TOPOLOGY-AWARE CONTRASTIVE ANCHORS FOR INCOMPLETE MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Incomplete multi-view clustering remains challenging because missing observations disrupt cross-view correspondence and distort the underlying structural relationships among samples. We propose **T**opology-**A**ware **C**ontrastive **A**nchors (**TACA**), a reconstruction-free framework that learns a unified representation directly from incomplete multi-view data through anchor-based contrastive modeling. Instead of explicitly recovering missing features, TACA constructs view-specific anchor representations and aligns their semantic information across views, thereby reducing the dependence on unreliable feature imputation. To further preserve structural information beyond a single anchor resolution, we introduce a topology-aware cross-scale consistency mechanism that coordinates multi-scale anchor graphs by encouraging agreement among their latent relational structures. Moreover, incomplete observations may generate semantically ambiguous negative pairs and weaken contrastive supervision. TACA addresses this issue through a negative sample decoupling strategy that suppresses unreliable negatives according to their semantic similarity. Experiments on five benchmark datasets under different missing rates demonstrate that TACA consistently achieves competitive clustering performance against incomplete multi-view clustering methods while scaling linearly with the number of samples under fixed anchor budgets.

Code: `https://anonymous.4open.science/status/TACA-A4B8`

1 INTRODUCTION

Multi-view clustering seeks to discover shared cluster structure from complementary representations without requiring labels. Its effectiveness depends on exploiting both cross-view agreement and complementary information (Yan et al., 2024; Dong et al., 2023). In practice, acquisition costs and sensor failures often leave samples observed in only a subset of views, motivating incomplete multi-view clustering (IMVC) (Tang & Liu, 2022). Missing observations limit the evidence available for recovering reliable sample relationships (Wen et al., 2023). Meanwhile, large datasets demand efficient representations: scalable tensor formulations (Long et al., 2024) and bipartite graph learning (Xia et al., 2022) provide alternatives to dense sample graphs. For IMVC, compact prototype and consensus bipartite graphs reduce computational costs (Li et al., 2022; Wang et al., 2022a), while diverse view-shared anchors capture information at multiple resolutions (Yu et al., 2024). These developments make anchors a useful basis for scalable learning, but their clustering utility still depends on the reliability of the relationships they encodeLiu et al. (2025b); Jiang et al. (2025).

Existing anchor-based approaches investigate view-independent representations (Liu et al., 2022a), corresponding anchor construction (Yu et al., 2023), and efficient consensus learning (Liu et al., 2022b; Wang et al., 2022b). Cluster-aware constraints further connect anchor learning to the desired partition (Chen et al., 2025b). Under incomplete observations, projective modeling (Deng et al., 2023) and distribution-consistent anchor imputation (Li et al., 2023) recover information through different reconstruction strategies. Cross-view reconstruction also explicitly addresses anchor shift (Liu et al., 2024). Alongside these approaches, partial sample and prototype alignment (Jin et al., 2023) and prototype matching (Yuan et al., 2025) strengthen cross-view correspondence. More recent methods explore diffusion-based generation (Zhang et al., 2025) and distribution recovery (Jin et al., 2025). Nevertheless, recovering features or matching individual prototypes does not by itself ensure that anchor graphs at different resolutions encode compatible sample relationships. This motivates explicit structural coordination during anchor representation learning.

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.

Contrastive learning offers another source of supervision by encouraging agreement across views. Confidence-guided IMVC (Chao et al., 2024) and robust graph contrastive learning (Zhuang et al., 2025) investigate its reliability under incomplete observations, while consensus semantic learning demonstrates an imputation-free alternative (Dai et al., 2025). Structure-aware contrastive objectives (Wang & Feng, 2024; He et al., 2025), multi-scale high-order modeling (Chen et al., 2025a), and manifold consistency (Wang et al., 2024) further emphasize relational information. Cross-view affinity learning likewise supports consensus construction (Liu et al., 2025a), and decoupled contrastive learning explores high-order relationships through random walks (Lu et al., 2024). Building on these advances, we focus on two coupled difficulties in anchor-based IMVC. First, pairwise anchor agreement does not guarantee structural consistency across anchor scales. Second, nonmatching anchors may represent similar cluster semantics, making indiscriminate negative-pair repulsion counterproductive. Incomplete observations further restrict the evidence available to resolve this ambiguity. Effective anchor learning therefore requires coordinating relational structure while selectively handling uncertain negativesDong et al. (2025); Wang et al. (2026).

We propose **Topology-Aware Contrastive Anchors (TACA)**, an IMVC framework that jointly models cross-view anchor agreement, cross-scale structural consistency, and negative-pair reliability. TACA learns compact anchor representations from available observations and applies anchor-level contrastive supervision without explicitly imputing missing input features. A cross-scale consistency objective encourages agreement among the relational structures encoded by multi-scale anchor graphs, complementing individual anchor alignment. A similarity-based negative sample decoupling mechanism excludes highly similar nonmatching anchors from the negative set, reducing potentially conflicting repulsion. These components jointly encourage coherent anchor relationships across views and scales while retaining discriminative supervision. Our contributions are summarized as follows:

- We propose TACA, a unified anchor-based IMVC framework that couples cross-view contrastive alignment with cross-scale structural consistency to learn a shared clustering representation without explicit missing-feature imputation.
- We develop cross-scale consistency and similarity-based negative decoupling objectives to address complementary sources of unreliable supervision: disagreement among multi-scale anchor graphs and potentially false-negative anchor pairs.
- We evaluate TACA through clustering comparisons, component ablations, and efficiency analyses under varying missing rates. We analyze its computational and memory costs, which scale linearly with the sample count per optimization iteration when anchor counts, view counts, and representation dimensions are fixed.

2 Related Work

2.1 Anchor-Based Incomplete Multi-View Clustering

Anchor-based IMVC represents sample relationships through compact bipartite graphs, reducing the computational burden of dense graph learning (Wang et al., 2022a; Li et al., 2022). Existing methods investigate view-independent anchors (Liu et al., 2022a), cross-view anchor correspondence (Yu et al., 2023), and distribution-consistent anchor imputation (Li et al., 2023). Beyond a single anchor configuration, DVSAI learns diverse view-shared anchors and integrates multi-scale anchor graphs (Yu et al., 2024). Other approaches address anchor shift through cross-view reconstruction (Liu et al., 2024) or improve semantic correspondence through prototype matching (Yuan et al., 2025). TACA builds on this compact representation paradigm. Its distinction lies in explicitly coupling cross-scale structural consistency with contrastive anchor learning: multiple anchor scales provide relational supervision during learning, rather than serving only as representations to be integrated.

2.2 Contrastive and Structure-Aware Multi-View Learning

Contrastive multi-view clustering learns shared information by aligning representations across views. CPSPAN combines partial sample and prototype alignment for incomplete observations (Jin et al., 2023), while ICMVC introduces high-confidence guidance (Chao et al., 2024). FreeCSL pursues consensus semantic learning without explicit imputation or alignment (Dai et al., 2025).

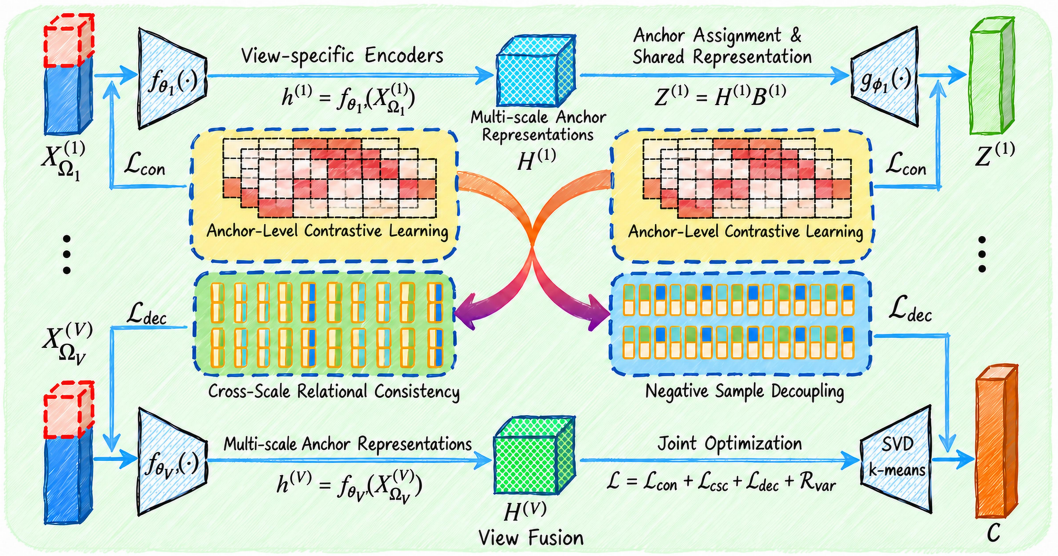

Figure 1: The overall framework of the proposed TACA pipeline. It extracts robust semantic signals via anchor-level contrastive learning, cross-scale consistency, and negative sample decoupling.

Structural extensions exploit cross-view topology (Dong et al., 2023), high-order random walks (Lu et al., 2024), and multi-scale contrastive diffusion (Chen et al., 2025a). These studies establish that contrastive alignment benefits from relational information and reliable supervision. TACA follows this direction but focuses on their interaction within compact anchor graphs under missing views. Cross-scale consistency coordinates the encoded relationships, while similarity-based negative decoupling filters potentially false-negative anchor pairs. Thus, its contribution concerns the joint treatment of structural disagreement and ambiguous anchor supervision, rather than contrastive learning or multi-scale modeling alone.

3 Method

TACA combines identity-linked anchor alignment, cross-scale relational consistency, and selective negative supervision. All views share known sample identities; missing views are unavailable observations, not unknown correspondences.

3.1 Observation-Aware Anchor Construction

Let $\omega_{rv} \in \{0,1\}$ indicate whether sample r is observed in view v, $\Omega_v = \{r : \omega_{rv} = 1\}$, and $n_v = |\Omega_v|$. We assume $\sum_v \omega_{rv} \geq 1$ for every sample. At scale $t \in \{1,\dots,T\}$, select m_t distinct sample identities $\mathcal{S}_t = \{\ell_{ti}\}_{i=1}^{m_t}$. The same identities define anchors in every view; encoders $f_{\theta_v} : \mathbb{R}^{d_v} \to \mathbb{R}^d$ remain view-specific:

$$
\begin{aligned}
h_r^{(v)} &= \frac{f_{\theta_v}(x_r^{(v)})}{\sqrt{\|f_{\theta_v}(x_r^{(v)})\|_2^2 + \epsilon_0^2}}, \quad r \in \Omega_v, \quad \epsilon_0 > 0, \\
a_{ti}^{(v)} &= h_{\ell_{ti}}^{(v)}, \qquad\qquad i \in \mathcal{J}_t^{(v)} := \{i : \ell_{ti} \in \Omega_v\}.
\end{aligned}
\tag{1}
$$

Thus, $\|h_r^{(v)}\|_2 \leq 1$, and positive anchors refer to the same observed entity rather than independently indexed dictionary entries. No anchor is generated for an unavailable observation.

For each scale, landmark selection uses the observation mask to ensure $|\mathcal{J}_t^{(v)}| \geq 2$ and a connected view graph, with edges defined by shared observed landmarks. Remaining landmarks are sampled uniformly with a fixed seed. If this coverage is infeasible, the anchor budget must increase; disconnected views cannot be aligned through this supervision alone.

For $r \in \Omega_v$, sample-to-anchor assignments use temperature $\kappa > 0$. Missing rows and unavailable anchor columns are set to zero:

$$q_{tri}^{(v)} = \begin{cases} \dfrac{\exp\left(h_r^{(v)\top} a_{ti}^{(v)}/\kappa\right)}{\sum_{j\in\mathcal{J}_t^{(v)}} \exp\left(h_r^{(v)\top} a_{tj}^{(v)}/\kappa\right)}, & r \in \Omega_v,\ i \in \mathcal{J}_t^{(v)}, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Consequently, $\sum_i q_{tri}^{(v)} = \omega_{rv}$. Average only available views and normalize the resulting bipartite graph:

$$\begin{gathered} Q_t[r,i] = \frac{\sum_{v=1}^{V} q_{tri}^{(v)}}{\sum_{v=1}^{V} \omega_{rv}}, \qquad D_t = \operatorname{diag}(Q_t^\top \mathbf{1}_n), \\ B_t = Q_t D_t^{-1/2} \in \mathbb{R}^{n\times m_t}, \qquad K_t = B_t B_t^\top \in \mathbb{R}^{n\times n}. \end{gathered} \tag{3}$$

Every diagonal entry of D_t is positive. Moreover, $Q_t \mathbf{1}_{m_t} = \mathbf{1}_n$ and $K_t \mathbf{1}_n = \mathbf{1}_n$. The sample affinity K_t is conceptual; computation uses B_t.

