RISK-CERTIFIED TOPOLOGY GUIDANCE FOR INCOMPLETE MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Incomplete multi-view clustering (IMVC) seeks shared clusters despite missing views, but existing relation-transfer methods often produce inaccurate predictions and conflict with representation learning. To address this, we propose the **R**isk-certified **T**opology **G**uidance framework (**RTG-IMVC**). **First**, we construct an admissible topology set for each missing sample via donor substitution, degree-preserving edge perturbation, and neighborhood bootstrapping. **Second**, we formulate a transfer-risk certificate combining prediction diameter and topology-cover radius to generate risk-normalized weights for a logarithmic opinion pool, bounding each view's influence on the consensus teacher. **Then**, we project the joint gradient onto a half-space to preserve the descent of instance- and cluster-level contrastive objectives, theoretically proving teacher robustness and one-step objective decrease. **Finally**, extensive experiments demonstrate that RTG-IMVC achieves superior clustering performance, significantly outperforming state-of-the-art methods across eight benchmark datasets while making high-confidence guidance testable and strictly risk-controlled. Code: `https://anonymous.4open.science/r/RTG-IMVC-DB37/`.

1 INTRODUCTION

Multi-view data describe the same objects through heterogeneous visual, textual, audio, and sensor features. Multi-view clustering (MVC) exploits cross-view agreement and complementarity to recover a common partition without supervision (Chao et al., 2021; Chen et al., 2022; Kumar et al., 2011; Xu et al., 2022b). Sensor failures, transmission errors, privacy restrictions, and asynchronous collection can leave views unavailable (Wen et al., 2022). Incomplete multi-view clustering (IMVC) therefore seeks cluster structure in partially observed data (Li et al., 2014; Hu & Chen, 2019; Wen et al., 2019; 2018; Jin et al., 2023). Cross-view relations support information sharing (Liu et al., 2024a), yet missingness obscures both feature measurements and the relations linking observations.

Existing IMVC methods broadly adopt reconstruction or imputation-free learning. Reconstruction recovers missing features, representations, or graphs (Lin et al., 2021; Tang & Liu, 2022a; Li et al., 2023; Wang et al., 2026b; Zhao et al., 2026; Wang et al., 2026a), but inaccurate recovery can introduce semantic inconsistency. Imputation-free methods instead learn from available views through representation or structural consensus (Hu & Chen, 2019; Wen et al., 2018; 2023a; Xu et al., 2024; Gu et al., 2026; Dai et al., 2025). ICMVC transfers observed-view similarities to missing views, propagates representations through graph convolution, and constructs high-confidence contrastive targets from view-specific and fused predictions (Chao et al., 2024). However, prediction confidence does not establish transferred-topology reliability. Graph neural networks can be sensitive to structural perturbations (Günnemann et al., 2019); incorrect transferred edges may therefore distort representations and cluster predictions. Moreover, confident neural predictions can be miscalibrated (Guo et al., 2017), so sharply peaked consensus targets need not be trustworthy. The resulting guidance may also conflict with contrastive representation learning. Safe-learning and gradient-manipulation methods address related optimization conflicts (Tang & Liu, 2022b; Sener & Koltun, 2018; Yu et al., 2020; Liu et al., 2021; Navon et al., 2022), but do not explicitly couple transferred-topology uncertainty with teacher influence and base-objective preservation.

Developing reliable topology guidance raises two central challenges.

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.

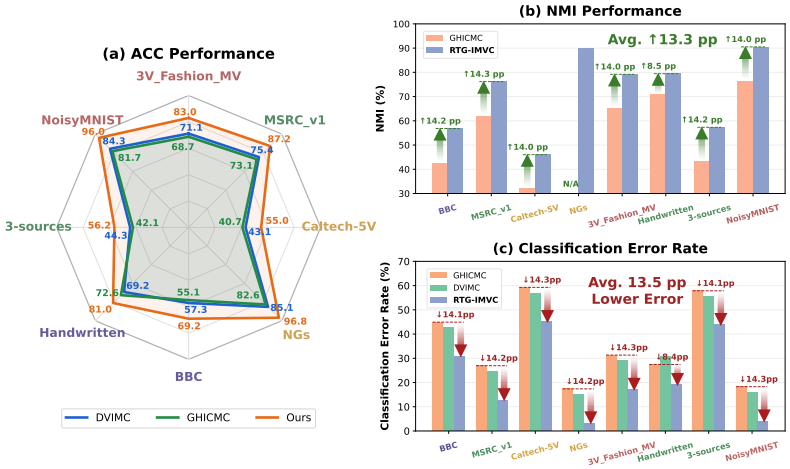


Figure 1: **Performance and reliability analysis of RTG-IMVC.** (a) Mean clustering accuracy (ACC) across eight benchmark datasets at a missing rate of $\eta = 0.1$. (b) NMI performance comparison between the GHICMC baseline and RTG-IMVC at $\eta = 0.1$. Green arrows indicate the absolute improvement in percentage points (pp). (c) Classification error rates of different methods at the same missing rate. Red arrows indicate the error reduction achieved by RTG-IMVC compared with GHICMC.

(1) Transfer-reliability certification. The true missing-view topology is unavailable, so its discrepancy from the transferred relation cannot be measured directly. Reliability thus be evaluated over plausible topology alternatives. Moreover, a finite collection of sampled alternatives must be connected to a broader admissible topology set to support a meaningful certificate.

(2) Guidance–objective compatibility. Even when topology risk can be estimated, that information regulate both consensus formation and optimization. An unreliable prediction should have limited influence on the teacher, while the resulting auxiliary gradient should not reverse progress on the representation-learning objective.

To address these challenges, we propose **RTG-IMVC**, a risk-certified topology-guidance framework for incomplete multi-view clustering. **First**, RTG-IMVC constructs candidate topologies per missing sample–view pair via donor substitution, degree-preserving edge replacement, and neighborhood bootstrap. Their prediction diameter, combined with a topology-cover correction, yields a sample-level transfer-risk certificate. **Second**, the certificate determines risk-normalized expert weights in a logarithmic opinion pool (Heskes, 1997). The same reliability controls guidance-loss strength, reducing the influence of topology-sensitive predictions. **Finally**, RTG-IMVC applies a minimum-change projection to the joint gradient. It preserves a prescribed amount of descent for the protected contrastive objective while retaining compatible guidance information. Under explicit coverage, boundedness, local continuity, and smoothness, the framework provides conditional guarantees on teacher sensitivity and one-step objective descent. RTG-IMVC significantly improves clustering accuracy, NMI, and error rates (Fig. 1).

The main contributions of this work are summarized as follows:

- **Risk-certified topology transfer.** We introduce a reliability-aware formulation for relation-transfer-based IMVC, explicitly separating predictive confidence from robustness to uncertain transferred topologies.
- **Sample-level transfer certificate.** We derive a topology-transfer risk certificate over an admissible topology set, combining finite-probe prediction variation with a coverage correction to quantify missing-view transfer sensitivity.
- **Certificate-constrained consensus.** We develop a risk-normalized logarithmic opinion pool that converts transfer certificates into expert-level consensus weights and sample-adaptive guidance strengths, suppressing topology-sensitive experts.

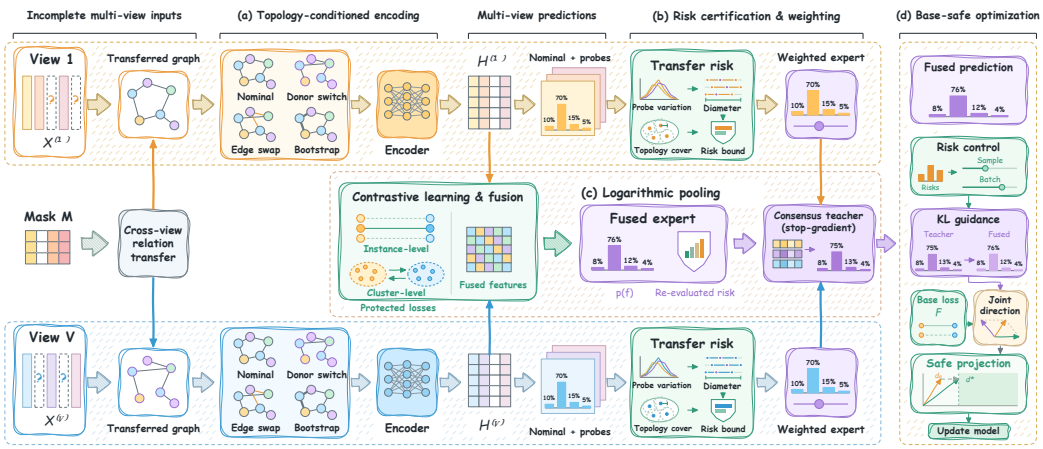


Figure 2: Overview of RTG-IMVC. (a) Shared encoders process topology probes within each view for contrastive learning and fusion. (b) Prediction variation and topology coverage determine transfer risk. (c) Inverse-risk logarithmic pooling combines view-specific and fused experts into a detached teacher. (d) Combined risk-scaled KL and protected-loss gradients undergo base-safe projection for $F = \mathcal{L}_{\text{ins}} + \mathcal{L}_{\text{clu}}$. Filled and dashed inputs denote observed and missing views.

- **Base-safe optimization with guarantees.** We derive a minimum-change constrained update that preserves descent of the protected representation objective while retaining compatible guidance, with guarantees on topology stability, expert influence, and one-step objective descent.

2 Related Work

2.1 Incomplete Multi-View Clustering

Incomplete multi-view clustering aims to recover cluster structure from partially observed views, with SURE (Yang et al., 2022) further considering cross-view misalignment. A major line of research reconstructs missing features, representations, or relational structures. COMPLETER (Lin et al., 2021) jointly learns representations and recovers missing views by maximizing mutual information while minimizing conditional entropy. AGDIMC (Pu et al., 2024) couples adaptive imputation with latent graph learning, whereas DMVG (Wen et al., 2024) introduces diffusion-based recovery. BRIDGE (Jiang et al., 2025) unifies transfer and imputation under non-IID missingness. More recently, ARSL-IMVC (Wang et al., 2026b) combines recovery with tensorial subspace learning, CBDSIMVC (Zhao et al., 2026) imposes block-diagonal constraints on graph completion.

Another line avoids explicit missing-feature synthesis and instead learns shared structure directly from incomplete observations. DIMVC (Xu et al., 2022a) exploits cluster-level complementarity, while consensus bipartite graph learning (Wang et al., 2022) supports scalable clustering. SIMVC-SA (Wen et al., 2023b) aligns cross-view anchors, and SCSL (Liu et al., 2024b) jointly models cross-view similarities and consensus representations. DVIMC (Xu et al., 2024) aggregates variational posteriors using a product of experts, whereas CAMEL (Gu et al., 2026) jointly learns embeddings and low-rank cross-view tensor structures. ICMVC (Chao et al., 2024) transfers observed-view similarity graphs to missing views for contrastive graph convolutional learning, but does not explicitly characterize how prediction changes under errors in the transferred topology. In contrast, RTG-IMVC explicitly evaluates sensitivity over admissible topology variations and uses the resulting certificates to regulate cross-view teaching through reliability-aware weighting and base-safe optimization.

2.2 Reliable Guidance and Safe Optimization

Reliability-aware IMVC considers both incomplete-view recovery and the trustworthiness of transferred or imputed information. DSIMVC (Tang & Liu, 2022b) dynamically imputes missing views

and selects reliable samples through bi-level safe learning with risk guarantees, while ICMVC (Chao et al., 2024) builds auxiliary supervision from confident view-specific and fused predictions. Beyond IMVC, Pro-GNN (Jin et al., 2020) jointly optimizes graph structure and model parameters for robustness. Related certification studies analyze graph propagation under admissible perturbations (Bojchevski & Günnemann, 2019) and certify GCN predictions against structural changes (Zügner & Günnemann, 2020). These methods highlight structural reliability, but do not directly quantify risk from cross-view topology transfer. RTG-IMVC instead derives a sample-dependent transfer-risk certificate from prediction variation over admissible topologies, with a coverage correction under stated assumptions.

A separate line of work studies conflicts between primary and auxiliary objectives. MGDA seeks Pareto-stationary solutions (Sener & Koltun, 2018), whereas PCGrad and CAGrad mitigate gradient conflicts (Yu et al., 2020; Liu et al., 2021). Auxiliary-gradient decomposition separates directions that benefit or interfere with the primary task (Dery et al., 2021), while GEM constrains updates to reduce interference with previous objectives (Lopez-Paz & Ranzato, 2017). More recently, Bloop (Hsieh et al., 2024) prioritizes the primary objective through orthogonal auxiliary-gradient projection stabilized by an exponential moving average. In contrast, RTG-IMVC couples primary-objective protection with topology-dependent reliability guidance. Its half-space projection enforces descent of the protected base objective while retaining compatible auxiliary updates, yielding one-step objective preservation under the stated assumptions.

3 METHODOLOGY

Fig. 2 motivates our formulation of incomplete multi-view learning, certified consensus teaching, and base-safe descent.

3.1 PROBLEM FORMULATION

Let $\mathcal{X} = \{\boldsymbol{X}^{(v)}\}_{v=1}^{V}$ contain N samples from V views and C clusters, where $\boldsymbol{X}^{(v)} \in \mathbb{R}^{d_v \times N}$. Given the observation mask $\boldsymbol{M} \in \{0,1\}^{N \times V}$, the observed and missing views of sample i are

$$\mathcal{O}_i = \{v : M_{iv} = 1\}, \qquad \mathcal{M}_i = \{v : M_{iv} = 0\}, \qquad \mathcal{O}_i \neq \varnothing. \tag{1}$$

For a missing pair (i, v), relation transfer produces a normalized adjacency row $\widetilde{\boldsymbol{a}}_i^{(v)}$. Replacing it by $\boldsymbol{a}$ gives

$$\boldsymbol{p}_{i,\theta}^{(v)}(\boldsymbol{a}) = \mathrm{softmax}\Big(\boldsymbol{z}_{i,\theta}^{(v)}(\boldsymbol{a})/T_p\Big), \qquad \boldsymbol{\ell}_{i,\theta}^{(v)}(\boldsymbol{a}) = \log\Big[\mathrm{clip}\Big(\boldsymbol{p}_{i,\theta}^{(v)}(\boldsymbol{a}), \varepsilon_p, 1\Big)\Big], \tag{2}$$

where $T_p > 0$ is a temperature and $\varepsilon_p > 0$ prevents divergent logarithms. Eq. (2) provides the log-probabilities used to compare predictions across alternative topologies. The network also outputs the nominal view prediction $\boldsymbol{p}_i^{(v)}$ and fused prediction $\boldsymbol{p}_i^{(f)}$. Its protected objective is

$$F(\theta) = \mathcal{L}_{\mathrm{ins}}(\theta) + \mathcal{L}_{\mathrm{clu}}(\theta), \tag{3}$$

where $\mathcal{L}_{\mathrm{ins}}$ and $\mathcal{L}_{\mathrm{clu}}$ are the instance- and cluster-level contrastive losses. We use F as the baseline when adding consensus guidance. Appendix A summarizes notation.

