclosure of Funding

This work was supported in part by Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China under Grant JYB2025XDXM610; in part by NSFC under Grants 62606342, 624B2099 and U24B20174; in part by the Fundamental Research Funds for the Central Universities under Grant CJ202303, 20826041K4049; in part by Sichuan Science and Technology Planning Project under Grant 2024NSFTD0038; and in part by the Postdoctoral Fellowship Program and China Postdoctoral Science Foundation under Grant Number BX2026302.

References

[1] Massih R. Amini, Nicolas Usunier, and Cyril Goutte. Learning from multiple partially observed views: An application to multilingual text categorization. In *Advances in Neural Information Processing Systems*, volume 22, 2009.

[2] Devansh Arpit, Stanisław Jastrzębski, Nicolas Ballas, David Krueger, Emmanuel Bengio, Maxinder S. Kanwal, Tegan Maharaj, Asja Fischer, Aaron Courville, Yoshua Bengio, and Simon Lacoste-Julien. A closer look at memorization in deep networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 233–242. PMLR, 2017.

[3] Man-Sheng Chen, Ling Huang, Chang-Dong Wang, and Dong Huang. Multi-view clustering in latent embedding space. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 3513–3520, 2020.

[4] Man-Sheng Chen, Chang-Dong Wang, Dong Huang, Jian-Huang Lai, and Philip S Yu. Efficient orthogonal multi-view subspace clustering. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pages 127–135, 2022.

[5] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. *arXiv preprint arXiv:2002.05709*, 2020.

[6] 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*, pages 5071–5081, 2025.

[7] Shide Du, Zihan Fang, Yanchao Tan, Changwei Wang, Shiping Wang, and Wenzhong Guo. Openviewer: Openness-aware multi-view learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 39, pages 16389–16397, 2025.

[8] Robert Duin. Multiple features. UCI Machine Learning Repository, 1998.

[9] Zihan Fang, Shide Du, Zhiling Cai, Shiyang Lan, Chunming Wu, Yanchao Tan, and Shiping Wang. Representation learning meets optimization-derived networks: From single-view to multi-view. *IEEE Transactions on Multimedia*, 26:8889–8901, 2024.

[10] Li Fei-Fei and Pietro Perona. A bayesian hierarchical model for learning natural scene categories. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 524–531, 2005.

[11] Lele Fu, Zhaoliang Chen, Yongyong Chen, and Shiping Wang. Unified low-rank tensor learning and spectral embedding for multi-view subspace clustering. *IEEE Transactions on Multimedia*, 25:4972–4985, 2022.

[12] Ruiming Guo, Mouxing Yang, Yijie Lin, Xi Peng, and Peng Hu. Robust contrastive multi-view clustering against dual noisy correspondence. *Advances in Neural Information Processing Systems*, 37, 2024.

[13] Xiaoshuai Hao, Zhihui Zhang, Yingbo Tang, Lingfeng Zhang, Peng Hao, Yunfeng Diao, Guangyin Jin, and Yu Liu. Synergistic prompting for complementarity and consistency in incomplete multi-view clustering. *IEEE Transactions on Image Processing*, 2026.

[14] Changhao He, Di Xue, Shuxian Li, Yanji Hao, Xi Peng, and Peng Hu. Bootstrapping multi-view learning for test-time noisy correspondence. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, June 2026.

[15] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9729–9738, 2020.

[16] Menglei Hu and Songcan Chen. Doubly aligned incomplete multi-view clustering. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 2262–2268. International Joint Conferences on Artificial Intelligence Organization, 2018.

[17] Xin Huang, Yuxin Peng, and Mingkuan Yuan. Mhtn: Modal-adversarial hybrid transfer network for cross-modal retrieval. *IEEE transactions on cybernetics*, 50(3):1047–1059, 2018.

[18] Zhenyu Huang, Joey Tianyi Zhou, Xi Peng, Changqing Zhang, Hongyuan Zhu, and Jiancheng Lv. Multi-view spectral clustering network. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 2563–2569, 2019.

[19] Yu-Gang Jiang, Guangnan Ye, Shih-Fu Chang, Daniel P. W. Ellis, and Alexander C. Loui. Consumer video understanding: A benchmark database and an evaluation of human and machine performance. In *Proceedings of the ACM International Conference on Multimedia Retrieval*, page 29, 2011.

[20] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proceedings of the International Conference on Learning Representations*, 2015.

[21] Haobin Li, Yijie Lin, Peng Hu, Mouxing Yang, and Xi Peng. Community-aware multi-view representation learning with incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026.

[22] Junnan Li, Richard Socher, and Steven C.H. Hoi. Dividemix: Learning with noisy labels as semi-supervised learning. In *International Conference on Learning Representations*, 2020.

[23] Yeqing Li, Feiping Nie, Heng Huang, and Junzhou Huang. Large-scale multi-view spectral clustering via bipartite graph. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2750–2756, 2015.

[24] Ke Liang, Lingyuan Meng, Hao Li, Jun Wang, Long Lan, Miaomiao Li, Xinwang Liu, and Huaimin Wang. From concrete to abstract: Multi-view clustering on relational knowledge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(10):9043–9060, 2025.

[25] 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 (CVPR)*, June 2021.

[26] 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.

[27] Jiyuan Liu, Xinwang Liu, Yuexiang Yang, Li Liu, Siqi Wang, Weixuan Liang, and Jiangyong Shi. One-pass multi-view clustering for large-scale data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12344–12353, 2021.

[28] Jiyuan Liu, Xinwang Liu, Yuexiang Yang, Qing Liao, and Yuanqing Xia. Contrastive multi-view kernel learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8): 9552–9566, 2023.

[29] Suyuan Liu, Ke Liang, Zhibin Dong, Siwei Wang, Xihong Yang, Sihang Zhou, En Zhu, and Xinwang Liu. Learn from view correlation: An anchor enhancement strategy for multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26151–26161, 2024.

[30] Xinwang Liu, Miaomiao Li, Chang Tang, Jingyuan Xia, Jian Xiong, Li Liu, Marius Kloft, and En Zhu. Efficient and effective regularized incomplete multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(8):2634–2646, 2021.

