CORE: COUPLED REPRESENTATION AND CLUSTER-AWARE TOPOLOGY FOR SCALABLE MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

As multi-view data continue to proliferate in real-world applications, anchor-based clustering provides an efficient paradigm for latent cluster discovery. Despite recent progress, existing methods still encounter the following limitations. First, they primarily emphasize the consistency among views, whereas the learned structures shared by diverse views lack sufficient cluster discriminability. Second, the interaction between sample-anchor representation learning and cluster-aware anchor topology optimization remains insufficiently explored, limiting mutual refinement during optimization. To address these issues, we propose Coupled Representation and Cluster-Aware Topology (CORE), a scalable multi-view clustering framework designed to leverage cross-view consistency to learn cluster-discriminative structures through unified optimization. Specifically, CORE jointly learns view-specific sample-anchor representations through robust reconstruction and imposes a low-rank tensor structure to capture high-order consistency across views. Meanwhile, an adaptive anchor topology is inferred from representations refined by cross-view consistency and regularized by spectral connectivity constraints to capture cluster-oriented structural information. Accordingly, representation learning and topology optimization are coupled to enable mutual refinement, yielding more discriminative structures for clustering. By performing structural learning in the compact anchor space, CORE avoids expensive sample-wise graph construction while maintaining favorable scalability on large-scale datasets. Extensive experiments on real-world benchmarks demonstrate the effectiveness of CORE across diverse clustering scenarios.

1 INTRODUCTION

By exploiting consistency and complementarity across multiple sources, multi-view clustering achieves more reliable and robust results than single-view methods (Chao et al., 2021; Yang et al., 2025; Liu et al., 2026). With the increasing scale of data collected in real-world applications, conventional methods suffer from substantial computational and storage costs due to sample-wise similarity learning or self-representation (Yang et al., 2024; Liu et al., 2025; Ji & Feng, 2025). To this end, anchor-based clustering constructs structural representations with a compact anchor set, substantially reducing the computational burden in large-scale settings (Guo et al., 2026; Tang et al., 2026; Wei et al., 2026). Nevertheless, the heterogeneity across views typically leads to inconsistent representations and structural discrepancies, making it challenging to effectively exploit shared information for reliable clustering (Zhang et al., 2024a; Wu et al., 2024a; Chen et al., 2026a).

Accordingly, to leverage complementary information from multiple views, existing methods generally seek to capture informative cross-view structures and can be broadly categorized into three groups: representation-based methods, graph-based methods, and tensor-based methods. Specifically, representation-based methods project view-specific features into latent spaces and integrate them for clustering (Gan et al., 2025; He et al., 2025; Zhang et al., 2025). For example, Wan et al. (2023) proposed AWMVC to learn coefficient representations across multiple latent spaces to derive a consensus representation, while Liu et al. (2021) combined multi-view matrix factorization with partition generation within a unified framework to learn latent representations. In contrast, graph-based methods explicitly characterize relationships among samples or anchors and exploit

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.

