QUALITY-TO-RELATION LEARNING FOR MULTI-VIEW CLUSTERING UNDER HETEROGENEOUS NOISE

Anonymous authors
Paper under double-blind review

ABSTRACT

Quality-aware multi-view clustering (QAMVC) has recently been explored to address heterogeneous observation reliability by estimating data quality and using it to regulate representation learning and multi-view fusion. However, existing QAMVC methods often provide coarse-grained quality estimates based on limited quality cues, and mainly use them to adjust feature contributions or fusion weights, leaving the relational influence of unreliable observations insufficiently regulated. Moreover, weak relations may be suppressed or discarded even when they receive partial support from other views. To address these issues, we propose Cross-Granularity Quality-to-Relation Learning (CGQRL), a unified framework that establishes a path from sample-view quality to relational supervision. CGQRL first estimates dynamic sample-view quality by jointly considering reconstruction fidelity, cross-view consistency, and cluster compatibility. The resulting quality scores are used to adaptively weight view-specific representations for consensus fusion and, more importantly, transferred to the relational level to characterize the reliability of sample neighborhood relations and cross-view assignment relations. Based on these quality-aware relations, we further develop a reliability-preserving structural alignment strategy that provides soft structural supervision: reliable relations receive stronger supervision, partially supported relations are down-weighted rather than discarded, and relations unsupported across views are treated as explicit negatives. By jointly modeling observation quality and relational influence, CGQRL suppresses unreliable structural information while preserving potentially useful weak relations. Extensive experiments on multiple multi-view clustering benchmarks demonstrate the effectiveness of CGQRL under heterogeneous noise.

1 INTRODUCTION

In recent years, multi-view data have emerged in various applications, providing complementary information about the same objects from different perspectives Wang et al. (2013); Yan et al. (2021); Zhao et al. (2017). Multi-view clustering (MVC) aims to discover the latent cluster structure of data by integrating complementary and consistent information across different views Xu et al. (2013); Kong et al. (2024); Zheng & Tang (2024), which has attracted widespread attention. However, most existing MVC methods do not explicitly account for the reliability of individual sample-view observations during representation learning and multi-view fusion Xu et al. (2021; 2022); Trosten et al. (2021), making them less capable of handling the heterogeneous reliability that naturally arises across views and samples under heterogeneous noise. This limitation becomes particularly pronounced under heterogeneous noise, where different sample-view observations can be affected to different degrees and thus exhibit substantially different reliability Wu et al. (2026). Without explicitly modeling such reliability differences, unreliable observations may exert disproportionate influence on the learned consensus representation, while the complementary information from reliable observations may not be fully exploitedDong et al. (2026); Zhao et al. (2026).

Quality-aware multi-view clustering (QAMVC) characterizes the reliability of multi-view data and uses quality estimates to regulate representation learning or view fusion. Existing methods mainly differ in the granularity at which quality is assessed. At the view level, Shapley-value-based methods (Wang et al., 2024) quantify the marginal contribution of individual views, while other approaches measure view consistency with the global consensus representation using maximum mean discrepancy (Cheng et al., 2027) or assess structural validity through clusterability-based

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

criteria (Jiang et al., 2026). Recent studies have moved toward finer-grained quality assessment. At the sample level, reconstruction-based methods estimate sample reliability to reweight contrastive pairs (Dong et al., 2026), while uncertainty refinement is used to reduce the influence of unreliable sample pairs in cross-view learning (Hu et al., 2025). More recently, sample-view-level quality has been explicitly modeled through information-bottleneck reconstruction discrepancy, enabling adaptive quality-aware contrastive learning and consensus fusion (Wu et al., 2026).

Despite the promising performance of existing QAMVC methods, three limitations remain. **First**, quality estimation is coarse and incomplete. Most existing methods focus on view-level quality while overlooking quality variations across samples within the same view. Moreover, they typically rely on single or limited quality cues, such as reconstruction error, consistency, or uncertainty. These cues assess reliability from isolated perspectives and may not adequately capture the actual contribution of a sample-view observation to the current representation and clustering structure. **Second**, quality-aware weighting does not adequately regulate relational influence. Existing methods mainly use estimated quality to adjust the feature contributions or fusion weights of views or samples, which cannot prevent unreliable observations from affecting clustering through neighborhood relations or cross-view assignment transfers induced by their representations. **Third**, weak relations are often treated too aggressively. Existing methods may regard weak or uncertain relations as unreliable and suppress or discard them indiscriminately. However, a weak relation in one view may still receive partial support from other views and thus provide useful structural information. Instead of making a hard keep-or-remove decision, a more flexible strategy should preserve such relations with appropriately reduced influence according to their aggregated support and cross-view agreement.

To address these limitations, we propose Cross-Granularity Quality-to-Relation Learning (CGQRL), which establishes a unified path from observation quality to relational supervision for multi-view clustering. As illustrated in Figure 1, CGQRL first estimates the quality of each sample-view observation by jointly considering reconstruction fidelity, cross-view consistency, and cluster compatibility, yielding a dynamic quality score that evolves with the learned representations and cluster structure. The quality scores are then used in two complementary ways: they adaptively weight view-specific representations for consensus fusion and, more importantly, are transferred to the relational level to characterize the reliability of sample neighborhood relations and cross-view assignment relations. Finally, these quality-aware relations provide soft structural supervision for the unified representation, where reliable relations receive stronger supervision, partially supported relations are down-weighted rather than discarded, and pairs unsupported by any view are treated as explicit negatives. In this way, CGQRL jointly models observation quality and its relational influence, while preserving potentially valuable information carried by weak relations.The main contributions of this work are summarized as follows:

- We develop a dynamic sample-view quality estimator that progressively combines reconstruction fidelity, cross-view consistency, and cluster compatibility. The resulting quality scores are progressively refined during training and provide a fine-grained characterization of observation reliability across both samples and views.
- We propose a cross-granularity quality-to-relation learning mechanism that transfers sample-view quality to two types of relational structures, namely sample neighborhood relations and cross-view assignment relations. This enables observation reliability to regulate relational influence beyond conventional feature-level or fusion-level quality weighting.
- We introduce a reliability-preserving structural alignment strategy that uses quality-aware relations as soft structural supervision. Reliable relations receive stronger supervision, while weak or partially supported relations are down-weighted rather than discarded, and pairs unsupported by all views are explicitly treated as negatives. This allows CGQRL to preserve potentially useful relational information while suppressing unreliable structural influence.

2 Related Work

2.1 Deep Multi-View Clustering

Deep multi-view clustering (DMVC) methods leverage end-to-end representation learning to capture complex cross-view structures, evolving from reconstruction-based representation learning toward semantic and structural alignment across views. Early approaches primarily relied on reconstruction

to learn view-specific representations and shared clustering structures. For example, a collaborative reconstruction framework learns view-specific representations while optimizing a shared clustering structure Ke et al. (2024). From the perspective of semantic alignment, positive and negative pairs promote cross-view instance-level alignment, improving both representation consistency and discriminability Yang et al. (2023). Contrastive learning further extends to multi-level correspondence by incorporating graph structural information, modeling semantic consistency jointly at the instance and cluster levels Pan & Kang (2021). From the perspective of structural coordination, constructing view-specific topological structures and coordinating their shared neighborhood patterns preserves cross-view structural consistency while exploiting complementary neighborhood information Dong et al. (2023). Furthermore, maximizing the mutual information between view-specific and shared representations enhances shared semantics while preserving informative view-specific variations Zhang et al. (2023).

2.2 Quality-aware Multi-View Clustering

Quality-aware multi-view clustering incorporates reliability and informativeness at different levels to guide the learning process. At the view and feature levels, a mutual-calibration mechanism reduces the propagation of unreliable information across views Xiao et al. (2026), while composite attention emphasizes informative features by modeling intra-view dependencies and inter-view interactions Du et al. (2024). Quality-aware teacher-student guidance further combines with adaptive fusion to exploit informative views while suppressing noisy information Zhao et al. (2026). At the sample level, estimating sample confidence from the entropy of multi-view assignments and progressively introducing reliable samples reduces the impact of uncertain pseudo-labels Li et al. (2026a). Measuring sample quality from cross-view inconsistency adapts contrastive learning and representation fusion accordingly Chen et al. (2026). Deriving quality weights from cross-view structural discrepancies further favors more reliable representations Li et al. (2026b). For noisy cross-view correspondences, estimating pairwise reliability and replacing unreliable counterparts with semantically matched alternatives handles the corruption Liu et al. (2026).

3 Method

The overall architecture of CGQRL is illustrated in Figure 1. CGQRL first estimates the reliability of each sample-view observation from reconstruction fidelity, cross-view agreement, and cluster compatibility, and then propagates this quality to two relation granularities. At the sample level, observation quality and cross-view structural agreement determine relation reliability for consensus geometry learning. At the cross-view level, relation reliability regulates the strength of structural transfer through RelScale. These quality-aware relations are jointly optimized with the base objective to learn a unified representation.

3.1 Preliminaries

Let $\mathcal{X} = \{X^{(v)}\}_{v=1}^{V}$ be a multi-view dataset containing N unlabeled samples from V views. The v-th view is denoted by $X^{(v)} = [x_1^{(v)}, \ldots, x_N^{(v)}]^\top \in \mathbb{R}^{N \times d_v}$, where d_v is the feature dimension of view v. The goal of multi-view clustering is to partition the N samples into K clusters based on the complementary information provided by multiple views, without access to ground-truth labels. To this end, we learn a unified representation $H \in \mathbb{R}^{N \times d}$. During training, neither the locations nor the magnitudes of the noise are assumed to be known. The core notation is summarized in Appendix Table 4.

3.2 Dynamic Observation Quality Learning

Under heterogeneous Gaussian noise, different sample–view observations may be corrupted to substantially different degrees, making their contributions to representation learning inherently unequal. A reliable quality measure should therefore reflect not only whether an observation can be well reconstructed, but also whether it agrees with the information from other views and the emerging cluster structure. Motivated by this, CGQRL evaluates each sample–view observation from three complementary aspects: reconstruction fidelity, cross-view agreement, and cluster compatibility.