[31] Xinwang Liu, Li Liu, Qing Liao, Siwei Wang, Yi Zhang, Wenxuan Tu, Chang Tang, Jiyuan Liu, and En Zhu. One pass late fusion multi-view clustering. In *Proceedings of the International Conference on Machine Learning*, volume 139, pages 6850–6859, 2021.

[32] Yiding Lu, Yijie Lin, Mouxing Yang, Dezhong Peng, Peng Hu, and Xi Peng. Decoupled contrastive multi-view clustering with high-order random walks. *Thirty-Eighth AAAI Conference on Artificial Intelligence*, 2024.

[33] Yalan Qin, Chuan Qin, Xinpeng Zhang, Donglian Qi, and Guorui Feng. Nim-nets: Noise-aware incomplete multi-view learning networks. *IEEE Transactions on Image Processing*, 32:175–189, 2022.

[34] Yalan Qin, Chuan Qin, Xinpeng Zhang, and Guorui Feng. Dual consensus anchor learning for fast multi-view clustering. *IEEE Transactions on Image Processing*, 33:5298–5311, 2024.

[35] Nikhil Rasiwasia, Jose Costa Pereira, Emanuele Coviello, Gabriel Doyle, Gert RG Lanckriet, Roger Levy, and Nuno Vasconcelos. A new approach to cross-modal multimedia retrieval. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 251–260, 2010.

[36] Yazhou Ren, JunLong Ke, Zichen Wen, Tianyi Wu, Yang Yang, Xiaorong Pu, and Lifang He. Multi-view graph clustering via node-guided contrastive encoding.

[37] Yazhou Ren, Jingyu Pu, Chenhang Cui, Yan Zheng, Xinyue Chen, Xiaorong Pu, and Lifang He. Dynamic weighted graph fusion for deep multi-view clustering. IJCAI, 2024.

[38] Yazhou Ren, Zichen Wen, Junlong Ke, Chenhang Cui, Yonghao Huang, Xinyue Chen, Philip S Yu, and Lifang He. Enhancing multi-view clustering: A sufficient information-theoretic approach for consistency acquisition and redundancy elimination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026.

[39] Yuan Sun, Yang Qin, Yongxiang Li, Dezhong Peng, Xi Peng, and Peng Hu. Robust multi-view clustering with noisy correspondence. *IEEE Transactions on Knowledge and Data Engineering*, 36(12):9150–9162, 2024.

[40] Yuan Sun, Yongxiang Li, Zhenwen Ren, Guiduo Duan, Dezhong Peng, and Peng Hu. Roll: Robust noisy pseudo-label learning for multi-view clustering with noisy correspondence. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 30732–30741, 2025.

[41] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[42] Siwei Wang, Xinwang Liu, En Zhu, Chang Tang, Jiyuan Liu, Jingtao Hu, Jingyuan Xia, and Jianping Yin. Multi-view clustering via late fusion alignment maximization. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 3778–3784, 2019.

[43] Siwei Wang, Xinwang Liu, Li Liu, Wenxuan Tu, Xinzhong Zhu, Jiyuan Liu, Sihang Zhou, and En Zhu. Highly-efficient incomplete large-scale multi-view clustering with consensus bipartite graph. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9776–9785, 2022.

[44] Jie Wen, Yong Xu, and Hong Liu. Incomplete multiview spectral clustering with adaptive graph learning. *IEEE transactions on cybernetics*, 50(4):1418–1429, 2018.

[45] Jie Wen, Zheng Zhang, Zhao Zhang, Zhihao Wu, Lunke Fei, Yong Xu, and Bob Zhang. Dimc-net: Deep incomplete multi-view clustering network. In *Proceedings of the 28th ACM international conference on multimedia*, pages 3753–3761, 2020.

[46] 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.

[47] Weiqing Yan, Yuanyang Zhang, Chenlei Lv, Chang Tang, Guanghui Yue, Liang Liao, and Weisi Lin. Gcfagg: Global and cross-view feature aggregation for multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19863–19872, June 2023.

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

[49] Xihong Yang, Jiaqi Jin, Siwei Wang, Ke Liang, Yue Liu, Yi Wen, Suyuan Liu, Sihang Zhou, Xinwang Liu, and En Zhu. DealMVC: Dual contrastive calibration for multi-view clustering. In *Proceedings of the ACM International Conference on Multimedia*, pages 337–346, 2023.

[50] Xihong Yang, Siwei Wang, Jiaqi Jin, Fangdi Wang, Tianrui Liu, Yueming Jin, Xinwang Liu, En Zhu, and Kunlun He. Generalized deep multi-view clustering via causal learning with partially aligned cross-view correspondence. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1990–1999, 2025.

[51] Hanning Yuan, Zhihui Zhang, Qi Guo, Lianhua Chi, Sijie Ruan, Wei Zhou, Jinhui Pang, and Xiaoshuai Hao. Dwcl: Dual-weighted contrastive learning for robust multi-view clustering. *Engineering Applications of Artificial Intelligence*, 165:113532, 2026.

[52] Haochen Zhou, Guofeng Ding, Mouxing Yang, Peng Hu, Yijie Lin, and Xi Peng. Uncover underlying correspondence for robust multi-view clustering. In *The Fourteenth International Conference on Learning Representations*, 2026.

A Related Work

In this section, we briefly review two most related topics to this work, *i.e.*, multi-view clustering with incomplete views and multi-view clustering with noisy correspondence.

A.1 Multi-View Clustering with Incomplete Views

According to whether the missing views are explicitly recovered, existing IV-oriented MvC methods [3, 4, 11, 16, 25, 32, 33, 46] can be roughly categorized into two groups, *i.e.*, recovery-free and recovery-based methods. In brief, recovery-free methods aim to learn a common representation from the available views and then perform clustering in the shared representation space. For example, DAIMC [16] aligns view-specific matrices and latent representations via weighted non-negative matrix factorization to learn a consensus representation from incomplete data. APADC [46] performs adaptive feature projection and distribution alignment to reduce the cross-view discrepancy among available observations. Different from them, recovery-based methods explicitly infer the unavailable views. COMPLETER [25] recovers missing representations through cross-view contrastive prediction, while DCP [26] further unifies consistency learning and missing-view recovery into a dual contrastive prediction framework. DIVIDE [32] employs cross-view decoders to recover missing samples and performs decoupled contrastive learning to preserve both cross-view consistency and view-specific information. Although these methods have achieved promising performance under incomplete views, they generally rely on reliable cross-view correspondences to learn cross-view recovery. Once the observed correspondences are noisy, the learned recovery could be seriously deviated from the desired one, thus leading to erroneous recovery.