3.2 TOPOLOGY-TRANSFER RISK CERTIFICATE

For each (i, v) with $v \in \mathcal{M}_i$, we construct candidates by donor substitution, degree-preserving edge replacement, and neighborhood bootstrap. Let $\mathcal{N}(\boldsymbol{a}) = \{j : a_j > 0\}$ and $d_{\boldsymbol{a}} = |\mathcal{N}(\boldsymbol{a})|$. For replacement ratio ρ, equal-sized sets $\mathcal{R}_{\rho,r} \subset \mathcal{N}(\boldsymbol{a})$ and $\mathcal{C}_{\rho,r} \subset [N] \setminus \mathcal{N}(\boldsymbol{a})$ remove and insert neighbors. For bootstrap strength β, draw $\boldsymbol{c}_b \sim \mathrm{Multinomial}(d_{\boldsymbol{a}}, \boldsymbol{a})$. The operators and candidate set are

$$\begin{aligned} \mathsf{Rep}_{\rho,r}(\boldsymbol{a}) &= \boldsymbol{a} \odot \mathbf{1}_{\mathcal{R}_{\rho,r}^c} + \bar{a}_{\rho,r} \mathbf{1}_{\mathcal{C}_{\rho,r}}, \quad \bar{a}_{\rho,r} = |\mathcal{R}_{\rho,r}|^{-1} \sum_{j \in \mathcal{R}_{\rho,r}} a_j, \\ \mathsf{Boot}_{\beta,b}(\boldsymbol{a}) &= (1-\beta)\boldsymbol{a} + \beta \boldsymbol{c}_b / d_{\boldsymbol{a}}, \qquad \mathcal{D}_i^{(v)} = \{\widetilde{\boldsymbol{a}}_i^{(u)} : u \in \mathcal{O}_i\}. \end{aligned} \tag{4}$$

Here $\mathbf{1}_{\mathcal{A}}$ denotes the indicator vector of set $\mathcal{A}$. Let $\mathcal{E}_i^{(v)}$ and $\mathcal{B}_i^{(v)}$ denote the edge-replacement and neighborhood-bootstrap candidate sets, respectively. The complete candidate set is

$$\mathcal{S}_i^{(v)} = \{\widetilde{\boldsymbol{a}}_i^{(v)}\} \cup \mathcal{D}_i^{(v)} \cup \mathcal{E}_i^{(v)} \cup \mathcal{B}_i^{(v)}. \tag{5}$$

Eq. (4) generates perturbed topologies. Eq. (5) includes these candidates and the nominal topology, allowing comparison with the transferred relation used by the model. We denote by K the candidate-topology for each missing sample–view pair. The grids and draw counts for ρ, β, r, and b are specified in the experimental settings. We evaluate Eq. (2) over $\mathcal{S}_i^{(v)}$ to estimate transfer risk.

Using $d_{\mathcal{A}}(\boldsymbol{a}, \boldsymbol{b}) = \frac{1}{2}\|\boldsymbol{a} - \boldsymbol{b}\|_1$, we define the finite-set prediction diameter and transfer risk as

$$\widehat{D}_i^{(v)} = \max_c \left[\max_{\boldsymbol{a} \in \mathcal{S}_i^{(v)}} \ell_{ic}^{(v)}(\boldsymbol{a}) - \min_{\boldsymbol{a} \in \mathcal{S}_i^{(v)}} \ell_{ic}^{(v)}(\boldsymbol{a}) \right], \qquad R_i^{(v)} = \widehat{D}_i^{(v)} + 2L_{\mathcal{A},i}^{(v)} \epsilon_i^{(v)}. \tag{6}$$

In Eq. (6), the first term measures prediction variation over the finite candidates of Eq. (5); the second accounts for admissible topologies outside this set. Under the assumptions below, the correction extends the finite-probe measurement to the admissible topology set.

Theorem 1 (**Certified topology stability**). *For fixed θ, assume that $\mathcal{S}_i^{(v)} \subseteq \mathcal{U}_i^{(v)}$ is an $\epsilon_i^{(v)}$-cover under $d_{\mathcal{A}}$, and $\boldsymbol{\ell}_{i,\theta}^{(v)}$ is $L_{\mathcal{A},i}^{(v)}$-Lipschitz on $\mathcal{U}_i^{(v)}$ in the ℓ_∞ norm. Then, for any $\boldsymbol{a}, \boldsymbol{a}' \in \mathcal{U}_i^{(v)}$,*

$$\|\boldsymbol{\ell}_{i,\theta}^{(v)}(\boldsymbol{a}) - \boldsymbol{\ell}_{i,\theta}^{(v)}(\boldsymbol{a}')\|_\infty \le R_i^{(v)}. \tag{7}$$

When the cover radius and Lipschitz constant are estimated without certified upper bounds, $R_i^{(v)}$ in Eq. (6) is an empirical sensitivity score (proof in Appendix C.1).

3.3 Certificate-Constrained Consensus Teacher

Let $\mathcal{P}_i = \{1, \dots, V, f\}$ index the view-specific and fused predictions. The risk of expert s is

$$\overline{R}_i^{(s)} = \begin{cases} 0, & s \in \mathcal{O}_i, \\ R_i^{(s)}, & s \in \mathcal{M}_i, \\ R_i^{(f)}, & s = f, \end{cases} \tag{8}$$

where $R_i^{(f)}$ is obtained by re-evaluating fusion under the candidates in Eq. (5) and applying Eq. (6). Eq. (8) assigns zero transfer risk to observed views while retaining the estimates for missing and fused experts. With risk floor $\varepsilon_R > 0$, teacher temperature $T_q > 0$, and $Z_i = \sum_{s \in \mathcal{P}_i} (\overline{R}_i^{(s)} + \varepsilon_R)^{-1}$, the expert weights and logarithmic opinion pool (Heskes, 1997) are

$$\omega_i^{(s)} = \frac{(\overline{R}_i^{(s)} + \varepsilon_R)^{-1}}{Z_i}, \qquad Q_{ic} = \frac{\exp[T_q^{-1} \sum_{s \in \mathcal{P}_i} \omega_i^{(s)} \ell_{ic}^{(s)}]}{\sum_{k=1}^{C} \exp[T_q^{-1} \sum_{s \in \mathcal{P}_i} \omega_i^{(s)} \ell_{ik}^{(s)}]}. \tag{9}$$

The inverse-risk weights in Eq. (9), computed from the risks in Eq. (8), reduce the contribution of uncertain transferred relations to the consensus teacher. To quantify their influence, let $\boldsymbol{Q}_i'$ be the teacher distribution computed from perturbed predictions with unchanged weights, and define $\Delta \log(Q_{ic}/Q_{ik}) := \log(Q_{ic}'/Q_{ik}') - \log(Q_{ic}/Q_{ik})$.

Theorem 2 (**Bounded expert influence**). *Fix the risks and expert weights. Let $\mathcal{J} \subseteq \mathcal{M}_i \cup \{f\}$ index the perturbed experts, whose log-predictions satisfy $\|\boldsymbol{\ell}_i'^{(s)} - \boldsymbol{\ell}_i^{(s)}\|_\infty \le \overline{R}_i^{(s)}$ for $s \in \mathcal{J}$. With $\chi_i = Z_i^{-1}$,*

$$\left| \Delta \log \frac{Q_{ic}}{Q_{ik}} \right| \le \frac{2}{T_q} \sum_{s \in \mathcal{J}} \omega_i^{(s)} \overline{R}_i^{(s)} \le \frac{2|\mathcal{J}|\chi_i}{T_q}. \tag{10}$$

The log-odds form of Eq. (9) removes the teacher's normalization term, yielding the bound from weighted prediction perturbations. Risks, weights, and the teacher are detached during each network update. The proof is provided in Appendix C.2.

3.4 Overall Objective of RTG-IMVC

Let $\mathcal{I}_{\mathrm{m}} = \{i : \mathcal{M}_i \neq \varnothing\}$ and let $\mathrm{sg}(\cdot)$ denote stop-gradient. After obtaining $\boldsymbol{Q}_i$ from Eq. (9), RTG-IMVC uses the risk values to regulate consensus guidance:

$$\begin{aligned} \min_{\theta}\ \mathcal{L}_{\mathrm{RTG}}(\theta) &= \mathcal{L}_{\mathrm{ins}}(\theta) + \mathcal{L}_{\mathrm{clu}}(\theta) + \lambda_t |\mathcal{I}_{\mathrm{m}}|^{-1} \sum_{i\in\mathcal{I}_{\mathrm{m}}} \kappa_i D_{\mathrm{KL}}\Big(\mathrm{sg}(\boldsymbol{Q}_i) \,\|\, \boldsymbol{p}_i^{(f)}(\theta)\Big), \\ \text{with}\quad \kappa_i &= \exp\left[-\tau^{-1} \sum_{s\in\mathcal{P}_i} \omega_i^{(s)} \overline{R}_i^{(s)}\right], \qquad \lambda_t = \lambda_0 |\mathcal{I}_{\mathrm{m}}|^{-1} \sum_{i\in\mathcal{I}_{\mathrm{m}}} \kappa_i. \end{aligned} \tag{11}$$

where $\tau > 0$ controls reliability decay and $\lambda_0 > 0$ is the maximum guidance coefficient. In Eq. (11), κ_i attenuates guidance for uncertain samples, while λ_t adjusts its strength across the batch. If $\mathcal{I}_{\mathrm{m}} = \varnothing$, the guidance term and λ_t are set to zero.

3.5 Alternating Base-Safe Optimization

At iteration t, RTG-IMVC performs three steps. First, fixing θ_t, it recomputes candidate predictions and refreshes

$$R_{i,t}^{(v)} = \widehat{D}_{i,t}^{(v)} + 2L_{\mathcal{A},i,t}^{(v)} \epsilon_{i,t}^{(v)}, \quad \omega_{i,t}^{(s)} = \frac{(\overline{R}_{i,t}^{(s)} + \varepsilon_R)^{-1}}{\sum_{r\in\mathcal{P}_i} (\overline{R}_{i,t}^{(r)} + \varepsilon_R)^{-1}}, \quad \boldsymbol{Q}_{i,t} = \mathrm{softmax}\left(T_q^{-1} \sum_s \omega_{i,t}^{(s)} \boldsymbol{\ell}_{i,t}^{(s)}\right). \tag{12}$$

Eq. (12) refreshes the certificate and teacher at θ_t using Eqs. (6) and (9). The resulting risks also determine $\kappa_{i,t}$ and λ_t through Eq. (11). Second, fixing these detached quantities, define

$$\boldsymbol{g}_{F,t} = \nabla_\theta F(\theta_t), \qquad \boldsymbol{g}_{G,t} = \nabla_\theta G_t(\theta_t), \qquad \boldsymbol{d}_{0,t} = \boldsymbol{g}_{F,t} + \lambda_t \boldsymbol{g}_{G,t}, \tag{13}$$

where $G_t(\theta) = |\mathcal{I}_{\mathrm{m}}|^{-1} \sum_{i\in\mathcal{I}_{\mathrm{m}}} \kappa_{i,t} D_{\mathrm{KL}}(\mathrm{sg}(\boldsymbol{Q}_{i,t}) \,\|\, \boldsymbol{p}_i^{(f)}(\theta))$ is the guidance-only term, with $G_t = 0$ when $\mathcal{I}_{\mathrm{m}} = \varnothing$. The direction $\boldsymbol{d}_{0,t}$ in Eq. (13) combines the protected and guidance gradients before projection.

Third, for $0 < \mu \leq 1$, the minimum-change base-safe direction and network update are

$$\begin{aligned} \boldsymbol{d}_t^\star &= \arg \min_{\langle \boldsymbol{g}_{F,t}, \boldsymbol{d}\rangle \geq \mu \|\boldsymbol{g}_{F,t}\|_2^2} \tfrac{1}{2}\|\boldsymbol{d} - \boldsymbol{d}_{0,t}\|_2^2, \\ \alpha_t &= \frac{[\mu\|\boldsymbol{g}_{F,t}\|_2^2 - \langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_{0,t}\rangle]_+}{\|\boldsymbol{g}_{F,t}\|_2^2}, \qquad \boldsymbol{d}_t^\star = \boldsymbol{d}_{0,t} + \alpha_t \boldsymbol{g}_{F,t}, \qquad \theta_{t+1} = \theta_t - \eta_t \boldsymbol{d}_t^\star. \end{aligned} \tag{14}$$

Here, $[x]_+ = \max(x, 0)$, $\eta_t > 0$, and $\alpha_t = 0$ if $\boldsymbol{g}_{F,t} = \boldsymbol{0}$. The correction in Eq. (14) activates only when the unconstrained direction violates the base-safe constraint. The updated parameters then refresh Eq. (12).

Theorem 3 (**One-step base-objective descent**). *Assume that F is L_F-smooth with $L_F > 0$, $\boldsymbol{g}_{F,t} \neq \boldsymbol{0}$, and $0 < \mu \leq 1$. Let $\boldsymbol{d}_t^\star$ be the exact projection in Eq. (14). If $\theta_{t+1} = \theta_t - \eta_t \boldsymbol{d}_t^\star$ and*

$$0 < \eta_t \leq \frac{\mu\|\boldsymbol{g}_{F,t}\|_2^2}{L_F\|\boldsymbol{d}_t^\star\|_2^2}, \tag{15}$$

then $F(\theta_{t+1}) \leq F(\theta_t) - \frac{1}{2}\eta_t \mu \|\boldsymbol{g}_{F,t}\|_2^2$.

The step-size bound in Eq. (15) controls the smoothness remainder in the protected objective. The proof is provided in Appendix C.3. Training stops at the iteration limit or when the relative change in $\mathcal{L}_{\mathrm{RTG}}$ falls below $\varepsilon_{\mathrm{stop}} > 0$ (an empirical criterion; see Appendix B for pseudocode).

4 Experiment

4.1 Experimental Settings

Metrics. We evaluate clustering performance using clustering accuracy (ACC), normalized mutual information (NMI), and adjusted Rand index (ARI). Higher values indicate better clustering performance for all three metrics.

Datasets. We conduct experiments on eight benchmark datasets: NGs, NoisyMNIST, MSRC_v1, 3V_Fashion_MV, Caltech-5V, BBC, Handwritten, and 3-sources. Fixed stratified subsets of 5,000 samples are used for NoisyMNIST and 3V_Fashion_MV. Detailed statistics are reported in Table 1.

Table 1: Statistics of the eight benchmark datasets.

Dataset	Samples	Views	Classes
NGs	500	3	5
NoisyMNIST	5,000	2	10
MSRC_v1	210	5	7
3V_Fashion_MV	5,000	3	10
Caltech-5V	1,400	5	7
BBC	685	4	5
Handwritten	2,000	4	10
3-sources	169	3	6

Baselines. We compare RTG-IMVC with seven representative incomplete multi-view clustering methods: ICMVC (Chao et al., 2024), DSIMVC (Tang & Liu, 2022b), DVIMC (Xu et al., 2024), GHICMC (Chao et al., 2025), SIIHPC (Yu et al., 2025), HSACC (Ding et al., 2026), and DDR-IMVC (Wang et al., 2026a). Public implementations and recommended settings are used whenever available. All methods use consistent views, preprocessing, incomplete-view masks, and random seeds.

Parameter Settings. We investigate the candidate-topology K in Section 4.3. Section 4.4 jointly evaluates the edge-replacement ratio ρ and teacher temperature T_q over a two-dimensional parameter grid, and separately analyzes the base-descent margin μ. All remaining settings are held fixed in each analysis.

4.2 Experimental Results and Analysis

Table 2 presents the clustering results under missing-view rates $\eta \in \{0.1, 0.3, 0.5\}$, where the reported values are the mean $\pm$ standard deviation over ten random seeds. Fig. 3 further shows the ACC and NMI trends under missing rates from 0.1 to 0.8.

The following observations can be drawn:

- **Quantitative superiority.** RTG-IMVC achieves the best mean in 68 of 72 dataset–missing-rate–metric combinations. On NGs at $\eta = 0.1$, it improves ACC over DDR-IMVC from 88.46 to 96.80, a gain of 8.34 percentage points.
- **Cross-dataset generalization.** At $\eta = 0.1$, RTG-IMVC improves ACC by 8.31 percentage points over DSIMVC on MSRC_v1, by 8.37 and 8.45 over DDR-IMVC on 3V_Fashion_MV and BBC, respectively, and by 8.46 over HSACC on Caltech-5V, demonstrating effective cross-view topology guidance across heterogeneous datasets.
- **Robustness to missing views.** RTG-IMVC leads across all evaluated metrics and missing rateCs on NGs, MSRC_v1, 3V_Fashion_MV, BBC, and Handwritten, and in most settings on the remaining datasets. On NGs, it retains an ACC of 89.96% and an NMI of 73.50% at $\eta = 0.5$, compared with 96.80% and 89.84% at $\eta = 0.1$. Fig. 3 visualizes performance trends over missing rates 0.1–0.8; additional profiles and training dynamics appear in Appendix D.5.