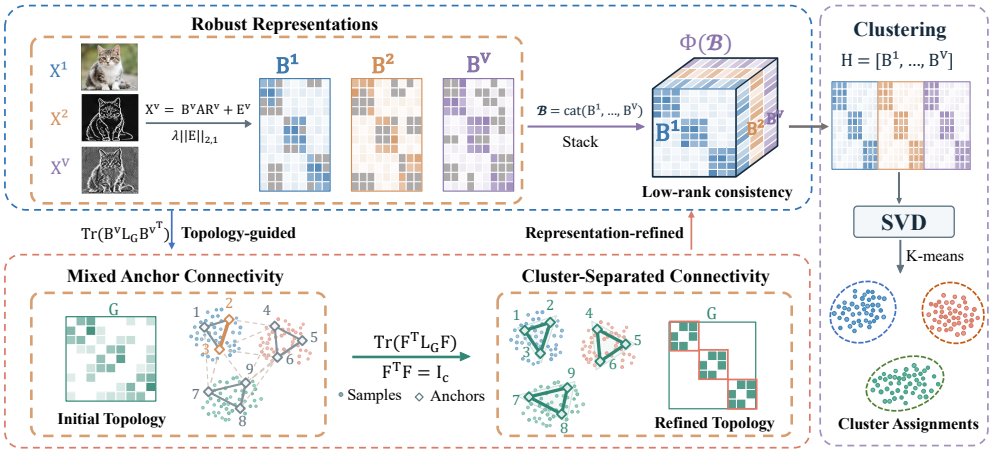


Figure 1: Schematic illustration of CORE. Specifically, CORE learns view-specific sample-anchor representations through robust reconstruction and tensor regularization to capture high-order cross-view consistency. An adaptive anchor topology is further optimized with spectral connectivity, enabling mutual refinement between representation and topology learning. Finally, the optimized representations are concatenated and transformed by SVD for subsequent k-means clustering.

similarity structures to uncover latent clustering information (Lu et al., 2025; Tang et al., 2025). Representative methods, including PGAL and HRFAL, construct compact sample-anchor structures and refine them through structural regularization, enabling cross-view relational information to be effectively integrated (Zeng et al., 2026; Wu et al., 2024b). By replacing full sample-wise graphs with compact anchor-based structures, these methods reduce computational costs while preserving informative structural relationships (Zhang et al., 2024b; Chen et al., 2026b). Despite these advances, representation-based and graph-based methods primarily rely on matrix-level representations, limiting their ability to fully capture complex dependencies across multiple views.

To further capture high-order structural information across views, tensor-based methods organize view-specific representations into tensors and impose low-rank constraints to more effectively exploit cross-view dependencies (Du et al., 2025; Sun et al., 2024; Gu et al., 2024). In this way, structural correlations among multiple views can be jointly characterized beyond separate matrix-level representations, enabling more comprehensive preservation of shared structures. For example, Xu et al. (2025) constructed slim tensors from latent representations to explore high-order dependencies across views. Ji & Feng (2025) combined anchor representation learning with non-convex low-rank tensor regularization to enhance cross-view structural consistency. Recently, Wang et al. (2026) introduced multi-rank constraints on anchor graph tensors to induce block-diagonal structures and mitigate anchor misalignment across views. Although these methods effectively exploit tensor structures, they mainly focus on preserving shared information across views while providing limited guidance for cluster-discriminative structure learning. Consequently, the learned representations remain consistent across views but lack explicit cluster-oriented structures.

Motivated by the aforementioned issues, we propose a Coupled Representation and Cluster-Aware Topology (CORE) framework for scalable multi-view clustering. The overall schematic of CORE is illustrated in Fig. 1, and the main contributions are summarized as follows:

- A unified scalable multi-view clustering framework is proposed by coupling sample-anchor representation learning with adaptive topology optimization, enabling cross-view consistency and clustering-oriented structural information to be jointly exploited within a unified learning process.
- A high-order representation learning mechanism is developed by integrating robust reconstruction with non-convex low-rank tensor regularization on view-specific sample-anchor representations to suppress corruptions and capture consistent high-order structures shared across multiple views.
- A cluster-aware topology learning strategy is introduced to adaptively construct anchor relationships from the learned representations and enforce spectral connectivity, thereby guiding representation refinement toward more discriminative cluster structures.

2 PRELIMINARIES

Notation: Throughout this paper, matrices, vectors, and tensors are represented by bold uppercase, bold lowercase, and calligraphic letters, respectively. For a matrix $\mathbf{M}$, $\mathbf{m}_i$ and m_{ij} denote its i-th row and the (i,j)-th element, respectively. The operators $\mathrm{Tr}(\cdot)$, $\|\cdot\|_F$ and $\|\cdot\|_{2,1}$ denote the trace, Frobenius norm and $\ell_{2,1}$-norm, respectively. The key notations are summarized in Appendix A.