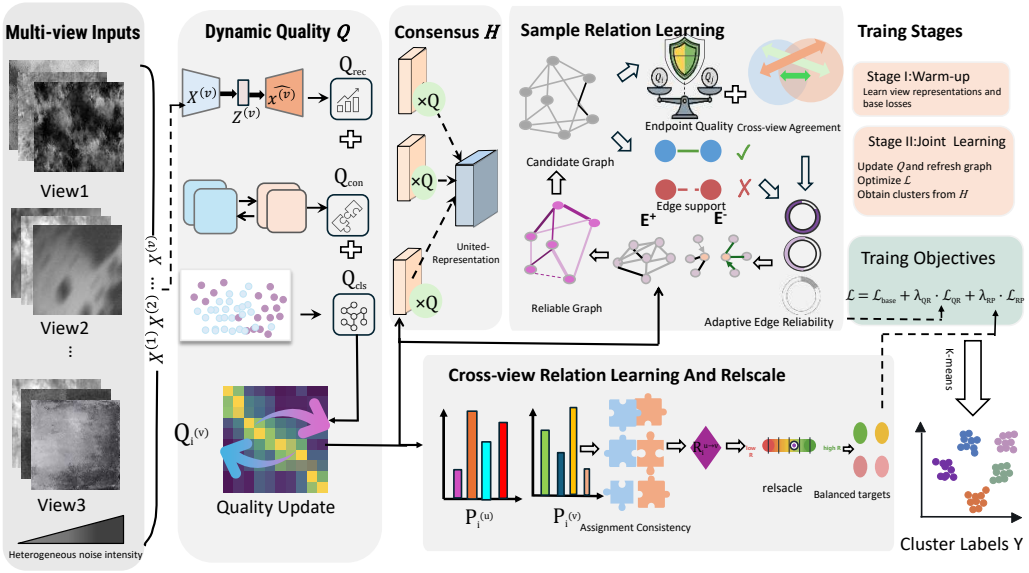


Figure 1: Overview of the proposed cross-granularity quality-to-relation learning framework.

Specifically, reconstruction fidelity measures how well an observation can be explained by its compact representation, providing a direct signal of observation-level corruption. Cross-view agreement further examines whether the observation is consistent with the information available from the other views, while cluster compatibility evaluates whether it supports the current cluster structure. These cues capture different aspects of reliability and are therefore complementary. Their detailed formulations are provided in Appendix A.2.

As the representation and cluster structure evolve during training, the relative informativeness of these cues may also change. We thus dynamically combine them after warm-up to obtain the instantaneous quality estimate

$$\widetilde{Q}_{i,t}^{(v)} = \mathcal{S}_\nu\left(a_t Q_{\mathrm{rec},i}^{(v)} + b_t Q_{\mathrm{con},i}^{(v)} + c_t Q_{\mathrm{cls},i}^{(v)}\right), \qquad a_t + b_t + c_t = 1, \tag{1}$$

where a_t, b_t, and c_t determine the contributions of the three quality cues at training step t. To suppress fluctuations caused by the evolving representation, we further smooth the estimate over iterations:

$$Q_{i,t}^{(v)} = \xi Q_{i,t-1}^{(v)} + (1-\xi)\widetilde{Q}_{i,t}^{(v)}. \tag{2}$$

The resulting $Q_{i,t}^{(v)}$ provides a dynamic estimate of the reliability of each sample–view observation, which is subsequently propagated to different relation granularities for structural learning.

3.3 Reliability-Aware Sample Relation Learning

The key challenge is to transfer observation-level quality to relations between samples: a reliable relation should be supported by the current neighborhood structure, reflect the reliability of its endpoint observations, and agree across views. CGQRL therefore separates relation construction from relation reliability: the former determines which sample pairs are retained, while the latter determines how strongly they influence the consensus geometry.

At relation-update step t, we construct a candidate graph from the current view-specific representations. Let $\mathcal{N}_{i,t}^{(v)}$ denote the k_n nearest neighbours of sample i in view v under cosine similarity. We retain reciprocal neighbours and combine the candidates across views:

$$\mathcal{U}_{i,t} = \bigcup_{v=1}^{V} \left\{ j \mid j \in \mathcal{N}_{i,t}^{(v)},\ i \in \mathcal{N}_{j,t}^{(v)} \right\}. \tag{3}$$

The candidate graph is initialized after warm-up and refreshed periodically throughout joint optimization. At each refresh step, we include an edge whenever its two endpoints form a reciprocal k_n-nearest-neighbour pair in at least one of the current view-specific representations, and we prune it once this reciprocal-neighbour condition no longer holds in any view. Thus, the evolving neighborhood structure determines which sample pairs are eligible for structural learning.

For each candidate pair (i, j), we first bound the edge reliability by the qualities of its two endpoint observations, such that an edge cannot receive high trust when either endpoint is substantially corrupted. We then fuse this endpoint-level quality with view-specific structural support to weight each view's contribution, while comparing the resulting evidence across views so that reliable and mutually corroborated structures are favored. Next, we aggregate the view-wise evidence into an edge-support estimate, whose residual disagreement with individual views we use to calibrate the final edge confidence. We defer the detailed definitions of these intermediate quantities to Appendix A.2. Rather than discarding weakly supported relations outright, we soften them through calibrated confidence, which preserves partial evidence without allowing noise to dominate the resulting structure.

The aggregated relation strength and cross-view agreement are combined into the final sample-relation reliability:

$$R^{\mathrm{s}}_{ij} = r_0 + (1 - r_0) C_{ij} \widehat{p}_{ij}, \qquad R^{\mathrm{s}}_{ij} \in [r_0, 1], \tag{4}$$

where $\widehat{p}_{ij}$ denotes the aggregated structural support and C_{ij} the residual agreement among views. A larger R^{s}_{ij} indicates stronger structural support and greater consistency with the reliability of the underlying observations; the floor r_0 prevents uncertain but potentially informative relations from being completely discarded.

Whenever the candidate graph is refreshed, the relation reliability is re-estimated using the latest observation quality and multi-view structural evidence. Hence, graph updating determines whether a sample pair participates in structural learning, while $R^{\mathrm{s}}_{ij,t}$ determines its relative influence.

Let $\mathcal{E}^+_t$ contain the currently supported reciprocal relations and $\mathcal{E}^-_t$ contain pairs unsupported by every view. The quality-aware structural objective is

$$\begin{aligned} \mathcal{L}_{\mathrm{QR}} = & \frac{\sum_{(i,j)\in\mathcal{E}^+} \mathrm{stopgrad}(R^{\mathrm{s}}_{ij}) \left(1 - \mathrm{sim}(h_i, h_j)\right)^2}{\sum_{(i,j)\in\mathcal{E}^+} \mathrm{stopgrad}(R^{\mathrm{s}}_{ij}) + \varepsilon} \\ & + \lambda_- \frac{\sum_{(i,j)\in\mathcal{E}^-} \mathrm{stopgrad}(R^{-}_{ij}) \left[\mathrm{sim}(h_i, h_j) - m\right]^2_+}{\sum_{(i,j)\in\mathcal{E}^-} \mathrm{stopgrad}(R^{-}_{ij}) + \varepsilon}. \end{aligned} \tag{5}$$

The positive term pulls supported pairs together according to their reliability, while the negative term separates pairs unsupported by all views. Consequently, the periodically refreshed graph determines which relations are learned, and the relation reliability determines how strongly each retained relation shapes the consensus geometry.

3.4 Cross-View Relation Reliability and RelScale

While sample-relation reliability weights how strongly two samples should agree in consensus geometry, cross-view relation reliability decides whether the cluster assignment of one view can reliably supervise another view for the same sample. For sample i in view v, we first compute its prototype-based assignment:

$$P^{(v)}_i(k) = \frac{\exp\Big(\mathrm{sim}(z^{(v)}_i, c_k)/\tau_p\Big)}{\sum_{r=1}^{K} \exp\Big(\mathrm{sim}(z^{(v)}_i, c_r)/\tau_p\Big)}. \tag{6}$$

For a directed transfer from source view u to target view v, we define the instance-specific cross-view reliability as

$$R^{u\to v}_i = \sqrt{Q^{(u)}_i Q^{(v)}_i} \left[1 - \frac{\mathrm{JS}\Big(P^{(u)}_i, P^{(v)}_i\Big)}{\log 2}\right]. \tag{7}$$

The geometric mean requires both endpoint observations to be reliable, while the Jensen–Shannon term measures the compatibility of their cluster assignments; a transfer thus receives high reliability only when both views are trustworthy and structurally consistent.

Directly using the soft assignments $P^{(u)}$ as supervision can yield imbalanced targets dominated by a few prototypes, so we apply the Sinkhorn–Knopp algorithm (Sinkhorn, 1967) to obtain approximately balanced targets $T^{(v)}$, alleviating prototype domination and assignment collapse. For the directed transfer $u \to v$, the balanced assignment of the source view is used as the target:

$$\ell_i^{u\to v} = -\sum_{k=1}^{K} \mathrm{stopgrad}\Big(T_i^{(u)}(k)\Big) \log P_i^{(v)}(k). \tag{8}$$

A conventional reliability-normalized objective would divide the weighted transfer losses by the sum of their reliability values. Although this preserves the relative importance of different transfers, it removes the absolute reliability scale: under heterogeneous noise, a batch of mostly unreliable relations could still produce a transfer loss of similar magnitude after normalization, amplifying uncertain supervision. We instead preserve the reliability scale directly through RelScale:

$$\mathcal{L}_{\mathrm{RP}} = \frac{1}{BV(V-1)} \sum_{u\neq v} \sum_{i\in B} \mathrm{stopgrad}(R_i^{u\to v})\, \ell_i^{u\to v}. \tag{9}$$

The denominator is fixed to the total number of directed sample-level transfers rather than the sum of their reliabilities, so each transfer contributes in proportion to its reliability and a global decrease in cross-view reliability directly reduces the total supervision. CGQRL thus automatically attenuates cross-view guidance when heterogeneous corruption makes the assignments collectively unreliable, while allowing reliable cross-view structure to exert stronger supervision when available.

3.5 Overall Objective and Optimization

The preceding components establish a unified path from observation quality to clustering structure. The observation quality provides the reliability evidence from which sample-relation reliability R^s and cross-view reliability $R_i^{u\to v}$ are derived. The former regulates the geometry of the consensus space, whereas the latter controls assignment transfer across views. These components are jointly optimized through

$$\mathcal{L} = \mathcal{L}_{\mathrm{base}} + \lambda_{\mathrm{QR}}\mathcal{L}_{\mathrm{QR}} + \lambda_{\mathrm{RP}}\mathcal{L}_{\mathrm{RP}}, \tag{10}$$

where $\mathcal{L}_{\mathrm{base}}$ learns the view-specific and consensus representations, $\mathcal{L}_{\mathrm{QR}}$ introduces reliable sample relations into the consensus geometry, and $\mathcal{L}_{\mathrm{RP}}$ preserves reliable cross-view structural supervision through RelScale. The complete formulation of $\mathcal{L}_{\mathrm{base}}$ is provided in Appendix A.2.