3.2 Anchor-Level Contrastive Supervision

Define the directed positive-pair set $\mathcal{I} = \{(t,v,w,i) : v \neq w,\ i \in \mathcal{J}_t^{(v)} \cap \mathcal{J}_t^{(w)}\}$. For $\xi = (t,v,w,i)$, let $z_{\xi j} = a_{ti}^{(v)\top} a_{tj}^{(w)}/\tau$, where $\tau > 0$. The anchor contrastive objective is

$$\mathcal{L}_{\text{con}} = \frac{1}{|\mathcal{I}|} \sum_{\xi\in\mathcal{I}} \ell_\xi^{\text{all}}, \qquad \ell_\xi^{\text{all}} = -\log \frac{\exp(z_{\xi i})}{\sum_{j\in\mathcal{J}_t^{(w)}} \exp(z_{\xi j})}, \qquad \xi = (t,v,w,i). \tag{4}$$

Known identity establishes positive correspondence, but does not imply that different identities belong to different clusters.

To discourage constant representations, introduce an observed-sample variance penalty with $0 < \sigma^2 < 1/d$:

$$\begin{gathered} \mathcal{R}_{\text{var}} = \frac{1}{Vd} \sum_{v=1}^{V} \sum_{k=1}^{d} \left[\sigma^2 - \frac{1}{n_v} \sum_{r\in\Omega_v} \left(h_{rk}^{(v)} - \bar{h}_k^{(v)} \right)^2 \right]_+^2, \\ \bar{h}^{(v)} = \frac{1}{n_v} \sum_{r\in\Omega_v} h_r^{(v)}, \qquad [u]_+ = \max\{u, 0\}. \end{gathered} \tag{5}$$

Theorem 3.1 (Variance-based exclusion of representation collapse)**.** Let $\mathcal{R}_{\text{var}}$ be defined in Eq. (5), with $\sigma^2 > 0$. For every view v and coordinate k, $\operatorname{Var}_{r\in\Omega_v}(h_{rk}^{(v)}) \geq [\sigma^2 - \sqrt{Vd\mathcal{R}_{\text{var}}}]_+$. Moreover, the mean coordinate variance satisfies $\frac{1}{dn_v} \sum_{r\in\Omega_v} \|h_r^{(v)} - \bar{h}^{(v)}\|_2^2 \geq [\sigma^2 - \sqrt{V\mathcal{R}_{\text{var}}}]_+$. Consequently, $\mathcal{R}_{\text{var}} < \sigma^4/V$ excludes a constant representation in every view.

Proof. See Appendix D.1. The certificate is evaluated from empirical variances; it does not require the total contrastive objective to approach zero.

This is a conditional certificate, not a guarantee that optimization attains the required objective value or recovers the true clusters.

3.3 Cross-Scale Relational Consistency

The matrices Q_t have different column dimensions and cannot be subtracted directly. Their affinities K_t instead describe the same n sample identities. For $T \geq 2$ and $P = T(T-1)/2$, define

$$\mathcal{L}_{\text{csc}} = \frac{1}{P} \sum_{1\leq t<u\leq T} \left(\|B_t^\top B_t\|_F^2 + \|B_u^\top B_u\|_F^2 - 2\|B_t^\top B_u\|_F^2 \right). \tag{6}$$

This equals $P^{-1} \sum_{t<u} \|K_t - K_u\|_F^2$ exactly. All products in Eq. (6) have anchor-sized dimensions; no $n \times n$ matrix is materialized. The objective encourages agreement in sample relations without requiring one-to-one correspondence between different anchor scales.

Theorem 3.2 (Cross-scale spectral stability)**.** If $\mathcal{L}_{\mathrm{csc}} \le \varepsilon$ and $\delta = \sqrt{P\varepsilon}$, then every scale pair satisfies $\|K_t - K_u\|_2 \le \delta$ and $|\lambda_j(K_t) - \lambda_j(K_u)| \le \delta$, where eigenvalues are ordered decreasingly. Let Π_t, Π_u project onto the leading $c < n$ eigenspaces. If $g = \lambda_c(K_t) - \lambda_{c+1}(K_t) > 2\delta$, then $\|\Pi_t - \Pi_u\|_F \le \sqrt{2c}\,\delta/(g - \delta)$.

Proof. The trace identity above bounds each affinity difference in Frobenius norm by δ, hence also in operator norm. The eigenvalue bound follows from the variational characterization. For a leading eigenvector of K_u, its projection onto the trailing eigenspace of K_t has norm at most $\delta/(g - \delta)$. Summing over c orthonormal eigenvectors and using $\|\Pi_t - \Pi_u\|_F = \sqrt{2}\|(I - \Pi_t)U_u\|_F$ gives the result. □

The eigengap condition makes explicit when relational agreement also controls the spectral embedding; exact topology recovery is not assumed.

3.4 Selective Negative Sample Decoupling

For $\xi = (t, v, w, i)$, let $\mathcal{N}_\xi = \mathcal{J}_t^{(w)} \setminus \{i\}$ and $s_{\xi j} = \tau z_{\xi j}$. A threshold $\eta \in [-1, 1]$ filters highly similar negatives. Retain the least-similar negative as a fallback, using sample identity to break ties:

$$j_\xi^\star = \arg\min_{j \in \mathcal{N}_\xi} \mathrm{sg}(s_{\xi j}), \qquad w_{\xi j} = \begin{cases} 1, & j = i, \\ \mathbf{1}\{\mathrm{sg}(s_{\xi j}) \le \eta \text{ or } j = j_\xi^\star\}, & j \in \mathcal{N}_\xi. \end{cases} \tag{7}$$

Here sg denotes stop-gradient. The mask is prescribed rather than freely optimized: $w_{\xi i} = 1$ and $\sum_{j \in \mathcal{N}_\xi} w_{\xi j} \ge 1$. Thus, neither an empty denominator nor a positive-only denominator is admissible. The filtered loss is

$$\mathcal{L}_{\mathrm{dec}} = \frac{1}{|\mathcal{I}|} \sum_{\xi \in \mathcal{I}} \ell_\xi^{\mathrm{keep}}, \qquad \ell_\xi^{\mathrm{keep}} = -\log \frac{\exp(z_{\xi i})}{\exp(z_{\xi i}) + \sum_{j \in \mathcal{N}_\xi} w_{\xi j} \exp(z_{\xi j})}. \tag{8}$$

To retain supervision from all negatives, use $0 \le \gamma < 1$ to interpolate between the two losses:

$$\begin{aligned} \ell_\xi^{\mathrm{mix}} &= (1 - \gamma)\ell_\xi^{\mathrm{all}} + \gamma \ell_\xi^{\mathrm{keep}}, \\ p_{\xi j} &= \frac{\exp(z_{\xi j})}{\sum_{k \in \mathcal{J}_t^{(w)}} \exp(z_{\xi k})}, \qquad r_\xi = \sum_{\substack{j \in \mathcal{N}_\xi \\ w_{\xi j} = 0}} p_{\xi j}. \end{aligned} \tag{9}$$

Theorem 3.3 (Controlled modification of contrastive gradients)**.** For fixed masks, $\ell_\xi^{\mathrm{keep}} \ge \log(1 + \exp(-2/\tau)) > 0$ and $\|\nabla_{z_\xi} \ell_\xi^{\mathrm{mix}} - \nabla_{z_\xi} \ell_\xi^{\mathrm{all}}\|_1 = 2\gamma r_\xi$. Holding target anchors fixed, the corresponding partial gradient with respect to the query anchor satisfies $\|\nabla_{a_{ti}^{(v)}} \ell_\xi^{\mathrm{mix}} - \nabla_{a_{ti}^{(v)}} \ell_\xi^{\mathrm{all}}\|_2 \le 2\gamma r_\xi / \tau$.

Proof. At least one negative is retained and $s_{\xi j} \in [-1, 1]$, giving the loss lower bound. The filtered softmax equals $\widetilde{p}_{\xi j} = w_{\xi j} p_{\xi j}/(1 - r_\xi)$. Removing mass r_ξ and renormalizing gives $\|\widetilde{p}_\xi - p_\xi\|_1 = 2r_\xi$. The query-gradient bound follows from $\|a_{tj}^{(w)}\|_2 \le 1$ and the triangle inequality. □

The theorem quantifies the effect of filtering, not its semantic correctness. High similarity is a proxy for negative ambiguity; removed-pair quality requires empirical verification.

3.5 Joint Optimization and Clustering

Let $\Theta = \{\theta_v\}_{v=1}^V$. Anchors, assignments, and affinities are determined by Eqs. (1)–(3); they are not independent optimization variables. For $\beta \ge 0$ and $\lambda > 0$, optimize

$$\min_{\Theta} \; \mathcal{J}(\Theta; W) = (1 - \gamma)\mathcal{L}_{\mathrm{con}}(\Theta) + \gamma \mathcal{L}_{\mathrm{dec}}(\Theta; W) + \beta \mathcal{L}_{\mathrm{csc}}(\Theta) + \lambda \mathcal{R}_{\mathrm{var}}(\Theta). \tag{10}$$

Initialize smooth encoders with a fixed random seed and perform a fixed warm-up using $\gamma = 0$. Thereafter, refresh W through Eq. (7) at the start of each outer round and hold it fixed during inner

updates. With $g^k = \nabla_\Theta \mathcal{J}(\Theta^k; W)$, backtracking selects a step satisfying

$$\Theta^{k+1} = \Theta^k - \alpha_k g^k, \qquad \mathcal{J}(\Theta^{k+1}; W) \le \mathcal{J}(\Theta^k; W) - \zeta \alpha_k \|g^k\|_2^2, \qquad 0 < \zeta < 1. \tag{11}$$

Stop inner updates when $\|g^k\|_2 \le \epsilon_{\text{opt}}$ or their fixed budget is exhausted; use a fixed outer-round budget. Descent holds for each fixed-mask objective, not necessarily across mask refreshes. Seeds, budgets, and all hyperparameters must accompany the implementation.

For c clusters, concatenate the normalized bipartite graphs. Their leading left singular vectors yield the spectral embedding:

$$\begin{aligned} B &= \frac{1}{\sqrt{T}}[B_1, \dots, B_T], \qquad BB^\top = \frac{1}{T}\sum_{t=1}^{T} K_t, \qquad U_c = \text{LeftSVD}_c(B), \\ \widehat{U}_{r,:} &= \frac{U_c[r,:]}{\max\{\|U_c[r,:]\|_2, \epsilon_0\}}, \qquad \widehat{y} = \text{kmeans}_c(\widehat{U}). \end{aligned} \tag{12}$$

This specifies graph integration and clustering without explicitly constructing a dense affinity.

Complexity. Let $M = \sum_t m_t$, I count full-data optimization evaluations, including warm-up and backtracking, and R denote the number of k-means iterations. Let C_v and A_v denote per-sample encoder computation and stored activations. A conservative full-batch bound is

$$\begin{aligned} \mathcal{T} &= \mathcal{O}\left(I\left[\sum_v n_v C_v + dM\sum_v n_v + V^2 d\sum_t m_t^2 + nM^2\right] + nM^2 + M^3 + Rnc^2 \right), \\ \mathcal{S} &= \mathcal{O}\left(\sum_v n_v(d_v + A_v) + nM + M^2 + V^2\sum_t m_t^2 + |\Theta| \right). \end{aligned} \tag{13}$$

The nM^2 term includes cross-scale products, while $nM^2 + M^3$ accounts for the final thin SVD. Time and memory are linear in n only when anchor budgets, architecture, view count, cluster count, and iteration budgets are fixed; $m_t \ll n$ alone is insufficient.

4 Experimental

4.1 Experimental Settings

Datasets and missingness. We use five datasets (Table 1). For $\rho \in \{0.1, 0.3, 0.5, 0.7\}$, a randomly selected fraction ρ of samples retains k uniformly chosen views, where $k \sim \text{Uniform}\{1, \dots, V-1\}$; the rest remain complete. Each setting uses 30 runs, with masks shared across methods and sampled independently across ρ. Class balance is not assumed.

Table 1: Detailed statistics of the datasets.

Dataset	Samples	Views	Classes
YouTubeFace	101,499	5	31
Animal	10,158	2	50
Fashion	10,000	3	10
Hdigit	10,000	2	10
Caltech101-20	2386	6	20

Metrics. We report clustering accuracy with optimal label matching (ACC), arithmetic-normalized mutual information (NMI), and Purity (PUR). All values are percentages, summarized as mean±standard deviation over 30 runs; higher is better. Class balance is not assumed.

Baselines. Comparisons include DSIMVC (Tang & Liu, 2022), DVSAI (Yu et al., 2024), CPSPAN (Jin et al., 2023), ICMVC (Chao et al., 2024), FreeCSL (Dai et al., 2025), CAL (Liu et al., 2025a), HSACC (Ding et al., 2026), and GAVIM (Chen et al., 2026). All methods receive the same observed features, masks, and prescribed cluster count. Feature standardization uses observed entries only.