Tensor Modeling: For a tensor $\mathcal{B} \in \mathbb{R}^{n\times m\times V}$, let $\widetilde{\mathcal{B}} = \mathscr{R}(\mathcal{B}) \in \mathbb{R}^{m\times V\times n}$ denote its rearranged form, with the corresponding Fourier representation given by $\widehat{\mathcal{B}} = \mathrm{fft}(\widetilde{\mathcal{B}}, [], 3)$. To promote low-rank structure of $\mathcal{B}$, the non-convex function $\phi_\delta(s) = e^{\delta^2} s/(s+\delta)$ is applied to the singular values of the Fourier-domain frontal slices Ji & Feng (2025), yielding $\Phi(\mathcal{B}) = \frac{1}{n}\sum_{k=1}^{n}\sum_j \phi_\delta(\sigma_j(\widehat{\mathcal{B}}^{(k)}))$, where $\sigma_j(\cdot)$ denotes the j-th singular value.

3 METHODOLOGY

3.1 CONSISTENT ANCHOR REPRESENTATION CONSTRUCTION

Under large-scale scenarios, directly exploring relationships among all samples incurs substantial computational and storage costs, limiting the applicability of multi-view clustering. Therefore, an anchor-based representation model is introduced to construct compact sample-anchor representations for efficient structural learning, which can be formulated as:

$$\mathbf{X}^v = \mathbf{B}^v\mathbf{A}\mathbf{R}^v + \mathbf{E}^v, \qquad \text{s.t. } \mathbf{A}\mathbf{A}^\top \preceq \mathbf{I}_m,\ \mathbf{R}^v\mathbf{R}^{v\top} = \mathbf{I}_r, \tag{1}$$

where $\mathbf{B}^v \in \mathbb{R}^{n\times m}$ denotes the view-specific sample-anchor representation, $\mathbf{A} \in \mathbb{R}^{m\times r}$ represents the shared anchor basis, $\mathbf{R}^v \in \mathbb{R}^{r\times d_v}$ maps the shared latent space to the v-th feature space and $\mathbf{E}^v \in \mathbb{R}^{n\times d_v}$ denotes the reconstruction residual. To avoid degenerate solutions and preserve the latent structure, a spectral-norm constraint is imposed on $\mathbf{A}$, while row orthogonality is imposed on $\mathbf{R}^v$. However, independently learned representations fail to adequately preserve structural consistency across multiple views, weakening the effective integration of shared information for subsequent clustering. To this end, the view-specific representations $\mathbf{B}^v$ are stacked into a third-order tensor and regularized with a low-rank tensor penalty to capture cross-view correlations:

$$\min_{\{\mathbf{B}^v\}_{v=1}^V,\mathbf{E}} \lambda\|\mathbf{E}\|_{2,1} + \Phi(\mathcal{B}), \quad \text{s.t. } \mathbf{X}^v = \mathbf{B}^v\mathbf{A}\mathbf{R}^v + \mathbf{E}^v, \mathcal{B} = \mathrm{cat}(\mathbf{B}^1, \mathbf{B}^2, \ldots, \mathbf{B}^V) \in \mathbb{R}^{n\times m\times V}, \tag{2}$$

where the reconstruction residuals from all views are concatenated as $\mathbf{E} = [\mathbf{E}^1, \mathbf{E}^2, \ldots, \mathbf{E}^V]$, and the $\ell_{2,1}$-norm is imposed to suppress the influence of corrupted samples. Specifically, the reconstruction term enhances the robustness of view-specific sample-anchor representations during representation learning, while the low-rank tensor regularization reveals dominant structures shared across views. Consequently, the learned representations jointly preserve robustness and cross-view consistency, exhibiting clearer and more coherent structures for clustering.

3.2 ADAPTIVE CLUSTER-AWARE TOPOLOGY LEARNING

To exploit information from multiple views, existing methods typically learn shared representations or preserve common structures across views, while overlooking the distinction between multi-view consistency and clustering discriminability. Considering that graph structures are effective in capturing intrinsic structural information and improving representation quality, an anchor-level graph is introduced to characterize the relationships among anchors while retaining computational efficiency. Specifically, the denoised and refined sample-anchor representations $\mathbf{B}^v$ are used to adaptively construct the anchor topology with structural guidance, formulated as follows:

$$\min_{\mathbf{G}} \gamma\sum_{v=1}^{V}\mathrm{Tr}(\mathbf{B}^v\mathbf{L}_G\mathbf{B}^{v\top}) + \beta\|\mathbf{G}\|_F^2, \quad \text{s.t. } \mathbf{G} \geq 0,\ \mathbf{G}\mathbf{1}_m = \mathbf{1}_m,\ \mathrm{diag}(\mathbf{G}) = 0, \tag{3}$$

where $\mathbf{G} \in \mathbb{R}^{m\times m}$ denotes the anchor topology matrix with g_{ij} measuring the relationship between the i-th and j-th anchors, and $\mathbf{L}_G = \mathbf{D}_G - \frac{\mathbf{G}+\mathbf{G}^\top}{2}$ denotes the corresponding graph Laplacian with $\mathbf{D}_G$ being the diagonal degree matrix. Therefore, Eq. (3) promotes structural consistency among

anchors with larger affinities, preserving reliable sample relationships and the structural information captured by the learned representations. According to Theorem 1 in Appendix B, to induce a topology with c connected components, the rank condition $\mathrm{rank}(\mathbf{L}_G) = m - c$ is introduced and further transformed into the following spectral formulation based on the Ky Fan theorem Mohar (1991):

$$\min_{\mathbf{F}^\top \mathbf{F} = \mathbf{I}_c} \mathrm{Tr}(\mathbf{F}^\top \mathbf{L}_G \mathbf{F}), \tag{4}$$

where $\mathbf{F} \in \mathbb{R}^{m \times c}$ denotes the spectral embedding of the anchor graph. Therefore, the learned topology is jointly shaped by the refined sample-anchor representations and spectral connectivity, which preserves local relational information and progressively enhances cluster-oriented structure to provide more discriminative guidance for clustering.

3.3 Overall Framework of CORE

By combining the consistent anchor representation in Eq. (2) with the adaptive anchor topology in Eq. (3) and spectral connectivity in Eq. (4), CORE unifies cross-view representation learning with cluster-aware topology optimization. Accordingly, the overall objective of CORE is formulated as:

$$\begin{aligned} \min_{\substack{\{\mathbf{B}^v, \mathbf{R}^v, \mathbf{E}^v\}_{v=1}^V, \mathbf{A}, \mathbf{G}, \mathbf{F} \\ \mathbf{G} \geq 0,\ \mathrm{diag}(\mathbf{G}) = 0}} & \underbrace{\lambda \|\mathbf{E}\|_{2,1} + \Phi(\mathcal{B})}_{\text{Robust Representation}} \quad \underbrace{+\gamma \sum_{v=1}^{V} \mathrm{Tr}(\mathbf{B}^v \mathbf{L}_G \mathbf{B}^{v\top})}_{\text{Structure Coupling}} \underbrace{+\beta \|\mathbf{G}\|_F^2 + \eta\, \mathrm{Tr}(\mathbf{F}^\top \mathbf{L}_G \mathbf{F})}_{\text{Cluster-Aware Topology}} \\ \text{s.t.} \quad & \mathbf{X}^v = \mathbf{B}^v \mathbf{A} \mathbf{R}^v + \mathbf{E}^v,\ \mathbf{A}\mathbf{A}^\top \preceq \mathbf{I}_m,\ \mathbf{R}^v \mathbf{R}^{v\top} = \mathbf{I}_r,\ \mathbf{G}\mathbf{1}_m = \mathbf{1}_m,\ \mathbf{F}^T \mathbf{F} = \mathbf{I}_c, \end{aligned} \tag{5}$$

where λ, γ, and β are hyperparameters, while η is adaptively updated during optimization to regulate the spectral connectivity constraint. Within Eq. (5), CORE couples anchor representation learning with topology optimization, allowing the refined representations to facilitate adaptive topology construction. Meanwhile, the learned topology in turn regularizes representation learning, progressively improving clustering discriminability while preserving cross-view consistency.