4 Experiments

4.1 Experimental Settings

Datasets. Six widely used multi-view datasets, including Hdigit, COIL20, HW2sources, Mfeat, MSRCV16view, and MNIST-4, are employed in our experiments. Details of these datasets are summarized in Table 1

Table 1: Summary of six datasets.

Dataset	Samples	Views	Clusters
MSRCV16view	210	6	7
COIL20	1440	4	20
HW2sources	2000	2	10
Mfeat	2000	6	10
MNIST-4	4000	3	4
Hdigit	10000	2	10

Heterogeneous-noise protocol. The main protocol uses heterogeneous Gaussian corruption with $\rho \in \{0.1, 0.3, 0.5, 0.7\}$, where ρ denotes the proportion of samples selected for corruption within each view and controls the prevalence, rather than the severity, of corruption. Severity is controlled separately by the sample-view-specific intensity $\alpha_i^{(v)} \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$: for a selected observation, $\widetilde{x}_i^{(v)} = (1-\alpha_i^{(v)})x_i^{(v)} + \alpha_i^{(v)}\epsilon_i^{(v)}$ with $\epsilon_i^{(v)} \sim \mathcal{N}(0, I)$. Unlike protocols that apply one noise level to an entire view or use a binary clean/corrupted state, this protocol samples the corruption status and intensity independently for each sample-view observation.For each dataset and corruption

ratio, we generate five fixed noisy sets using seeds $\{0, 1, 2, 3, 4\}$. Seed s determines the corrupted sample-view pairs, their intensities, and the Gaussian draws, and is paired with training seed s. The resulting inputs are stored once and shared by all methods. We report the mean over the five paired runs; seed-level results and standard deviations are retained in the supplementary files. Additional results under partial correspondence mismatch and its combination with heterogeneous Gaussian corruption are reported in Appendix C.

Comparison methods. Seven representative MVC methods are used for comparison: EAMC Zhou & Shen (2020), DCP Lin et al. (2023), DSMVC Tang & Liu (2022), SiMVC and CoMVC Trosten et al. (2021), DCMVC Cui et al. (2024), and GLGC He et al. (2026).

Evaluation metrics. Clustering performance is evaluated using four standard metrics: clustering accuracy (ACC), normalized mutual information (NMI), adjusted Rand index (ARI), and F1 score. Higher values indicate better clustering performance.

Implementation details. All views are standardized independently. CGQRL is trained with Adam for 150 epochs, including 50 warm-up epochs. The learning rate and batch size are 10^{-3} and 256, respectively.

4.2 Comparison Results

Tables 2 reports the performance over $\rho \in \{0.1, 0.3, 0.5, 0.7\}$. The best result is shown in bold, and the second-best is underlined. Overall, CGQRL achieves superior or competitive performance in noisy multi-view clustering across most dataset-metric pairs. As the noise ratio increases from 10% to 70%, baseline methods degrade significantly, while CGQRL remains stable, validating the effectiveness of the proposed quality-to-relation framework in suppressing noisy observations and preserving clustering structures. On HW2sources, CGQRL consistently achieves optimal results under all noise settings, with ACC/NMI/ARI/F1 scores declining from 94.65/88.47/88.54/94.66 ($\rho = 0.1$) to 81.55/65.37/63.62/81.56 ($\rho = 0.7$). At the highest noise level, it outperforms the strongest baseline by 10.68/8.89/12.12/11.95 points. Similar improvements are observed on MSRCV16view, especially under moderate and high noise. The performance advantage of CGQRL becomes more prominent with increasing noise, as mild corruption yields barely perceptible sample-view quality discrepancies whereas severe noise sharpens these differences and enables our model to identify unreliable observations and protect clustering structures against contamination from noisy relations.

4.3 Ablation Studies

We evaluate five variants to examine the contributions of dynamic quality estimation, sample relations, cross-view relations, and RelScale under the same training protocol. As shown in Table 3, dynamic quality estimation provides the foundation for reliable relation learning. Without this component, the ARI on HW2sources drops by 22%, indicating that accurately characterizing sample-view quality is crucial under heterogeneous noise. Building on the estimated quality, the sample relation branch further incorporates quality information into local structural modeling, whose absence reduces the ARI from 63.62% to 58.45% on HW2sources. At the cross-view level, the corresponding relation branch yields a 4.98-point ARI improvement on MSRCV16view, demonstrating the benefit of coordinating structural information across views. Finally, RelScale regulates the strength of such cross-view interactions according to their reliability; disabling it decreases the ARI from 69.98% to 66.37%. Overall, the results suggest a progressive role of the four components, from estimating reliability to modeling local relations, coordinating cross-view structures, and controlling relation strength, with their combination providing complementary benefits for robust multi-view clustering.

4.4 Convergence Analysis

Figure 3 shows the ACC, NMI, and ARI curves of CGQRL on HW2sources, where curves and shaded regions denote the mean and standard deviation over five runs. The clustering metrics gradually stabilize as training proceeds, while the transition around epoch 50 corresponds to the activation of quality and relation learning after the warm-up stage.The convergence curves further show that CGQRL optimizes stably across independent runs and reaches a consistent performance plateau

Table 2: Clustering performance (%) under four heterogeneous Gaussian corruption ratios.

Dataset	Method	$\rho = 0.1$				$\rho = 0.3$				$\rho = 0.5$				$\rho = 0.7$			
		ACC	NMI	ARI	F1	ACC	NMI	ARI	F1	ACC	NMI	ARI	F1	ACC	NMI	ARI	F1
Hdigit	EAMC	51.53	55.46	39.04	46.37	39.37	32.87	21.44	34.24	35.29	24.22	15.31	31.65	29.09	15.25	9.18	25.34
	DCP	75.65	85.69	69.19	71.14	74.54	79.83	67.55	71.73	64.27	69.00	56.37	60.62	59.41	60.32	49.22	55.44
	DSMVC	69.82	65.19	55.32	69.90	59.62	51.22	41.67	60.02	47.97	38.25	28.11	48.44	42.00	30.42	21.03	42.21
	SiMVC	52.36	43.34	33.11	52.25	48.83	36.86	27.42	48.65	36.54	25.00	16.20	36.84	32.32	19.87	12.51	32.47
	CoMVC	75.28	71.57	63.23	75.43	63.16	56.98	47.65	63.40	68.53	56.90	49.73	68.82	48.10	35.57	27.22	48.46
	DCMVC	91.56	90.96	88.23	90.55	83.87	80.79	76.21	82.64	75.05	67.42	60.53	73.73	64.89	55.31	45.47	62.38
	GLGC	**97.41**	**92.93**	**94.33**	**97.41**	92.08	81.25	83.18	92.10	82.50	64.38	64.89	82.56	69.35	46.09	43.73	69.52
	Ours	96.92	92.38	93.33	96.92	**94.04**	**86.14**	**87.31**	**94.04**	**89.74**	**77.97**	**78.81**	**89.75**	**82.88**	**67.10**	**66.12**	**82.93**
COIL20	EAMC	46.42	51.92	33.70	42.20	38.53	37.59	22.45	35.19	26.75	25.28	11.42	25.08	23.57	20.84	8.92	22.13
	DCP	48.26	70.27	35.52	44.21	42.31	65.97	30.51	37.58	39.22	60.51	29.64	32.46	34.46	54.90	25.75	27.43
	DSMVC	67.57	79.44	62.08	66.02	68.72	77.37	60.23	67.41	61.99	69.87	50.27	60.99	54.40	63.33	42.59	52.96
	SiMVC	64.11	75.59	56.60	63.04	63.71	74.08	54.93	63.01	60.83	68.76	49.59	59.97	52.06	59.94	39.20	51.36
	CoMVC	65.39	76.50	57.31	64.34	61.56	73.13	53.26	60.16	59.83	68.97	49.13	59.01	52.26	60.73	39.59	51.70
	DCMVC	**74.74**	**82.85**	**68.69**	**74.38**	75.07	**82.94**	**69.06**	74.71	73.51	**81.98**	**67.05**	72.64	73.96	**81.10**	**67.03**	73.31
	GLGC	57.33	70.21	47.21	57.45	59.25	71.23	49.39	59.75	60.62	71.29	50.63	60.73	63.35	70.92	52.88	63.25
	Ours	74.33	81.83	67.77	73.82	**75.17**	82.32	68.04	**74.90**	**73.85**	81.00	66.04	**73.55**	**73.96**	81.00	66.44	**73.49**
HW2sources	EAMC	38.24	30.50	21.61	35.05	33.73	21.67	14.98	30.83	28.24	14.21	9.61	25.79	25.84	11.00	7.40	23.49
	DCP	66.67	79.72	60.32	59.20	57.22	68.01	49.52	48.47	50.97	53.43	35.77	44.37	29.29	28.14	13.91	22.14
	DSMVC	54.48	50.23	37.97	53.34	50.74	43.62	31.53	49.73	44.55	31.83	23.02	43.82	36.82	23.22	15.60	35.61
	SiMVC	57.12	50.02	38.38	56.95	49.74	41.53	30.73	49.50	39.47	28.68	19.02	39.52	33.63	20.68	13.14	33.54
	CoMVC	58.44	53.68	41.68	57.77	50.06	42.91	32.11	49.48	42.17	30.69	20.90	42.37	34.64	21.81	13.65	34.61
	DCMVC	90.28	83.99	81.91	89.94	84.91	76.19	73.70	84.46	77.14	66.40	62.52	75.67	70.87	56.48	51.50	69.61
	GLGC	93.83	86.04	86.78	93.84	86.99	73.49	73.28	87.03	74.07	55.54	52.07	74.28	62.34	41.21	36.42	62.59
	Ours	**94.65**	**88.47**	**88.54**	**94.66**	**90.78**	**81.45**	**80.78**	**90.78**	**86.10**	**72.83**	**71.86**	**86.10**	**81.55**	**65.37**	**63.62**	**81.56**
Mfeat	EAMC	50.50	55.04	42.00	43.10	47.01	40.75	30.80	42.53	34.48	22.38	15.24	31.33	26.42	13.28	8.44	24.52
	DCP	85.87	86.32	79.91	84.45	78.72	78.23	70.40	77.39	72.19	68.05	58.23	70.94	44.37	47.91	28.96	42.90
	DSMVC	83.91	85.97	78.86	83.19	80.52	81.75	74.44	79.21	70.71	68.63	57.74	69.57	64.75	59.42	49.40	63.38
	SiMVC	91.75	88.52	85.58	91.63	75.35	76.45	66.93	74.65	74.57	72.29	64.56	73.26	64.13	57.41	47.95	63.57
	CoMVC	90.33	87.72	83.71	90.16	75.95	77.74	68.64	75.04	75.69	73.55	66.34	74.47	71.83	61.74	54.56	71.72
	DCMVC	92.26	**90.29**	**87.94**	91.88	91.95	**89.18**	**87.05**	91.51	90.86	**87.63**	**85.50**	90.59	86.38	**83.71**	**79.91**	85.79
	GLGC	76.23	74.40	63.51	76.42	79.80	77.45	69.60	79.82	80.19	75.87	68.75	80.14	79.11	73.49	67.84	79.17
	Ours	**93.66**	87.48	86.47	**93.70**	**93.51**	87.15	86.21	**93.53**	**92.82**	85.42	84.80	**92.83**	**90.34**	81.13	79.87	**90.37**
MSRCV16view	EAMC	33.14	16.85	9.62	32.15	32.00	14.43	7.14	31.21	27.24	9.23	3.26	27.00	26.19	7.39	1.97	25.69
	DCP	39.62	40.18	17.91	33.44	34.29	36.42	13.53	27.20	27.62	22.33	3.97	22.92	22.10	14.73	2.25	15.50
	DSMVC	43.90	28.50	18.81	42.53	39.43	24.04	15.15	38.05	30.19	13.54	5.93	29.90	32.48	14.45	7.28	31.65
	SiMVC	52.76	42.00	29.83	52.08	42.86	29.72	18.68	41.83	39.81	27.85	17.11	38.71	33.62	17.61	9.14	33.00
	CoMVC	53.24	42.75	30.46	52.55	42.57	29.70	18.32	41.59	39.14	27.52	16.61	37.87	33.33	17.33	8.73	32.85
	DCMVC	85.05	76.77	71.47	84.43	86.38	76.32	71.96	85.88	80.29	70.22	63.38	79.81	78.57	67.70	59.92	78.25
	GLGC	84.19	75.15	69.78	84.21	71.71	58.85	49.85	71.75	67.05	54.89	45.60	66.96	57.52	44.56	33.85	57.79
	Ours	**90.76**	**84.27**	**80.19**	**90.45**	**91.52**	**84.71**	**81.56**	**91.35**	**88.57**	**79.04**	**75.82**	**88.34**	**85.62**	**73.89**	**69.98**	**85.20**
MNIST-4	EAMC	66.28	44.77	43.00	62.94	56.81	26.32	25.18	55.06	53.78	24.15	23.09	52.23	44.90	11.71	11.27	43.27
	DCP	70.58	51.39	38.53	70.91	74.56	53.26	46.65	74.22	76.31	51.16	46.67	76.29	67.38	39.67	29.82	68.02
	DSMVC	77.74	57.10	56.29	77.67	78.56	50.48	52.93	78.57	66.72	33.52	35.16	66.75	57.88	21.83	23.17	57.79
	SiMVC	73.50	48.36	47.96	73.62	68.83	43.09	42.74	68.56	71.78	37.41	40.39	71.69	60.35	23.87	25.21	60.16
	CoMVC	75.66	52.41	51.21	75.66	76.67	52.40	52.67	76.59	71.28	37.62	39.90	71.26	64.22	27.10	28.82	64.19
	DCMVC	80.81	68.36	66.85	77.61	88.11	**72.33**	**74.35**	**89.04**	85.95	**67.00**	**69.53**	**86.84**	82.21	**56.91**	59.32	82.07
	GLGC	60.14	33.39	30.11	60.34	63.69	34.79	33.79	63.72	67.63	35.42	36.38	67.58	70.65	38.68	40.43	70.81
	Ours	**89.64**	**74.07**	**75.64**	**89.61**	**88.44**	70.31	72.83	88.39	**86.21**	65.00	67.96	86.17	**82.23**	56.01	**59.70**	**82.16**