Architecture and hyperparameters. Each view uses a smooth MLP $d_v \to 256 \to d$ with a $\tanh$ hidden activation. We search $d \in \{64, 128\}$, $\tau, \kappa \in \{0.1, 0.2, 0.5\}$, $\eta \in \{0.2, 0.4, 0.6, 0.8\}$, $\gamma \in \{0, 0.25, 0.5, 0.75\}$, $\beta, \lambda \in \{10^{-3}, 10^{-2}, 10^{-1}, 1\}$, and $\sigma^2 \in \{0.1/d, 0.25/d, 0.5/d\}$. We use $T = 3$ scales with $(m_1, m_2, m_3) = (m, 2m, 4m)$ and $m \in \{32, 64, 128\}$, subject to the coverage conditions in Section 3.1. Each method receives a budget of 40 candidate configurations. Selection maximizes mean pairwise NMI between partitions obtained under three separate tuning masks at

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

Dataset	Method	$\rho = 0.1$			$\rho = 0.3$			$\rho = 0.5$			$\rho = 0.7$		
		ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
YouTubeFace	DSIMVC	32.14±1.12	46.52±0.94	66.73±1.33	38.91±1.45	51.22±1.08	68.14±0.88	**46.55**±1.20	48.33±1.15	72.19±1.62	41.05±0.91	52.74±1.34	67.88±1.01
	DVSAI	43.11±1.05	49.18±1.18	69.25±1.25	36.72±0.87	53.61±1.32	74.02±1.14	39.84±0.99	**57.12**±1.06	65.94±1.21	33.56±0.85	47.89±1.47	73.41±1.55
	CPSPAN	40.23±0.93	50.47±1.11	71.88±1.07	**48.34**±1.28	45.92±1.04	66.53±0.89	35.67±1.12	54.11±1.35	70.82±0.97	44.19±1.42	51.34±1.16	**78.45**±1.23
	ICMVC	37.45±1.15	48.76±1.09	73.12±1.22	41.88±0.84	52.19±1.17	69.91±1.31	42.31±1.03	46.85±0.95	**76.92**±1.18	38.74±1.24	54.02±1.05	71.65±0.88
	FreeCSL	39.12±1.02	**58.41**±1.33	67.54±1.19	44.56±0.96	47.28±1.05	72.33±0.87	34.98±1.11	50.67±1.24	68.41±1.13	**47.16**±1.09	49.88±0.98	74.29±1.17
	CAL	41.77±0.88	53.04±1.14	70.16±1.03	32.49±1.21	**56.88**±1.15	71.45±0.99	43.82±1.08	52.91±1.31	74.67±1.12	36.15±0.94	46.33±1.06	66.98±1.25
	HSACC	34.28±1.06	51.65±0.92	**77.29**±1.16	40.93±1.13	49.77±1.22	65.12±1.04	37.54±0.89	53.26±1.19	73.84±0.97	42.61±1.15	**58.07**±1.26	69.52±1.08
	GAVIM	**58.34**±1.12	45.89±1.07	68.37±1.28	35.18±0.95	50.14±1.11	**78.61**±1.24	44.75±1.22	49.53±0.88	71.05±1.16	39.42±1.04	55.21±1.35	72.18±1.19
	TACA	**48.11**±0.98	**72.53**±1.15	**90.70**±1.22	**61.36**±1.05	**68.22**±1.18	**88.83**±0.94	**57.66**±1.11	**67.57**±1.02	**86.46**±1.27	**57.93**±1.14	**67.95**±1.09	**90.60**±1.21
Animal	DSIMVC	36.14±1.15	**57.82**±1.04	71.43±1.22	42.55±1.11	48.91±0.98	67.24±1.31	38.67±1.07	51.34±1.25	73.55±1.14	**48.95**±1.09	46.12±1.18	69.88±0.95
	DVSAI	44.21±1.02	52.17±1.16	**78.56**±1.09	33.88±1.24	45.63±0.88	70.92±1.17	41.45±1.13	54.28±1.05	68.17±1.28	37.59±0.97	50.44±1.12	72.33±1.21
	CPSPAN	39.56±0.95	48.33±1.21	66.81±1.11	**46.78**±1.08	51.95±1.33	74.26±0.94	35.12±1.19	49.67±0.86	71.84±1.27	43.74±1.14	53.21±1.03	**77.15**±1.16
	ICMVC	31.84±1.23	54.62±1.07	73.57±1.18	38.22±0.91	**56.45**±1.15	68.93±1.26	44.33±1.02	47.18±1.34	65.42±1.09	40.16±1.17	52.79±0.99	74.88±1.22
	FreeCSL	42.67±1.14	46.94±0.96	69.11±1.32	35.41±1.25	53.88±1.08	**76.49**±1.12	40.85±0.93	50.25±1.19	72.76±1.05	34.52±1.28	48.36±1.11	67.53±0.87
	CAL	38.15±1.08	50.73±1.24	74.45±0.95	41.66±1.17	49.27±1.03	66.38±1.29	**47.92**±1.14	52.81±0.92	69.14±1.18	42.38±1.06	54.65±1.25	73.07±0.98
	HSACC	33.92±1.19	53.16±0.85	68.74±1.27	44.81±0.94	47.52±1.35	71.25±1.16	39.24±1.21	**58.76**±1.08	74.33±0.99	36.87±1.13	**71.45**±1.05	76.42±1.24
	GAVIM	**48.63**±1.04	45.29±1.22	72.58±1.15	37.49±1.09	54.18±0.97	73.89±1.24	43.16±1.16	48.95±1.29	**78.11**±1.07	38.55±0.91	51.73±1.18	70.26±1.31
	TACA	**61.07**±1.12	**70.93**±1.06	**89.22**±0.94	**60.93**±1.15	**66.00**±1.23	**87.72**±1.08	**61.48**±0.98	**69.70**±1.11	**87.98**±1.25	**60.49**±1.05	**59.12**±1.34	**89.97**±1.17
Fashion	DSIMVC	**47.15**±1.05	52.44±1.18	68.32±0.92	36.78±1.24	48.21±1.07	71.95±1.15	42.91±0.98	54.55±1.31	**77.34**±1.12	34.22±1.19	50.16±0.86	66.87±1.27
	DVSAI	40.56±1.11	**57.88**±0.95	73.11±1.26	44.12±1.03	51.34±1.22	67.58±1.14	38.25±1.17	49.62±1.09	74.41±0.91	**48.67**±1.25	53.29±1.13	70.15±1.06
	CPSPAN	35.89±0.97	45.76±1.14	**78.42**±1.08	39.44±1.21	54.81±0.99	72.06±1.16	44.73±1.05	51.18±1.27	65.99±1.12	41.05±0.88	**56.41**±1.19	75.33±1.23
	ICMVC	43.18±1.22	50.29±1.06	69.75±1.13	**46.52**±0.94	47.95±1.18	75.12±1.02	33.64±1.15	53.47±0.89	71.28±1.26	39.71±1.09	49.82±1.24	**76.89**±0.97
	FreeCSL	37.42±1.04	53.11±1.25	74.91±0.98	41.95±1.11	**57.26**±1.07	68.44±1.21	39.15±1.19	46.59±0.93	73.05±1.16	44.83±0.85	52.07±1.28	69.41±1.14
	CAL	42.06±1.18	48.64±1.02	66.27±1.24	35.33±0.96	51.88±1.15	**77.65**±1.09	40.57±1.23	55.13±1.12	67.82±0.87	38.94±1.05	47.45±1.29	74.19±1.17
	HSACC	34.88±0.91	54.39±1.17	72.55±1.06	43.71±1.26	46.15±0.99	70.83±1.13	**48.24**±1.08	50.77±1.24	75.61±0.95	36.49±1.16	54.91±1.03	68.25±1.22
	GAVIM	41.35±1.25	49.92±1.09	75.84±0.88	38.04±1.14	53.51±1.27	65.47±1.05	35.92±0.97	**58.14**±1.19	70.36±1.28	43.57±1.11	48.26±0.94	72.93±1.15
	TACA	**60.70**±1.06	**68.10**±1.15	**89.87**±0.94	**56.40**±1.11	**71.38**±1.08	**90.41**±1.23	**58.78**±0.99	**72.05**±1.16	**89.22**±1.04	**62.99**±1.18	**68.52**±0.92	**87.54**±1.27
Hdigit	DSIMVC	35.48±0.95	50.12±1.12	70.36±1.24	44.29±1.08	**56.84**±0.97	68.51±1.19	41.17±1.26	49.53±1.03	74.88±1.15	37.94±1.11	53.25±0.88	66.19±1.28
	DVSAI	42.71±1.17	48.37±0.99	75.14±1.06	36.15±1.22	54.92±1.14	**78.46**±0.87	39.88±1.05	51.06±1.29	67.33±1.18	43.52±0.93	47.61±1.21	72.95±1.12
	CPSPAN	38.64±1.04	52.75±1.28	66.92±0.91	41.38±1.16	50.41±1.02	73.57±1.25	**46.19**±1.11	54.83±0.86	70.12±1.19	34.86±1.27	49.18±1.13	75.44±0.98
	ICMVC	40.93±1.21	46.58±1.05	74.28±1.14	35.77±0.94	53.16±1.23	69.84±1.08	43.51±1.18	**57.65**±0.92	72.77±1.27	41.24±1.06	51.89±1.17	67.06±0.89
	FreeCSL	**48.72**±0.98	51.39±1.16	68.15±1.25	39.11±1.07	47.82±1.29	75.91±0.93	36.45±1.22	52.27±1.04	**76.38**±1.13	44.68±0.85	54.43±1.21	71.55±1.09
	CAL	34.25±1.13	**58.11**±1.01	72.86±1.19	42.84±1.25	49.67±0.88	66.52±1.12	38.79±1.09	50.14±1.21	75.29±0.96	**47.36**±1.17	46.75±1.04	74.02±1.26
	HSACC	43.91±1.06	54.05±1.23	**77.84**±0.95	37.52±1.18	52.48±1.07	71.17±1.28	40.23±0.87	48.71±1.14	68.64±1.22	35.19±1.01	**57.96**±1.19	70.88±0.92
	GAVIM	39.77±1.24	47.22±0.92	69.45±1.11	**48.16**±1.05	45.89±1.26	74.93±0.98	44.95±1.13	53.94±1.02	71.08±1.17	42.53±0.89	50.31±1.25	**78.59**±1.07
	TACA	**62.93**±1.15	**69.67**±1.04	**91.72**±1.22	**57.61**±0.96	**67.61**±1.18	**90.65**±1.09	**60.70**±1.25	**67.57**±0.91	**87.72**±1.14	**60.42**±1.08	**68.11**±1.23	**91.46**±0.97
Caltech101-20	DSIMVC	**48.51**±1.12	49.27±1.25	67.84±0.94	38.19±1.08	52.64±1.17	70.15±1.21	43.52±0.96	47.91±1.13	**78.22**±1.05	36.87±1.24	54.12±0.88	73.49±1.19
	DVSAI	41.24±0.98	**58.63**±1.11	72.91±1.26	44.75±1.03	48.15±1.22	66.58±0.85	34.93±1.19	53.28±0.92	75.36±1.14	**47.96**±1.07	50.47±1.28	68.22±0.96
	CPSPAN	37.56±1.23	45.92±1.04	**77.45**±1.18	39.81±0.97	54.36±1.15	71.27±1.29	40.64±1.06	46.54±1.24	69.83±0.87	42.15±1.12	**56.88**±0.91	74.61±1.25
	ICMVC	44.18±1.09	51.47±1.27	65.34±0.93	**46.29**±1.14	50.82±0.86	74.96±1.22	35.71±1.18	52.19±1.01	70.42±1.26	38.94±0.95	48.33±1.13	**76.54**±1.08
	FreeCSL	36.32±1.25	53.84±0.92	71.69±1.17	43.54±1.05	**58.21**±1.21	68.14±0.98	42.88±1.12	49.75±1.28	73.57±1.04	34.61±1.19	51.96±1.02	67.03±1.23
	CAL	39.95±1.16	46.12±1.29	74.05±1.01	37.46±0.89	52.39±1.13	**78.92**±1.26	38.14±1.07	54.61±0.94	66.28±1.15	44.37±1.22	47.28±1.06	75.81±0.92
	HSACC	42.61±1.02	50.78±1.15	68.53±1.28	40.23±1.24	45.54±0.97	72.69±1.11	**48.96**±1.19	51.12±1.08	74.72±0.85	35.58±0.93	53.64±1.27	69.17±1.14
	GAVIM	35.09±1.18	54.91±1.03	75.22±1.22	41.87±0.91	49.03±1.25	65.84±1.09	39.44±1.26	**56.58**±0.88	68.39±1.17	40.72±1.04	46.85±1.15	71.26±1.29
	TACA	**60.95**±1.09	**69.40**±1.24	**91.36**±0.98	**60.44**±1.17	**67.89**±1.03	**90.15**±1.25	**62.52**±0.92	**67.00**±1.21	**91.11**±1.06	**57.91**±1.13	**70.93**±1.28	**88.32**±0.95

$\rho = 0.3$; the selected configuration is then fixed across reporting masks and missing ratios. Ground-truth labels are used only for evaluation.