Different from existing IV-oriented methods [13, 38] that mainly handle missing views under reliable cross-view correspondences, this study addresses a more practical problem where incomplete views and noisy correspondences coexist. Rather than reconstructing missing views or explicitly re-aligning correspondences, our method models the desired counterpart of each instance as a latent variable and infers its posterior distribution by jointly exploiting instance-level reliability and prototype-level semantic transport. This enables unavailable and unreliable counterparts to be handled in a unified probabilistic framework.

A.2 Multi-View Clustering with Noisy Correspondence

Besides instance completeness, another key assumption in conventional MvC is view consistency, which treats cross-view counterparts as correctly aligned. Different from partially view-aligned problem (PVP), where a subset of aligned instances is known during training and can serve as reliable anchors [21, 48], NC-oriented MvC assumes that the reliability of observed counterparts is unknown. Existing methods mainly improve clustering by suppressing the adverse effect of unreliable pairs during representation learning [39, 40], handling false positive and false negative correspondences in contrastive supervision [12], or estimating soft correspondence distributions under a generative framework [52].

However, these methods mainly focus on unreliable observed counterparts while assuming that all views are available. As a result, instances with unavailable views are usually excluded from their formulation, and the information from unavailable views cannot be exploited during training. Therefore, existing NC-oriented MvC methods mainly target unreliable counterparts, but are not designed for the setting considered in this work, where unavailable views and unreliable counterparts coexist.

B More Experiment Details

B.1 Details of Used Datasets

- **Scene15** [10] contains 4,485 images from 15 indoor and outdoor scene categories. Following [48], we use the PHOG and GIST features as two views.
- **LandUse21** [23] consists of 2,100 satellite images from 21 categories. Following [26], we use the PHOG and LBP features as two views.
- **Reuters** [1] is a multilingual news dataset with 18,758 samples from 6 categories. Similar to [18], the English and French texts are used as two views.
- **CCV20** [19] contains 6,773 web videos from 20 semantic categories. Here we use two commonly adopted views, *i.e.*, Scale-Invariant Feature Transform (SIFT) visual features and Mel-Frequency Cepstral Coefficients (MFCC) audio features.
- **HandWritten** [8] contains 2,000 handwritten digit samples from 10 categories, corresponding to digits 0 through 9. We use Pixel features and Fourier coefficients as two views.

B.2 More Implementation Details

We implement PLCI in PyTorch and optimize it using Adam [20] optimizer with a learning rate of 0.002. The model is trained for 200 epochs with a batch size of 1024, of which 20 epochs are used for warmup. All experiments are conducted on Ubuntu 20.04 with NVIDIA 3090 GPUs. The temperature parameter τ_p in Eq. 6 is set to 0.07 across all experiments. For unavailable counterparts, the transported prototype posterior is sparsified by retaining k_{top} entries before KL regularization. We set $\alpha = 0.2$ and $k_{\text{top}} = 3$ across all experiments.

C Hyperparameter Sensitivity Analysis

We investigate the sensitivity of PLCI to α and k_{top} on Scene15 and LandUse21 under IIR$= 0.5$. The coefficient α controls the weight of the incomplete-view objective $\mathcal{L}_{\text{IV}}$ relative to $\mathcal{L}_{\text{NC}}$, while k_{top} specifies the number of retained entries in the transported prototype assignment before KL regularization. We vary one parameter at a time while fixing the other at its default value, namely $\alpha = 0.2$ or $k_{\text{top}} = 3$.

Sensitivity to α. We vary α from 0.1 to 0.9 in increments of 0.1 while fixing $k_{\text{top}} = 3$. As shown in Table 8, PLCI maintains relatively stable performance over the tested values in $[0.1, 0.5]$. For larger values, performance generally declines on both datasets, suggesting that excessive emphasis on $\mathcal{L}_{\text{IV}}$ can impair the balance between the two objectives. Although the best value varies across datasets and metrics, $\alpha = 0.2$ provides competitive performance and is used as the common default.

Table 8: Sensitivity to α at IIR$= 0.5$ with $k_{\text{top}} = 3$. Clustering results are reported in %. The default setting is shaded, and the best result in each column is in **bold**.

α	Scene15			LandUse21		
	ACC	NMI	ARI	ACC	NMI	ARI
0.1	**42.76**	**41.57**	**25.22**	30.35	34.67	**16.33**
0.2	42.17	41.39	25.00	30.09	34.58	16.13
0.3	41.90	41.01	24.65	30.49	35.22	16.31
0.4	42.03	41.31	24.82	30.48	35.37	16.32
0.5	40.99	41.10	24.32	**30.63**	**35.58**	**16.33**
0.6	41.35	40.88	24.45	29.61	34.64	15.79
0.7	40.39	40.35	23.52	29.85	34.82	15.97
0.8	40.26	40.30	23.48	29.58	34.00	15.45
0.9	40.16	40.44	23.53	28.91	33.71	14.79

Sensitivity to k_{top}. With $\alpha = 0.2$ fixed, we evaluate $k_{\text{top}} \in \{1, 3, 5, 7\}$. Table 9 shows that retaining three or more prototype candidates outperforms retaining only one across all three metrics on both datasets. Moreover, performance varies moderately over $k_{\text{top}} \in \{3, 5, 7\}$, indicating that PLCI is relatively insensitive to the precise number of retained candidates within this range. These results suggest that retaining multiple high-confidence candidates helps preserve useful semantic information. We use $k_{\text{top}} = 3$ as the common default, which remains competitive despite not being optimal for every dataset and metric.