after the warm-up and quality-to-relation learning stages, confirming the convergence and practical optimization stability of the proposed method.

4.5 Sensitivity Analysis

The sensitivity of λ_{RP} and τ_p is evaluated over all combinations of $\lambda_{\mathrm{RP}} \in \{0.001, 0.005, 0.01, 0.02, 0.05\}$ and $\tau_p \in \{0.10, 0.20, 0.50, 1.00\}$, with all other parameters fixed. As shown in Figure 2, the ACC values remain relatively stable across the tested configurations on both datasets, with only minor performance fluctuations. This suggests that CGQRL is robust to a broad range of parameter choices without requiring careful tuning.

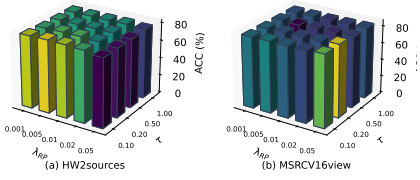


Figure 2: Sensitivity analysis of λ_{RP} and τ_p under $\rho = 0.7$.

Table 3: Ablation studies on the main components of CGQRL under heterogeneous Gaussian noise with $\rho = 0.7$.

Variants	HW2sources				MSRCV16view				Mfeat			
	ACC	NMI	ARI	F1	ACC	NMI	ARI	F1	ACC	NMI	ARI	F1
w/o Quality Learning	65.79	47.10	41.62	65.82	77.33	63.99	57.54	76.20	83.86	70.44	67.69	83.85
w/o Relation Learning	78.50	60.70	58.45	78.45	81.43	68.94	63.30	80.55	87.55	77.01	74.29	87.61
w/o Cross-view Relation	81.12	65.18	62.95	81.15	82.10	70.90	65.00	81.89	89.33	79.39	77.85	89.37
w/o RelScale	81.39	65.07	63.30	81.41	82.38	71.74	66.37	81.85	89.10	79.23	77.34	89.16
Our model	**81.55**	**65.37**	**63.62**	**81.56**	**85.62**	**73.89**	**69.98**	**85.20**	**90.34**	**81.13**	**79.87**	**90.37**

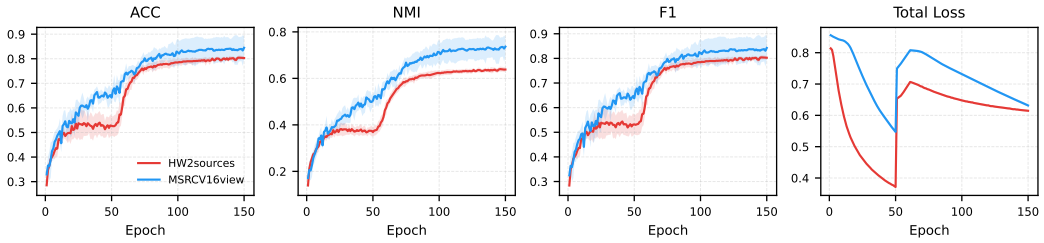


Figure 3: Convergence analysis on HW2sources.

4.6 Visualization

We use t-SNE to visualize noisy inputs and learned consensus representations on the HW2sources and MSRCV16view datasets. Each row corresponds to a corruption ratio $\rho \in \{0.1, 0.3, 0.5, 0.7\}$, with the left and right columns showing the noisy input features and learned consensus representations, respectively. We report the results for HW2sources in Figure 4 and for MSRCV16view in Appendix Figure 5. The same corrupted samples and random seed 0 are used for both representations. As illustrated in Figure 4, increasing ρ makes the noisy inputs progressively more scattered and overlapping, whereas the learned consensus representations remain relatively compact and well separated. These results suggest that CGQRL can reduce the influence of unreliable observations while preserving discriminative structural information in the consensus space.

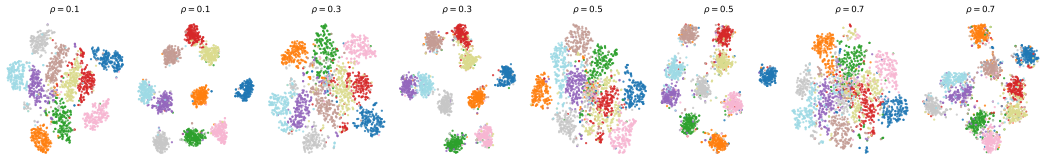


Figure 4: t-SNE visualization of noisy input features and learned consensus representations under four corruption ratios.

5 Conclusion

This paper presented CGQRL for multi-view clustering under heterogeneous observation noise. Rather than using observation quality only for feature fusion, CGQRL transfers sample–view reliability to sample relations and instance-specific cross-view relations. RelScale further preserves the absolute reliability of structural supervision during cross-view alignment. Experiments under four corruption ratios show that CGQRL maintains stable clustering performance as noise increases, while the ablation studies verify the roles of quality estimation, relation learning, and reliability-preserving alignment. Collectively, these components build a quality-to-relation learning pipeline that improves the robustness of consensus representation learning.The resulting framework adapts structural supervision to heterogeneous reliability patterns across samples and views.

REPRODUCIBILITY STATEMENT

We provide sufficient details to facilitate the reproducibility of our results. The method is described in Section 3, while the model architecture, implementation details, and experimental settings are provided in Appendix A.2. The source code will be publicly released upon acceptance.

AI USE STATEMENT

The language of this manuscript was refined with the assistance of generative AI tools. Generative AI tools were also used to assist with code implementation and table formatting. They were not used to generate research ideas or design the proposed method. All experiments, analyses, and conclusions were designed and conducted by the authors. All AI-assisted outputs were reviewed by the authors, who take full responsibility for the final content of this work.

REFERENCES

Peixia Chen, Huazhu Chen, and Chenping Zhao. End-to-end sample-quality-adaptive contrastive fusion for deep multi-view clustering. *The Journal of Supercomputing*, 82(13):653, 2026. doi: 10.1007/s11227-026-08801-1.

Li Cheng, Zhengzhong Zhu, Jiangping Zhu, Lanxi Bai, Jianbo Liu, and Pei Zhou. Quality-aware contrastive learning: Resolving view imbalance and false negative issues in multi-view clustering. In *Advanced Intelligent Computing Technology and Applications*, pp. 533–544, 2027. doi: 10.1007/978-981-92-3403-5_42.

Jinrong Cui, Yuting Li, Han Huang, and Jie Wen. Dual contrast-driven deep multi-view clustering. *IEEE Transactions on Image Processing*, 33:4753–4764, 2024. doi: 10.1109/TIP.2024.3444269.