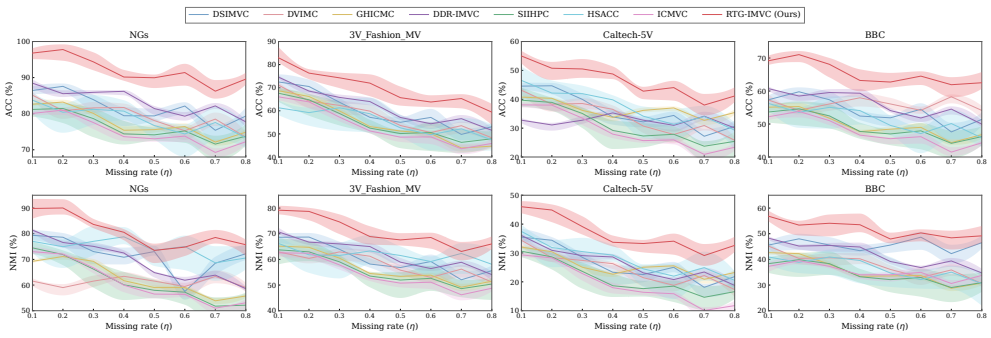


Figure 3: Clustering results on NGs, 3V_Fashion_MV, Caltech-5V, and BBC with different missing rates. The upper row reports ACC and the lower row reports NMI.

Table 2: Performance comparison under different missing rates on 8 datasets.

Missing Rate	Method	NGs			NoisyMNIST			MSRC_v1			3V_Fashion_MV		
		ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
0.1	DSIMVC	86.43 ± 0.69	79.36 ± 1.79	/	85.52 ± 0.70	/	80.90 ± 0.71	78.93 ± 1.19	67.75 ± 1.43	65.12 ± 1.21	72.49 ± 1.46	68.62 ± 1.70	62.01 ± 1.48
	DVIMC	85.13 ± 0.77	/	80.49 ± 1.68	84.27 ± 0.55	78.69 ± 0.71	79.65 ± 0.55	75.37 ± 1.41	64.55 ± 1.64	61.56 ± 1.43	71.11 ± 1.84	/	/
	ICMVC	80.08 ± 0.61	73.02 ± 1.67	75.43 ± 1.45	79.21 ± 0.38	/	74.60 ± 0.39	70.31 ± 1.21	59.49 ± 1.42	56.50 ± 1.20	66.05 ± 1.61	62.54 ± 1.85	55.57 ± 1.63
	SIIHPC	81.18 ± 0.52	74.49 ± 1.56	76.53 ± 1.34	/	75.05 ± 0.70	76.02 ± 0.54	71.63 ± 1.77	/	57.83 ± 1.77	67.61 ± 1.25	63.73 ± 1.47	57.12 ± 1.25
	GHICMC	82.62 ± 0.80	/	77.97 ± 1.73	81.72 ± 0.83	76.49 ± 0.40	77.46 ± 0.83	73.07 ± 1.37	61.90 ± 1.58	59.27 ± 1.37	68.69 ± 1.64	65.17 ± 1.85	58.20 ± 1.64
	HSACC	83.75 ± 0.49	77.04 ± 1.50	79.09 ± 1.28	83.25 ± 0.84	77.05 ± 0.43	78.62 ± 0.84	/	/	60.53 ± 1.83	/	66.20 ± 1.68	59.59 ± 1.46
	DDR-IMVC	88.46 ± 0.77	81.39 ± 1.89	83.80 ± 1.67	87.59 ± 0.44	82.00 ± 0.60	82.96 ± 0.44	77.07 ± 1.65	66.26 ± 1.88	63.27 ± 1.67	74.58 ± 1.29	70.71 ± 1.50	64.10 ± 1.28
	Ours	**96.80 ± 0.49**	**89.84 ± 1.77**	**92.15 ± 1.55**	**96.00 ± 0.20**	**90.51 ± 0.33**	**91.38 ± 0.41**	**87.24 ± 1.83**	**76.17 ± 1.25**	**73.44 ± 1.83**	**82.95 ± 1.12**	**79.18 ± 1.36**	**72.47 ± 1.14**
0.3	DSIMVC	83.76 ± 0.87	72.25 ± 0.44	75.59 ± 0.87	82.42 ± 0.89	74.26 ± 0.46	75.41 ± 1.82	76.23 ± 1.44	67.76 ± 1.68	62.89 ± 1.46	62.52 ± 1.71	63.64 ± 1.16	53.19 ± 1.73
	DVIMC	/	/	74.23 ± 0.36	81.42 ± 0.72	73.54 ± 0.89	73.78 ± 1.61	/	64.58 ± 1.10	59.48 ± 1.68	62.28 ± 1.28	61.48 ± 1.51	/
	ICMVC	77.14 ± 0.78	66.31 ± 0.37	68.47 ± 0.80	76.01 ± 0.55	68.13 ± 0.72	/	/	59.26 ± 1.67	54.67 ± 1.45	57.08 ± 1.86	57.46 ± 1.28	46.01 ± 1.86
	SIIHPC	77.54 ± 0.70	/	70.47 ± 0.70	77.60 ± 0.71	69.06 ± 0.89	70.05 ± 1.61	69.23 ± 1.21	60.59 ± 1.44	/	57.74 ± 1.48	59.01 ± 1.72	48.29 ± 1.50
	GHICMC	80.03 ± 0.40	68.15 ± 0.57	71.01 ± 0.41	78.62 ± 0.41	71.31 ± 0.59	71.76 ± 1.20	70.67 ± 1.60	61.57 ± 1.83	/	/	59.32 ± 1.31	49.28 ± 1.89
	HSACC	81.21 ± 0.66	/	72.03 ± 0.67	81.35 ± 0.42	72.57 ± 0.60	72.01 ± 1.22	72.02 ± 1.26	63.14 ± 1.50	/	61.19 ± 1.69	61.62 ± 1.12	50.02 ± 1.69
	DDR-IMVC	85.71 ± 0.36	73.08 ± 0.52	77.54 ± 0.36	83.39 ± 0.61	76.85 ± 0.78	77.05 ± 1.46	74.07 ± 1.11	66.83 ± 1.34	61.87 ± 1.12	65.21 ± 1.51	64.03 ± 1.75	55.83 ± 1.53
	Ours	**94.36 ± 0.43**	**83.63 ± 0.93**	**86.39 ± 0.99**	**93.30 ± 0.64**	**85.53 ± 0.78**	**85.91 ± 1.64**	**85.14 ± 1.27**	**76.34 ± 1.50**	**71.41 ± 1.29**	**74.28 ± 1.35**	**74.60 ± 1.58**	**64.05 ± 1.37**
0.5	DSIMVC	79.39 ± 0.45	/	65.93 ± 1.26	51.75 ± 1.25	/	37.68 ± 1.10	56.35 ± 1.67	48.65 ± 1.11	36.14 ± 1.69	55.03 ± 1.17	56.83 ± 1.38	44.09 ± 1.17
	DVIMC	78.09 ± 0.53	61.54 ± 1.58	64.65 ± 1.36	50.50 ± 1.84	/	35.75 ± 1.84	52.79 ± 1.12	45.45 ± 1.33	32.58 ± 1.11	53.65 ± 1.52	55.81 ± 1.76	42.71 ± 1.54
	ICMVC	73.04 ± 0.38	56.48 ± 1.35	59.59 ± 1.13	45.44 ± 1.62	44.99 ± 1.85	30.70 ± 1.63	47.73 ± 1.68	40.39 ± 1.13	27.52 ± 1.68	48.59 ± 1.30	50.75 ± 1.53	37.65 ± 1.32
	SIIHPC	74.14 ± 0.87	57.95 ± 1.24	60.69 ± 1.82	46.88 ± 1.84	/	32.12 ± 1.83	49.05 ± 1.46	41.72 ± 1.67	28.85 ± 1.46	50.15 ± 1.73	51.94 ± 1.15	39.20 ± 1.73
	GHICMC	75.58 ± 0.58	59.02 ± 1.63	62.13 ± 1.41	/	47.85 ± 1.64	33.56 ± 1.43	50.49 ± 1.84	/	30.29 ± 1.84	51.23 ± 1.32	53.38 ± 1.56	40.28 ± 1.34
	HSACC	76.71 ± 0.84	/	/	46.49 ± 1.45	49.91 ± 1.68	34.72 ± 1.46	51.75 ± 1.51	44.06 ± 1.73	31.55 ± 1.51	52.62 ± 1.13	/	41.67 ± 1.15
	DDR-IMVC	81.42 ± 0.53	64.85 ± 1.57	67.96 ± 1.36	53.82 ± 1.69	53.36 ± 1.12	**39.06 ± 1.69**	54.49 ± 1.35	47.16 ± 1.57	34.29 ± 1.35	57.12 ± 1.76	58.92 ± 1.18	46.18 ± 1.76
	Ours	**89.96 ± 0.48**	**73.50 ± 1.45**	**76.51 ± 1.24**	**64.62 ± 1.87**	**62.07 ± 1.32**	38.00 ± 1.27	**64.86 ± 1.52**	**57.27 ± 1.73**	**44.66 ± 1.51**	**65.69 ± 1.60**	**67.59 ± 1.83**	**54.75 ± 1.61**

Missing Rate	Method	Caltech-5V			BBC			Handwritten			3-sources		
		ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
0.1	DSIMVC	43.45 ± 1.69	35.27 ± 1.13	25.50 ± 0.80	/	/	40.32 ± 1.20	70.23 ± 1.76	69.20 ± 0.43	62.68 ± 1.78	44.28 ± 1.12	46.65 ± 1.36	30.21 ± 1.14
	DVIMC	43.12 ± 1.69	34.24 ± 1.13	/	57.32 ± 1.49	45.24 ± 0.82	38.16 ± 1.51	69.20 ± 1.86	67.43 ± 0.48	60.56 ± 1.85	44.28 ± 1.24	45.53 ± 1.47	29.11 ± 1.25
	ICMVC	38.16 ± 1.47	29.23 ± 1.70	19.20 ± 0.64	52.10 ± 1.29	/	34.06 ± 1.29	/	62.30 ± 0.92	55.43 ± 1.65	38.33 ± 1.80	40.72 ± 1.25	25.35 ± 1.82
	SIIHPC	39.57 ± 1.46	30.56 ± 1.69	20.60 ± 0.63	53.52 ± 1.76	/	35.46 ± 1.76	65.31 ± 1.53	63.75 ± 0.85	56.76 ± 1.55	40.40 ± 1.70	41.60 ± 1.12	26.22 ± 1.70
	GHICMC	40.64 ± 1.84	32.01 ± 1.29	21.82 ± 0.92	55.04 ± 1.35	42.54 ± 0.70	36.88 ± 1.35	72.56 ± 1.15	70.93 ± 0.54	63.99 ± 1.14	42.05 ± 1.30	43.04 ± 1.51	26.20 ± 1.29
	HSACC	46.53 ± 1.30	37.46 ± 1.53	27.70 ± 0.51	56.21 ± 1.12	/	38.03 ± 1.12	68.07 ± 1.46	66.44 ± 0.79	59.59 ± 1.48	43.18 ± 1.63	43.08 ± 1.87	27.18 ± 1.65
	DDR-IMVC	/	36.01 ± 1.23	26.29 ± 0.88	60.72 ± 1.42	48.27 ± 0.75	42.61 ± 1.42	71.11 ± 1.65	69.40 ± 0.92	62.43 ± 1.65	48.52 ± 1.88	48.67 ± 1.30	33.51 ± 1.88
	Ours	**54.99 ± 1.39**	**46.02 ± 1.62**	**36.16 ± 0.96**	**69.17 ± 1.59**	**56.83 ± 0.85**	**51.07 ± 1.39**	**80.98 ± 1.34**	**79.45 ± 0.71**	**72.41 ± 1.36**	**56.21 ± 1.10**	**57.31 ± 1.34**	**41.11 ± 1.12**
0.3	DSIMVC	39.86 ± 1.15	28.49 ± 1.36	20.02 ± 1.14	57.25 ± 1.43	45.59 ± 1.66	39.09 ± 1.44	/	68.06 ± 1.43	59.62 ± 1.22	47.02 ± 1.37	46.06 ± 1.60	32.21 ± 1.39
	DVIMC	37.42 ± 1.15	26.07 ± 1.36	18.27 ± 1.15	54.02 ± 1.74	/	37.25 ± 1.74	66.75 ± 1.29	64.56 ± 1.53	58.47 ± 1.31	47.32 ± 1.49	45.68 ± 1.70	32.53 ± 1.48
	ICMVC	32.37 ± 1.71	/	13.37 ± 1.71	51.02 ± 1.52	36.13 ± 1.75	32.02 ± 1.53	62.21 ± 1.86	61.01 ± 1.30	53.58 ± 1.87	41.11 ± 1.26	41.23 ± 1.49	26.57 ± 1.28
	SIIHPC	35.08 ± 1.71	23.22 ± 1.13	/	52.26 ± 1.19	38.55 ± 1.43	35.43 ± 1.21	64.52 ± 1.78	62.74 ± 1.22	55.53 ± 1.80	43.45 ± 1.14	42.21 ± 1.37	28.82 ± 1.16
	GHICMC	35.05 ± 1.30	25.04 ± 1.53	15.17 ± 1.32	/	/	35.27 ± 1.60	72.03 ± 1.37	70.23 ± 1.61	62.53 ± 1.39	44.52 ± 1.52	43.05 ± 1.76	29.36 ± 1.54
	HSACC	41.74 ± 1.55	30.23 ± 1.76	**22.05 ± 1.54**	54.01 ± 1.35	40.68 ± 1.58	36.52 ± 1.36	67.42 ± 1.69	/	58.24 ± 1.71	45.01 ± 1.88	44.29 ± 1.32	30.48 ± 1.11
	DDR-IMVC	/	28.54 ± 1.46	20.15 ± 1.25	59.44 ± 1.65	45.52 ± 1.88	40.10 ± 1.67	70.56 ± 1.88	68.37 ± 1.32	61.47 ± 1.11	50.28 ± 1.32	49.06 ± 1.55	35.12 ± 1.34
	Ours	**50.50 ± 1.63**	**39.14 ± 1.87**	20.66 ± 1.65	**68.09 ± 1.23**	**54.08 ± 1.46**	**49.76 ± 1.25**	**80.73 ± 1.59**	**78.65 ± 1.80**	**71.45 ± 1.59**	**59.05 ± 1.35**	**57.62 ± 1.56**	**43.93 ± 1.35**
0.5	DSIMVC	32.16 ± 1.37	22.53 ± 1.61	11.47 ± 1.39	51.88 ± 1.65	/	31.14 ± 1.67	/	53.18 ± 1.68	39.61 ± 1.47	44.08 ± 1.60	43.51 ± 1.83	27.25 ± 1.62
	DVIMC	30.73 ± 1.38	21.45 ± 1.61	12.04 ± 1.39	/	36.09 ± 1.41	29.88 ± 1.20	56.17 ± 1.54	/	38.46 ± 1.54	42.52 ± 1.71	42.03 ± 1.16	25.67 ± 1.73
	ICMVC	25.67 ± 1.15	16.39 ± 1.38	4.98 ± 1.17	45.55 ± 1.76	34.01 ± 1.19	24.82 ± 1.76	51.12 ± 1.31	46.61 ± 1.55	33.40 ± 1.33	37.76 ± 1.51	37.18 ± 1.72	20.91 ± 1.50
	SIIHPC	27.28 ± 1.14	17.65 ± 1.38	6.60 ± 1.16	46.99 ± 1.44	32.11 ± 1.66	26.26 ± 1.44	52.44 ± 1.24	47.94 ± 1.45	34.72 ± 1.23	39.41 ± 1.39	37.26 ± 1.60	22.30 ± 1.38
	GHICMC	/	/	7.68 ± 1.55	48.44 ± 1.83	33.55 ± 1.27	27.70 ± 1.83	59.53 ± 1.62	55.03 ± 1.84	41.82 ± 1.62	40.65 ± 1.77	38.70 ± 1.22	/
	HSACC	34.14 ± 1.77	24.51 ± 1.22	13.46 ± 1.79	49.58 ± 1.60	34.70 ± 1.81	28.85 ± 1.59	55.14 ± 1.15	50.64 ± 1.38	37.43 ± 1.16	41.58 ± 1.34	41.84 ± 1.55	24.64 ± 1.34
	DDR-IMVC	32.75 ± 1.48	23.11 ± 1.71	**18.06 ± 1.49**	54.07 ± 1.11	39.18 ± 1.32	33.33 ± 1.10	58.08 ± 1.34	/	40.36 ± 1.34	46.12 ± 1.57	45.53 ± 1.78	**29.27 ± 1.56**
	Ours	**42.80 ± 1.86**	**33.27 ± 1.30**	10.12 ± 1.88	**62.72 ± 1.48**	**47.94 ± 1.51**	**41.99 ± 1.49**	**68.15 ± 1.52**	**63.75 ± 1.26**	**50.44 ± 1.44**	**54.91 ± 1.38**	**54.07 ± 1.61**	26.07 ± 1.50

Table 3: Risk-formulation ablation on Caltech-5V.