Table 9: Sensitivity to k_{top} at IIR$= 0.5$ with $\alpha = 0.2$. Clustering results are reported in %. The default setting is shaded, and the best result in each column is in **bold**.

k_{top}	Scene15			LandUse21		
	ACC	NMI	ARI	ACC	NMI	ARI
1	41.73	40.97	24.47	29.11	32.89	15.32
3	42.17	41.39	25.00	30.09	34.58	**16.13**
5	**42.69**	**41.60**	**25.14**	**30.28**	**34.85**	16.05
7	42.48	41.45	24.99	30.09	34.58	16.07

D Performance under Only IV or NC

Although our proposed method PLCI targets the joint IV+NC setting, we also evaluate it under incomplete view only and noisy correspondence only scenarios. As shown in Tables 10 and 11, PLCI remains competitive when only one challenge exists. Under NC-only settings, PLCI achieves the best average ACC, NMI, and ARI at NCR$= 0.2$ and NCR$= 0.5$, and still obtains the best average ACC and second-best average NMI and ARI at NCR$= 0.8$. Under IV-only settings, PLCI consistently achieves the best average performance across all incomplete view rates. These results show that reliability-aware counterpart rectification and prototype-guided counterpart substitution are both effective in their corresponding single-degradation scenarios, while their unified design further enables PLCI to handle the more practical IV+NC problem.

E Limitations of Two-Stage Combination of IV- and NC-Oriented Methods

A straightforward solution to imperfect information is to combine an IV-oriented method and an NC-oriented method. To examine this strategy, we sequentially combine DIVIDE and CorreGen in two orders. D+C means that it first uses DIVIDE to recover the incomplete views and then

Table 10: Clustering performance under NC-only settings with IVR fixed to 0. The best and second-best results are shown in **bold** and underline, respectively.

Setting	Method	Scene15			LandUse21			Reuters			CCV20			HandWritten			Average		
		ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
NCR=0.2	SURE	36.94	36.67	20.30	26.31	27.56	11.81	36.82	18.88	18.81	19.28	14.19	5.49	62.83	55.84	44.45	36.44	30.63	20.17
	CANDY	40.14	39.75	23.62	30.70	34.71	15.47	50.15	**32.95**	25.68	20.49	16.51	6.93	88.48	79.64	77.02	45.99	40.71	29.74
	ROLL	**45.87**	**45.35**	**28.46**	28.85	32.60	14.62	46.35	22.04	21.48	21.74	16.14	6.65	80.68	66.35	62.67	44.70	36.50	26.78
	CorreGen	42.94	42.94	26.41	**31.87**	**38.04**	**17.64**	40.36	24.74	16.58	21.81	18.43	7.34	92.31	**86.2**	84.17	45.86	42.07	30.43
	CAMERA	42.58	40.30	23.84	26.49	28.72	12.08	49.56	27.32	23.44	20.34	15.83	6.26	79.55	64.89	62.61	43.70	35.41	25.65
	Ours	44.92	44.92	28.10	30.99	35.13	16.52	**53.02**	32.44	**26.69**	**24.46**	**19.64**	**8.40**	**92.58**	84.86	**84.27**	**49.19**	**43.40**	**32.80**
NCR=0.5	SURE	36.04	35.92	20.04	23.42	23.49	9.07	38.34	16.70	13.20	15.50	11.10	3.74	61.32	52.66	41.58	34.92	27.97	17.53
	CANDY	40.54	39.04	23.62	29.56	32.93	14.84	43.79	24.91	19.04	18.24	14.30	5.49	83.06	74.29	69.57	43.04	37.09	26.51
	ROLL	41.27	36.32	22.62	28.75	32.63	14.42	45.72	20.78	20.77	18.33	10.93	4.22	82.57	69.63	65.90	43.33	34.06	25.59
	CorreGen	**42.37**	41.10	**24.94**	**31.99**	**37.95**	**18.17**	40.22	21.80	14.51	21.78	18.64	7.43	88.86	**83.67**	80.26	45.04	40.63	29.06
	CAMERA	34.40	34.88	18.33	20.53	30.25	6.89	43.86	18.36	15.64	17.93	13.07	4.84	63.94	40.73	37.11	36.13	27.46	16.56
	Ours	42.27	**41.26**	24.39	31.40	35.74	17.01	**49.70**	**27.84**	**23.35**	**23.64**	**19.37**	**8.02**	**91.13**	82.42	**81.47**	**47.63**	**41.33**	**30.85**
NCR=0.8	SURE	26.40	28.61	13.64	18.59	18.15	6.01	36.31	16.29	13.40	12.41	9.56	2.79	60.50	51.46	40.00	30.84	24.81	15.17
	CANDY	39.24	36.54	21.50	27.61	30.52	13.22	39.45	15.93	13.71	17.56	12.46	4.39	53.98	48.05	34.46	35.57	28.70	17.46
	ROLL	31.23	21.24	12.58	26.93	29.23	12.92	38.28	16.12	14.94	11.57	3.90	0.40	76.92	69.78	62.31	36.99	28.05	20.63
	CorreGen	39.42	**38.64**	22.50	**31.71**	**37.50**	**17.67**	35.72	16.97	9.76	21.03	**17.88**	6.99	**85.42**	**77.89**	**73.68**	42.66	**37.78**	**26.12**
	CAMERA	34.62	33.93	18.13	15.32	13.73	3.56	34.67	11.59	9.49	14.97	10.27	3.47	33.29	12.79	9.20	26.57	16.46	8.77
	Ours	**40.25**	38.36	**22.59**	29.58	33.38	15.43	**43.34**	**21.60**	**17.16**	**21.82**	17.06	**7.19**	79.03	69.26	64.15	**42.80**	35.93	25.30

Table 11: Clustering performance under IV-only settings with NCR fixed to 0. The best and second-best results are shown in **bold** and underline, respectively.