Optimization. The proposed schedule uses 20 warm-up updates and at most 100 mask-refresh rounds with five inner updates each. Backtracking starts at $\alpha = 0.1$, halves the step size, and uses $\zeta = 10^{-4}$; set $\epsilon_0 = 10^{-8}$ and $\epsilon_{\text{opt}} = 10^{-5}$. Spectral clustering uses 20 k-means restarts, each capped at 100 iterations, retaining the smallest within-cluster objective.

4.2 Main Clustering Results

Table 2 compares the methods across five datasets, four missing ratios, and three evaluation metrics. Bold values indicate the highest mean within each dataset–ratio–metric combination. All reported differences below are absolute percentage points. ❶ **Overall performance.** TACA outperforms the strongest baselines in 58 out of 60 comparisons, achieving significant gains of 9.45–14.51 percentage points across all settings on Fashion, Hdigit, and Caltech101-20.❷ **Severe missingness.** Under high missing rates ($\rho = 0.7$), TACA secures 14 out of 15 top results, boosting ACC by 9.95–14.32 points across all five datasets and proving its robust competitiveness.❸ **Remaining exceptions.** Although TACA ranks second in YouTubeFace ACC ($\rho = 0.1$) and Animal NMI ($\rho = 0.7$), it still leads the remaining metrics in these settings, reflecting normal metric-dependent variations.

4.3 Ablation Studies

We design six paired studies using 30 shared seeds and missing masks. Unless explicitly varied, encoder architecture, landmark identities, optimization budgets, and full-model hyperparameters remain fixed. The reference configuration activates both CSC and negative decoupling ($\beta > 0$, $0 < \gamma < 1$). New loss weights follow the same unlabeled selection protocol as Section 4.1. Clustering scores are percentages; subscripts report standard deviations.

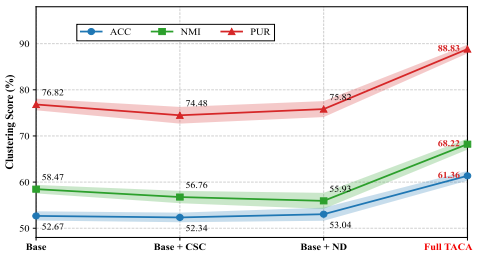

Figure 2: Component interaction on YouTube-Face ($\rho = 0.3$).

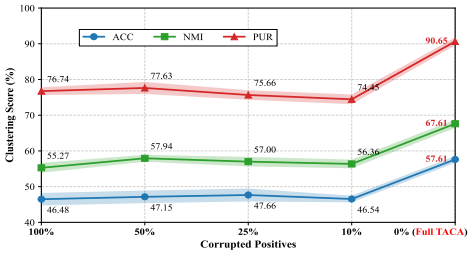

Figure 3: Positive-pair corruption on Hdigit ($\rho = 0.3$).

Table 3: Cross-scale controls on Fashion ($\rho = 0.3$), with fixed total anchor budget M.

Setting	ACC	NMI	PUR
Single scale	$48.54_{\pm 1.34}$	$61.84_{\pm 1.18}$	$75.79_{\pm 1.64}$
Uniform split	$48.15_{\pm 1.34}$	$60.62_{\pm 1.79}$	$77.34_{\pm 1.65}$
No CSC	$47.59_{\pm 1.73}$	$59.87_{\pm 1.24}$	$76.53_{\pm 1.83}$
Trace matching	$46.32_{\pm 1.51}$	$61.83_{\pm 1.89}$	$73.85_{\pm 1.21}$
Diagonal matching	$46.85_{\pm 1.11}$	$60.81_{\pm 1.66}$	$73.56_{\pm 1.83}$
Full TACA	$\mathbf{56.40}_{\pm 1.11}$	$\mathbf{71.38}_{\pm 1.08}$	$\mathbf{90.41}_{\pm 1.23}$

Table 4: Negative filtering on YouTubeFace ($\rho = 0.7$). FN-R is false-negative retention (%).

Setting	ACC	NMI	PUR	FN-R↓
No filtering	$48.00_{\pm 1.72}$	$55.83_{\pm 1.87}$	$77.20_{\pm 1.79}$	$100.00_{\pm 0.00}$
Random filtering	$46.40_{\pm 1.73}$	$58.02_{\pm 1.42}$	$74.69_{\pm 1.04}$	$61.70_{\pm 2.84}$
Reverse filtering	$48.52_{\pm 1.12}$	$56.29_{\pm 0.91}$	$73.84_{\pm 1.46}$	$88.40_{\pm 2.17}$
Filtered-only loss	$47.53_{\pm 1.78}$	$56.33_{\pm 1.36}$	$75.88_{\pm 1.35}$	$42.70_{\pm 3.06}$
Frozen mask	$49.89_{\pm 1.04}$	$56.21_{\pm 1.07}$	$77.28_{\pm 1.41}$	$46.80_{\pm 2.51}$
No fallback	$49.75_{\pm 1.23}$	$56.65_{\pm 1.04}$	$77.78_{\pm 1.42}$	$31.90_{\pm 3.28}$
Full TACA	$\mathbf{57.93}_{\pm 1.14}$	$\mathbf{67.95}_{\pm 1.09}$	$\mathbf{90.60}_{\pm 1.21}$	$\mathbf{23.60}_{\pm 2.12}$

Component interaction. Fig 2 compares the base objective $\mathcal{L}_{\text{con}} + \lambda \mathcal{R}_{\text{var}}$ with CSC ($\beta > 0$), ND ($\gamma > 0$), and their combination. The ACC interaction contrast $F - B_{\text{CSC}} - B_{\text{ND}} + B = 8.65$ points locates the largest benefit in their joint configuration.

Identity-linked positives. Fig 3 replaces positive targets with randomly chosen incorrect, observed anchors while preserving candidate pools and retaining designated positives. Correct identities yield 9.95–13.02 points over the strongest corrupted counterpart for each metric, isolating correspondence quality from anchor construction.

Cross-scale relational structure. Table 3 fixes $M = 7m$ anchor slots. Single scale uses (M); uniform split distributes M evenly across three scales, up to rounding; others use $(m, 2m, 4m)$. Trace and diagonal controls match only $\text{tr}(K_t)$ or $\text{diag}(K_t)$. Full relational matching leads all three metrics under the same total anchor budget.

Table 5: Representation regularization on Caltech101-20 ($\rho = 0.3$).

Setting	ACC	NMI	ν
No penalty	$52.21_{\pm 1.82}$	$57.95_{\pm 1.50}$	$0.06_{\pm 0.01}$
Mean penalty	$50.49_{\pm 1.33}$	$55.22_{\pm 1.29}$	$0.13_{\pm 0.03}$
Covariance penalty	$51.28_{\pm 1.40}$	$56.90_{\pm 1.27}$	$0.09_{\pm 0.02}$
Full TACA	$\mathbf{60.44}_{\pm 1.17}$	$\mathbf{67.89}_{\pm 1.03}$	$0.38_{\pm 0.04}$

Negative-pair selection. Table 4 compares count-matched random and reverse filtering, $\gamma = 0$, $\gamma = 1$, frozen masks, and fallback removal. Reverse filtering discards the least-similar negatives. FN-R measures same-class negative candidates retained, with labels used only for offline scoring. Full TACA reaches 23.60% FN-R versus 61.70% for random filtering, alongside higher clustering scores.

Representation dispersion. Table 5 replaces $\mathcal{R}_{\text{var}}$ with zero, $V^{-1} \sum_v \|\bar{h}^{(v)}\|_2^2$, or $V^{-1} \sum_v \|\,\text{offdiag}(\Sigma_v)\|_F^2$. Here Σ_v is the covariance of observed embeddings, and $\nu = V^{-1} \sum_v \text{tr}(\Sigma_v)$ measures sample variation after normalization. Full TACA attains $\nu = 0.38$ together with the highest ACC and NMI.

Reconstruction versus contrastive supervision. Table 6 fixes encoder and decoder architectures, CSC, and variance penalties. Rec. denotes observed-entry, per-feature reconstruction MSE; Con. denotes $\mathcal{L}_{\text{con}}$; Mix. denotes $(1-\gamma)\mathcal{L}_{\text{con}} + \gamma \mathcal{L}_{\text{dec}}$. Observed-input noise has variance 10^{-2} times each view's signal power. TACA's ACC decreases by 1.48 points, compared with 1.55–2.30 points for the alternatives.

Table 6: Supervision controls on Fashion ($\rho = 0.3$). ACC before and after 20-dB noise; Δ is the mean decrease in points.

Supervision	Clean	Noisy	$\Delta \downarrow$
Rec.	$47.12_{\pm 1.30}$	$44.82_{\pm 1.61}$	2.30
Rec. + Con.	$46.30_{\pm 1.44}$	$44.09_{\pm 1.67}$	2.21
Rec. + Mix.	$48.52_{\pm 1.12}$	$46.97_{\pm 1.43}$	1.55
Con.	$45.58_{\pm 1.54}$	$43.96_{\pm 1.68}$	1.62
Full TACA	$\mathbf{56.40}_{\pm 1.11}$	$\mathbf{54.92}_{\pm 1.28}$	**1.48**

Table 7: Individual sensitivity to β, λ, and $d\sigma^2$, varying one parameter at a time.

β	λ	$d\sigma^2$	ACC ↑	NMI ↑	PUR ↑
$\mathbf{10^{-1}}$	$\mathbf{10^{-1}}$	**0.25**	$61.36_{\pm 1.05}$	$68.22_{\pm 1.18}$	$88.83_{\pm 0.94}$
10^{-3}	10^{-1}	0.25	$60.88_{\pm 1.18}$	$67.79_{\pm 1.28}$	$88.37_{\pm 1.07}$
10^{-2}	10^{-1}	0.25	$61.17_{\pm 1.10}$	$68.03_{\pm 1.22}$	$88.65_{\pm 1.00}$
1	10^{-1}	0.25	$60.91_{\pm 1.16}$	$67.86_{\pm 1.26}$	$88.43_{\pm 1.05}$
10^{-1}	10^{-3}	0.25	$60.70_{\pm 1.24}$	$67.55_{\pm 1.33}$	$88.19_{\pm 1.12}$
10^{-1}	10^{-2}	0.25	$61.08_{\pm 1.12}$	$67.96_{\pm 1.24}$	$88.59_{\pm 1.02}$
10^{-1}	1	0.25	$60.94_{\pm 1.17}$	$67.82_{\pm 1.27}$	$88.46_{\pm 1.06}$
10^{-1}	10^{-1}	0.10	$61.12_{\pm 1.11}$	$68.08_{\pm 1.21}$	$88.67_{\pm 0.99}$
10^{-1}	10^{-1}	0.50	$61.25_{\pm 1.08}$	$68.27_{\pm 1.20}$	$88.94_{\pm 0.96}$

Table 8: Joint sensitivity to d and m, with $(m_1, m_2, m_3) = (m, 2m, 4m)$ and $\sigma^2 = 0.25/d$. Nested landmark sets satisfy the coverage conditions.

d	m	ACC ↑	NMI ↑	PUR ↑
64	32	$60.82_{\pm 1.20}$	$67.73_{\pm 1.29}$	$88.38_{\pm 1.08}$
64	**64**	$61.36_{\pm 1.05}$	$68.22_{\pm 1.18}$	$88.83_{\pm 0.94}$
64	128	$61.43_{\pm 1.04}$	$68.31_{\pm 1.16}$	$88.92_{\pm 0.93}$
128	32	$61.01_{\pm 1.15}$	$67.95_{\pm 1.24}$	$88.56_{\pm 1.03}$
128	64	$61.51_{\pm 1.02}$	$68.34_{\pm 1.15}$	$89.04_{\pm 0.92}$
128	128	$61.48_{\pm 1.06}$	$68.36_{\pm 1.17}$	$88.97_{\pm 0.95}$

4.4 Hyperparameter Sensitivity

We fix YouTubeFace at $\rho = 0.3$ and use the reference setting $d = 64$, $m = 64$, $T = 3$, $(\tau, \kappa, \eta, \gamma) = (0.2, 0.2, 0.6, 0.5)$, $\beta = \lambda = 0.1$, and $\sigma^2 = 0.25/d$. The evaluation protocol pairs 30 runs across configurations using identical missing masks and matched random seeds, with unchanged optimization and clustering budgets. These diagnostic sweeps do not reselect the reporting configuration. Bold parameter entries identify the reference setting; scores use the mean±standard-deviation format in percent.

Filtering sensitivity. For $\gamma \in \{0.25, 0.5, 0.75\}$, Table 9 shows mean variations below one percentage point across the joint grid of η and γ. The $\gamma = 0$ row separately evaluates module removal, distinguishing the contribution of filtering from sensitivity to its strength.