No.	Variants	$\eta = 0.3$		$\eta = 0.7$	
		ACC(%) ↑	NMI(%) ↑	ACC(%) ↑	NMI(%) ↑
#1	Cover-only risk	42.17±2.54	30.10±2.80	22.11±1.46	12.38±0.65
#2	Diameter-only risk	40.42±2.68	29.10±3.03	21.65±1.90	11.50±1.58
#3	Entropy/JS heuristic	39.49±2.14	28.26±3.08	22.23±1.64	12.41±1.49
#4	**Full risk (Ours)**	**50.50±1.63**	**39.14±1.87**	**31.62±1.82**	**21.34±1.50**

Table 4: Consensus-teacher ablation on Caltech-5V.

No.	Variants	$\eta = 0.3$		$\eta = 0.7$	
		ACC(%) ↑	NMI(%) ↑	ACC(%) ↑	NMI(%) ↑
#1	Hard-max + square	41.69±2.93	29.88±4.46	20.66±1.78	11.81±1.33
#2	Unw. arithmetic mean	38.73±2.74	26.50±3.34	21.59±1.95	11.65±1.20
#3	Unw. LogOP	39.41±3.74	28.56±3.60	21.37±0.91	11.75±0.93
#4	**Certified LogOP (Ours)**	**50.50±1.63**	**39.14±1.87**	**31.62±1.82**	**21.34±1.50**

4.3 Ablation Study

Risk formulation. As shown in Table 3, Full risk achieves the highest ACC and NMI at both $\eta = 0.3$ and $\eta = 0.7$, reaching $50.50/39.14$ and $31.62/21.34$, respectively. Compared with **#1** at $\eta = 0.3$ and **#3** at $\eta = 0.7$, it improves ACC/NMI by $8.33/9.04$ and $9.39/8.93$ percentage points, respectively. These results indicate that combining finite-set prediction variation with the coverage correction is beneficial under both missing rates.

Consensus-teacher design. As shown in Table 4 ("Unw." denotes unweighted), Certified LogOP (**#4**) consistently achieves the best ACC and NMI under both missing-view rates. At $\eta = 0.3$, compared with Hard-max + square (**#1**), it improves ACC and NMI by 8.81 and 9.26 percentage points, respectively. These results demonstrate that reliability-aware consensus weighting provides more effective guidance than the alternative teacher designs, with the advantage remaining evident as the missing-view rate increases. Further results appear in Appendix D.2.

Candidate-topology set. As shown in Fig. 4, the Full set consistently outperformed the other topology variants under both missing-view rates. It attained peak ACC/NMI scores of $52.40/40.06$ at $\eta = 0.3$ and $32.60/23.00$ at $\eta = 0.7$. With the Full set, RTG-IMVC maintains a clear advantage over the other candidate-topology variants, even as ACC and NMI fluctuate.

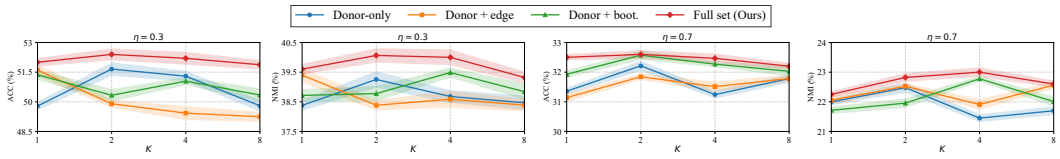


Figure 4: Ablation results of different candidate-topology sets on Caltech-5V.

4.4 Hyperparameter Analysis

Edge-replacement ratio ρ and teacher temperature T_q. The parameters ρ and T_q govern edge replacement in Eq. (4) and teacher sharpness in Eq. (9), respectively. On NoisyMNIST at $\eta = 0.3$, we vary $\rho \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ and $T_q \in \{0.25, 0.5, 0.75, 1.0, 2.0\}$. The experimental results indicate that RTG-IMVC is insensitive to these parameter variations, maintaining consistently stable performance (see Appendix E.1).

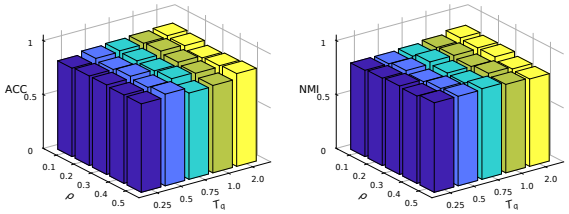


Figure 5: Hyperparameter sensitivity of RTG-IMVC on NoisyMNIST at a missing-view rate of $\eta = 0.3$.

Base-descent margin μ. μ controls the descent constraint in Eq. (14). Table 5 shows limited ACC/NMI variation under both missing-view rates, with peak performance at $\mu = 0.4$, demonstrating robustness to the choice of μ. Additional sweeps appear in Appendix D.3.

Table 5: Sensitivity to μ on NoisyMNIST.

μ	$\eta = 0.3$		$\eta = 0.7$	
	ACC	NMI	ACC	NMI
0.02	93.42 ± 0.48	86.20 ± 0.69	40.80 ± 4.67	41.19 ± 1.71
0.05	93.45 ± 0.53	86.01 ± 0.71	40.38 ± 3.71	41.26 ± 1.80
0.10	93.38 ± 0.71	86.64 ± 1.17	40.08 ± 1.43	41.32 ± 1.32
0.20	93.56 ± 0.76	86.19 ± 0.98	40.20 ± 4.06	41.36 ± 1.43
0.40	$\mathbf{93.62 \pm 0.74}$	$\mathbf{86.97 \pm 1.02}$	$\mathbf{40.81 \pm 3.94}$	$\mathbf{41.63 \pm 1.23}$

4.5 Running Time Comparison and Convergence Analysis

Fig. 6(a,b) shows convergence on MSRC_v1 and NoisyMNIST at $\eta = 0.1$. Loss decreases rapidly initially, while ACC and NMI rise and stabilize (see Appendix E.2 for additional curves). Fig. 6(c,d) compares per-iteration runtime on four datasets. RTG-IMVC achieves the lowest per-iteration runtime on all four datasets, complementing its rapid and stable convergence. See Appendices D.1 and E.3 for further timings.

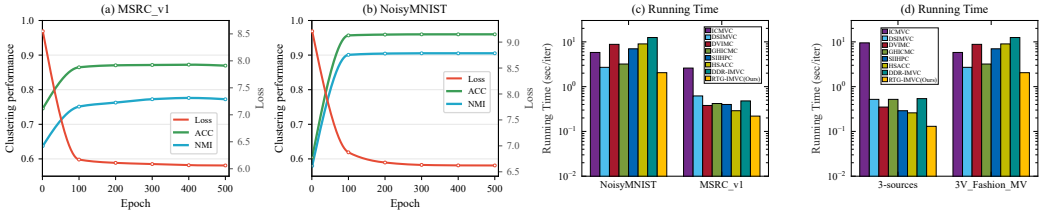


Figure 6: (a,b) RTG-IMVC convergence on MSRC_v1 and NoisyMNIST at $\eta = 0.1$. (c,d) Per-iteration running time of all compared methods on four datasets.

5 Conclusion

We presented RTG-IMVC to address unreliable relation transfer in IMVC. It estimates sample-level topology risk, regulates consensus guidance, and constrains updates through base-safe gradient projection. Under stated assumptions, we establish guarantees on topology sensitivity, expert influence, and one-step descent. Experiments across eight benchmarks demonstrate consistent gains and improved stability under topology perturbations. These results support reliability-aware topology guidance for IMVC. Future work will explore tighter coverage estimation and corrupted observed views.

REFERENCES

Aleksandar Bojchevski and Stephan Günnemann. Certifiable robustness to graph perturbations. *Advances in Neural Information Processing Systems*, 32, 2019. 4

Guoqing Chao, Shiliang Sun, and Jinbo Bi. A survey on multiview clustering. *IEEE transactions on artificial intelligence*, 2(2):146–168, 2021. 1

Guoqing Chao, Yi Jiang, and Dianhui Chu. Incomplete contrastive multi-view clustering with high-confidence guiding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 11221–11229, 2024. 1, 3, 4, 7

Guoqing Chao, Kaixin Xu, Xijiong Xie, and Yongyong Chen. Global graph propagation with hierarchical information transfer for incomplete contrastive multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 39, pp. 15713–15721, 2025. 7

Man-Sheng Chen, Jia-Qi Lin, Xiang-Long Li, Bao-Yu Liu, Chang-Dong Wang, Dong Huang, and Jian-Huang Lai. Representation learning in multi-view clustering: A literature review. *Data Science and Engineering*, 7(3):225–241, 2022. 1

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 *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5071–5081. IEEE, 2025. 1

Lucio M Dery, Yann Dauphin, and David Grangier. Auxiliary task update decomposition: The good, the bad and the neutral. *arXiv preprint arXiv:2108.11346*, 2021. 4

Xiaojian Ding, Lin Zhao, Xian Li, and Xiaoying Zhu. Incomplete multi-view clustering via hierarchical semantic alignment and cooperative completion. *Advances in Neural Information Processing Systems*, 38:52175–52195, 2026. 7

Zhibin Gu, Zhenhao Zhong, Xi Zhang, and Bing Li. Constant degree matrix-driven incomplete multi-view clustering via connectivity-structure and embedding tensor learning. In *International Conference on Learning Representations*, volume 2026, pp. 84649–84671, 2026. 1, 3

S Günnemann et al. Adversarial attacks on graph neural networks via meta learning. In *7th International Conference on Learning Representations, ICLR 2019*, 2019. 1

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *International conference on machine learning*, pp. 1321–1330. PMLR, 2017. 1

Tom Heskes. Selecting weighting factors in logarithmic opinion pools. *Advances in neural information processing systems*, 10, 1997. 2, 5

Yu-Guan Hsieh, James Thornton, Eugene Ndiaye, Michal Klein, Marco Cuturi, and Pierre Ablin. Careful with that scalpel: improving gradient surgery with an ema. *arXiv preprint arXiv:2402.02998*, 2024. 4

Menglei Hu and Songcan Chen. Doubly aligned incomplete multi-view clustering. *arXiv preprint arXiv:1903.02785*, 2019. 1

Xiaorui Jiang, Buyun He, Peng Yuan Zhou, Xinyue Chen, Jingcai Guo, Jie Xu, and Yong Liao. A unified framework to bridge complete and incomplete deep multi-view clustering under non-iid missing patterns. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 594–603. IEEE, 2025. 3

Jiaqi Jin, Siwei Wang, Zhibin Dong, Xinwang Liu, and En Zhu. Deep incomplete multi-view clustering with cross-view partial sample and prototype alignment. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11600–11609. IEEE, 2023. 1

Wei Jin, Yao Ma, Xiaorui Liu, Xianfeng Tang, Suhang Wang, and Jiliang Tang. Graph structure learning for robust graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 66–74, 2020. 4

Abhishek Kumar, Piyush Rai, and Hal Daume. Co-regularized multi-view spectral clustering. *Advances in neural information processing systems*, 24, 2011. 1

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

Shao-Yuan Li, Yuan Jiang, and Zhi-Hua Zhou. Partial multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 28, 2014. 1

Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. Completer: Incomplete multi-view clustering via contrastive prediction. In *2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 11169–11178. IEEE, 2021. 1, 3

Bo Liu, Xingchao Liu, Xiaojie Jin, Peter Stone, and Qiang Liu. Conflict-averse gradient descent for multi-task learning. *Advances in neural information processing systems*, 34:18878–18890, 2021. 1, 4

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 *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 26151–26161. IEEE, 2024a. 1

Suyuan Liu, Junpu Zhang, Yi Wen, Xihong Yang, Siwei Wang, Yi Zhang, En Zhu, Chang Tang, Long Zhao, and Xinwang Liu. Sample-level cross-view similarity learning for incomplete multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 14017–14025, 2024b. 3

David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. *Advances in neural information processing systems*, 30, 2017. 4

Aviv Navon, Aviv Shamsian, Idan Achituve, Haggai Maron, Kenji Kawaguchi, Gal Chechik, and Ethan Fetaya. Multi-task learning as a bargaining game. *arXiv preprint arXiv:2202.01017*, 2022. 1

Jingyu Pu, Chenhang Cui, Xinyue Chen, Yazhou Ren, Xiaorong Pu, Zhifeng Hao, Philip S Yu, and Lifang He. Adaptive feature imputation with latent graph for deep incomplete multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 14633–14641, 2024. 3

Ozan Sener and Vladlen Koltun. Multi-task learning as multi-objective optimization. *Advances in neural information processing systems*, 31, 2018. 1, 4

Huayi Tang and Yong Liu. Deep incomplete multi-view clustering with bi-level optimization framework. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 202–211, 2022a. 1

Huayi Tang and Yong Liu. Deep safe incomplete multi-view clustering: Theorem and algorithm. In *International conference on machine learning*, pp. 21090–21110. PMLR, 2022b. 1, 3, 7

Qianqian Wang, Xurui Liao, Wei Feng, and Quanxue Gao. Debiased and denoised representation learning for incomplete multi-view clustering. In *International Conference on Learning Representations*, volume 2026, pp. 73338–73350, 2026a. 1, 7

Siwei Wang, Xinwang Liu, Li Liu, Wenxuan Tu, Xinzhong Zhu, Jiyuan Liu, Sihang Zhou, and En Zhu. Highly-efficient incomplete largescale multiview clustering with consensus bipartite graph. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9766–9775. IEEE, 2022. 3

Youqing Wang, Yu Cao, Jinlu Wang, Xiang Xu, Jiapu Wang, Tengfei Liu, Junbin Gao, and Jipeng Guo. Aligning collaborative view recovery and tensorial subspace learning via latent representation for incomplete multi-view clustering. In *The Fourteenth International Conference on Learning Representations*, 2026b. 1, 3

Jie Wen, Yong Xu, and Hong Liu. Incomplete multiview spectral clustering with adaptive graph learning. *IEEE transactions on cybernetics*, 50(4):1418–1429, 2018. 1

Jie Wen, Zheng Zhang, Yong Xu, Bob Zhang, Lunke Fei, and Hong Liu. Unified embedding alignment with missing views inferring for incomplete multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 5393–5400, 2019. 1

Jie Wen, Zheng Zhang, Lunke Fei, Bob Zhang, Yong Xu, Zhao Zhang, and Jinxing Li. A survey on incomplete multiview clustering. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(2):1136–1149, 2022. 1

Jie Wen, Chengliang Liu, Gehui Xu, Zhihao Wu, Chao Huang, Lunke Fei, and Yong Xu. Highly confident local structure based consensus graph learning for incomplete multi-view clustering. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15712–15721. IEEE, 2023a. 1

Jie Wen, Shijie Deng, Waikeung Wong, Guoqing Chao, Chao Huang, Lunke Fei, and Yong Xu. Diffusion-based missing-view generation with the application on incomplete multi-view clustering. In *Forty-first international conference on machine learning*, 2024. 3

Yi Wen, Siwei Wang, Ke Liang, Weixuan Liang, Xinhang Wan, Xinwang Liu, Suyuan Liu, Jiyuan Liu, and En Zhu. Scalable incomplete multi-view clustering with structure alignment. In *Proceedings of the 31st ACM international conference on multimedia*, pp. 3031–3040, 2023b. 3

Gehui Xu, Jie Wen, Chengliang Liu, Bing Hu, Yicheng Liu, Lunke Fei, and Wei Wang. Deep variational incomplete multi-view clustering: Exploring shared clustering structures. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 16147–16155, 2024. 1, 3, 7

Jie Xu, Chao Li, Yazhou Ren, Liang Peng, Yujie Mo, Xiaoshuang Shi, and Xiaofeng Zhu. Deep incomplete multi-view clustering via mining cluster complementarity. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 8761–8769, 2022a. 3

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. IEEE, 2022b. 1

Mouxing Yang, Yunfan Li, Peng Hu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Robust multi-view clustering with incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):1055–1069, 2022. 3

Shengju Yu, Zhibin Dong, Siwei Wang, Pei Zhang, Yi Zhang, Xinwang Liu, Thomas Guan, Tiejun Li, and Yiu-ming Cheung. Simple yet effective incomplete multi-view clustering: Similarity-level imputation and intra-view hybrid-group prototype construction. In *International Conference on Learning Representations*, volume 2025, pp. 6971–6997, 2025. 7

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. *Advances in neural information processing systems*, 33: 5824–5836, 2020. 1, 4

Shuping Zhao, Chen Yulong, Jie Wen, Lunke Fei, Jinrong Cui, and Tingting Chai. Confident block diagonal structure-aware invariable graph completion for incomplete multi-view clustering. In *International Conference on Learning Representations*, volume 2026, pp. 99860–99870, 2026. 1, 3

Daniel Zügner and Stephan Günnemann. Certifiable robustness of graph convolutional networks under structure perturbations. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 1656–1665, 2020. 4

SUMMARY OF THE APPENDIX

This appendix provides supplementary material for our paper, *Risk-Certified Topology Guidance for Incomplete Multi-View Clustering*. It is organized as follows:

- §A summarizes the symbols used throughout the paper and their meanings.
- §B presents two algorithms detailing the procedures of RTG-IMVC.
- §C provides proofs of the theoretical results stated in the main text.
- §D reports supplementary experiments and analyses.
- §E extends the visual analyses presented in the main text to additional datasets.
- §F limitations and Visual Analysis.