Setting	Method	Scene15			LandUse21			Reuters			CCV20			HandWritten			Average		
		ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
IVR=0.2	SURE	41.16	42.51	24.79	26.13	30.00	12.48	45.65	27.65	21.16	19.45	16.52	6.16	69.01	61.33	52.83	40.28	35.60	23.48
	GCFAgg	39.73	40.46	22.57	25.70	29.05	12.12	31.69	14.68	5.76	18.31	15.30	5.41	86.65	76.61	72.74	40.42	35.22	23.72
	DIVIDE	**46.39**	**46.27**	**29.43**	30.82	**37.51**	16.93	53.73	**36.70**	27.59	23.02	17.60	7.29	92.73	85.55	84.77	49.34	44.73	33.20
	FreeCSL	40.25	43.27	24.81	28.77	34.07	15.32	49.52	30.89	27.53	21.13	17.36	6.89	85.64	82.68	77.56	45.06	41.65	30.42
	CAMERA	45.98	45.73	27.91	28.18	34.18	14.53	54.92	34.66	29.86	22.50	19.07	7.69	89.37	83.83	80.39	48.19	43.49	32.08
	Ours	45.17	46.00	29.21	**32.13**	37.26	**18.11**	**58.36**	36.02	**31.83**	**25.37**	**21.11**	**9.22**	**93.72**	**87.88**	**86.67**	**50.95**	**45.65**	**35.01**
IVR=0.5	SURE	39.45	41.32	23.32	25.19	28.45	11.88	46.31	27.87	22.90	19.64	16.24	6.07	68.07	60.53	51.65	39.73	34.88	23.16
	GCFAgg	36.72	38.49	20.96	24.23	26.51	10.64	34.46	16.46	6.74	18.00	14.83	5.22	77.86	70.24	62.58	38.25	33.31	21.23
	DIVIDE	44.08	**44.57**	**27.59**	30.33	35.54	16.25	49.77	**35.93**	26.30	22.15	17.27	6.98	90.28	81.88	80.24	47.32	43.04	31.47
	FreeCSL	39.39	40.86	23.06	27.76	32.21	14.13	47.20	29.03	24.77	20.01	16.56	6.33	77.12	78.08	80.62	42.30	39.35	29.78
	CAMERA	**44.85**	43.73	26.42	26.81	31.85	13.22	**53.43**	32.84	28.35	21.98	18.17	7.46	87.52	79.51	76.07	46.92	41.22	30.30
	Ours	42.50	44.55	27.13	**30.64**	**37.39**	**17.22**	53.21	33.01	**28.94**	**24.97**	**21.33**	**9.16**	**92.99**	**86.75**	**85.21**	**48.86**	**44.61**	**33.53**
IVR=0.8	SURE	38.85	38.05	21.94	22.03	23.28	9.13	48.71	31.21	25.33	19.15	15.14	5.73	61.89	54.40	44.28	38.13	32.42	21.28
	GCFAgg	38.68	37.46	21.98	18.20	18.41	5.95	35.01	19.29	8.47	16.13	12.29	4.11	59.99	54.35	38.64	33.60	28.36	15.83
	DIVIDE	**43.12**	**42.88**	**26.49**	29.1	32.65	14.88	**53.37**	**34.82**	**30.52**	21.01	16.22	6.06	84.4	73.42	70.07	46.20	40.00	29.60
	FreeCSL	38.08	38.06	21.89	26.12	28.46	12.19	46.94	28.48	24.73	20.31	15.87	6.26	74.00	79.56	62.31	41.09	38.09	25.48
	CAMERA	40.33	40.42	22.68	22.97	27.49	9.74	51.79	32.44	27.13	19.51	16.18	6.35	83.22	71.59	67.73	43.56	37.62	26.73
	Ours	41.26	42.86	26.17	**30.62**	**37.80**	**16.54**	48.56	28.38	24.95	**24.03**	**20.57**	**8.31**	**91.58**	**84.39**	**82.41**	**47.21**	**42.80**	**31.68**

applies CorreGen to re-align the recovered multi-view data for clustering. Conversely, C+D first uses CorreGen to correct noisy correspondences and then applies DIVIDE to recover the remaining missing views. The results are reported under two settings, *i.e.*, IIR=0.5 and IIR=0.8.

As shown in Table 12, sequentially combining IV- and NC-oriented methods does not provide a reliable solution to imperfect information. Under IIR=0.5, C+D generally outperforms D+C, suggesting that correcting noisy correspondences before incomplete view recovery is less harmful in the moderate setting. However, C+D is only competitive on Scene15 and remains clearly inferior to PLCI on LandUse21, CCV20, and HandWritten. Under the more severe setting, *i.e.*, IIR=0.8, both two-stage combinations degrade substantially, and their relative performance becomes order-dependent. In contrast, PLCI achieves the best results on most datasets and metrics. These results indicate that incomplete view recovery and correspondence re-alignment are tightly coupled, and simply cascading existing IV- and NC-oriented methods cannot effectively handle unavailable and unreliable counterparts simultaneously.

The reason is that the two subproblems are tightly coupled. View recovery requires reliable cross-view correspondences, while correspondence correction requires sufficiently complete instances. In a two-stage pipeline, errors produced in the first stage are propagated to the second stage, leading to

suboptimal clustering. In contrast, PLCI models unavailable and unreliable counterparts in a unified latent counterpart inference framework, which avoids the error accumulation of sequential recovery and correction.

Table 12: Comparison of two-stage combinations and PLCI under joint IV and NC. C+D denotes applying DIVIDE after CorreGen, while D+C denotes applying CorreGen after DIVIDE.

Setting	Method	Scene15			LandUse21			CCV20			HandWritten		
		ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
IIR=0.5	C+D	41.58	41.34	**25.05**	29.51	33.12	14.80	19.33	14.98	5.87	85.66	73.48	71.57
	D+C	38.63	35.14	20.57	26.50	27.78	12.29	16.39	9.54	3.45	76.71	59.86	57.09
	Ours	**42.17**	**41.39**	25.00	**30.09**	**34.58**	**16.13**	**22.97**	**18.78**	**8.06**	**86.94**	**76.08**	**74.05**
IIR=0.8	C+D	30.47	22.22	12.23	20.34	18.99	6.54	11.37	5.11	1.43	49.03	31.04	23.89
	D+C	26.71	31.16	14.58	21.33	23.31	8.26	16.29	10.78	3.69	45.55	41.48	27.54
	Ours	**37.78**	**36.71**	**22.21**	**26.66**	**30.68**	**11.68**	**16.47**	**12.39**	**4.35**	**72.54**	**63.07**	**57.13**