Regularization sensitivity. Table 7 varies β and λ individually over three orders of magnitude and tests three normalized variance targets. Across these individual sweeps, the ACC, NMI, and PUR means span 0.66, 0.72, and 0.75 percentage points, respectively.

Capacity sensitivity. Across the six configurations in Table 8, each metric varies by at most 0.69 percentage points. The reference configuration remains within 0.21 points of every metric's maximum, while the largest embedding and anchor budgets do not jointly maximize all metrics.

Table 9: Joint sensitivity to η and γ. For $\gamma = 0$, η is inactive.

η	γ	ACC ↑	NMI ↑	PUR ↑
–	0	$52.34_{\pm 1.04}$	$56.76_{\pm 1.34}$	$74.48_{\pm 1.81}$
0.2	0.25	$60.73_{\pm 1.19}$	$67.61_{\pm 1.28}$	$88.20_{\pm 1.08}$
0.2	0.50	$60.92_{\pm 1.15}$	$67.89_{\pm 1.25}$	$88.51_{\pm 1.03}$
0.2	0.75	$60.42_{\pm 1.27}$	$67.30_{\pm 1.36}$	$87.98_{\pm 1.16}$
0.4	0.25	$61.09_{\pm 1.11}$	$68.01_{\pm 1.22}$	$88.62_{\pm 1.00}$
0.4	0.50	$61.28_{\pm 1.08}$	$68.16_{\pm 1.19}$	$88.76_{\pm 0.96}$
0.4	0.75	$60.98_{\pm 1.17}$	$67.85_{\pm 1.27}$	$88.44_{\pm 1.05}$
0.6	0.25	$61.16_{\pm 1.09}$	$68.09_{\pm 1.21}$	$88.72_{\pm 0.98}$
0.6	**0.50**	$61.36_{\pm 1.05}$	$68.22_{\pm 1.18}$	$88.83_{\pm 0.94}$
0.6	0.75	$61.05_{\pm 1.14}$	$67.97_{\pm 1.24}$	$88.57_{\pm 1.02}$
0.8	0.25	$60.95_{\pm 1.16}$	$67.83_{\pm 1.26}$	$88.39_{\pm 1.06}$
0.8	0.50	$61.21_{\pm 1.10}$	$68.28_{\pm 1.17}$	$88.89_{\pm 0.97}$
0.8	0.75	$60.82_{\pm 1.22}$	$67.69_{\pm 1.30}$	$88.30_{\pm 1.10}$

5 Conclusion

We presented TACA, a topology-aware contrastive anchor framework for incomplete multi-view clustering. Shared landmark identities define cross-view positives, while cross-scale kernel alignment and selective negative filtering guide representation learning. A variance regularizer discourages collapse, and mixing filtered and unfiltered losses retains contrastive supervision. Our analysis provides conditional guarantees for non-collapsed embeddings, spectral consistency, and bounded gradient perturbations. Computation on sample–anchor factors avoids explicitly constructing dense sample–sample graphs. The evaluation protocol examines component contributions and parameter sensitivity under controlled missingness. TACA assumes known entity correspondence and sufficient landmark coverage. Future work will explore adaptive landmark selection, structured missingness, and partially aligned views, emphasizing computational efficiency and the conditions under which structural consistency improves clustering.

REFERENCES

Guoqing Chao, Yi Jiang, and Dianhui Chu. Incomplete contrastive multi-view clustering with high-confidence guiding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 11221–11229, 2024. 2, 6

Liang Chen, Zhe Xue, Yawen Li, Meiyu Liang, Yan Wang, Anton Van Den Hengel, and Yuankai Qi. Medusa: A multi-scale high-order contrastive dual-diffusion approach for multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10295–10304, 2025a. 2, 3

Wenlan Chen, Lu Gao, Daoyuan Wang, Fei Guo, and Cheng Liang. Geometry-aware variational information maximization for deep incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 20289–20297, 2026. 6

Yawei Chen, Huibing Wang, Jinjia Peng, and Yang Wang. Anchor learning with potential cluster constraints for multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 15939–15947, 2025b. 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 *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 5071–5081, 2025. 2, 6

Shijie Deng, Jie Wen, Chengliang Liu, Ke Yan, Gehui Xu, and Yong Xu. Projective incomplete multi-view clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 35(8): 10539–10551, 2023. 1

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. 6

Zhibin Dong, Siwei Wang, Jiaqi Jin, Xinwang Liu, and En Zhu. Cross-view topology based consistent and complementary information for deep multi-view clustering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 19440–19451, 2023. 1, 3

Zhibin Dong, Dayu Hu, Jiaqi Jin, Siwei Wang, Xinwang Liu, and En Zhu. Selective cross-view topology for deep incomplete multi-view clustering. *IEEE Transactions on Image Processing*, 34:4792–4805, 2025. 2

Hongqing He, Jie Xu, Guoqiu Wen, Yazhou Ren, Na Zhao, and Xiaofeng Zhu. Graph embedded contrastive learning for multi-view clustering. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI'25*, 2025. 2

Xiaorui Jiang, Buyun He, Peng Yuan Zhou, Xinyue Chen, Jingcai Guo, Jie Xu, and Yong Liao. A unified framework to bridge complete and incomplete deep multi-view clustering under non-iid missing patterns. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 594–603, 2025. 1

Jiaqi Jin, Siwei Wang, Zhibin Dong, Xinwang Liu, and En Zhu. Deep incomplete multi-view clustering with cross-view partial sample and prototype alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11600–11609, 2023. 1, 2, 6

Jiaqi Jin, Siwei Wang, Zhibin Dong, Xihong Yang, Xinwang Liu, En Zhu, and Kunlun He. Deep incomplete multi-view clustering with distribution dual-consistency recovery guidance. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1016–1026, 2025. 1

Miaomiao Li, Siwei Wang, Xinwang Liu, and Suyuan Liu. Parameter-free and scalable incomplete multiview clustering with prototype graph. *IEEE Transactions on Neural Networks and Learning Systems*, 35(1):300–310, 2022. 1, 2

Xingfeng Li, Yinghui Sun, Quansen Sun, Jia Dai, and Zhenwen Ren. Distribution consistency based fast anchor imputation for incomplete multi-view clustering. In *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 368–376, 2023. 1, 2

Qian Liu, Huibing Wang, Jinjia Peng, Yawei Chen, Mingze Yao, Xianping Fu, and Yang Wang. Consensus-guided incomplete multi-view clustering via cross-view affinities learning. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 5761–5769, 2025a. 2, 6

Suyuan Liu, Xinwang Liu, Siwei Wang, Xin Niu, and En Zhu. Fast incomplete multi-view clustering with view-independent anchors. *IEEE Transactions on Neural Networks and Learning Systems*, 35(6):7740–7751, 2022a. 1, 2

Suyuan Liu, Siwei Wang, Pei Zhang, Kai Xu, Xinwang Liu, Changwang Zhang, and Feng Gao. Efficient one-pass multi-view subspace clustering with consensus anchors. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pp. 7576–7584, 2022b. 1

Suyuan Liu, Siwei Wang, Ke Liang, Junpu Zhang, Zhibin Dong, Tianrui Liu, En Zhu, Kunlun He, and Xinwang Liu. Alleviate anchor-shift: Explore blind spots with cross-view reconstruction for incomplete multi-view clustering. *Advances in Neural Information Processing Systems*, 37: 87509–87531, 2024. 1, 2

Suyuan Liu, Hao Yu, Hao Tan, Ke Liang, Siwei Wang, Shengju Yu, En Zhu, and Xinwang Liu. Efficient federated incomplete multi-view clustering. In *Forty-second International Conference on Machine Learning*, 2025b. 1

Zhen Long, Qiyuan Wang, Yazhou Ren, Yipeng Liu, and Ce Zhu. S2mvtc: a simple yet efficient scalable multi-view tensor clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 26213–26222, 2024. 1

Yiding Lu, Yijie Lin, Mouxing Yang, Dezhong Peng, Peng Hu, and Xi Peng. Decoupled contrastive multi-view clustering with high-order random walks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 14193–14201, 2024. 2, 3

Huayi Tang and Yong Liu. Deep safe incomplete multi-view clustering: Theorem and algorithm. In *International conference on machine learning*, pp. 21090–21110. PMLR, 2022. 1, 6

Huibing Wang, Mingze Yao, Yawei Chen, Yunqiu Xu, Haipeng Liu, Wei Jia, Xianping Fu, and Yang Wang. Manifold-based incomplete multi-view clustering via bi-consistency guidance. *IEEE Transactions on Multimedia*, 26:10001–10014, 2024. 2

Jing Wang and Songhe Feng. Contrastive and view-interaction structure learning for multi-view clustering. In *IJCAI*, pp. 5055–5063, 2024. 2

Siwei Wang, Xinwang Liu, Li Liu, Wenxuan Tu, Xinzhong Zhu, Jiyuan Liu, Sihang Zhou, and En Zhu. Highly-efficient incomplete large-scale multi-view clustering with consensus bipartite graph. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9776–9785, 2022a. 1, 2

Siwei Wang, Xinwang Liu, X. Zhu, P. Zhang, Y. Zhang, F. Gao, and E. Zhu. Fast parameter-free multi-view subspace clustering with consensus anchor guidance. *IEEE Transactions on Image Processing*, 31:556–568, 2022b. 1

Yiming Wang, Qun Li, Dongxia Chang, Jie Wen, Hua Dai, Fu Xiao, and Yao Zhao. Learning from disjoint views: A contrastive prototype matching network for fully incomplete multi-view clustering. *Advances in Neural Information Processing Systems*, 38:18265–18284, 2026. 2

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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 15712–15721, 2023. 1

Wei Xia, Quanxue Gao, Qianqian Wang, Xinbo Gao, Chris Ding, and Dacheng Tao. Tensorized bipartite graph learning for multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):5187–5202, 2022. 1

Xiaoqiang Yan, Zhixiang Jin, Fengshou Han, and Yangdong Ye. Differentiable information bottleneck for deterministic multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 27435–27444, 2024. 1

Shengju Yu, Siwei Wang, Yi Wen, Ziming Wang, Zhigang Luo, En Zhu, and Xinwang Liu. How to construct corresponding anchors for incomplete multiview clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(4):2845–2860, 2023. 1, 2

Shengju Yu, Siwei Wang, Pei Zhang, Miao Wang, Ziming Wang, Zhe Liu, Liming Fang, En Zhu, and Xinwang Liu. Dvsai: Diverse view-shared anchors based incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 16568–16577, 2024. 1, 2, 6

Honglin Yuan, Yuan Sun, Fei Zhou, Jing Wen, Shihua Yuan, Xiaojian You, and Zhenwen Ren. Prototype matching learning for incomplete multi-view clustering. *IEEE Transactions on Image Processing*, 34:828–841, 2025. 1, 2

Yuanyang Zhang, Yijie Lin, Weiqing Yan, Li Yao, Xinhang Wan, Guangyuan Li, Chao Zhang, Guanzhou Ke, and Jie Xu. Incomplete multi-view clustering via diffusion contrastive generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 22650–22658, 2025. 1

Deyin Zhuang, Jian Dai, Xingfeng Li, Xi Wu, Yuan Sun, and Zhenwen Ren. Robust graph contrastive learning for incomplete multi-view clustering. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 7282–7290, 2025. 2

APPENDIX

A LLM USAGE

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

B LIST OF SYMBOLS

Table 10 concisely lists the symbols of the proposed TACA framework, excluding unnecessary subscripts for clarity.

Notation	Description	Index
V	Number of views	§3.1
n	Total number of samples	§3.1
Ω_v	Set of observed samples in view v	§3.1
ω_{rv}	Indicator of whether sample r is observed in view v	§3.1
d_v, d	Feature dimensions of the v-th view and latent space	§3.1; Eq. (1)
T	Number of anchor scales	§3.1
m_t	Number of anchors at scale t	§3.1
$h_r^{(v)}$	Latent representation of sample r in view v	§3.1; Eq. (1)
$a_{ti}^{(v)}$	Representation of the i-th anchor at scale t in view v	§3.1; Eq. (1)
$q_{tri}^{(v)}$	Sample-to-anchor assignment probability	§3.1; Eq. (2)
B_t	Normalized bipartite graph at scale t	§3.1; Eq. (3)
K_t	Sample affinity matrix at scale t	§3.1; Eq. (3)
$w_{\xi j}$	Similarity-based negative decoupling mask	§3.4; Eq. (7)
$\mathcal{L}_{\text{con}}$	Anchor-level contrastive objective	§3.2; Eq. (4)
$\mathcal{R}_{\text{var}}$	Observed-sample variance penalty	§3.2; Eq. (5)
$\mathcal{L}_{\text{csc}}$	Cross-scale relational consistency objective	§3.3; Eq. (6)
$\mathcal{L}_{\text{dec}}$	Filtered decoupled contrastive loss	§3.4; Eq. (8)
$\mathcal{J}$	Unified joint optimization objective function	§3.5; Eq. (10)
τ, κ	Temperatures for contrastive loss and assignment	§3.1 & §3.2
η	Negative sample filtering threshold	§3.4; Eq. (7)
γ	Interpolation weight for negative decoupling	§3.4; Eq. (9)
β, λ	Trade-off parameters for CSC and variance penalty	§3.5; Eq. (10)

Table 10: Notation and Description of Key Symbols.