A LIST OF SYMBOLS

A.1 LLM USAGE

We employ LLMs solely to refine the language and wording for improved fluency and precision, with no further application beyond this purpose.

A.2 LIST OF SYMBOLS

Table 6 summarizes the main symbols used in the paper.

Notation	Description	Index
$\mathcal{X},\ \boldsymbol{X}^{(v)}$	Multi-view data and view-v feature matrix	§3.1
$N,\ V,\ C,\ d_v$	Sample/view/cluster counts; feature dimension	§3.1
$\boldsymbol{M}$	Binary observation mask	§3.1
$\mathcal{O}_i,\ \mathcal{M}_i$	Observed and missing view sets	Eq. (1)
θ	Network parameters	§3.1
$\widetilde{\boldsymbol{a}}_i^{(v)}$	Nominal transferred adjacency row	§3.1
$\boldsymbol{z}_{i,\theta}^{(v)}(\boldsymbol{a})$	Topology-conditioned cluster logits	Eq. (2)
$\boldsymbol{p}_{i,\theta}^{(v)}(\boldsymbol{a})$	Topology-conditioned cluster probabilities	Eq. (2)
$\boldsymbol{\ell}_{i,\theta}^{(v)}(\boldsymbol{a})$	Logarithms of clipped cluster probabilities	Eq. (2)
$T_p,\ \varepsilon_p$	Prediction temperature and clipping floor	Eq. (2)
$\boldsymbol{p}_i^{(v)},\ \boldsymbol{p}_i^{(f)}$	Nominal view-specific and fused predictions	§3.1
$\rho,\ \beta$	Edge-replacement ratio and bootstrap strength	Eq. (4)
$\mathcal{D}_i^{(v)}, \mathcal{E}_i^{(v)}, \mathcal{B}_i^{(v)}$	Donor, replacement, and bootstrap candidate sets	§3.2
$\mathcal{S}_i^{(v)},\ \mathcal{U}_i^{(v)}$	Candidate and admissible topology sets	§3.2
$d_{\mathcal{A}}(\boldsymbol{a}, \boldsymbol{b})$	Topology distance, $\frac{1}{2}\lVert\boldsymbol{a}-\boldsymbol{b}\rVert_1$	§3.2
$\widehat{D}_i^{(v)}$	Finite-candidate log-prediction diameter	Eq. (6)
$\epsilon_i^{(v)},\ L_{\mathcal{A},i}^{(v)}$	Cover radius and log-prediction Lipschitz constant	Eq. (6)
$R_i^{(v)}$	Topology-transfer risk score	Eq. (6)
$\mathcal{P}_i,\ \mathcal{J}$	All expert indices and perturbed expert indices	§3.3
$\overline{R}_i^{(s)},\ R_i^{(f)}$	Assigned expert risk and fused-expert risk	Eq. (8)
$\varepsilon_R,\ T_q$	Risk floor and teacher temperature	Eq. (9)
$Z_i,\ \chi_i$	Weight normalizer and its reciprocal	§3.3
$\omega_i^{(s)}$	Normalized inverse-risk expert weight	Eq. (9)
$\boldsymbol{Q}_i$	Consensus-teacher distribution	Eq. (9)
$\mathcal{L}_{\mathrm{ins}},\ \mathcal{L}_{\mathrm{clu}}$	Instance- and cluster-level contrastive losses	Eq. (3)
$F(\theta),\ G_t(\theta)$	Protected loss and fixed-teacher guidance loss	§3.5
$\mathcal{I}_{\mathrm{m}}$	Sample indices with missing views	§3.4
$\kappa_i,\ \tau$	Sample guidance weight and reliability-decay scale	Eq. (11)
$\lambda_t,\ \lambda_0$	Adaptive guidance coefficient and its upper limit	Eq. (11)
$\mathcal{L}_{\mathrm{RTG}}$	Overall training objective	Eq. (11)
$\mathrm{sg}(\cdot)$	Stop-gradient operator	Eq. (11)
$\boldsymbol{g}_{F,t},\ \boldsymbol{g}_{G,t}$	Protected-loss and guidance gradients	Eq. (13)
$\boldsymbol{d}_{0,t},\ \boldsymbol{d}_t^{\star}$	Joint direction and its base-safe projection	Eq. (14)
$\mu,\ \alpha_t$	Descent margin and projection coefficient	Eq. (14)
$\eta_t,\ L_F$	Learning rate and smoothness constant of F	Eq. (15)
$\varepsilon_{\mathrm{stop}}$	Relative-loss-change stopping threshold	§3.5

i, v, s, t: sample, view, expert, and iteration; f: fused expert. Primes indicate perturbations; η is the missing-view rate, distinct from η_t.

Table 6: Notation and description of key symbols in RTG-IMVC.

B ALGORITHM DETAILS

Algorithm 1 constructs the risk-weighted teacher; Algorithm 2 uses it to train RTG-IMVC.

B.1 RISK AND TEACHER CONSTRUCTION

Algorithm 1 converts topology-induced prediction variation into expert risks, which determine the consensus teacher and guidance strength. Missing-view and fused predictions are evaluated under topology probes, while observed views receive zero topology-transfer risk. The returned teachers and weights are detached for the subsequent network update.

Algorithm 1: Risk-weighted teacher construction

Input : Current parameters θ_t; data $\mathcal{X}$; view sets $\{\mathcal{O}_i, \mathcal{M}_i\}$; nonempty $\mathcal{I}_{\mathrm{m}}$; prediction/probe settings and risk-estimation quantities; $\varepsilon_R, T_q, \tau, \lambda_0$.
Output : Detached teachers $\{\boldsymbol{Q}_{i,t}\}$, sample weights $\{\kappa_{i,t}\}$, and coefficient λ_t.
foreach $i \in \mathcal{I}_{\mathrm{m}}$ **do**
  `// Topology probing and risk`
  Construct $\mathcal{S}_i^{(v)}$ for each $v \in \mathcal{M}_i$ using Eqs. (4)–(5)
  Evaluate nominal expert predictions and candidate predictions, including fusion
  Compute missing-view and fused risks by Eq. (6)
  Assign $\overline{R}_{i,t}^{(s)}$ by Eq. (8), with zero risk for observed views
  `// Consensus and guidance weights`
  Compute $\omega_{i,t}^{(s)}$ and $\boldsymbol{Q}_{i,t}$ by Eq. (9)
  Compute $\kappa_{i,t}$ by Eq. (11)
Compute λ_t by Eq. (11)
return detached $\{\boldsymbol{Q}_{i,t}, \kappa_{i,t}\}$ and λ_t

B.2 OVERALL TRAINING PROCEDURE

Algorithm 2 alternates teacher refresh with network optimization. At each iteration, Algorithm 1 provides fixed targets and weights for the guidance loss. The joint gradient is then projected to satisfy the base-descent constraint before updating the network. When no views are missing, the guidance term is disabled.

Algorithm 2: Training RTG-IMVC

Input : Incomplete data $\mathcal{X}$; mask $\boldsymbol{M}$; cluster number C; initial parameters θ_0; iteration limit T; learning rates $\{\eta_t\}$; model, probe, and risk-estimation settings.
Output : Trained parameters θ_t.
`// Initialization`
Set $t = 0$ and obtain $\mathcal{O}_i$, $\mathcal{M}_i$, and $\mathcal{I}_{\mathrm{m}}$ from $\boldsymbol{M}$
while $t < T$ **do**
  `// Teacher refresh`
  if $\mathcal{I}_{\mathrm{m}} \neq \varnothing$ **then**
    Obtain $\boldsymbol{Q}_{i,t}$, $\kappa_{i,t}$, and λ_t using Algorithm 1
    Form G_t with these quantities fixed, as in Section 3.5
  else
    Set $G_t = 0$ and $\lambda_t = 0$
  `// Base-safe update`
  Compute $\boldsymbol{g}_{F,t}$, $\boldsymbol{g}_{G,t}$, and $\boldsymbol{d}_{0,t}$ by Eq. (13)
  Compute $\boldsymbol{d}_t^\star$ by Eq. (14), taking $\alpha_t = 0$ if $\boldsymbol{g}_{F,t} = \boldsymbol{0}$
  Update $\theta_{t+1} = \theta_t - \eta_t \boldsymbol{d}_t^\star$; set $t \leftarrow t + 1$
  if *the stopping rule in Section 3.5 is met* **then**
    break
return θ_t

C THEORETICAL PROOFS

We prove the three results in Section 3 using the main-text notation. Predictions within each certificate share the same parameters; in Section C.1, the fixed θ is suppressed in $\boldsymbol{\ell}_i^{(v)}(\boldsymbol{a})$.

C.1 PROOF OF CERTIFIED TOPOLOGY STABILITY

Fix a missing sample–view pair (i, v) and θ. Theorem 1 assumes that the finite, nonempty probe set $\mathcal{S}_i^{(v)} \subseteq \mathcal{U}_i^{(v)}$ covers the admissible topologies and that the log-prediction map is Lipschitz on this set. More explicitly,

$$\begin{aligned} &\sup_{\boldsymbol{a}\in\mathcal{U}_i^{(v)}} \min_{\boldsymbol{b}\in\mathcal{S}_i^{(v)}} d_{\mathcal{A}}(\boldsymbol{a},\boldsymbol{b}) \le \epsilon_i^{(v)}, \qquad d_{\mathcal{A}}(\boldsymbol{a},\boldsymbol{b}) = \tfrac{1}{2}\|\boldsymbol{a}-\boldsymbol{b}\|_1, \\ &\|\boldsymbol{\ell}_i^{(v)}(\boldsymbol{a}) - \boldsymbol{\ell}_i^{(v)}(\boldsymbol{b})\|_\infty \le L_{\mathcal{A},i}^{(v)} d_{\mathcal{A}}(\boldsymbol{a},\boldsymbol{b}), \qquad \boldsymbol{a},\boldsymbol{b} \in \mathcal{U}_i^{(v)}. \end{aligned} \tag{16}$$

The first condition connects finite probes to the entire admissible set; the second controls prediction changes between a topology and a nearby probe.

Proof. **Finite-probe diameter.** For any fixed coordinate c, its maximum absolute difference over the probes equals its range. Since both the coordinate set and probe set are finite, the maxima can be interchanged, giving the exact identity

$$\begin{aligned} \widehat{D}_i^{(v)} &= \max_{1\le c\le C}\left[\max_{\boldsymbol{b}\in\mathcal{S}_i^{(v)}} \ell_{ic}^{(v)}(\boldsymbol{b}) - \min_{\boldsymbol{b}'\in\mathcal{S}_i^{(v)}} \ell_{ic}^{(v)}(\boldsymbol{b}')\right] \\ &= \max_{1\le c\le C} \max_{\boldsymbol{b},\boldsymbol{b}'\in\mathcal{S}_i^{(v)}} \left|\ell_{ic}^{(v)}(\boldsymbol{b}) - \ell_{ic}^{(v)}(\boldsymbol{b}')\right| \\ &= \max_{\boldsymbol{b},\boldsymbol{b}'\in\mathcal{S}_i^{(v)}} \max_{1\le c\le C} \left|\ell_{ic}^{(v)}(\boldsymbol{b}) - \ell_{ic}^{(v)}(\boldsymbol{b}')\right| = \max_{\boldsymbol{b},\boldsymbol{b}'\in\mathcal{S}_i^{(v)}} \|\boldsymbol{\ell}_i^{(v)}(\boldsymbol{b}) - \boldsymbol{\ell}_i^{(v)}(\boldsymbol{b}')\|_\infty. \end{aligned} \tag{17}$$

Thus, the coordinate-wise computation in Eq. (6) recovers the full ℓ_∞ diameter on the probe set, even if different coordinates attain their extrema at different probes.

Extension to admissible topologies. Take arbitrary $\boldsymbol{a}, \boldsymbol{a}' \in \mathcal{U}_i^{(v)}$. The covering condition is equivalently expressed as

$$\begin{aligned} &\sup_{\boldsymbol{u}\in\mathcal{U}_i^{(v)}} \min_{\boldsymbol{b}\in\mathcal{S}_i^{(v)}} d_{\mathcal{A}}(\boldsymbol{u},\boldsymbol{b}) \le \epsilon_i^{(v)} \\ &\quad\iff\quad \forall \boldsymbol{u} \in \mathcal{U}_i^{(v)},\ \exists \boldsymbol{b} \in \mathcal{S}_i^{(v)} : d_{\mathcal{A}}(\boldsymbol{u},\boldsymbol{b}) \le \epsilon_i^{(v)}. \end{aligned} \tag{18}$$

Choose such probes $\boldsymbol{b}$ and $\boldsymbol{b}'$ for $\boldsymbol{a}$ and $\boldsymbol{a}'$, respectively. Adding and subtracting the two probe predictions yields, for each c,

$$\begin{aligned} &\left|\ell_{ic}^{(v)}(\boldsymbol{a}) - \ell_{ic}^{(v)}(\boldsymbol{a}')\right| \\ &= \left|\ell_{ic}^{(v)}(\boldsymbol{a}) - \ell_{ic}^{(v)}(\boldsymbol{b}) + \ell_{ic}^{(v)}(\boldsymbol{b}) - \ell_{ic}^{(v)}(\boldsymbol{b}') + \ell_{ic}^{(v)}(\boldsymbol{b}') - \ell_{ic}^{(v)}(\boldsymbol{a}')\right| \\ &\le \left|\ell_{ic}^{(v)}(\boldsymbol{a}) - \ell_{ic}^{(v)}(\boldsymbol{b})\right| + \left|\ell_{ic}^{(v)}(\boldsymbol{b}) - \ell_{ic}^{(v)}(\boldsymbol{b}')\right| + \left|\ell_{ic}^{(v)}(\boldsymbol{b}') - \ell_{ic}^{(v)}(\boldsymbol{a}')\right| \\ &\le L_{\mathcal{A},i}^{(v)} d_{\mathcal{A}}(\boldsymbol{a},\boldsymbol{b}) + \widehat{D}_i^{(v)} + L_{\mathcal{A},i}^{(v)} d_{\mathcal{A}}(\boldsymbol{b}',\boldsymbol{a}') \\ &\le L_{\mathcal{A},i}^{(v)}\epsilon_i^{(v)} + \widehat{D}_i^{(v)} + L_{\mathcal{A},i}^{(v)}\epsilon_i^{(v)} = \widehat{D}_i^{(v)} + 2L_{\mathcal{A},i}^{(v)}\epsilon_i^{(v)} = R_i^{(v)}. \end{aligned} \tag{19}$$