F Performance under Different IVR and NCR

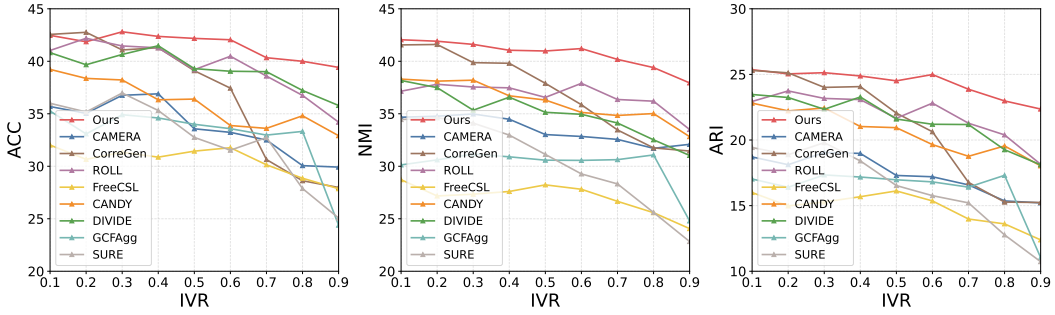


Figure 4: Clustering performance under different IVR with a fixed NCR=0.5 on Scene15.

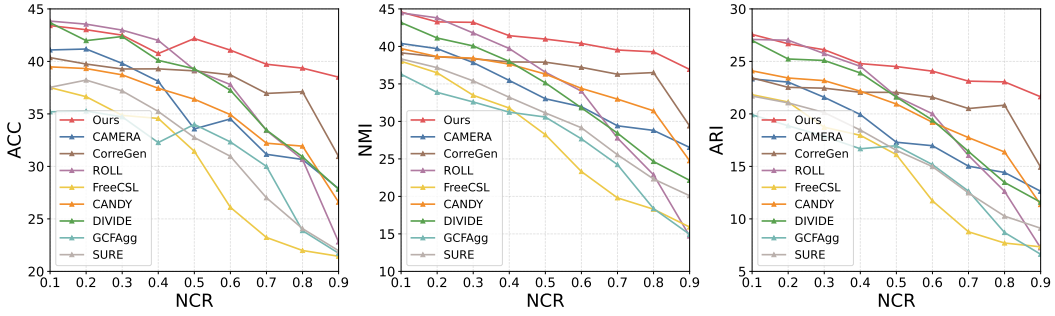


Figure 5: Clustering performance under different NCR with a fixed IVR=0.5 on Scene15.

To further verify the robustness of PLCI, we conduct additional experiments by varying one degradation rate from 0.1 to 0.9 with an interval of 0.1 while fixing the other at 0.5. Specifically, we first vary IVR while fixing NCR at 0.5, and then vary NCR while fixing IVR at 0.5. These experiments provide a more fine-grained robustness analysis beyond the paired NCR/IVR settings reported in Table 1.

Fig. 4 reports the results on Scene15 under different IVR settings with NCR fixed at 0.5. PLCI maintains highly competitive performance across almost all IVR settings in terms of ACC, NMI, and ARI, whereas most baselines exhibit clear performance degradation as IVR increases. This indicates that PLCI is less sensitive to the reduction of available cross-view counterparts, benefiting from its prototype-level semantic transport for unavailable counterparts. Fig. 5 shows the results on

Scene15 under different NCR settings with IVR fixed at 0.5. When NCR is small, PLCI already performs competitively with the strongest baselines. As NCR increases, PLCI becomes consistently superior, while most competing methods degrade sharply. This demonstrates that PLCI can effectively mitigate the negative effect of unreliable counterparts by adaptively reducing the reliance on corrupted observed counterparts.

To further examine the generality of this robustness, we provide additional results on LandUse21, Reuters, CCV20, and HandWritten under the same settings. The results are shown in Fig. 7, Fig. 8, Fig. 9, and Fig. 10. We observe that PLCI is consistently more robust than baselines across different datasets and degradation settings. Moreover, the performance gap between PLCI and competing methods generally widens as the degradation rate increases, indicating that PLCI is particularly effective in handling severe imperfect information.

Overall, these experiments verify the robustness of PLCI under both increasing IVR and increasing NCR. The stable performance further demonstrates the effectiveness of inferring latent counterparts for both unavailable and unreliable cross-view counterparts.

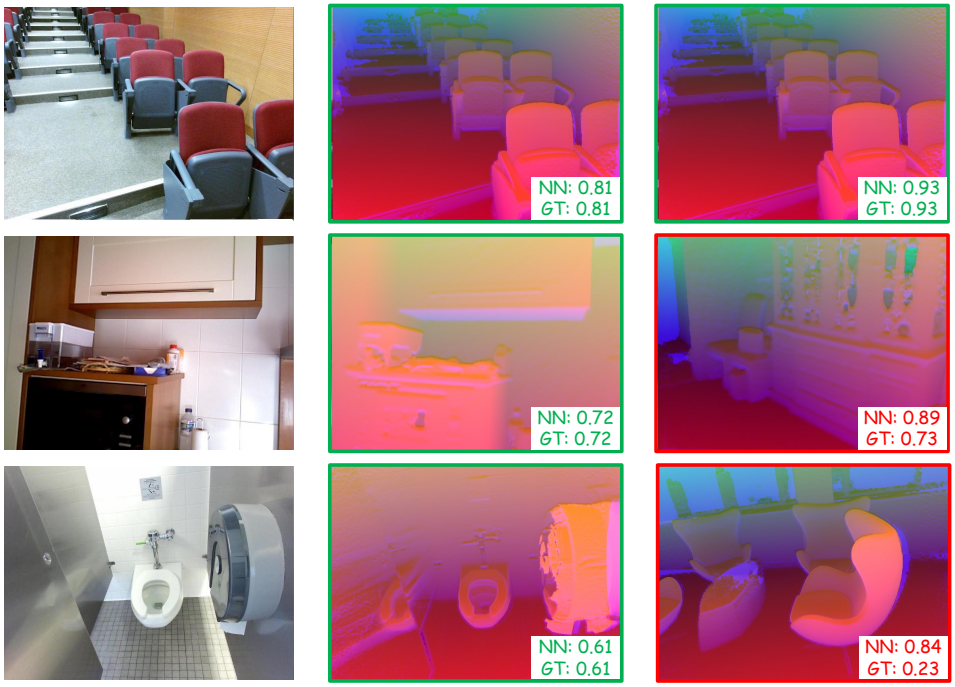
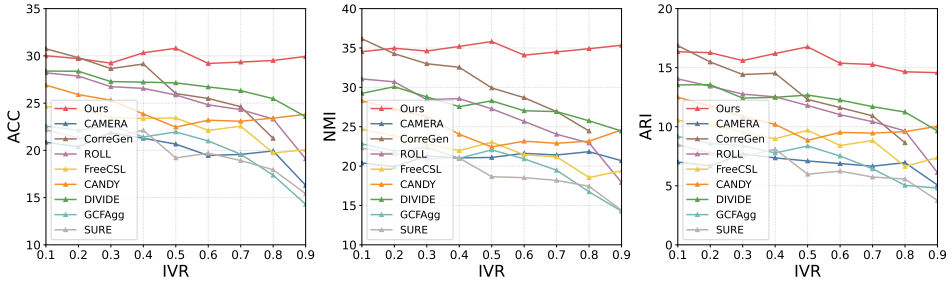