Shihao Dong, Yue Liu, Xiaotong Zhou, Yuhui Zheng, Huiying Xu, and Xinzhong Zhu. RAC-DMVC: Reliability-aware contrastive deep multi-view clustering under multi-source noise. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 20835–20843, 2026. doi: 10.1609/aaai.v40i25.39223.

Zhibin Dong, Siwei Wang, Jiaqi Jin, Xinwang Liu, and En Zhu. Cross-view topology based consistent and complementary information for deep multi-view clustering. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 19383–19394, 2023. doi: 10.1109/ICCV51070.2023.01781.

Tingting Du, Wei Zheng, and Xingang Xu. Composite attention mechanism network for deep contrastive multi-view clustering. *Neural Networks*, 176:106361, 2024. doi: 10.1016/j.neunet.2024.106361.

Hongqing He, Jie Xu, Wenyuan Yang, Yonghua Zhu, Guoqiu Wen, and Xiaofeng Zhu. Global-graph guided and local-graph weighted contrastive learning for unified clustering on incomplete and noise multi-view data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 24813–24822, 2026. doi: 10.48550/arXiv.2512.21516.

Shizhe Hu, Binyan Tian, Weibo Liu, and Yangdong Ye. Self-supervised trusted contrastive multi-view clustering with uncertainty refined. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 17305–17313, 2025. doi: 10.1609/aaai.v39i16.33902.

Mudi Jiang, Jiahui Zhou, Xinying Liu, Zengyou He, and Zhikui Chen. Clusterability-based assessment of potentially noisy views for multi-view clustering. *arXiv preprint arXiv:2604.18024*, 2026. doi: 10.48550/arXiv.2604.18024.

Guanzhou Ke, Guoqing Chao, Xiaoli Wang, Chenyang Xu, Yongqi Zhu, and Yang Yu. A clustering-guided contrastive fusion for multi-view representation learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(4):2056–2069, 2024. doi: 10.1109/TCSVT.2023.3300319.

Zhiwen Kong, Zhen Fu, Dongdong Chang, Yiming Wang, and Yao Zhao. One for all: A novel dual-space co-training baseline for large-scale multi-view clustering. *arXiv preprint arXiv:2401.15691*, 2024. doi: 10.48550/arXiv.2401.15691.

Zeke Li, Rui Zhai, Xiang Yin, Yitong Shang, Chunhui Zeng, and Longge Wang. CG-MVC: Structure-aware multi-view clustering with entropy-quantile curriculum learning. *Multimedia Systems*, 32 (9):560, 2026a. doi: 10.1007/s00530-026-02632-z.

Ziqi Li, Junjie Liao, Yonghong Zhang, Jun Sun, and Tianyang Xu. Hausdorff-weighted contrastive fusion for multi-view clustering. *Knowledge-Based Systems*, 339:115658, 2026b. doi: 10.1016/j.knosys.2026.115658.

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. doi: 10.1109/TPAMI.2022.3197238.

Xin Liu, Lican Dai, and Boyuan Zheng. Reliability-aware gaussian residual counterpart generation for robust multi-view clustering with noisy correspondence. *Computation*, 14(8):185, 2026. doi: 10.3390/computation14080185.

ErLin Pan and Zhao Kang. Multi-view contrastive graph clustering. In *Advances in Neural Information Processing Systems*, volume 34, pp. 2148–2159, 2021. doi: 10.48550/arXiv.2110.11842.

R. Sinkhorn. Diagonal equivalence to matrices with prescribed row and column sums. *American Mathematical Monthly*, 74:402, 1967. doi: 10.2307/2314570.

Huayi Tang and Yong Liu. Deep safe multi-view clustering: Reducing the risk of clustering performance degradation caused by view increase. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 202–211, 2022. doi: 10.1109/CVPR52688.2022.00030.

Daniel J. Trosten, Sigurd Lokse, Robert Jenssen, and Michael Kampffmeyer. Reconsidering representation alignment for multi-view clustering. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1255–1265, 2021. doi: 10.1109/CVPR46437.2021.00131.

Fangdi Wang, Jiaqi Jin, Jingtao Hu, Suyuan Liu, Xihong Yang, Siwei Wang, Xinwang Liu, and En Zhu. Evaluate then cooperate: Shapley-based view cooperation enhancement for multi-view clustering. In *Advances in Neural Information Processing Systems*, volume 37, pp. 135355–135379, 2024. doi: 10.52202/079017-4302.

Hua Wang, Feiping Nie, Heng Huang, and Chris Ding. Heterogeneous visual features fusion via sparse multimodal machine. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3097–3102, 2013. doi: 10.1109/CVPR.2013.398.

Peihan Wu, Guanjie Cheng, Yufei Tong, Meng Xi, and Shuiguang Deng. Quality-aware robust multi-view clustering for heterogeneous observation noise. *arXiv preprint arXiv:2602.22568*, feb 2026. doi: 10.48550/arXiv.2602.22568.

Yuang Xiao, Chang Tang, Xiao Zheng, Weiqi Yan, Yuanyuan Liu, and Xinwang Liu. Mutual calibration network for multi-view clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 36(1):393–405, 2026. doi: 10.1109/TCSVT.2025.3588889.

Chang Xu, Dacheng Tao, and Chao Xu. A survey on multi-view learning. *arXiv preprint arXiv:1304.5634*, 2013. doi: 10.48550/arXiv.1304.5634.

Jie Xu, Yazhou Ren, Guofeng Li, Lili Pan, Ce Zhu, and Zenglin Xu. Deep embedded multi-view clustering with collaborative training. *Information Sciences*, 573:279–290, 2021. doi: 10.1016/j.ins.2020.12.073.

Jie Xu, Huayi Tang, Yazhou Ren, Liang Peng, Xiaofeng Zhu, and Lifang He. Multi-level feature learning for contrastive multi-view clustering. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16030–16039, 2022. doi: 10.1109/CVPR52688.2022.01558.

Xiaoqiang Yan, Shizhe Hu, Yiqiao Mao, Yangdong Ye, and Hui Yu. Deep multi-view learning methods: A review. *Neurocomputing*, 448:106–129, 2021. doi: 10.1016/j.neucom.2021.03.090.

Mouxing Yang, Yunfan Li, Peng Hu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Robust multi-view clustering with incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):1055–1069, 2023. doi: 10.1109/TPAMI.2022.3155499.

Lei Zhang, Lele Fu, Tong Wang, Chuan Chen, and Chuanfu Zhang. Mutual information-driven multi-view clustering. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pp. 3268–3277, 2023. doi: 10.1145/3583780.3614986.

Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun. Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, 38:43–54, 2017. doi: 10.1016/j.inffus.2017.02.007.

Xiaojie Zhao, Xueying Niu, Xianhua Zeng, and Jifu Zhang. Adaptive deep multi-view clustering via quality-aware representation learning. *Knowledge-Based Systems*, 339:115619, 2026. doi: 10.1016/j.knosys.2026.115619.

Qinghai Zheng and Haoyu Tang. Twin reciprocal completion for incomplete multi-view clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(12):13201–13212, 2024. doi: 10.1109/TCSVT.2024.3437756.

Runwu Zhou and Yi-Dong Shen. End-to-end adversarial-attention network for multi-modal clustering. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14607–14616, 2020. doi: 10.1109/CVPR42600.2020.01463.

A Implementation and Reproducibility Details

A.1 Optimization Algorithm

For completeness, Algorithm 1 gives the complete optimization procedure used by CGQRL. It is provided in the appendix so that the main text can focus on the motivation and the individual learning objectives.

Algorithm 1 Optimization of CGQRL

Require: Multi-view data $\{X^{(v)}\}_{v=1}^{V}$; cluster number K; maximum epoch T; warm-up length T_w; relation-update interval T_r

Ensure: Cluster labels $\widehat{Y}$

Initialize the view-specific encoders, decoders, IB branches, and quality memory Q

▷ Stage I: representation and reconstruction-quality warm-up

for $t = 1, \ldots, T_w$ **do**

  for each mini-batch B **do**

    Encode all views and compute the terms in $\mathcal{L}_{\mathrm{base}}$; see Appendix A.2

    Compute $Q_{\mathrm{rec}}^{(v)}$ for every sample–view observation and update the warm-up quality memory

    Optimize $\mathcal{L}_{\mathrm{base}}$ and update the representation and reconstruction parameters

  end for

end for

Extract the full-data view-specific representations and current quality estimates

Construct the quality-aware consensus representation H using Eq. 24

Initialize the cluster prototypes $\{c_k\}_{k=1}^{K}$ by K-means on H

Construct the initial reciprocal-neighbour candidate graph $\mathcal{E}_0^{+}$ and the eligible negative set $\mathcal{E}_0^{-}$

Compute $p_{ij,0}^{(v)}$ and initialize $R_{ij,0}^{\mathrm{s}}$ using the formulas in Appendix A.2; detach and cache the graph quantities.

▷ Stage II: joint quality-to-relation learning

for $t = T_w + 1, \ldots, T$ **do**

  if $t = T_w + 1$ or $(t - T_w - 1) \bmod T_r = 0$ **then**

    Extract the current full-data view-specific representations and latest observation-quality estimates

    Reconstruct $\mathcal{E}_t^{+}$ and $\mathcal{E}_t^{-}$ from the current reciprocal-neighbour structures

    Recalculate $p_{ij,t}^{(v)}$ for all refreshed candidate edges

    Re-estimate $R_{ij,t}^{\mathrm{s}}$ from the current endpoint quality and multi-view structural evidence

    Detach and cache the refreshed graph masks and relation reliabilities

  else

    Reuse the most recently cached graph and sample-relation reliabilities

  end if

  for each mini-batch B **do**

    Encode the batch and compute $Q_{\mathrm{rec}}^{(v)}$, $Q_{\mathrm{con}}^{(v)}$, and $Q_{\mathrm{cls}}^{(v)}$

    Compute $\widetilde{Q}_t^{(v)}$ by Eq. 1 and update $Q_t^{(v)}$ by Eq. 2

    Obtain $w_i^{(v)}$ and h_i from Eq. 24

    Retrieve the cached batch subgraph and compute $\mathcal{L}_{\mathrm{QR}}$ by Eq. 5

    Compute the prototype assignments $P_i^{(v)}$ by Eq. 6 and obtain detached balanced targets $T_i^{(v)}$ using Sinkhorn–Knopp

    Construct $R_i^{u \to v}$ by Eq. 7 and compute $\mathcal{L}_{\mathrm{RP}}$ by Eq. 9

    Optimize the complete objective in Eq. 10; update the model parameters and cluster prototypes

  end for

end for

Extract the final quality-aware consensus representation H

return $\widehat{Y} \leftarrow \mathrm{KMeans}(H, K)$

Table 4: Core notation. The superscript s denotes sample relations, whereas $u \to v$ denotes directed cross-view transfer.