The first inequality is the scalar triangle inequality, and the second uses the Lipschitz bound and the exact probe diameter. Taking the maximum over coordinates and the supremum over admissible pairs yields

$$\begin{aligned} \sup_{\boldsymbol{a},\boldsymbol{a}'\in\mathcal{U}_i^{(v)}} \|\boldsymbol{\ell}_i^{(v)}(\boldsymbol{a}) - \boldsymbol{\ell}_i^{(v)}(\boldsymbol{a}')\|_\infty &= \sup_{\boldsymbol{a},\boldsymbol{a}'\in\mathcal{U}_i^{(v)}} \max_{1\le c\le C} \left|\ell_{ic}^{(v)}(\boldsymbol{a}) - \ell_{ic}^{(v)}(\boldsymbol{a}')\right| \\ &\le \widehat{D}_i^{(v)} + 2L_{\mathcal{A},i}^{(v)}\epsilon_i^{(v)} = R_i^{(v)}. \end{aligned} \tag{20}$$

This proves Theorem 1. □

Interpretation of the cover correction. The factor 2 accounts for moving both endpoints to the probe set. For comparison with the nominal topology, which belongs to $\mathcal{S}_i^{(v)}$ by Eq. (5), only one endpoint needs this correction:

$$\begin{aligned}\|\boldsymbol{\ell}_i^{(v)}(\boldsymbol{a}) - \boldsymbol{\ell}_i^{(v)}(\widetilde{\boldsymbol{a}}_i^{(v)})\|_\infty &\le \|\boldsymbol{\ell}_i^{(v)}(\boldsymbol{a}) - \boldsymbol{\ell}_i^{(v)}(\boldsymbol{b})\|_\infty + \|\boldsymbol{\ell}_i^{(v)}(\boldsymbol{b}) - \boldsymbol{\ell}_i^{(v)}(\widetilde{\boldsymbol{a}}_i^{(v)})\|_\infty \\ &\le L_{\mathcal{A},i}^{(v)}\epsilon_i^{(v)} + \widehat{D}_i^{(v)} \le R_i^{(v)}, \qquad \boldsymbol{a} \in \mathcal{U}_i^{(v)}.\end{aligned} \tag{21}$$

The main-text certificate uses the uniform two-endpoint bound. If the admissible set is exactly the probe set, one may take $\epsilon_i^{(v)} = 0$ and the certificate equals the exact diameter. For a larger admissible set, finite probes alone do not certify the cover radius or Lipschitz constant. Without valid upper bounds on both quantities, $R_i^{(v)}$ remains an empirical sensitivity score. In either case, sensitivity control does not establish clustering correctness.

C.2 Proof of Bounded Expert Influence

Fix sample i, temperature $T_q > 0$, and the risks $\{\overline{R}_i^{(s)}\}_{s\in\mathcal{P}_i}$. The weights $\omega_i^{(s)}$, their normalizer Z_i, and $\chi_i = Z_i^{-1}$ therefore remain unchanged. Let $\mathcal{J} \subseteq \mathcal{M}_i \cup \{f\}$ contain all perturbed experts, so that

$$\|\boldsymbol{\ell}_i'^{(s)} - \boldsymbol{\ell}_i^{(s)}\|_\infty \le \overline{R}_i^{(s)} \quad (s \in \mathcal{J}), \quad \boldsymbol{\ell}_i'^{(s)} = \boldsymbol{\ell}_i^{(s)} \quad (s \notin \mathcal{J}). \tag{22}$$

For missing-view experts, these bounds follow from Theorem 1 under its assumptions. If fusion changes, the fused expert must also be included in $\mathcal{J}$, with a valid bound for its own log-predictions.

Proof. **Cancellation of the teacher normalizer.** Clipping makes expert log-predictions finite, and the exponential pool in Eq. (9) therefore has strictly positive coordinates. For any clusters c, k, the softmax denominator cancels exactly:

$$\begin{aligned}\frac{Q_{ic}}{Q_{ik}} &= \frac{\exp\left[T_q^{-1}\sum_{s\in\mathcal{P}_i}\omega_i^{(s)}\ell_{ic}^{(s)}\right]}{\exp\left[T_q^{-1}\sum_{s\in\mathcal{P}_i}\omega_i^{(s)}\ell_{ik}^{(s)}\right]} \\ &= \exp\left[\frac{1}{T_q}\sum_{s\in\mathcal{P}_i}\omega_i^{(s)}\big(\ell_{ic}^{(s)} - \ell_{ik}^{(s)}\big)\right], \qquad \log\frac{Q_{ic}}{Q_{ik}} = \frac{1}{T_q}\sum_{s\in\mathcal{P}_i}\omega_i^{(s)}\big(\ell_{ic}^{(s)} - \ell_{ik}^{(s)}\big).\end{aligned} \tag{23}$$

The same identity applies to $\boldsymbol{Q}_i'$ with the original weights. Subtracting the two expressions, and then removing unchanged experts, gives

$$\begin{aligned}\Delta\log\frac{Q_{ic}}{Q_{ik}} &= \log\frac{Q'_{ic}}{Q'_{ik}} - \log\frac{Q_{ic}}{Q_{ik}} \\ &= \frac{1}{T_q}\sum_{s\in\mathcal{P}_i}\omega_i^{(s)}\left[\big(\ell_{ic}'^{(s)} - \ell_{ik}'^{(s)}\big) - \big(\ell_{ic}^{(s)} - \ell_{ik}^{(s)}\big)\right] \\ &= \frac{1}{T_q}\sum_{s\in\mathcal{J}}\omega_i^{(s)}\left[\big(\ell_{ic}'^{(s)} - \ell_{ic}^{(s)}\big) - \big(\ell_{ik}'^{(s)} - \ell_{ik}^{(s)}\big)\right].\end{aligned} \tag{24}$$

Thus, the comparison reduces to a weighted sum of coordinate differences; no separate bound on either teacher normalizer is needed.

From expert perturbations to teacher influence. Since $\overline{R}_i^{(s)} \geq 0$ and $\varepsilon_R > 0$, all weights are positive and sum to one. Applying the triangle inequality before the coordinate-wise perturbation bound yields

$$\begin{aligned}\left|\Delta \log \frac{Q_{ic}}{Q_{ik}}\right| &\leq \frac{1}{T_q}\sum_{s\in\mathcal{J}} \omega_i^{(s)} \left|\left(\ell_{ic}'^{(s)} - \ell_{ic}^{(s)}\right) - \left(\ell_{ik}'^{(s)} - \ell_{ik}^{(s)}\right)\right| \\ &\leq \frac{1}{T_q}\sum_{s\in\mathcal{J}} \omega_i^{(s)} \left(\left|\ell_{ic}'^{(s)} - \ell_{ic}^{(s)}\right| + \left|\ell_{ik}'^{(s)} - \ell_{ik}^{(s)}\right|\right) \\ &\leq \frac{2}{T_q}\sum_{s\in\mathcal{J}} \omega_i^{(s)} \|\boldsymbol{\ell}_i'^{(s)} - \boldsymbol{\ell}_i^{(s)}\|_\infty \leq \frac{2}{T_q}\sum_{s\in\mathcal{J}} \omega_i^{(s)} \overline{R}_i^{(s)}.\end{aligned} \tag{25}$$

The factor 2 arises because an expert can change both coordinates in the log-odds. The inverse-risk rule then bounds each expert's weighted contribution through the exact decomposition

$$\begin{aligned}\omega_i^{(s)}\overline{R}_i^{(s)} &= \frac{1}{Z_i}\frac{\overline{R}_i^{(s)}}{\overline{R}_i^{(s)} + \varepsilon_R} = \chi_i\left(1 - \frac{\varepsilon_R}{\overline{R}_i^{(s)} + \varepsilon_R}\right) = \chi_i - \varepsilon_R\omega_i^{(s)}, \\ \sum_{s\in\mathcal{J}} \omega_i^{(s)}\overline{R}_i^{(s)} &= |\mathcal{J}|\chi_i - \varepsilon_R\sum_{s\in\mathcal{J}}\omega_i^{(s)} \leq |\mathcal{J}|\chi_i.\end{aligned} \tag{26}$$

Combining the last two displays proves the theorem:

$$\left|\Delta \log \frac{Q_{ic}}{Q_{ik}}\right| \leq \frac{2}{T_q}\sum_{s\in\mathcal{J}} \omega_i^{(s)}\overline{R}_i^{(s)} = \frac{2}{T_q}\left(|\mathcal{J}|\chi_i - \varepsilon_R\sum_{s\in\mathcal{J}}\omega_i^{(s)}\right) \leq \frac{2|\mathcal{J}|\chi_i}{T_q}. \tag{27}$$

For $\mathcal{J} = \varnothing$, the teachers coincide. This proves Theorem 2. □

Equivalent odds-ratio statement. Exponentiation preserves order, so the same conclusion can be expressed without logarithms:

$$\begin{aligned}&\left|\log \frac{Q'_{ic}Q_{ik}}{Q'_{ik}Q_{ic}}\right| \leq \frac{2}{T_q}\sum_{s\in\mathcal{J}} \omega_i^{(s)}\overline{R}_i^{(s)} \\ \iff \quad &\exp\left[-\frac{2}{T_q}\sum_{s\in\mathcal{J}} \omega_i^{(s)}\overline{R}_i^{(s)}\right] \leq \frac{Q'_{ic}/Q'_{ik}}{Q_{ic}/Q_{ik}} \leq \exp\left[\frac{2}{T_q}\sum_{s\in\mathcal{J}} \omega_i^{(s)}\overline{R}_i^{(s)}\right].\end{aligned} \tag{28}$$

This is a multiplicative control on relative cluster preference. It does not assert that either teacher predicts the correct cluster. Under the main-text convention $\overline{R}_i^{(s)} = 0$ for observed views and $\mathcal{O}_i \neq \varnothing$, the normalizer further satisfies

$$Z_i = \frac{|\mathcal{O}_i|}{\varepsilon_R} + \sum_{s\in\mathcal{M}_i\cup\{f\}} \frac{1}{\overline{R}_i^{(s)} + \varepsilon_R} \geq \frac{|\mathcal{O}_i|}{\varepsilon_R} \quad \Longrightarrow \quad \frac{2|\mathcal{J}|\chi_i}{T_q} \leq \frac{2|\mathcal{J}|\varepsilon_R}{T_q|\mathcal{O}_i|}. \tag{29}$$

This consequence uses the assigned topology risks and the assumption that observed experts are unchanged; zero assigned risk is not a statement of semantic accuracy. The entire comparison holds the risks and weights fixed. Recomputing them after perturbation changes the weighted log-predictions as well, introducing a weight-variation term outside Theorem 2. Stop-gradient during an update does not control such changes between weight refreshes.

C.3 Proof of One-Step Base-Objective Descent

Fix iteration t, let $\boldsymbol{g}_{F,t} = \nabla F(\theta_t) \neq \mathbf{0}$, and take $0 < \mu \leq 1$. Assume that ∇F is L_F-Lipschitz, with $L_F > 0$, on a region containing the entire update segment. We first derive the minimum-change projection in Eq. (14) and then establish the finite-step decrease. All quantities below use the detached risks and teacher of the current iteration.

Proof. **Equivalent projection constraint.** Writing the correction directly as $\boldsymbol{d} - \boldsymbol{d}_{0,t}$, feasibility is equivalent to

$$\langle \boldsymbol{g}_{F,t}, \boldsymbol{d} \rangle \geq \mu \|\boldsymbol{g}_{F,t}\|_2^2 \quad \Longleftrightarrow \quad \langle \boldsymbol{g}_{F,t}, \boldsymbol{d} - \boldsymbol{d}_{0,t} \rangle \geq \mu \|\boldsymbol{g}_{F,t}\|_2^2 - \langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_{0,t} \rangle. \tag{30}$$

The feasible set is a nonempty closed half-space: $\boldsymbol{d} = \boldsymbol{g}_{F,t}$ is feasible because $\mu \leq 1$. The squared-distance objective is coercive and strictly convex, so it has a unique minimizer on this set.

If $\langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_{0,t} \rangle \geq \mu \|\boldsymbol{g}_{F,t}\|_2^2$, the original direction is already feasible and attains the minimum value zero. Hence $\boldsymbol{d}_t^\star = \boldsymbol{d}_{0,t}$ and $\alpha_t = 0$. Otherwise, every feasible direction must satisfy

$$\begin{aligned} 0 < \mu \|\boldsymbol{g}_{F,t}\|_2^2 - \langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_{0,t} \rangle \leq \langle \boldsymbol{g}_{F,t}, \boldsymbol{d} - \boldsymbol{d}_{0,t} \rangle \leq \|\boldsymbol{g}_{F,t}\|_2 \, \|\boldsymbol{d} - \boldsymbol{d}_{0,t}\|_2, \\ \frac{1}{2} \|\boldsymbol{d} - \boldsymbol{d}_{0,t}\|_2^2 \geq \frac{\big(\mu \|\boldsymbol{g}_{F,t}\|_2^2 - \langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_{0,t} \rangle\big)^2}{2 \|\boldsymbol{g}_{F,t}\|_2^2}. \end{aligned} \tag{31}$$

The last inequality is Cauchy–Schwarz. Equality is attained by a correction parallel to $\boldsymbol{g}_{F,t}$ that places the direction on the boundary. Combining the feasible and infeasible cases recovers exactly

$$\alpha_t = \frac{[\mu \|\boldsymbol{g}_{F,t}\|_2^2 - \langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_{0,t} \rangle]_+}{\|\boldsymbol{g}_{F,t}\|_2^2}, \qquad \boldsymbol{d}_t^\star = \boldsymbol{d}_{0,t} + \alpha_t \boldsymbol{g}_{F,t}. \tag{32}$$

In particular, direct substitution gives the exact protected-gradient alignment:

$$\begin{aligned} \langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_t^\star \rangle &= \langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_{0,t} \rangle + [\mu \|\boldsymbol{g}_{F,t}\|_2^2 - \langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_{0,t} \rangle]_+ \\ &= \max\big\{ \langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_{0,t} \rangle, \mu \|\boldsymbol{g}_{F,t}\|_2^2 \big\} \geq \mu \|\boldsymbol{g}_{F,t}\|_2^2 > 0. \end{aligned} \tag{33}$$

This also proves $\boldsymbol{d}_t^\star \neq \boldsymbol{0}$, so the step-size bound in Theorem 3 has a positive denominator.

Verification of minimum change. The solution can also be checked directly against every feasible $\boldsymbol{d}$, without introducing a new optimization variable or Lagrange multiplier. Equation (32) implies

$$\alpha_t \geq 0, \qquad \langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_t^\star \rangle - \mu \|\boldsymbol{g}_{F,t}\|_2^2 \geq 0, \qquad \alpha_t \big(\langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_t^\star \rangle - \mu \|\boldsymbol{g}_{F,t}\|_2^2 \big) = 0. \tag{34}$$

Expanding the squared distances and using this complementary relation yields

$$\begin{aligned} &\tfrac{1}{2} \|\boldsymbol{d} - \boldsymbol{d}_{0,t}\|_2^2 - \tfrac{1}{2} \|\boldsymbol{d}_t^\star - \boldsymbol{d}_{0,t}\|_2^2 \\ &= \tfrac{1}{2} \|\boldsymbol{d} - \boldsymbol{d}_t^\star\|_2^2 + \langle \boldsymbol{d} - \boldsymbol{d}_t^\star, \boldsymbol{d}_t^\star - \boldsymbol{d}_{0,t} \rangle \\ &= \tfrac{1}{2} \|\boldsymbol{d} - \boldsymbol{d}_t^\star\|_2^2 + \alpha_t \big(\langle \boldsymbol{g}_{F,t}, \boldsymbol{d} \rangle - \langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_t^\star \rangle \big) \\ &= \tfrac{1}{2} \|\boldsymbol{d} - \boldsymbol{d}_t^\star\|_2^2 + \alpha_t \big(\langle \boldsymbol{g}_{F,t}, \boldsymbol{d} \rangle - \mu \|\boldsymbol{g}_{F,t}\|_2^2 \big) \geq \tfrac{1}{2} \|\boldsymbol{d} - \boldsymbol{d}_t^\star\|_2^2 \geq 0. \end{aligned} \tag{35}$$

Equality of the objective values forces $\boldsymbol{d} = \boldsymbol{d}_t^\star$, verifying both global optimality and uniqueness of the stated projection.