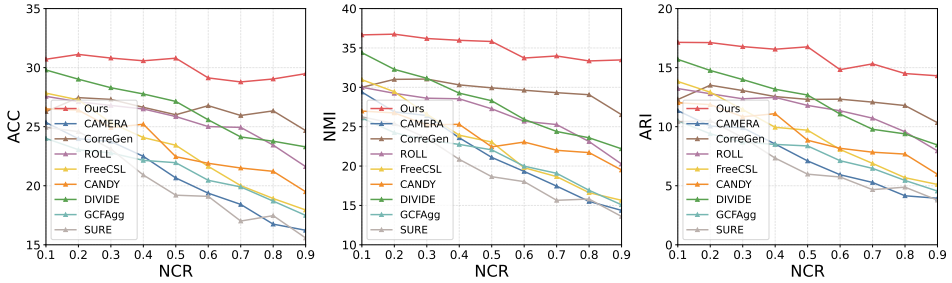
Figure 6: Case study on SUN RGB-D under IIR$= 0.5$. The first column shows the observed RGB views. The second and third columns show the depth maps whose features have the highest cosine similarity to the recovered depth-view features produced by PLCI and DIVIDE, respectively. From top to bottom, the ground-truth categories are *lecture theatre*, *dining room*, and *bathroom*. NN denotes the cosine similarity between the recovered depth-view feature and its nearest-neighbor depth feature, while GT denotes the cosine similarity between the recovered depth-view feature and the ground-truth depth counterpart.

G Case study

We conduct a case study on SUN RGB-D under IIR$= 0.5$. Given an observed RGB image, each method recovers the missing depth-view feature. We then retrieve the depth map whose feature is most similar to the recovered feature. As shown in Fig. 6, PLCI retrieves the ground-truth depth counterpart in all cases. By contrast, DIVIDE succeeds only in the first case, retrieving a same-category but incorrect depth map for *dining room* and even a wrong-category depth map for *bathroom*. These results show that PLCI recovers more reliable missing-view counterparts under imperfect information.

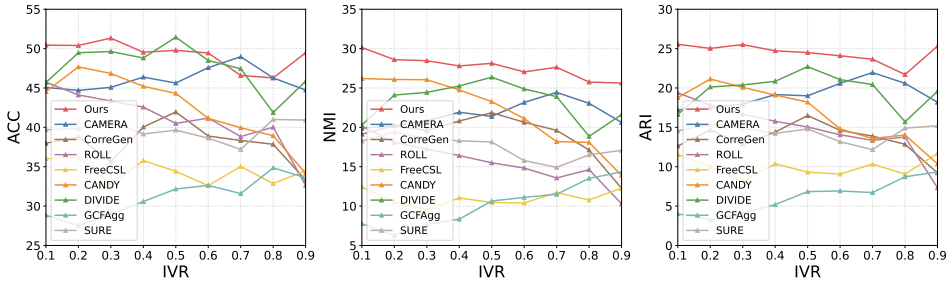


(a) Different IVR with NCR fixed at 0.5.

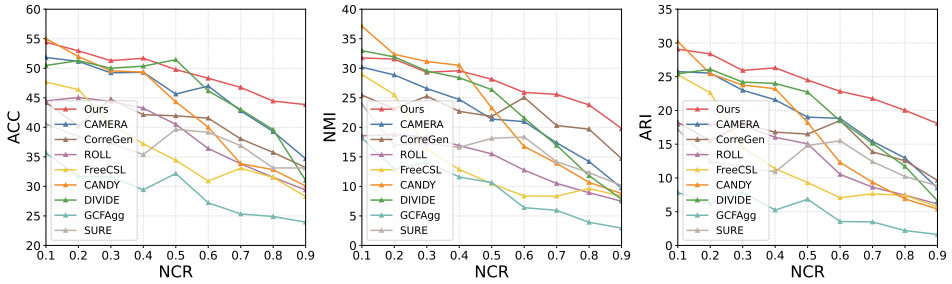


(b) Different NCR with IVR fixed at 0.5.

Figure 7: Additional robustness analysis on LandUse21 under different IVR and NCR settings.

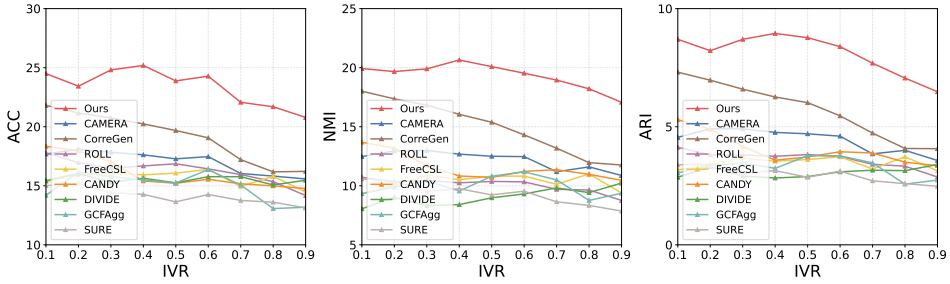


(a) Different IVR with NCR fixed at 0.5.