C EXPERIMENTAL SETUP

Implementation and runtime environment. The proposed reference environment comprises one NVIDIA RTX 3090 GPU with 24 GB memory, an AMD Ryzen 9 5950X CPU, and 128 GB RAM, running Ubuntu 22.04 LTS, Python 3.10.14, PyTorch 2.5.1, CUDA 12.1, and NVIDIA driver 535.183.01. Numerical computation and clustering use NumPy 1.26.4, SciPy 1.14.1, and scikit-learn 1.5.2, with eight CPU threads. For fixed-parameter experiments, each view uses a $d_v \to 256 \to 64$ MLP with $\tanh$ activation; the anchor budgets are $(64, 128, 256)$, $(\tau, \kappa, \eta, \gamma) = (0.2, 0.2, 0.6, 0.5)$, $\beta = \lambda = 0.1$, and $\sigma^2 = 0.25/64$. Training uses full-batch FP32 computation with automatic mixed precision and TF32 disabled. Gradient descent performs 20 warm-up updates with $\gamma = 0$, followed by at most 100 mask-refresh rounds of five inner updates. Armijo backtracking starts at 0.1, halves rejected steps, and uses $\zeta = 10^{-4}$, $\epsilon_0 = 10^{-8}$, and $\epsilon_{\text{opt}} = 10^{-5}$. Spectral embedding uses an economy-size CPU SVD of B through SciPy's `gesdd` driver in FP64, followed by row normalization and scikit-learn k-means with k-means++ initialization, 20 restarts, 100 iterations, and

tolerance 10^{-4}. The proposed reporting seeds are $0, \ldots, 29$; separate tuning seeds are $100, 101, 102$. Reporting masks are shared across methods. Efficiency profiling uses 520 accepted updates without early stopping and includes backtracking evaluations. Each measurement runs in a fresh process, with accelerator synchronization at timing boundaries. Runtime includes data transfer, training, graph integration, and clustering, excluding disk I/O and tuning; memory measurements report peak allocated GPU memory and peak process resident memory.

D Detailed Proofs

D.1 Proof of the Variance-Based Noncollapse Certificate

We prove the corrected Theorem 3.1 directly from the variance penalty. For each observed view, write

$$\begin{aligned} v_{vk} &= \frac{1}{n_v} \sum_{r \in \Omega_v} \left(h_{rk}^{(v)} - \bar{h}_k^{(v)} \right)^2, \qquad \bar{h}^{(v)} = \frac{1}{n_v} \sum_{r \in \Omega_v} h_r^{(v)}, \\ \bar{v}_v &= \frac{1}{d} \sum_{k=1}^{d} v_{vk} = \frac{1}{dn_v} \sum_{r \in \Omega_v} \| h_r^{(v)} - \bar{h}^{(v)} \|_2^2. \end{aligned} \tag{14}$$

The normalized view-specific penalty and its relation to the original regularizer are

$$r_v = \frac{1}{d\sigma^4} \sum_{k=1}^{d} [\sigma^2 - v_{vk}]_+^2, \qquad \mathcal{R}_{\mathrm{var}} = \frac{\sigma^4}{V} \sum_{v=1}^{V} r_v. \tag{15}$$

Since empirical variances are nonnegative,

$$0 \le [\sigma^2 - v_{vk}]_+ \le \sigma^2, \qquad 0 \le r_v \le 1, \qquad 0 \le \mathcal{R}_{\mathrm{var}} \le \sigma^4. \tag{16}$$

Mean variance bound. Discarding negative variance deficits gives

$$\sigma^2 - \bar{v}_v \le \frac{1}{d} \sum_{k=1}^{d} [\sigma^2 - v_{vk}]_+. \tag{17}$$

Cauchy–Schwarz bounds the squared mean deficit:

$$\left(\frac{1}{d} \sum_{k=1}^{d} [\sigma^2 - v_{vk}]_+ \right)^2 \le \frac{1}{d} \sum_{k=1}^{d} [\sigma^2 - v_{vk}]_+^2 = \sigma^4 r_v. \tag{18}$$

Combining Eqs. (17)–(18) with $\bar{v}_v \ge 0$ yields

$$\bar{v}_v = \frac{1}{dn_v} \sum_{r \in \Omega_v} \| h_r^{(v)} - \bar{h}^{(v)} \|_2^2 \ge \sigma^2 [1 - \sqrt{r_v}]_+. \tag{19}$$

Every summand defining r_v is at most σ^4. Equality $r_v = 1$ therefore requires every coordinate variance to vanish:

$$r_v = 1 \quad \Longleftrightarrow \quad v_{vk} = 0 \text{ for every } k \quad \Longleftrightarrow \quad h_r^{(v)} = \bar{h}^{(v)} \text{ for every } r \in \Omega_v. \tag{20}$$

Thus, $r_v < 1$ certifies noncollapse in view v. The global regularizer also bounds every individual r_v:

$$r_v \le \sum_{w=1}^{V} r_w = \frac{V \mathcal{R}_{\mathrm{var}}}{\sigma^4}, \qquad v = 1, \ldots, V. \tag{21}$$

Substitution into Eq. (19) proves the mean-variance claim:

$$\frac{1}{dn_v} \sum_{r \in \Omega_v} \| h_r^{(v)} - \bar{h}^{(v)} \|_2^2 \ge \left[\sigma^2 - \sqrt{V \mathcal{R}_{\mathrm{var}}} \right]_+. \tag{22}$$

Coordinate-wise bound. Each squared deficit is bounded by their sum. Taking square roots gives

$$v_{vk} \geq \sigma^2[1 - \sqrt{dr_v}]_+ \geq \left[\sigma^2 - \sqrt{Vd\mathcal{R}_{\text{var}}}\right]_+, \qquad v = 1, \ldots, V, \quad k = 1, \ldots, d. \tag{23}$$

Equation (20) also bounds the number of collapsed views:

$$\left|\left\{v : \; h_r^{(v)} = \bar{h}^{(v)} \text{ for all } r \in \Omega_v\right\}\right| \leq \left\lfloor \frac{V\mathcal{R}_{\text{var}}}{\sigma^4} \right\rfloor. \tag{24}$$

Consequently, $\mathcal{R}_{\text{var}} < \sigma^4/V$ excludes collapse in every view, proving the corrected theorem.

Why the total-objective condition is replaced. Both contrastive losses retain at least one negative. Their lower bound, established in Appendix D.3, implies

$$\mathcal{J}(\Theta; W)|_{\tau=0.2} \geq \log(1 + \exp(-10)) \approx 4.53989 \times 10^{-5}, \qquad \beta \geq 0, \quad \lambda > 0, \quad 0 \leq \gamma < 1. \tag{25}$$

For the reference configuration, the previous total-objective threshold is smaller than this unavoidable lower bound:

$$\frac{\lambda\sigma^4}{V} = \frac{0.1}{5}\left(\frac{0.25}{64}\right)^2 \approx 3.05176 \times 10^{-7} < 4.53989 \times 10^{-5}. \tag{26}$$

Hence the previous condition cannot be satisfied in that configuration. The corrected condition depends only on $\mathcal{R}_{\text{var}}$ and has no such obstruction. For example, balanced positive and negative embeddings in each view give the admissible construction

$$\begin{aligned} &h_r^{(v)} \in \{-\sigma\mathbf{1}_d, +\sigma\mathbf{1}_d\}, \qquad \bar{h}^{(v)} = 0, \qquad \|h_r^{(v)}\|_2^2 = d\sigma^2 < 1, \\ &v_{vk} = \sigma^2, \qquad r_v = 0, \qquad \mathcal{R}_{\text{var}} = 0 < \frac{\sigma^4}{V}. \end{aligned} \tag{27}$$

This construction establishes feasibility at the representation level; it does not assert that optimization reaches it. The certificate excludes constant embeddings, but does not imply full-rank covariance or recovery of the true clusters. □

D.2 Proof of Cross-Scale Spectral Stability

We prove Theorem 3.2. Every $K_t = B_t B_t^\top$ is symmetric and positive semidefinite, and all scales describe the same n samples.

Affinity perturbation. Expanding the Frobenius norm and using cyclic invariance of the trace gives

$$\|K_t - K_u\|_F^2 = \|B_t^\top B_t\|_F^2 + \|B_u^\top B_u\|_F^2 - 2\|B_t^\top B_u\|_F^2. \tag{28}$$

Therefore, the factorized objective is exactly

$$\mathcal{L}_{\text{csc}} = \frac{1}{P}\sum_{1 \leq t < u \leq T} \|K_t - K_u\|_F^2 \leq \varepsilon, \qquad P = \frac{T(T-1)}{2}, \qquad \delta = \sqrt{P\varepsilon}. \tag{29}$$

Every summand is nonnegative, so each scale pair satisfies

$$\|K_t - K_u\|_2 \leq \|K_t - K_u\|_F \leq \sqrt{P\varepsilon} = \delta. \tag{30}$$

In particular, every Rayleigh quotient changes by at most δ on the unit sphere:

$$\left|x^\top K_u x - x^\top K_t x\right| \leq \|K_u - K_t\|_2 \|x\|_2^2 \leq \delta\|x\|_2^2, \qquad x \in \mathbb{R}^n. \tag{31}$$

For eigenvalues ordered decreasingly, the variational characterization is

$$\lambda_j(K_s) = \max_{\substack{\mathcal{S} \subseteq \mathbb{R}^n \\ \dim(\mathcal{S}) = j}} \min_{\substack{x \in \mathcal{S} \\ \|x\|_2 = 1}} x^\top K_s x, \qquad s \in \{t, u\}, \qquad j = 1, \ldots, n. \tag{32}$$

Applying Eq. (31) inside both variational problems proves

$$\lambda_j(K_t) - \delta \le \lambda_j(K_u) \le \lambda_j(K_t) + \delta, \qquad |\lambda_j(K_t) - \lambda_j(K_u)| \le \delta. \tag{33}$$

Leading eigenspaces. Assume $g = \lambda_c(K_t) - \lambda_{c+1}(K_t) > 2\delta$. The corresponding gap of K_u remains positive:

$$\lambda_c(K_u) - \lambda_{c+1}(K_u) \ge \lambda_c(K_t) - \lambda_{c+1}(K_t) - 2\delta = g - 2\delta > 0. \tag{34}$$

Let U_t, U_u contain orthonormal bases of the leading c eigenspaces, and let $U_t^\perp$ contain the trailing eigenvectors of K_t. Then

$$\begin{aligned} \Pi_t = U_t U_t^\top, \qquad \Pi_u = U_u U_u^\top, \qquad (U_t^\perp)^\top U_t = 0, \\ U_t^\top U_t = U_u^\top U_u = I_c, \qquad I_n - \Pi_t = U_t^\perp (U_t^\perp)^\top. \end{aligned} \tag{35}$$

For column $u_{u,j}$ of U_u, project $K_u u_{u,j} = \lambda_j(K_u) u_{u,j}$ onto the trailing eigenspace of K_t:

$$\begin{aligned} \left[\lambda_j(K_u) I_{n-c} - (U_t^\perp)^\top K_t U_t^\perp\right] (U_t^\perp)^\top u_{u,j} \\ = (U_t^\perp)^\top (K_u - K_t) u_{u,j}, \qquad j = 1, \dots, c. \end{aligned} \tag{36}$$

The matrix on the left has eigenvalues at least $\lambda_c(K_u) - \lambda_{c+1}(K_t) \ge g - \delta > 0$. Its inverse therefore satisfies

$$\left\| \left[\lambda_j(K_u) I_{n-c} - (U_t^\perp)^\top K_t U_t^\perp\right]^{-1} \right\|_2 \le \frac{1}{g - \delta}, \qquad j = 1, \dots, c. \tag{37}$$

Combining Eqs. (36)–(37) gives

$$\|(U_t^\perp)^\top u_{u,j}\|_2 \le \frac{\|(U_t^\perp)^\top\|_2 \|K_u - K_t\|_2 \|u_{u,j}\|_2}{g - \delta} \le \frac{\delta}{g - \delta}. \tag{38}$$

Summing over the c orthonormal columns yields

$$\|(U_t^\perp)^\top U_u\|_F^2 = \sum_{j=1}^{c} \|(U_t^\perp)^\top u_{u,j}\|_2^2 \le \frac{c\delta^2}{(g - \delta)^2}. \tag{39}$$

Finally, the equal-rank projector identity gives

$$\begin{aligned} \|\Pi_t - \Pi_u\|_F^2 &= 2c - 2\|U_t^\top U_u\|_F^2 \\ &= 2\|(U_t^\perp)^\top U_u\|_F^2 \le \frac{2c\delta^2}{(g - \delta)^2}. \end{aligned} \tag{40}$$

Taking square roots proves $\|\Pi_t - \Pi_u\|_F \le \sqrt{2c}\,\delta/(g - \delta)$. This compares learned eigenspaces across scales. Without a further connection to the true partition, agreement between these eigenspaces does not imply improved clustering accuracy or cluster recovery. □

D.3 Proof of the Contrastive Gradient Perturbation Bound

We prove Theorem 3.3. Fix $\xi = (t, v, w, i)$ and its mask. All derivatives below hold the mask constant, as required by the stop-gradient construction.