Smoothness along the actual update. For $\theta_{t+1} = \theta_t - \eta_t \boldsymbol{d}_t^\star$, the fundamental theorem of calculus and the Lipschitz-gradient assumption give

$$\begin{aligned} &F(\theta_{t+1}) - F(\theta_t) + \eta_t \langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_t^\star \rangle \\ &= - \int_0^{\eta_t} \langle \nabla F(\theta_t - u \boldsymbol{d}_t^\star) - \nabla F(\theta_t), \boldsymbol{d}_t^\star \rangle \, du \\ &\leq \int_0^{\eta_t} \|\nabla F(\theta_t - u \boldsymbol{d}_t^\star) - \nabla F(\theta_t)\|_2 \, \|\boldsymbol{d}_t^\star\|_2 \, du \\ &\leq \int_0^{\eta_t} L_F u \|\boldsymbol{d}_t^\star\|_2^2 \, du = \frac{L_F \eta_t^2}{2} \|\boldsymbol{d}_t^\star\|_2^2. \end{aligned} \tag{36}$$

Here u is only an integration variable. Rearranging and substituting Eq. (33) gives the descent chain

$$\begin{aligned} F(\theta_{t+1}) &\leq F(\theta_t) - \eta_t \langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_t^\star \rangle + \frac{L_F \eta_t^2}{2} \|\boldsymbol{d}_t^\star\|_2^2 \\ &\leq F(\theta_t) - \eta_t \mu \|\boldsymbol{g}_{F,t}\|_2^2 + \frac{L_F \eta_t^2}{2} \|\boldsymbol{d}_t^\star\|_2^2 \\ &= F(\theta_t) - \eta_t \left(\mu \|\boldsymbol{g}_{F,t}\|_2^2 - \frac{L_F \eta_t}{2} \|\boldsymbol{d}_t^\star\|_2^2 \right). \end{aligned} \tag{37}$$

For $\eta_t > 0$, the theorem's learning-rate condition is equivalent to requiring that the quadratic remainder consume at most half of the guaranteed linear decrease:

$$\begin{aligned} \eta_t \le \frac{\mu\|\boldsymbol{g}_{F,t}\|_2^2}{L_F\|\boldsymbol{d}_t^\star\|_2^2} \quad &\iff \quad L_F\eta_t\|\boldsymbol{d}_t^\star\|_2^2 \le \mu\|\boldsymbol{g}_{F,t}\|_2^2 \\ &\iff \quad \frac{L_F\eta_t^2}{2}\|\boldsymbol{d}_t^\star\|_2^2 \le \frac{\eta_t\mu}{2}\|\boldsymbol{g}_{F,t}\|_2^2. \end{aligned} \tag{38}$$

Consequently,

$$F(\theta_{t+1}) \le F(\theta_t) - \eta_t\mu\|\boldsymbol{g}_{F,t}\|_2^2 + \tfrac{1}{2}\eta_t\mu\|\boldsymbol{g}_{F,t}\|_2^2 = F(\theta_t) - \tfrac{1}{2}\eta_t\mu\|\boldsymbol{g}_{F,t}\|_2^2 < F(\theta_t), \tag{39}$$

This proves Theorem 3. □

When the correction is active. Substituting the main-text direction $\boldsymbol{d}_{0,t} = \boldsymbol{g}_{F,t} + \lambda_t\boldsymbol{g}_{G,t}$ into the projection formula gives an equivalent condition in terms of the protected and guidance gradients:

$$\begin{aligned} \alpha_t > 0 &\iff \langle \boldsymbol{g}_{F,t}, \boldsymbol{g}_{F,t} + \lambda_t\boldsymbol{g}_{G,t}\rangle < \mu\|\boldsymbol{g}_{F,t}\|_2^2 \\ &\iff \lambda_t\langle \boldsymbol{g}_{F,t}, \boldsymbol{g}_{G,t}\rangle < -(1-\mu)\|\boldsymbol{g}_{F,t}\|_2^2, \\ \alpha_t &= \frac{[-(1-\mu)\|\boldsymbol{g}_{F,t}\|_2^2 - \lambda_t\langle \boldsymbol{g}_{F,t}, \boldsymbol{g}_{G,t}\rangle]_+}{\|\boldsymbol{g}_{F,t}\|_2^2}. \end{aligned} \tag{40}$$

Thus, a nonnegative guidance alignment requires no correction. When $\mu = 1$, any negative weighted alignment activates the projection; when $\mu < 1$, a limited negative alignment is admissible. The correction changes only the component parallel to $\boldsymbol{g}_{F,t}$, as the following identity makes explicit:

$$\begin{aligned} \boldsymbol{d}_t^\star - \frac{\langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_t^\star\rangle}{\|\boldsymbol{g}_{F,t}\|_2^2}\boldsymbol{g}_{F,t} &= \boldsymbol{d}_{0,t} + \alpha_t\boldsymbol{g}_{F,t} - \left(\frac{\langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_{0,t}\rangle}{\|\boldsymbol{g}_{F,t}\|_2^2} + \alpha_t\right)\boldsymbol{g}_{F,t} \\ &= \boldsymbol{d}_{0,t} - \frac{\langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_{0,t}\rangle}{\|\boldsymbol{g}_{F,t}\|_2^2}\boldsymbol{g}_{F,t}. \end{aligned} \tag{41}$$

When the constraint is active, $\alpha_t = \mu - \langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_{0,t}\rangle/\|\boldsymbol{g}_{F,t}\|_2^2$. The norm entering the step-size condition can therefore be expanded without any additional notation:

$$\begin{aligned} \|\boldsymbol{d}_t^\star\|_2^2 &= \|\boldsymbol{d}_{0,t}\|_2^2 + 2\alpha_t\langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_{0,t}\rangle + \alpha_t^2\|\boldsymbol{g}_{F,t}\|_2^2 \\ &= \|\boldsymbol{d}_{0,t}\|_2^2 - \frac{\langle \boldsymbol{g}_{F,t}, \boldsymbol{d}_{0,t}\rangle^2}{\|\boldsymbol{g}_{F,t}\|_2^2} + \mu^2\|\boldsymbol{g}_{F,t}\|_2^2. \end{aligned} \tag{42}$$

The first two terms in the last line are the squared norm of the retained orthogonal component. Consequently, positive protected-gradient alignment alone does not control the finite-step remainder: a large orthogonal component can still require a smaller learning rate. This explains why the projection constraint and the step-size condition are both needed.

Scope of the descent guarantee. The result concerns one step of the stated update with the exact gradient $\nabla F(\theta_t)$, the exact projected direction, and a step size satisfying Eq. (38). It does not assert improvement in clustering metrics or descent of the refreshed guidance objective. If an optimizer subsequently transforms the direction, its actual parameter displacement must satisfy the alignment and smoothness-remainder conditions separately. When $\boldsymbol{g}_{F,t} = \boldsymbol{0}$, the half-space is all of parameter space and its projection is $\boldsymbol{d}_{0,t}$; the implementation convention $\alpha_t = 0$ is consistent with this projection, but strict descent is no longer guaranteed.

D Additional Experiments

D.1 Runtime Analysis

Tables 7 and 8 summarize computational cost, clustering performance, and the RTG-IMVC runtime breakdown on NoisyMNIST at $\eta = 0.3$.

Table 7: Computational cost and clustering performance on NoisyMNIST at $\eta = 0.3$.

Method	Time (s/iter) ↓	Mem. (GB) ↓	FLOPs (G/iter) ↓	ACC (%) ↑	NMI (%) ↑	ARI (%) ↑
ICMVC	1.42	2.36	8.74	76.01 ± 0.55	68.13 ± 0.72	–
DSIMVC	2.18	3.12	12.68	82.42 ± 0.89	74.26 ± 0.46	75.41 ± 1.82
DVIMC	1.86	2.85	10.92	81.42 ± 0.72	73.54 ± 0.89	73.78 ± 1.61
GHICMC	2.45	3.48	14.36	78.62 ± 0.41	71.31 ± 0.59	71.76 ± 1.20
SIIHPC	1.67	2.54	9.85	77.60 ± 0.71	69.06 ± 0.89	70.05 ± 1.61
HSACC	2.73	3.76	16.24	81.35 ± 0.42	72.57 ± 0.60	72.01 ± 1.22
DDR-IMVC	3.06	4.15	18.57	83.39 ± 0.61	76.85 ± 0.78	77.05 ± 1.46
RTG-IMVC	**0.82**	**1.94**	**6.83**	$\mathbf{93.30 \pm 0.64}$	$\mathbf{85.53 \pm 0.78}$	$\mathbf{85.91 \pm 1.64}$

D.1.1 Overall Computational Cost

The per-iteration timing should include topology probing, risk and teacher construction, and the complete network update. Table 7 reports runtime, peak GPU memory, and FLOPs alongside ACC, NMI, and ARI.

D.1.2 Topology-Conditioned Prediction

Nominal and candidate topologies are evaluated with shared parameters using Eqs. (2) and (5). The cost depends on the candidate budget K, the number of missing sample–view pairs, and the evaluation strategy.

Table 8: Per-iteration runtime breakdown of RTG-IMVC on NoisyMNIST at $\eta = 0.3$.

Component	Time (s) ↓	Share (%)
Topology-conditioned prediction	0.48	58.5
Risk and teacher construction	0.03	3.7
Gradient computation and update	0.31	37.8
Total	**0.82**	**100.0**

D.1.3 Risk and Teacher Construction

Candidate log-predictions determine topology risks, expert weights, the consensus teacher, and guidance coefficients through Eqs. (6), (9), and (11). This stage reuses the preceding predictions without an additional forward pass.

D.1.4 Gradient Computation and Base-Safe Updating

This stage computes both protected and guidance gradients, projects their combined direction, and updates the parameters using Eqs. (13) and (14). Given the gradients, the projection adds $\mathcal{O}(P)$ vector operations for P trainable parameters.

D.2 Generality of Risk-Aware Guidance

We evaluate the generality of risk-aware guidance through paired comparisons of DSIMVC, DVIMC, and DDR-IMVC with corresponding variants incorporating risk-aware guidance. The

selected host methods provide topology-conditioned predictions and view-specific and fused distributions in a shared cluster-label space. The variants use Eqs. (6) and (9) to construct the teacher and incorporate the guidance term in Eq. (11).

For each pair, we use the same preprocessing, missing-view masks, random seeds, training budgets, and the respective host optimizer, with ACC, NMI, and ARI averaged over ten runs. This comparison isolates the effect of risk-aware guidance, while the projected update in Eq. (14) is evaluated separately.

Table 9: Generality of risk-aware guidance on NoisyMNIST at a missing-view rate of $\eta = 0.3$. Each baseline is paired with a variant incorporating the risk-weighted consensus guidance of RTG-IMVC.

Method	ACC (%)↑	NMI (%)↑	ARI (%)↑
DSIMVC	82.42	74.26	75.41
DSIMVC + risk-aware guidance	**92.36**	**82.81**	**84.24**
DVIMC	81.42	73.54	73.78
DVIMC + risk-aware guidance	**91.18**	**81.54**	**85.02**
DDR-IMVC	83.39	76.85	77.05
DDR-IMVC + risk-aware guidance	**92.15**	**83.68**	**83.13**

As demonstrated in Table 9, incorporating risk-aware guidance yields ACC improvements of 9.94, 9.76, and 8.76 percentage points for DSIMVC, DVIMC, and DDR-IMVC, respectively, together with consistent improvements in NMI and ARI. These results indicate that the proposed risk-aware guidance mechanism can be integrated with different host methods and provide consistent performance gains under the same experimental protocol.

D.3 Additional Hyperparameter Analysis

Table 10: Sensitivity to additional hyperparameters on NoisyMNIST at a missing-view rate of $\eta = 0.3$. Panels (a)–(c) examine the risk floor ε_R, reliability-decay scale τ, and maximum guidance coefficient λ_0, respectively. Each parameter is expressed relative to its reference setting, denoted by the superscript ref. ACC and NMI are reported in percent.

(a) Risk floor ε_R

$\varepsilon_R/\varepsilon_R^{\mathrm{ref}}$	ACC↑	NMI↑
0.25	92.86	85.02
0.50	93.12	85.29
0.75	93.24	85.44
1.00	**93.30**	**85.53**
1.50	93.18	85.36
2.00	92.98	85.15

(b) Reliability scale τ

τ/τ^{ref}	ACC↑	NMI↑
0.25	92.41	84.62
0.50	92.91	85.12
0.75	93.17	85.39
1.00	**93.30**	**85.53**
1.50	93.21	85.46
2.00	92.95	85.18

(c) Guidance cap λ_0

$\lambda_0/\lambda_0^{\mathrm{ref}}$	ACC↑	NMI↑
0.25	92.58	84.80
0.50	93.00	85.21
0.75	93.19	85.41
1.00	**93.30**	**85.53**
1.50	93.13	85.32
2.00	92.76	84.97

We further extend Section 4.4 to analyze ε_R, τ, and λ_0. Specifically, a larger ε_R results in more uniform expert weights, a larger τ weakens reliability attenuation, and λ_0 caps the guidance strength (Eqs. (9) and (11)).

The parameter sweep scales each parameter by $\{0.25, 0.50, 0.75, 1.00, 1.50, 2.00\}$ times its reference value on NoisyMNIST at $\eta = 0.3$. All other settings, including the optimizer, candidate-topology , and training budget, are kept constant. To ensure fair comparisons, all configurations share identical missing-view masks and seeds, reporting ACC and NMI averaged across ten runs.

As shown in Table 10, ACC and NMI vary by at most 0.89 and 0.91 percentage points across the sweeps, respectively. This stability confirms that the model has limited sensitivity within the tested ranges.

D.4 Component Accumulation Ablation

Step 1 uses uniform LogOP weights, unscaled KL guidance ($\kappa_i = 1$, $\lambda_t = \lambda_0$), and an unconstrained update, without evaluating topology risk. Steps 2–4 successively add inverse-risk teacher weighting, risk-scaled KL guidance, and the base-safe projection.

Table 11: Component accumulation on NoisyMNIST at $\eta = 0.3$. ACC and NMI are percentages.

Step	Configuration	ACC ↑	NMI ↑
1	No certificate	81.92	72.41
2	+ Risk-weighted consensus teacher	86.74	78.36
3	+ Risk-scaled KL	91.52	83.71
4	**+ Base-safe projection (full RTG-IMVC)**	**93.30**	**85.53**

The four rows trace how reliability control enters RTG-IMVC. Risk weighting reduces the contribution of uncertain experts to the consensus teacher. Risk-scaled KL then attenuates guidance for samples with less reliable transfer, while base-safe projection constrains the update to maintain alignment with the gradient of the base objective.

D.5 Performance Profiles and Training Dynamics under Missing Views

We complement Table 2 with cross-dataset performance profiles and epoch-wise measurements at missing-view rates $\eta \in \{0.1, 0.3, 0.5\}$.

Cross-dataset performance profiles. Fig. 7 reorganizes the ACC and NMI means in Table 2 into dataset-specific profiles. The response to increasing missingness varies across datasets. For example, ACC decreases from 96.80% to 89.96% on NGs between $\eta = 0.1$ and 0.5, compared with 96.00% to 64.62% on NoisyMNIST. The separate dataset planes preserve these differences without averaging across benchmarks.