Symbol	Meaning	Symbol	Meaning
$z_i^{(v)}$	view-specific code	h_i	consensus code
$Q_i^{(v)}$	observation quality	$p_{ij}^{(v)}$	calibrated edge probability
$q_{ij}^{(v)}$	endpoint quality	$a_{ij}^{(v)}$	cross-view edge agreement
$\pi_{ij}^{(v)}$	view contribution to a candidate edge	p_{ij}^{Q}	reliability-aware edge estimate
p_{ij}^{0}	quality-independent edge estimate	g_{ij}	conflict gate
C_{ij}	residual cross-view edge consistency	$\widehat{p}_{ij}$	interpolated edge strength
$R_{ij,t}^{\mathrm{s}}$	sample-edge reliability at update step t	$R_i^{u\to v}$	directed cross-view reliability
$G_i^{(v)}$	Student-t quality assignment	$P_i^{(v)}$	QCSA softmax assignment
$T_i^{(v)}$	balanced QCSA target	r_0	sample-edge reliability floor
$\mathcal{E}_t^+, \mathcal{E}_t^-$	refreshed positive and negative edge sets	T_r	relation-update interval

A.2 Additional Method and Optimization Details

Bottleneck reconstruction and reconstruction quality. For each view, a separate bottleneck branch reconstructs the corrupted observation as

$$\widehat{x}_{\mathrm{IB},i}^{(v)} = g_v^{\mathrm{IB}}\left(f_v^{\mathrm{IB}}(\widetilde{x}_i^{(v)})\right). \tag{11}$$

Its parameters are optimized by

$$\mathcal{L}_{\mathrm{IB}} = \frac{1}{BV}\sum_{v=1}^{V}\sum_{i\in B}\omega_i^{(v)}\left\|\widetilde{x}_i^{(v)} - \widehat{x}_{\mathrm{IB},i}^{(v)}\right\|_2^2, \tag{12}$$

where B denotes the mini-batch size. The detached weight $\omega_i^{(v)}$ is the within-view mean-normalized value of $\max(Q_i^{(v)}, 0.05)$. Detaching this weight prevents the model from reducing reconstruction pressure by directly lowering its quality estimate, while the lower bound preserves a reconstruction gradient for every observation.

Let m_v and s_v denote the running median and median absolute deviation of the logarithmic, dimension-normalized bottleneck residual in view v. The reconstruction-quality component is

$$Q_{\mathrm{rec},i}^{(v)} = 1 - \sigma\left(\kappa\frac{\log\left(d_v^{-1}\|\widetilde{x}_i^{(v)} - \widehat{x}_{\mathrm{IB},i}^{(v)}\|_1 + \varepsilon\right) - m_v}{s_v + \varepsilon}\right), \tag{13}$$

where σ is the logistic sigmoid and κ controls the calibration slope. A smaller residual relative to the current view-wise error distribution therefore produces a larger reconstruction-quality value.

Cross-view consistency quality. All representation comparisons use cosine similarity,

$$\mathrm{sim}(a,b) = \frac{a^\top b}{\|a\|_2\|b\|_2 + \varepsilon}. \tag{14}$$

To evaluate view v without using its own contribution, we construct the leave-one-view-out reference

$$h_i^{-v} = \mathrm{normalize}\left(\sum_{u\neq v}\frac{\exp(Q_i^{(u)}/\tau_q)}{\sum_{r\neq v}\exp(Q_i^{(r)}/\tau_q)}z_i^{(u)}\right). \tag{15}$$

The global agreement with h_i^{-v} and the local agreement with the reference-space neighbourhood are combined as

$$Q_{\mathrm{con},i}^{(v)} = \frac{1}{2} + \frac{1}{4}\mathrm{sim}(z_i^{(v)}, h_i^{-v}) + \frac{1}{4}\mathrm{sim}\left(z_i^{(v)}, \bar{z}_{\mathcal{N}_i^{-v}}^{(v)}\right), \tag{16}$$

where $\bar{z}_{\mathcal{N}_i^{-v}}^{(v)}$ is the normalized mean, in view v, of the neighbours retrieved from the leave-one-view-out reference space.

Cluster-compatibility quality. After warm-up, K-means on the current consensus representation initializes the prototypes $\{c_k\}_{k=1}^{K}$. The Student-t assignment of observation (i, v) is

$$G_i^{(v)}(k) = \frac{(1 + \|z_i^{(v)} - c_k\|_2^2)^{-1}}{\sum_{r=1}^{K}(1 + \|z_i^{(v)} - c_r\|_2^2)^{-1}}. \tag{17}$$

The detached consensus reference and its view contributions are

$$\bar{h}_i = \mathrm{normalize}\left(\sum_v \bar{w}_i^{(v)} z_i^{(v)}\right), \qquad \bar{w}_i^{(v)} = \mathrm{stopgrad}(w_i^{(v)}). \tag{18}$$

Applying the same Student-t rule to $\bar{h}_i$ gives the detached consensus assignment $\bar{G}_i$. Assignment confidence, prototype proximity, and consensus agreement are then computed as

$$\begin{aligned} C_{\mathrm{ent},i}^{(v)} &= 1 - \frac{-\sum_k G_i^{(v)}(k) \log G_i^{(v)}(k)}{\log K}, \\ C_{\mathrm{dis},i}^{(v)} &= \exp\left(-\frac{\min_k \|z_i^{(v)} - c_k\|_2^2}{\bar{d} + \varepsilon}\right), \\ C_{\mathrm{agr},i}^{(v)} &= \left[\frac{\langle G_i^{(v)}, \mathrm{stopgrad}(\bar{G}_i)\rangle - 1/K}{1 - 1/K}\right]_0^1, \end{aligned} \tag{19}$$

and their joint compatibility is

$$Q_{\mathrm{cls},i}^{(v)} = \sqrt{C_{\mathrm{ent},i}^{(v)} C_{\mathrm{dis},i}^{(v)} C_{\mathrm{agr},i}^{(v)}}. \tag{20}$$

Quality schedule, sharpening, and temporal update. The sharpening operator in Eq. 1 is defined elementwise as

$$\mathcal{S}_\nu(q) = \frac{q^\nu}{q^\nu + (1-q)^\nu + \varepsilon}. \tag{21}$$

During warm-up, only reconstruction quality is used, so $(a_t, b_t, c_t) = (1, 0, 0)$. After warm-up, let

$$\gamma_t = \mathrm{clip}\left(\frac{t - T_w}{T_q}, 0, 1\right), \qquad T_q = 60. \tag{22}$$

With $\alpha_{\mathrm{rec}} = 0.3$, $\beta_{\mathrm{con}} = 0.7$, and $\gamma_{\mathrm{cls}} = 0.15$, the normalized coefficients are

$$(a_t, b_t, c_t) = \frac{(\alpha_{\mathrm{rec}}, \beta_{\mathrm{con}}, \gamma_t \gamma_{\mathrm{cls}})}{\alpha_{\mathrm{rec}} + \beta_{\mathrm{con}} + \gamma_t \gamma_{\mathrm{cls}}}. \tag{23}$$

Thus, the coefficients change from $(1, 0, 0)$ during warm-up to $(0.3, 0.7, 0)$ immediately afterwards and approach $(0.2609, 0.6087, 0.1304)$ after the cluster-compatibility ramp. We use $\nu = 1.1$ and clip operational quality values to $[0.05, 0.95]$. Before the temporal-refinement stage, the instantaneous estimate is used directly; once refinement begins, Eq. 2 applies momentum smoothing. Each post-warm-up update is additionally clipped to a maximum absolute change of 0.15 to avoid abrupt reversals in quality ordering.

Quality-aware consensus representation. The current observation quality is converted into sample-specific view contributions and used to construct the consensus code:

$$w_i^{(v)} = \frac{\exp(Q_i^{(v)}/\tau_q)}{\sum_{u=1}^{V} \exp(Q_i^{(u)}/\tau_q)}, \qquad h_i = \mathrm{normalize}\left(\sum_{v=1}^{V} w_i^{(v)} z_i^{(v)}\right). \tag{24}$$

Here, τ_q controls the sharpness of the quality-derived view contributions. The same quality values are subsequently propagated to the two relation granularities described below.

Local structural support. For every candidate edge in Eq. 3, the local structural code and the calibrated view-specific support are

$$\begin{aligned} z_{i,\mathrm{s}}^{(v)} &= \mathrm{normalize}\left(z_i^{(v)} + \eta \bar{z}^{(v)}_{\mathcal{N}_i^{(v)}} \right), \\ p_{ij}^{(v)} &= \sigma\left(\frac{\mathrm{sim}(z_{i,\mathrm{s}}^{(v)}, z_{j,\mathrm{s}}^{(v)}) - \frac{1}{2}(b_i^{(v)} + b_j^{(v)}) - \mu_v^{\mathrm{s}}}{\tau_s s_v^{\mathrm{s}} + \varepsilon} \right). \end{aligned} \tag{25}$$

Here, $b_i^{(v)}$ is the local similarity baseline, while μ_v^{s} and s_v^{s} are the median and interquartile range of the view-specific relation margins. These robust statistics are recomputed whenever the sample graph is refreshed.