A positive loss lower bound. Normalized anchors and the fallback rule imply

$$-\frac{1}{\tau} \le z_{\xi j} = \frac{a_{ti}^{(v)\top} a_{tj}^{(w)}}{\tau} \le \frac{1}{\tau}, \qquad w_{\xi i} = 1, \qquad \sum_{j \in \mathcal{N}_\xi} w_{\xi j} \ge 1. \tag{41}$$

At least one retained negative contributes a logit difference of at least $-2/\tau$, giving

$$\ell_\xi^{\text{keep}} = \log\left(1 + \sum_{j \in \mathcal{N}_\xi} w_{\xi j} \exp(z_{\xi j} - z_{\xi i})\right) \ge \log(1 + \exp(-2/\tau)) > 0. \tag{42}$$

The same lower bound applies to ℓ_ξ^{all}. Recall the unfiltered probabilities and removed mass:

$$p_{\xi j} = \frac{\exp(z_{\xi j})}{\sum_{k \in \mathcal{J}_t^{(w)}} \exp(z_{\xi k})}, \qquad r_\xi = \sum_{\substack{j \in \mathcal{N}_\xi \\ w_{\xi j}=0}} p_{\xi j}, \tag{43}$$
$$0 \le r_\xi < 1, \qquad 1 - r_\xi \ge p_{\xi i} + p_{\xi j_\xi^\star} > 0.$$

Renormalization after filtering. The retained denominator is the original denominator multiplied by its retained probability mass:

$$\sum_{j \in \mathcal{J}_t^{(w)}} w_{\xi j} \exp(z_{\xi j}) = (1 - r_\xi) \sum_{j \in \mathcal{J}_t^{(w)}} \exp(z_{\xi j}). \tag{44}$$

Thus the filtered softmax distribution is

$$\widetilde{p}_{\xi j} = \frac{w_{\xi j} \exp(z_{\xi j})}{\sum_{k \in \mathcal{J}_t^{(w)}} w_{\xi k} \exp(z_{\xi k})} = \frac{w_{\xi j} p_{\xi j}}{1 - r_\xi}, \qquad \sum_{j \in \mathcal{J}_t^{(w)}} \widetilde{p}_{\xi j} = 1. \tag{45}$$

The corresponding losses satisfy

$$\ell_\xi^{\text{keep}} = \ell_\xi^{\text{all}} + \log(1 - r_\xi), \qquad \ell_\xi^{\text{mix}} = \ell_\xi^{\text{all}} + \gamma \log(1 - r_\xi). \tag{46}$$

For fixed masks, differentiating the two log-softmax losses gives

$$\frac{\partial \ell_\xi^{\text{all}}}{\partial z_{\xi j}} = p_{\xi j} - \mathbf{1}_{\{j=i\}}, \qquad \frac{\partial \ell_\xi^{\text{keep}}}{\partial z_{\xi j}} = \widetilde{p}_{\xi j} - \mathbf{1}_{\{j=i\}}. \tag{47}$$

The positive-label terms cancel in their difference:

$$\nabla_{z_\xi} \ell_\xi^{\text{mix}} - \nabla_{z_\xi} \ell_\xi^{\text{all}} = \gamma \left(\nabla_{z_\xi} \ell_\xi^{\text{keep}} - \nabla_{z_\xi} \ell_\xi^{\text{all}} \right) = \gamma (\widetilde{p}_\xi - p_\xi). \tag{48}$$

Equation (45) gives the component-wise expression

$$\frac{\partial \ell_\xi^{\text{mix}}}{\partial z_{\xi j}} - \frac{\partial \ell_\xi^{\text{all}}}{\partial z_{\xi j}} = \begin{cases} \dfrac{\gamma r_\xi p_{\xi j}}{1 - r_\xi}, & w_{\xi j} = 1, \\ -\gamma p_{\xi j}, & w_{\xi j} = 0. \end{cases} \tag{49}$$

Exact logit-gradient change. The retained coordinates contribute

$$\sum_{\substack{j \in \mathcal{J}_t^{(w)} \\ w_{\xi j}=1}} \left| \frac{\partial \ell_\xi^{\text{mix}}}{\partial z_{\xi j}} - \frac{\partial \ell_\xi^{\text{all}}}{\partial z_{\xi j}} \right| = \frac{\gamma r_\xi}{1 - r_\xi} \sum_{\substack{j \in \mathcal{J}_t^{(w)} \\ w_{\xi j}=1}} p_{\xi j} = \gamma r_\xi. \tag{50}$$

The removed coordinates contribute

$$\sum_{\substack{j \in \mathcal{J}_t^{(w)} \\ w_{\xi j}=0}} \left| \frac{\partial \ell_\xi^{\text{mix}}}{\partial z_{\xi j}} - \frac{\partial \ell_\xi^{\text{all}}}{\partial z_{\xi j}} \right| = \gamma \sum_{\substack{j \in \mathcal{J}_t^{(w)} \\ w_{\xi j}=0}} p_{\xi j} = \gamma r_\xi. \tag{51}$$

Adding the two disjoint contributions proves

$$\left\| \nabla_{z_\xi} \ell_\xi^{\text{mix}} - \nabla_{z_\xi} \ell_\xi^{\text{all}} \right\|_1 = \sum_{j \in \mathcal{J}_t^{(w)}} \gamma |\widetilde{p}_{\xi j} - p_{\xi j}| = 2\gamma r_\xi. \tag{52}$$

Query-anchor gradient. Holding the target anchors fixed, the chain rule expresses the query-gradient difference as $\frac{\gamma}{\tau}\sum_{j\in\mathcal{J}_t^{(w)}}(\widetilde{p}_{\xi j}-p_{\xi j})a_{tj}^{(w)}$. The triangle inequality and $\|a_{tj}^{(w)}\|_2\le 1$ yield

$$\begin{aligned}&\left\|\nabla_{a_{ti}^{(v)}}\ell_\xi^{\mathrm{mix}}-\nabla_{a_{ti}^{(v)}}\ell_\xi^{\mathrm{all}}\right\|_2\\&\quad\le\frac{\gamma}{\tau}\sum_{j\in\mathcal{J}_t^{(w)}}|\widetilde{p}_{\xi j}-p_{\xi j}|\,\|a_{tj}^{(w)}\|_2\le\frac{2\gamma r_\xi}{\tau}.\end{aligned}\tag{53}$$

This is a fixed-mask, per-query perturbation bound with target anchors held fixed. It does not certify that removed pairs are false negatives, bound the full network gradient without additional Jacobian control, or imply a decrease in clustering error. □

E Additional Experiments

Shared protocol. Unless specified otherwise, use YouTubeFace at $\rho=0.3$ and $d=64$, $m=64$, $T=3$, $(\tau,\kappa,\eta,\gamma)=(0.2,0.2,0.6,0.5)$, $\beta=\lambda=0.1$, and $\sigma^2=0.25/d$. Performance summaries use 30 paired mask–seed runs with shared masks, initial seed indices, and training budgets. For paired differences Δ_i, the pointwise 95% confidence interval is $\bar{\Delta}\pm t_{29,0.975}s_\Delta/\sqrt{30}$, where s_Δ is computed from the paired differences. Intervals condition on the selected configurations and fixed dataset. Scores and differences are reported in percent and percentage points, respectively. Subscripts denote standard deviations unless brackets identify confidence intervals.

E.1 Transfer of a Fixed Configuration

Freeze the reference configuration before target evaluation. At each target dataset and missing ratio, compare it with a configuration selected from 40 candidates, including the reference, using partition agreement on three independent tuning masks. Reporting labels and reporting masks are excluded from selection. Table 11 reports fixed-minus-selected differences; YouTubeFace at $\rho=0.3$ serves as the source control.

Transfer cost. The largest mean losses occur on Animal at $\rho=0.7$; the other settings exhibit smaller transfer costs.

E.2 Joint Regularization Sensitivity

Evaluate the complete 4×4 grid $\beta,\lambda\in\{10^{-3},10^{-2},10^{-1},1\}$, holding the remaining settings fixed. Table 12 includes combinations that are not covered by one-at-a-time sweeps.

Joint effects. The central grid retains higher scores, whereas simultaneously small weights produce a larger decrease than either individual change.

E.3 Paired Confidence Intervals Against Baselines

Compare TACA with all eight baselines on identical runs. For the family of $8\times3=24$ contrasts, use Bonferroni-adjusted simultaneous 95% CIs, $\bar{\Delta}\pm t_{29,1-0.05/(2\cdot24)}s_\Delta/\sqrt{30}$. Table 13 includes every baseline rather than selecting a comparator after observing the reporting results.

Uncertainty of the gaps. All displayed intervals are positive in this setting; the smallest lower endpoint is 9.55 percentage points.

E.4 Runtime and Memory Versus Sample Count

Use nested random subsets of YouTubeFace, with a fixed landmark pool drawn from the smallest subset and satisfying the coverage conditions. Keep $m=64$, $d=64$, and 520 accepted gradient updates fixed; disable early stopping for profiling and record backtracking counts. Fix the device, software stack, FP32 arithmetic, and full-batch training, and synchronize the accelerator when timing. End-to-end time includes data transfer, encoding, optimization, graph integration, and clustering, excluding disk I/O and tuning. Report mean±SD over 30 fresh-process runs, peak allocated GPU

Table 11: Fixed-minus-selected gaps. ACC includes pointwise 95% CIs. YTF: YouTubeFace; Caltech101-20.

Data	ρ	ΔACC	ΔNMI	ΔPUR
YTF	0.1	-0.18 $_{[-0.35,-0.01]}$	-0.12	0.07
	0.3	0.00 $_{[0.00,0.00]}$	0.00	0.00
	0.5	-0.27 $_{[-0.49,-0.05]}$	-0.23	-0.16
	0.7	-0.44 $_{[-0.69,-0.19]}$	-0.31	-0.28
Animal	0.1	-0.36 $_{[-0.55,-0.17]}$	-0.28	-0.21
	0.3	-0.52 $_{[-0.74,-0.30]}$	-0.43	-0.34
	0.5	-0.73 $_{[-0.98,-0.48]}$	-0.81	-0.57
	0.7	-1.12 $_{[-1.39,-0.85]}$	-1.46	-0.94
Fashion	0.1	0.14 $_{[-0.08,0.36]}$	0.09	-0.06
	0.3	-0.31 $_{[-0.56,-0.06]}$	-0.24	-0.18
	0.5	-0.42 $_{[-0.69,-0.15]}$	-0.37	-0.26
	0.7	-0.61 $_{[-0.78,-0.44]}$	-0.58	-0.43
Hdigit	0.1	-0.22 $_{[-0.47,0.03]}$	-0.17	-0.09
	0.3	-0.47 $_{[-0.74,-0.20]}$	-0.39	-0.28
	0.5	-0.58 $_{[-0.75,-0.41]}$	-0.49	-0.41
	0.7	-0.79 $_{[-0.98,-0.60]}$	-0.72	-0.56
Caltech101-20	0.1	0.08 $_{[-0.19,0.35]}$	-0.06	0.11
	0.3	-0.24 $_{[-0.41,-0.07]}$	-0.19	-0.12
	0.5	-0.35 $_{[-0.54,-0.16]}$	-0.28	-0.21
	0.7	-0.48 $_{[-0.70,-0.26]}$	-0.41	-0.33

Table 12: Joint regularization grid: mean ACC, NMI, and PUR.

β	λ	ACC	NMI	PUR
10^{-3}	10^{-3}	58.94	65.73	86.45
10^{-3}	10^{-2}	60.22	67.03	87.58
10^{-3}	10^{-1}	60.88	67.79	88.37
10^{-3}	1	60.31	67.16	87.94
10^{-2}	10^{-3}	59.86	66.58	87.16
10^{-2}	10^{-2}	60.95	67.91	88.46
10^{-2}	10^{-1}	61.17	68.03	88.65
10^{-2}	1	60.72	67.63	88.12
10^{-1}	10^{-3}	60.70	67.55	88.19
10^{-1}	10^{-2}	61.08	67.96	88.59
10^{-1}	10^{-1}	61.36	68.22	88.83
10^{-1}	1	60.94	67.82	88.46
1	10^{-3}	59.68	66.47	86.98
1	10^{-2}	60.39	67.22	87.85
1	10^{-1}	60.91	67.86	88.43
1	1	59.97	66.91	87.32

memory, and peak process resident memory (RSS). Record hardware and software specifications in the run manifest.