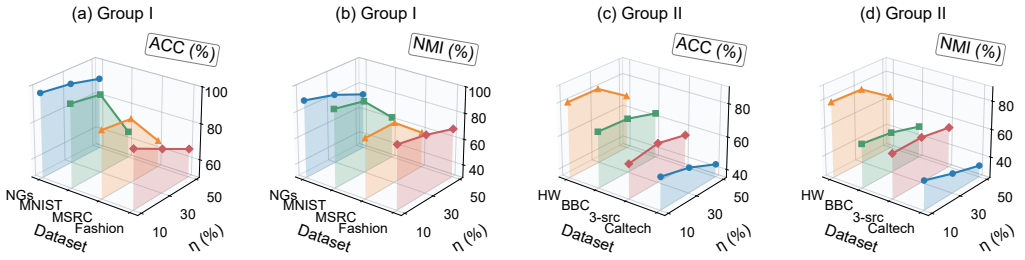


Figure 7: ACC and NMI profiles of RTG-IMVC across eight datasets. Panels (a,b) show NGs, NoisyMNIST (MNIST), MSRC_v1 (MSRC), and 3V_Fashion_MV (Fashion); panels (c,d) show Caltech-5V (Caltech), BBC, Handwritten (HW), and 3-sources (3-src). Markers reproduce the reported means in Table 2 at $\eta = 0.1, 0.3, 0.5$. Lines connect these settings; translucent planes aid visual separation and do not represent uncertainty.

Training dynamics and seed variability. Fig. 8 shows epoch-specific ACC and NMI measurements on NoisyMNIST and Caltech-5V, averaged over ten training runs. On NoisyMNIST, ACC and NMI rise rapidly and remain comparatively stable at $\eta = 0.1$ and 0.3. At $\eta = 0.5$, mean ACC decreases from 62.09% at epoch 100 to 54.58% at epoch 500, a drop of 7.51 percentage points. On Caltech-5V at the same missing rate, mean ACC decreases from 42.80% to 32.64%, a drop of 10.16 percentage points. These fixed-epoch measurements should be distinguished from selected-checkpoint results.

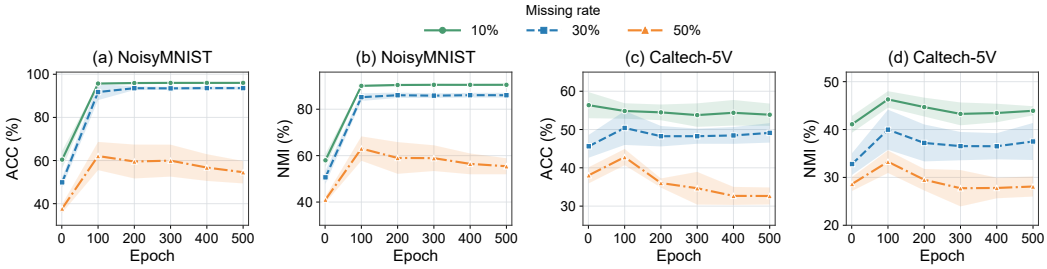


Figure 8: Training dynamics at three missing-view rates. Panels (a,b) show ACC and NMI on NoisyMNIST; panels (c,d) show the same metrics on Caltech-5V. Markers indicate means across ten seeds at evaluation epochs 1, 100, 200, 300, 400, and 500. Shaded bands denote $\pm$ one sample standard deviation across seeds, rather than confidence intervals. Straight segments connect the recorded evaluations. Values are epoch-specific and are not selected-checkpoint scores.

D.6 Similarity Structure of Cross-Dataset Robustness Profiles

Fig. 9 summarizes how RTG-IMVC responds across the eight benchmarks. 3V_Fashion_MV and Handwritten are closest for the joint ($S = 0.918$), ACC ($S = 0.910$), and NMI ($S = 0.921$) profiles, whereas MSRC_v1 and Handwritten are closest for ARI ($S = 0.943$) and degradation ($S = 0.754$).

The score-based matrices also share a stable block structure: in the joint panel, mean similarity is 0.791 within Ward groups versus 0.475 between groups. This margin and the consistent ACC, NMI, and ARI patterns indicate that RTG-IMVC maintains coherent robustness across metrics, while the weaker degradation blocks retain dataset-specific responses to missingness.

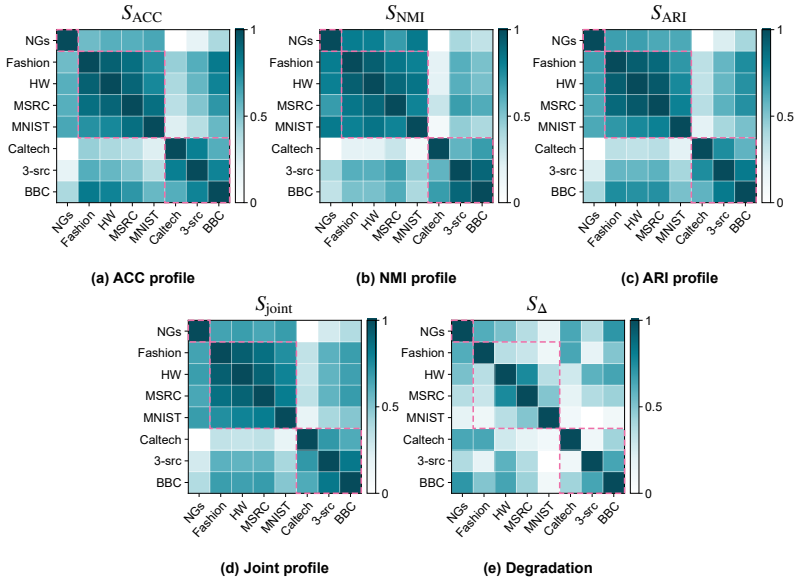
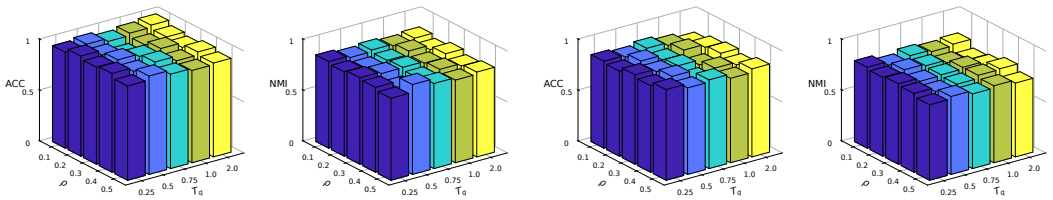


Figure 9: Similarity matrices of RTG-IMVC across eight datasets. Panels (a)–(c) show ACC, NMI, and ARI profiles, (d) their joint profile, and (e) degradation. Darker cyan indicates greater similarity; pink boxes mark Ward clusters.

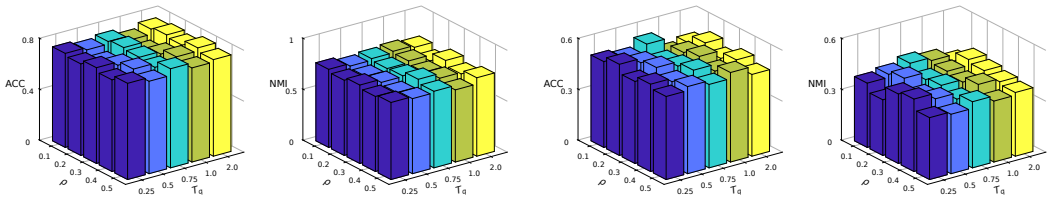
E Additional Visualizations Across Datasets

E.1 Hyperparameter Sensitivity Across Datasets

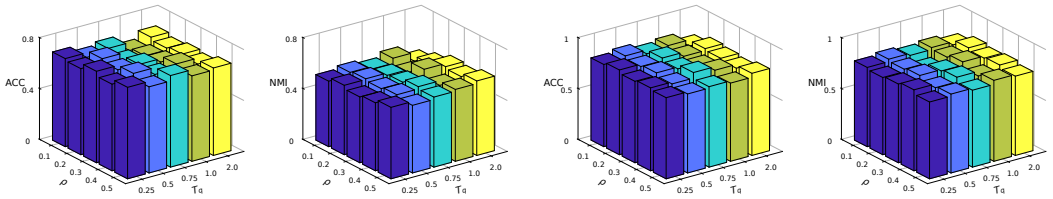
Fig. 10 extends the ρ–T_q analysis in Section 4.4 to the seven datasets other than NoisyMNIST, all at $\eta = 0.3$.



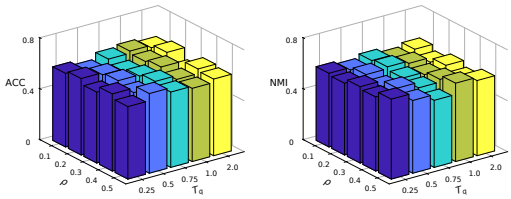
(a) NGs **(b) MSRC_v1**



(c) 3V_Fashion_MV **(d) Caltech-5V**



(e) BBC **(f) Handwritten**



(g) 3-sources

Figure 10: Sensitivity to the edge-replacement ratio ρ and teacher temperature T_q on seven additional datasets at $\eta = 0.3$. Each pair shows ACC (left) and NMI (right), with $\rho \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ and $T_q \in \{0.25, 0.5, 0.75, 1.0, 2.0\}$.

E.2 Convergence on Additional Datasets

Fig. 11 extends Section 4.5 to six additional datasets at $\eta = 0.1$. Loss decreases across the displayed checkpoints, while ACC and NMI follow dataset-dependent trajectories.

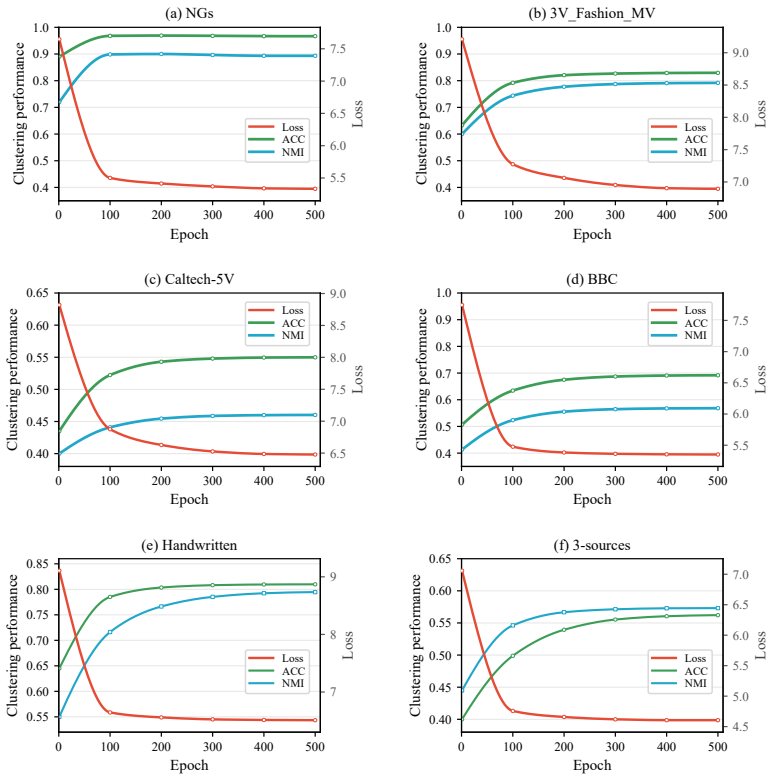


Figure 11: Convergence of RTG-IMVC on (a) NGs, (b) 3V_Fashion_MV, (c) Caltech-5V, (d) BBC, (e) Handwritten, and (f) 3-sources at $\eta = 0.1$. Hollow markers show means over ten runs at epochs 1, 100, 200, 300, 400, and 500; curves are shape-preserving cubic interpolations between these checkpoints. ACC and NMI use the left axis; loss uses a separate right axis in each panel.

E.3 Running Time on Additional Datasets

Fig. 12 complements Fig. 6(c,d) with per-iteration running times on four additional datasets. RTG-IMVC is the fastest method on all four.

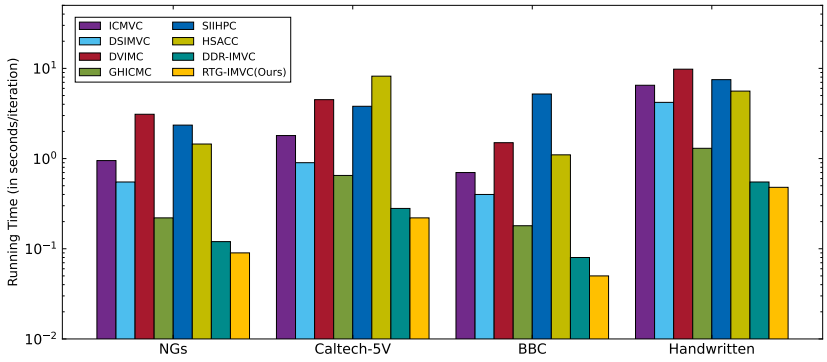


Figure 12: Per-iteration running time on NGs, Caltech-5V, BBC, and Handwritten. The vertical axis uses a logarithmic scale.

F LIMITATIONS AND VISUAL ANALYSIS

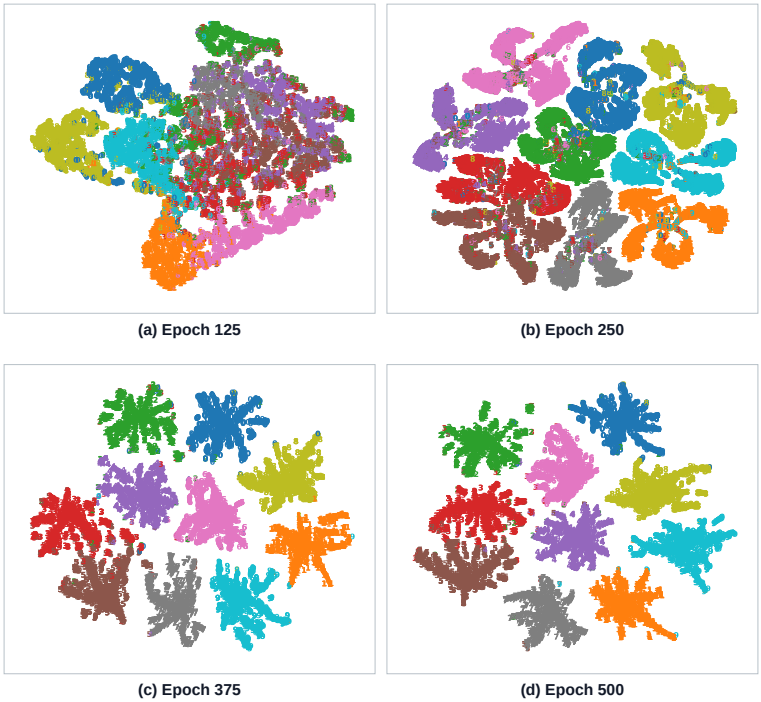
F.1 LIMITATIONS AND FUTURE WORK

While RTG-IMVC establishes a risk-certified framework to regulate transferred-topology influence, several theoretical and practical limitations open avenues for future research:

- **Limitation 1: Conditional certification vs. empirical sensitivity.** The theoretical stability bound (Theorem 1) guarantees risk control under valid topology coverage and local Lipschitz continuity. However, finite probing operators (donor substitution, edge replacement, and bootstrapping) only provide an empirical lower/upper estimate of sensitivity rather than an absolute mathematical guarantee. Developing tighter coverage diagnostics and data-adaptive perturbation bounds remains an important future direction.
- **Limitation 2: Susceptibility to severe corruption in observed views.** Eq. (8) assigns zero transfer risk to observed views under the assumption that observed views provide raw reference signals. If an observed view is corrupted by heavy feature noise or domain shift, it may improperly dominate the logarithmic opinion pool. Incorporating observation-quality estimation into the risk metric will be crucial for noisy-view robustness.

F.2 EVOLUTION OF LEARNED REPRESENTATIONS ON NOISYMNIST

We visualize the learned representation on NoisyMNIST, whose 5,000 samples span ten interpretable digit classes. In Fig. 13, epoch 125 shows broad inter-class overlap and bridges between emerging groups; coarse class-dependent regions appear by epoch 250. At epochs 375 and 500, most classes occupy distinct regions with less central mixing, although a few off-cluster samples remain. This progression qualitatively indicates that training establishes global organization before refining class-specific neighborhoods. As two-dimensional projections, the panels illustrate representation geometry rather than provide another quantitative clustering metric.



(a) Epoch 125 (b) Epoch 250

(c) Epoch 375 (d) Epoch 500

Figure 13: NoisyMNIST representations at epochs 125, 250, 375, and 500. Colors and numeric glyphs denote digit classes. All panels use the supplied checkpoints at a common display scale; later checkpoints show clearer separation.