Endpoint quality and cross-view corroboration. Observation quality is transferred to an edge using the harmonic mean

$$q_{ij}^{(v)} = \frac{2 Q_i^{(v)} Q_j^{(v)}}{Q_i^{(v)} + Q_j^{(v)} + \varepsilon}, \tag{26}$$

which remains high only when both endpoint observations are reliable. For each view, the support supplied by the remaining reliable views and the corresponding agreement are

$$\begin{aligned} \bar{p}_{ij}^{(-v,Q)} &= \frac{\sum_{u \neq v} q_{ij}^{(u)} p_{ij}^{(u)}}{\sum_{u \neq v} q_{ij}^{(u)} + \varepsilon}, \\ a_{ij}^{(v)} &= \exp\left(-\frac{|p_{ij}^{(v)} - \bar{p}_{ij}^{(-v,Q)}|}{\tau_a} \right). \end{aligned} \tag{27}$$

Reliability-aware aggregation and conflict interpolation. Endpoint quality and cross-view agreement determine the contribution of view v to the candidate edge:

$$\pi_{ij}^{(v)} = \frac{q_{ij}^{(v)} a_{ij}^{(v)}}{\sum_{u=1}^{V} q_{ij}^{(u)} a_{ij}^{(u)} + \varepsilon}. \tag{28}$$

The reliability-aware and quality-independent estimates are

$$p_{ij}^{Q} = \sum_{v=1}^{V} \pi_{ij}^{(v)} p_{ij}^{(v)}, \qquad p_{ij}^{0} = \frac{1}{V} \sum_{v=1}^{V} p_{ij}^{(v)}. \tag{29}$$

When the views agree, the uniform estimate provides a stable reference; as their evidence diverges, the quality-aware estimate becomes more informative. We therefore define

$$\begin{aligned} g_{ij} &= \frac{V^{-1} \sum_v (p_{ij}^{(v)} - p_{ij}^{0})^2}{V^{-1} \sum_v (p_{ij}^{(v)} - p_{ij}^{0})^2 + \tau_g}, \\ \widehat{p}_{ij} &= (1 - g_{ij}) p_{ij}^{0} + g_{ij} p_{ij}^{Q}. \end{aligned} \tag{30}$$

Finally, the residual agreement after aggregation is

$$C_{ij} = \exp\left[-\frac{\sum_v \pi_{ij}^{(v)} (p_{ij}^{(v)} - \widehat{p}_{ij})^2}{\tau_r} \right]. \tag{31}$$

Substituting $\widehat{p}_{ij}$ and C_{ij} into Eq. 4 gives the final sample-relation reliability.

Dynamic sample-graph and relation update. At the warm-up boundary, K-means initializes the prototypes, after which the initial reciprocal-neighbour graph and its relation reliabilities are constructed from the full dataset. During joint optimization, the full-data representations and quality scores are extracted every T_r epochs. The reciprocal k_n-nearest-neighbour sets are reconstructed in each view and their union forms $\mathcal{E}_t^{+}$. Pairs unsupported by every current view and satisfying the negative-pair eligibility criterion form $\mathcal{E}_t^{-}$. Newly supported reciprocal edges are added, whereas edges that lose reciprocal support in every view are removed.

After graph membership is updated, $R^{\mathrm{s}}_{ij,t}$ is re-estimated for every retained positive edge using the current observation quality and multi-view structural evidence. The positive and negative masks and their reliabilities are detached and cached on the CPU. Between two updates, each mini-batch retrieves the corresponding submatrices by sample identifier when evaluating $\mathcal{L}_{\mathrm{QR}}$. We use $T_r = 10$ in all experiments.

A.2.1 Cross-View Relation Reliability and RelScale

Normalized assignment agreement. The Jensen–Shannon divergence in Eq. 7 satisfies $0 \leq \mathrm{JS}(P_i^{(u)}, P_i^{(v)}) \leq \log 2$. Division by $\log 2$ therefore maps assignment disagreement to $[0, 1]$, so the bracketed term in Eq. 7 is a normalized assignment-agreement score. The geometric mean of $Q_i^{(u)}$ and $Q_i^{(v)}$ ensures that a directed transfer is reliable only when both the source and target observations are trustworthy.

Balanced transfer targets. The Sinkhorn–Knopp procedure alternately normalizes the rows and columns of the exponentiated assignment matrix to obtain $T^{(v)}$. Its batch-level prototype marginals are approximately balanced, preventing a small number of prototypes from dominating the structural targets. The resulting targets are detached before they are used in Eq. 8.

RelScale gradient interpretation. Index the $M = BV(V-1)$ directed mini-batch transfers by j. With detached reliability R_j, normalization by the reliability sum gives $\nabla_\theta \mathcal{L}_{\mathrm{norm}} = (\sum_j R_j + \varepsilon)^{-1} \sum_j R_j \nabla_\theta \ell_j$; multiplying all reliabilities by a common factor therefore leaves the gradient nearly unchanged. RelScale instead gives $\nabla_\theta \mathcal{L}_{\mathrm{RP}} = M^{-1} \sum_j R_j \nabla_\theta \ell_j$, so a common decline in reliability proportionally reduces the total cross-view structural gradient. When ε is negligible, $\mathcal{L}_{\mathrm{RP}} = \bar{R}\mathcal{L}_{\mathrm{norm}}$, where $\bar{R} = M^{-1} \sum_j R_j$.

Shared representation objective. The base objective in Eq. 10 is

$$\mathcal{L}_{\mathrm{base}} = \lambda_{\mathrm{rec}}\mathcal{L}_{\mathrm{rec}} + \lambda_{\mathrm{IB}}\mathcal{L}_{\mathrm{IB}} + \lambda_{\mathrm{con}}\mathcal{L}_{\mathrm{con}} + \lambda_{\mathrm{MI}}\mathcal{L}_{\mathrm{MI}}. \tag{32}$$

Here, $\mathcal{L}_{\mathrm{rec}}$ trains the view-specific autoencoders, $\mathcal{L}_{\mathrm{IB}}$ trains the bottleneck reconstruction branches, $\mathcal{L}_{\mathrm{con}}$ encourages cross-view consistency, and $\mathcal{L}_{\mathrm{MI}}$ preserves information between the view-specific and consensus representations.

Initialization, update order, and gradient paths. CGQRL is optimized in two stages. During warm-up, the representation networks and the reconstruction-based quality evidence are learned without relational supervision. At the warm-up boundary, K-means initializes the prototypes and the initial sample graph is constructed. During joint optimization, quality components are updated for each mini-batch, whereas the full sample graph and its relation reliabilities are refreshed every T_r epochs before mini-batch optimization begins.

Quality scores, graph membership, sample-relation reliabilities, cross-view reliabilities, and Sinkhorn targets are detached when used as structural supervision. Consequently, they regulate the representation gradients without providing shortcuts through which the network can reduce the objective by directly changing its reliability estimates. Between graph updates, the cached relations remain fixed while the representation parameters and prototypes continue to be optimized. At inference, the quality estimator remains active and the final partition is obtained by applying K-means to H.

Data preprocessing. For each view, NaN and infinite entries are first replaced by zero. Each feature dimension is then standardized independently as $x \leftarrow (x - \mu_v)/(\sigma_v + 10^{-6})$, where the statistics are computed from all samples of that view. Noise is injected after standardization. Sample indices remain aligned across views, and no label information is used during preprocessing or training.

Heterogeneous Gaussian corruption. For every view, $\lfloor \rho N \rceil$ samples are selected without replacement. Each selected sample-view observation independently receives an intensity from $\{0.2, 0.4, 0.6, 0.8, 1.0\}$ with equal probability and is corrupted using the corruption rule specified in Section 4.1. Gaussian noise is drawn independently from $\mathcal{N}(0, I)$ in the standardized feature

space. We use $\rho \in \{0.1, 0.3, 0.5, 0.7\}$ and fix the generated corruption for all methods under the same dataset, ratio, and seed.

Network and optimization. Each view-specific encoder contains two fully connected layers of width 256 with ReLU activations, followed by a 32-dimensional latent layer. Its decoder uses the symmetric two-layer structure. All view encoders have the same latent dimension. Models are optimized with Adam using a learning rate of 10^{-3} and a batch size of 256 for 150 epochs. The first 50 epochs are used for warm-up. At the warm-up boundary, K-means initializes the prototypes and the initial sample graph is constructed. Afterwards, quality and relation learning are activated, and the sample graph and its relation reliabilities are refreshed every $T_r = 10$ epochs according to the schedule described in Appendix A.2. Final predictions are obtained by K-means on the consensus representation with 20 initializations.

CGQRL hyperparameters. The loss weights are $\lambda_{\mathrm{rec}} = 1$, $\lambda_{\mathrm{IB}} = 0.5$, $\lambda_{\mathrm{con}} = 0.0125$, $\lambda_{\mathrm{MI}} = 0.01$, $\lambda_{\mathrm{QR}} = 0.05$, and $\lambda_{\mathrm{RP}} = 0.01$. The negative topology weight is 0.5 and its cosine margin is $m = 0.2$. The QCSA assignment temperature is $\tau_p = 0.20$, the Sinkhorn regularization coefficient is 0.05, and three Sinkhorn iterations are used. Structural alignment starts without an additional delay and is linearly increased over 20 epochs. The relation power is 1.0. For sample relations, $k_n = 10$, the local structural residual is $\eta = 0.5$, the local relation temperature is $\tau_s = 1.0$, and the leave-one-view-out agreement temperature is $\tau_a = 0.10$, the conflict temperature is $\tau_g = 0.02$, the confidence temperature is $\tau_r = 0.05$, and the conservative weight floor is $r_0 = 0.75$. The reliability floor r_0 applies to edges retained in the current candidate graph and prevents partially supported relations from losing all structural influence. Graph membership is updated separately through the periodically reconstructed reciprocal neighbour sets. The reported sensitivity analysis does not establish robustness to alternative values of r_0; throughout the experiments it is fixed at the stated value. For quality estimation, $\kappa = 3$, $\tau_q = 1$, the refinement momentum is $\xi = 0.9$, and the maximum post-warm-up quality change per update is 0.15. Semantic consistency is enabled at the warm-up boundary, while the clustering coefficient is ramped over 60 epochs. The quality sharpening power is $\nu = 1.1$, and operational scores are clipped to $[0.05, 0.95]$. After the quality ramp, the normalized reconstruction, semantic, and clustering coefficients are $(a, b, c) = (0.2609, 0.6087, 0.1304)$, obtained from the unnormalized values $(0.3, 0.7, 0.15)$. RelScale retains the unnormalized mean relation reliability in the loss magnitude. Cluster prototypes are used to compute cluster-level quality and balanced cross-view assignments. The complete model does not include an additional quality-anchor, prototype-separation, teacher-guidance, pseudo-center, or KNN auxiliary regularizer.