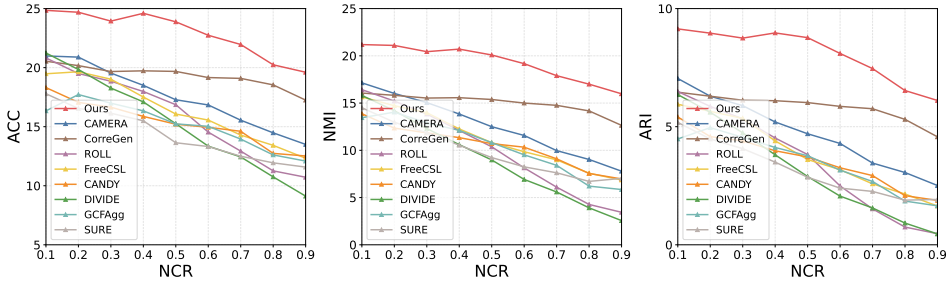


(b) Different NCR with IVR fixed at 0.5.

Figure 8: Additional robustness analysis on Reuters under different IVR and NCR settings.

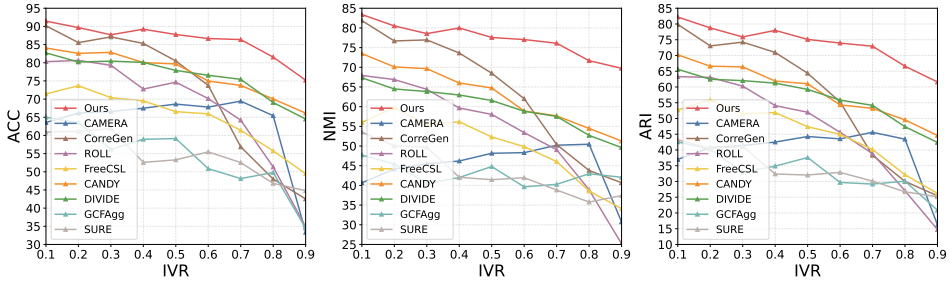


(a) Different IVR with NCR fixed at 0.5.

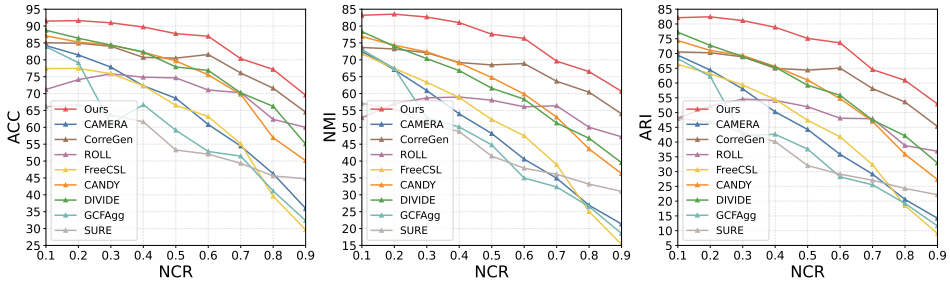


(b) Different NCR with IVR fixed at 0.5.

Figure 9: Additional robustness analysis on CCV20 under different IVR and NCR settings.



(a) Different IVR with NCR fixed at 0.5.

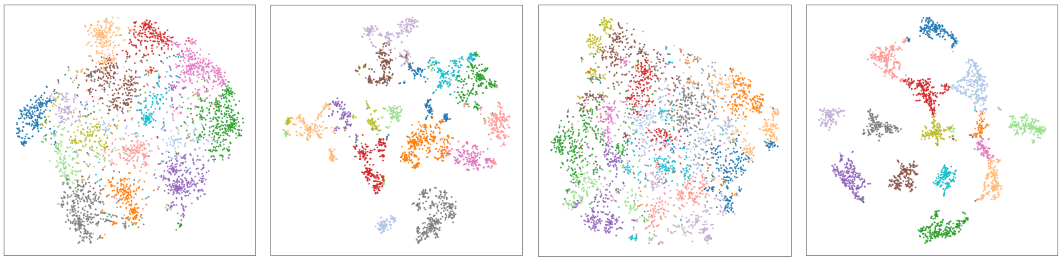


(b) Different NCR with IVR fixed at 0.5.

Figure 10: Additional robustness analysis on HandWritten under different IVR and NCR settings.

H Visualization

Fig. 11 shows the t-SNE visualization on Scene15. PLCI yields more compact and better separated clusters than the compared methods. This indicates that posterior-guided latent counterpart inference helps reduce the adverse influence of unavailable and unreliable counterparts, leading to more discriminative representations under imperfect information.



(a) DIVIDE (b) CANDY (c) CAMERA (d) PLCI

Figure 11: The t-SNE visualization on Scene15 under IIR$= 0.5$.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction accurately state the problem setting, scope, and main contribution of this work, *i.e.*, addressing the coexistence of incomplete views and noisy correspondences in MvC within a unified framework. These claims are supported by the proposed method and the empirical results reported in the paper.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The limitations are discussed in the Conclusion.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [N/A]

Justification: This paper is experimental instead of theoretical. It includes no theorems or proofs.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: This paper provides a detailed algorithm procedure and implementation details for the reproducibility of the proposed method.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The data and code of the proposed method will be uploaded in a zip file along with the supplementary material. All baseline methods and datasets evaluated in this paper are publicly available, and the main experimental results can be faithfully reproduced under the same experimental settings as the proposed method.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

Answer: [Yes]

Justification: This paper provides the details of data preprocessing steps, hyperparameters and optimization in Experiments section.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: Almost all compared baselines do not include the statistical significance in experiments and so we do not report it.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The experiments computer resources are stated in the Experiments section

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This paper conforms with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: The proposed algorithm has limited direct societal impact. All datasets used in this paper are publicly available and do not contain any personally identifiable information.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.

- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: All datasets used in this paper are publicly available.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Original papers and datasets are properly cited. The license and terms of use are explicitly mentioned and properly respected.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: This paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This work does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: This work does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The core method does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.