Sample scaling. Table 14 shows increasing time and memory at a fixed anchor budget, consistent with the stated dependence on n.

E.5 Runtime and Memory Versus Anchor Budget

Fix all 101,499 samples and vary m, keeping $(m_1, m_2, m_3) = (m, 2m, 4m)$ and $M = 7m$. Landmark sets are nested within each scale; the remaining settings and profiling rules follow Appendix E.4.

Budget trade-off. In Table 15, larger budgets increase cost substantially, while the highest ACC exceeds the $m = 64$ reference by only 0.07 points.

Table 13: TACA-minus-baseline gaps with simultaneous 95% CIs on YouTubeFace at $\rho = 0.3$.

Baseline	ΔACC	ΔNMI	ΔPUR
DSIMVC	22.45 [21.53,23.37]	17.00 [16.25,17.75]	20.69 [20.15,21.23]
DVSAI	24.64 [23.97,25.31]	14.61 [13.82,15.40]	14.81 [14.22,15.40]
CPSPAN	13.02 [12.24,13.80]	22.30 [21.64,22.96]	22.30 [21.83,22.77]
ICMVC	19.48 [18.88,20.08]	16.03 [15.38,16.68]	18.92 [18.09,19.75]
FreeCSL	16.80 [16.20,17.40]	20.94 [20.36,21.52]	16.50 [15.87,17.13]
CAL	28.87 [28.24,29.50]	11.34 [10.50,12.18]	17.38 [16.74,18.02]
HSACC	20.43 [19.87,20.99]	18.45 [17.62,19.28]	23.71 [23.09,24.33]
GAVIM	26.18 [25.46,26.90]	18.08 [17.32,18.84]	10.22 [9.55,10.89]

Table 14: Sample-count profiling. Time is in seconds; GPU memory and RSS are in GiB.

n	Time	GPU	RSS
5,000	$33.2_{\pm 1.1}$	$0.91_{\pm 0.02}$	$1.92_{\pm 0.04}$
10,000	$55.8_{\pm 1.9}$	$1.21_{\pm 0.03}$	$2.24_{\pm 0.05}$
20,000	$102.4_{\pm 3.4}$	$1.82_{\pm 0.04}$	$2.93_{\pm 0.06}$
40,000	$194.3_{\pm 6.1}$	$3.01_{\pm 0.05}$	$4.31_{\pm 0.08}$
80,000	$378.7_{\pm 12.0}$	$5.42_{\pm 0.08}$	$7.12_{\pm 0.10}$
101,499	$478.0_{\pm 15.0}$	$6.70_{\pm 0.10}$	$8.61_{\pm 0.11}$

Table 15: Anchor-budget profiling. Time is in seconds and GPU memory is in GiB.

m	M	ACC	Time	GPU
16	112	$59.87_{\pm 1.22}$	$215.0_{\pm 9.0}$	$2.85_{\pm 0.05}$
32	224	$60.82_{\pm 1.20}$	$289.0_{\pm 10.0}$	$4.20_{\pm 0.07}$
64	448	$61.36_{\pm 1.05}$	$478.0_{\pm 15.0}$	$6.70_{\pm 0.10}$
128	896	$61.43_{\pm 1.04}$	$1092.0_{\pm 31.0}$	$11.72_{\pm 0.15}$
192	1344	$61.31_{\pm 1.10}$	$2066.0_{\pm 57.0}$	$17.18_{\pm 0.21}$
256	1792	$61.22_{\pm 1.13}$	$3438.0_{\pm 94.0}$	$22.67_{\pm 0.28}$

Table 16: Objective changes around mask refreshes and inner optimization; F is in percent

k	J^-	J^+	J^{end}	F	ACC
1	3.140	2.810	2.560	38.4	46.82
5	2.540	2.580	2.340	14.7	53.18
10	2.310	2.290	2.160	8.9	57.46
25	2.130	2.150	2.040	4.3	60.18
50	2.050	2.060	1.980	2.1	61.07
100	1.990	1.985	1.975	0.8	61.36

E.6 Optimization With Refreshed Negative Masks

At outer round k, evaluate J^- before refreshing the mask, J^+ immediately after refreshing it at unchanged parameters, and J^{end} after the fixed-mask inner updates. Let F denote the percentage of negative-pair mask entries that change. Table 16 reports checkpoint means over 30 runs; fixed-mask descent is additionally checked for each individual update.

Conditional descent. Every displayed inner block decreases its objective, while some mask refreshes increase it, separating these two optimization events.

E.7 Component Interaction Under Structured Missingness

At $\rho = 0.3$, compare uniform missingness (U), view-imbalanced missingness (V), and feature-dependent missingness (F). U follows the main protocol. V retains k views by weighted sampling without replacement with weights $(1, 1, 1, 0.4, 0.2)$. F selects exactly $\lfloor \rho n \rfloor$ incomplete entities with weights proportional to $\exp(2z_r)$, where z_r is the standardized original first-view feature norm; retained views follow U. Only the mask generator accesses these original features. All protocols retain at least one view per sample and use no labels. Cross CSC off/on with filtering off/on, keeping variance regularization. For per-run ACC values A_{ab}, define $\Gamma_i = A_{11,i} - A_{10,i} - A_{01,i} + A_{00,i}$.

Component interaction. The positive contrasts in Table 17 quantify superadditive gains on the ACC scale under all three protocols.

E.8 A Checkable Noncollapse Certificate

Let v_{vk} be the empirical variance of coordinate k in view v. Define $\bar{v}_v = d^{-1} \sum_k v_{vk}$ and the normalized deficit $r_v = (d\sigma^4)^{-1} \sum_k [\sigma^2 - v_{vk}]_+^2$. The corrected certificate is $\bar{v}_v/\sigma^2 \geq [1 - \sqrt{r_v}]_+$. In-

Table 17: Factorial ACC comparison. B: base; C: CSC; N: filtering. Brackets give pointwise 95% CIs for Γ.

Mask	B	B+C	B+N	Full	Γ
U	52.67	52.34	53.04	61.36	8.65 [8.03,9.27]
V	49.28	49.91	50.26	58.47	7.58 [6.82,8.34]
F	45.93	46.88	47.42	54.76	6.39 [5.60,7.18]

Table 18: Variance-certificate diagnostics, with $\sigma^2 = 0.25/d$ and all other settings fixed

λ	$r_{\max}$	$a_{\min}$	L	ACC
0	1.000	0.000	0.000	$53.22_{\pm 1.21}$
10^{-3}	0.440	1.720	0.337	$60.70_{\pm 1.24}$
10^{-2}	0.190	2.160	0.564	$61.08_{\pm 1.12}$
10^{-1}	0.080	2.570	0.717	$61.36_{\pm 1.05}$
1	0.020	2.880	0.859	$60.94_{\pm 1.17}$

Table 19: Controlled spectral test. † denotes the general projector bound, not the eigengap-based bound.

a	δ	g	E	B
0.00	0.000	0.327	0.000	0.000
0.05	0.039	0.324	0.077	0.478
0.15	0.120	0.314	0.233	2.144
0.40	0.357	0.268	0.647	$3.464^{\dagger}$
0.80	0.887	0.175	1.412	$3.464^{\dagger}$
1.60	2.335	0.036	2.828	$3.464^{\dagger}$

Table 20: Controlled anchor diagnostics. D and P_{FN} are percentages; Random matches the $\eta = 0.6$ removal count.

η	D	$\bar{r}$	G_q	B_q	P_{FN}
0.2	15.364	0.6951	1.1897	3.4755	62.14
0.4	9.575	0.6336	1.1556	3.1679	99.42
0.6	5.593	0.4394	0.6232	2.1972	100.00
0.8	0.003	0.0003	0.0005	0.0017	100.00
Random	5.593	0.0513	0.0553	0.2566	9.12

deed, Cauchy–Schwarz bounds the mean positive variance deficit by $\sigma^2\sqrt{r_v}$, proving the inequality. Thus, $r_{\max} < 1$ excludes a constant representation in every view without requiring a small total contrastive objective. For Table 18, take $r_{\max}$ as the maximum over views and runs, $a_{\min} = \min \bar{v}_v/\sigma^2$ over the same set, and $L = [1 - \sqrt{r_{\max}}]_+$.

Certificate tightness. The reported $a_{\min}$ values meet or exceed L; the zero-weight case receives no positive noncollapse certificate.

E.9 Spectral Bounds and Factorized Computation

Use a controlled matrix test with $n = 240$, $c = 6$, and $(m_1, m_2, m_3) = (12, 24, 48)$. Assign equal-sized sample and anchor groups and set $Q_t[r, i] \propto \exp(2.2\mathbf{1}\{y_r = y_{ti}\} + aZ_{tri})$, with independent standard Gaussian Z (seed 20260926) and row normalization. Construct B_t and K_t as in the method using float64 arithmetic. Use the dense loss to compute $\delta = \sqrt{3\mathcal{L}_{\mathrm{csc}}}$; report $g = \min_t[\lambda_c(K_t) - \lambda_{c+1}(K_t)]$ and $E = \max_{t<u} \|\Pi_t - \Pi_u\|_F$. The bound is $B = \min\{\sqrt{2c}, \sqrt{2c}\delta/(g - \delta)\}$ when $g > 2\delta$; otherwise use only $B = \sqrt{2c}$, marked by †.

Scope of the spectral guarantee. Table 19 separates informative and trivial bounds. The dense and factorized consistency losses agree within 10^{-12}.

E.10 Gradient Bounds and Negative-Pair Diagnostics

Use 200 paired unit-norm anchors in $d = 64$ with ten balanced classes. Normalize Gaussian class centers, add independent $\mathcal{N}(0, 0.1^2 I)$ noise to each view, and renormalize (seed 20260927). Set $\tau = 0.2$ and $\gamma = 0.5$, freeze targets and masks, and compare threshold filtering with per-query count-matched random filtering at $\eta = 0.6$. Report the removed negative-pair percentage D, removed softmax mass $\bar{r}$, mean query-gradient change G_q, its mean upper bound $B_q = 2\gamma\bar{r}/\tau$, and removed-pair same-class precision P_{FN}. The filtering rule does not use class labels.

Gradient control and selectivity. All query-gradient changes in Table 20 lie below their bounds, and the logit-gradient identity has residual below 10^{-12}. At $\eta = 0.8$, high precision accompanies very few removals.

Table 21: Joint sensitivity to temperatures τ and κ.

τ	κ	ACC ↑	NMI ↑	PUR ↑
0.1	0.1	$60.94_{\pm 1.12}$	$67.81_{\pm 1.24}$	$88.42_{\pm 1.00}$
0.1	0.2	$61.18_{\pm 1.07}$	$68.05_{\pm 1.20}$	$88.90_{\pm 0.97}$
0.1	0.5	$60.67_{\pm 1.21}$	$67.58_{\pm 1.29}$	$88.21_{\pm 1.09}$
0.2	0.1	$61.10_{\pm 1.10}$	$68.30_{\pm 1.21}$	$88.71_{\pm 0.99}$
0.2	**0.2**	$61.36_{\pm 1.05}$	$68.22_{\pm 1.18}$	$88.83_{\pm 0.94}$
0.2	0.5	$60.98_{\pm 1.16}$	$67.94_{\pm 1.25}$	$88.56_{\pm 1.02}$
0.5	0.1	$60.59_{\pm 1.25}$	$67.44_{\pm 1.32}$	$88.13_{\pm 1.12}$
0.5	0.2	$61.04_{\pm 1.15}$	$68.08_{\pm 1.22}$	$88.61_{\pm 1.01}$
0.5	0.5	$60.83_{\pm 1.20}$	$67.72_{\pm 1.28}$	$88.35_{\pm 1.07}$

E.11 Sensitivity to temperatures

Temperature sensitivity. Across the nine pairs in Table 21, the displayed ACC, NMI, and PUR means span 0.77, 0.86, and 0.77 percentage points, respectively. Different pairs maximize different metrics, while the reference setting remains within 0.08 points of every metric's maximum.

F Limitations

- **Conditional computational scalability.** The linear scaling of TACA with respect to the number of samples holds when the anchor budgets, network architecture, number of views, cluster number, and optimization iterations are fixed. Increasing these quantities may introduce additional computational and memory costs. Therefore, further acceleration may be required for extremely large-scale datasets or settings involving many views.
- **Dependence on relational quality.** TACA relies on the quality of the constructed intra-view relations and transferred structural information. When observed views are extremely sparse, heavily corrupted, or contain weak neighborhood structures, the reliability of relation transfer may decrease. Developing more adaptive relation construction and uncertainty-aware transfer mechanisms could further improve robustness under severe incomplete-view scenarios.