For completeness, for a mini-batch of size B, the remaining base terms are

$$\mathcal{L}_{\mathrm{IB}} = \frac{1}{BV} \sum_{v,i} \omega_i^{(v)} \|\widetilde{x}_i^{(v)} - g_v^{\mathrm{IB}}(f_v^{\mathrm{IB}}(\widetilde{x}_i^{(v)}))\|_2^2, \tag{33}$$

$$\mathcal{L}_{\mathrm{con}} = \frac{1}{V(V-1)} \sum_{u \neq v} \frac{\sum_i \mathrm{sg}(\sqrt{Q_i^{(u)} Q_i^{(v)}})\, \ell_{i,\mathrm{NCE}}^{u \to v}}{\sum_i \mathrm{sg}(\sqrt{Q_i^{(u)} Q_i^{(v)}}) + \varepsilon}, \tag{34}$$

$$\mathcal{L}_{\mathrm{MI}} = \frac{1}{V} \sum_v \frac{1}{B} \sum_i -\log \frac{\exp(h_i^\top z_i^{(v)}/\tau)}{\sum_j \exp(h_i^\top z_j^{(v)}/\tau)}. \tag{35}$$

In Base, $\omega_i^{(v)}$ and the pair weight in $\mathcal{L}_{\mathrm{con}}$ are one. In the quality-aware variants, $\omega_i^{(v)}$ is the detached, within-view mean-normalised value of $\max(Q_i^{(v)}, 0.05)$, and the contrastive term uses detached endpoint quality as shown above.

Fixed noise and randomness. For every dataset and corruption ratio, we generate five fixed noise files indexed by $s \in \{0, 1, 2, 3, 4\}$. Each seed determines the corrupted sample-view pairs, their intensities, and the Gaussian values. All compared methods read the same file for seed s. Training seed s is used with noise file s and controls parameter initialization and mini-batch order. Thus, every condition contains five paired runs. ACC, NMI, ARI, and F1 are computed for each run and then summarized by their mean and standard deviation; the main tables show the means for readability. Existing noise files are reused when an interrupted experiment is resumed. Ground-truth labels are used only for evaluation.

B Auxiliary Experiment Protocols

B.1 Additional Ablation Results at $\rho = 0.5$

To complement the main ablation study, we report the corresponding results at $\rho = 0.5$ using the same variants and datasets. The numerical entries are from the independently completed $\rho = 0.5$ ablation runs.

Table 5: Ablation studies on the main components of CGQRL under heterogeneous Gaussian noise with $\rho = 0.5$.

Variants	HW2sources				MSRCV16view				Mfeat			
	ACC	NMI	ARI	F1	ACC	NMI	ARI	F1	ACC	NMI	ARI	F1
w/o Quality Learning	78.38	62.55	58.91	78.34	81.33	68.42	63.93	80.62	90.55	81.11	80.21	90.53
w/o Relation Learning	83.91	68.97	67.75	83.91	82.38	71.37	66.46	81.23	89.93	80.85	78.84	89.97
w/o Cross-view Relation	85.67	72.08	70.98	85.69	84.95	75.14	69.90	84.35	91.78	83.82	82.68	91.81
w/o RelScale	**86.46**	**73.27**	**72.44**	**86.46**	87.52	78.22	73.99	87.11	91.58	83.60	82.28	91.62
Our model	86.10	72.83	71.86	86.10	**88.57**	**79.04**	**75.82**	**88.34**	**92.82**	**85.42**	**84.80**	**92.83**

B.2 Additional t-SNE Visualization

The HW2sources visualization is presented in the main text. Here we provide the corresponding MSRCV16view result as an additional qualitative check. The noisy-input and learned representations use seed 0, the same corrupted observations, PCA preprocessing, and t-SNE initialization. This visualization does not isolate the contribution of an individual component.

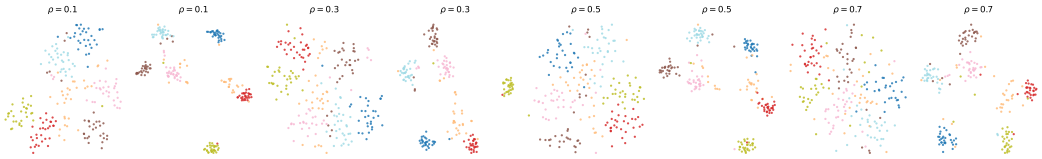


Figure 5: Additional t-SNE comparison on MSRCV16view under four corruption ratios.

C Additional Corruption Experiments

We further evaluate CGQRL under partial correspondence mismatch and a combined mismatch–Gaussian setting. These experiments use the same six datasets, corruption ratios, and five paired seeds as the main study. They assess the complete CGQRL model under additional corruptions; because no comparison-method results are included here, they should not be interpreted as comparative evidence against the baselines in the main tables.

Partial correspondence mismatch. View 1 defines the reference sample order. For each seed, a shared subset of $\lfloor \rho N \rceil$ samples is selected, and its indices are independently permuted without fixed points in every non-reference view. Thus, ρ is the fraction of samples whose cross-view correspondence is mismatched, while the reference view remains unchanged. Table 6 reports the resulting performance.

Combined mismatch and Gaussian corruption. We first generate partial correspondence mismatch as above. Heterogeneous Gaussian corruption is then applied independently at the same ratio using a separate deterministic seed. Its sample-view intensities are drawn from $\{0.2, 0.4, 0.6, 0.8, 1.0\}$ and follow the heterogeneous-noise rule described in Section 4. Table 7 reports all four ratios on all six datasets, including the completed MNIST-4 conditions.

Under both protocols, performance generally decreases as ρ increases. The combined setting is more demanding in most matched conditions, especially at larger corruption ratios. These results extend

the empirical scope beyond the feature-level Gaussian protocol used in the main comparison, while the absence of additional baselines limits them to a within-method robustness assessment.

Table 6: CGQRL under partial correspondence mismatch (mean ± standard deviation, %).

Dataset	ρ	ACC	NMI	ARI	F1
Hdigit	0.1	92.21± 0.31	81.94± 0.42	83.54± 0.60	92.21± 0.32
	0.3	81.33± 0.24	63.14± 0.49	63.16± 0.40	81.32± 0.25
	0.5	69.97± 0.48	47.22± 0.62	45.27± 0.68	69.88± 0.47
	0.7	57.63± 0.78	33.70± 0.66	29.87± 0.75	57.45± 0.83
COIL20	0.1	71.00± 1.16	72.56± 0.62	59.42± 1.13	70.83± 1.35
	0.3	58.50± 1.50	54.74± 1.06	40.31± 1.32	58.78± 1.70
	0.5	46.39± 1.84	41.71± 0.97	25.07± 0.85	47.63± 2.42
	0.7	34.22± 1.87	30.57± 0.82	14.38± 1.02	35.59± 1.93
HW2sources	0.1	90.92± 0.27	80.88± 0.71	81.01± 0.59	90.91± 0.26
	0.3	80.10± 0.62	62.68± 0.70	61.03± 0.97	80.10± 0.64
	0.5	67.32± 1.87	46.36± 1.91	42.28± 2.34	67.24± 1.80
	0.7	50.52± 3.76	33.33± 1.96	26.31± 2.29	50.30± 3.94
Mfeat	0.1	84.95± 5.04	73.88± 1.36	71.97± 3.71	85.10± 4.87
	0.3	65.85± 0.91	54.42± 0.87	45.22± 1.27	68.16± 0.87
	0.5	54.26± 0.86	37.61± 0.50	27.03± 0.47	56.87± 0.67
	0.7	41.33± 1.08	21.98± 1.13	14.51± 0.63	42.66± 0.92
MSRCV16view	0.1	84.00± 1.56	72.83± 2.31	67.63± 2.60	83.90± 1.55
	0.3	69.71± 4.03	52.57± 3.03	44.05± 4.07	70.18± 4.14
	0.5	54.67± 5.78	35.83± 2.86	25.89± 2.92	55.16± 6.30
	0.7	39.14± 3.15	21.00± 1.66	12.17± 1.28	39.17± 3.16
MNIST-4	0.1	85.96± 0.47	63.80± 1.11	67.31± 1.00	85.90± 0.48
	0.3	77.88± 0.39	46.53± 0.78	50.63± 0.81	77.81± 0.38
	0.5	68.89± 0.38	32.24± 0.56	35.04± 0.63	68.83± 0.37
	0.7	57.58± 0.96	18.10± 0.97	19.47± 1.17	57.48± 0.90

Table 7: CGQRL under combined partial correspondence mismatch and heterogeneous Gaussian corruption (mean ± standard deviation, %).

Dataset	ρ	ACC	NMI	ARI	F1
Hdigit	0.1	91.33± 0.30	80.23± 0.47	81.79± 0.58	91.33± 0.31
	0.3	77.34± 0.53	57.12± 0.59	56.40± 0.82	77.35± 0.53
	0.5	63.20± 0.66	38.41± 0.52	35.65± 0.74	63.17± 0.67
	0.7	45.29± 1.98	21.10± 0.84	17.21± 0.96	45.23± 2.01
COIL20	0.1	70.40± 1.26	71.91± 0.94	58.32± 1.67	70.46± 1.19
	0.3	58.36± 1.40	54.40± 0.96	40.47± 0.85	58.45± 1.74
	0.5	44.90± 1.97	39.33± 1.36	24.34± 1.50	45.60± 2.08
	0.7	32.86± 1.76	26.11± 1.22	12.73± 1.17	33.26± 1.83
HW2sources	0.1	89.86± 0.59	78.69± 0.92	78.89± 1.14	89.86± 0.59
	0.3	75.65± 0.86	55.77± 1.05	53.61± 1.30	75.65± 0.89
	0.5	53.86± 1.76	32.16± 0.58	26.97± 0.89	53.77± 1.82
	0.7	33.98± 1.34	15.21± 1.03	10.07± 0.79	34.16± 1.40
Mfeat	0.1	86.62± 0.50	73.60± 0.68	72.48± 0.94	86.71± 0.50
	0.3	64.99± 0.42	52.02± 1.02	43.30± 0.62	67.08± 0.63
	0.5	53.30± 2.50	33.42± 0.67	26.20± 1.08	54.51± 2.09
	0.7	35.78± 0.37	16.59± 1.11	11.17± 0.79	35.99± 0.41

Continued on next page

Table 7 (continued)

Dataset	ρ	ACC	NMI	ARI	F1
MSRCV16view	0.1	84.19± 2.14	72.51± 3.26	67.71± 3.82	84.03± 2.11
	0.3	67.81± 2.95	50.75± 3.74	41.66± 2.93	68.15± 3.38
	0.5	53.43± 5.96	33.80± 3.50	24.22± 3.57	53.61± 6.26
	0.7	38.29± 3.13	18.39± 1.91	10.44± 1.97	37.84± 3.04
MNIST-4	0.1	85.51± 0.41	62.69± 1.10	66.34± 0.92	85.46± 0.40
	0.3	75.25± 0.27	42.04± 0.38	45.83± 0.48	75.20± 0.23
	0.5	63.68± 0.52	25.07± 0.68	27.31± 0.74	63.63± 0.51
	0.7	48.92± 1.09	10.10± 1.01	10.60± 1.14	48.87± 1.04