3.4 Optimization

Considering that Eq. 5 is non-convex due to the coupling among multiple variables and the non-convex tensor regularization, an alternating optimization algorithm is developed to update one variable with the remaining variables fixed. Specifically, an auxiliary tensor $\mathcal{J} = \mathcal{B}$ is introduced, and Eq. 5 is reformulated as the following augmented Lagrangian problem (Bertsekas, 1982):

$$\begin{aligned} \mathcal{L} =& \lambda \|\mathbf{E}\|_{2,1} + \Phi(\mathcal{J}) + \gamma \sum_{v=1}^{V} \mathrm{Tr}(\mathbf{B}^v \mathbf{L}_G \mathbf{B}^{v\top}) + \beta \|\mathbf{G}\|_F^2 + \eta\, \mathrm{Tr}(\mathbf{F}^\top \mathbf{L}_\mathbf{G} \mathbf{F}) \\ &+ \sum_{v=1}^{V} \frac{\mu}{2} \|\mathbf{X}^v - \mathbf{B}^v \mathbf{A} \mathbf{R}^v - \mathbf{E}^v + \frac{\mathbf{\Lambda}^v}{\mu}\|_F^2 + \frac{\xi}{2} \|\mathcal{B} - \mathcal{J} + \frac{\mathcal{Q}}{\xi}\|_F^2. \end{aligned} \tag{6}$$

where $\{\mathbf{\Lambda}^v\}_{v=1}^V$ are the Lagrange multipliers for the reconstruction constraints, $\mathcal{Q}$ is associated with the auxiliary constraint $\mathcal{B} = \mathcal{J}$, μ and ξ are the penalty parameters for the reconstruction and tensor consistency constraints, respectively. The detailed optimization procedure is described as follows:

• **Updating $\mathbf{B}_v$:** By fixing other variables, the subproblem w.r.t $\mathbf{B}_v$ can be formulated as follows:

$$\min_{\mathbf{B}^v} \quad \frac{\mu}{2} \|\mathbf{X}^v - \mathbf{B}^v \mathbf{A} \mathbf{R}^v - \mathbf{E}^v + \frac{\mathbf{\Lambda}^v}{\mu}\|_F^2 + \frac{\xi}{2} \|\mathbf{B}^v - \mathbf{J}^v + \frac{\mathbf{Q}^v}{\xi}\|_F^2 + \gamma\, \mathrm{Tr}(\mathbf{B}^v \mathbf{L}_\mathbf{G} \mathbf{B}^{vT}). \tag{7}$$

where $\mathbf{J}^v$ and $\mathbf{Q}^v$ denote the v-th slices of $\mathcal{J}$ and $\mathcal{Q}$, respectively. Denoting $\widetilde{\mathbf{X}}^v = \mathbf{X}^v - \mathbf{E}^v + \mathbf{\Lambda}^v / \mu$, the derivative of Eq. 7 w.r.t $\mathbf{B}^v$ is set to zero, yielding:

$$\mu(\mathbf{B}^v \mathbf{A}\mathbf{A}^T - \widetilde{\mathbf{X}}^v \mathbf{R}^{vT} \mathbf{A}^T) + \xi(\mathbf{B}^v - \mathbf{J}^v + \frac{\mathbf{Q}^v}{\xi}) + 2\gamma \mathbf{B}^v \mathbf{L}_\mathbf{G} = \mathbf{0}, \tag{8}$$

Accordingly, the closed-form solution of $\mathbf{B}^v$ can be calculated as follows:

$$\mathbf{B}^v = (\mu\widetilde{\mathbf{X}}^v\mathbf{R}^{vT}\mathbf{A}^T + \xi\mathbf{J}^v - \mathbf{Q}^v)(\mu\mathbf{A}\mathbf{A}^T + \xi\mathbf{I}_m + 2\gamma\mathbf{L}_\mathbf{G})^{-1}. \tag{9}$$

Therefore, the update of $\mathbf{B}^v$ jointly incorporates residual-corrected reconstruction information, cross-view consistency, and structural guidance from the learned anchor topology, refining the sample-anchor representation toward a more discriminative structure.

• **Updating $\mathbf{E}$ and $\mathbf{\Lambda}^v$:** By fixing other variables, the subproblem w.r.t. $\mathbf{E}$ can be formulated as:

$$\min_{\mathbf{E}} \quad \lambda\|\mathbf{E}\|_{2,1} + \frac{\mu}{2}\sum_{v=1}^{V}\|\mathbf{E}^v - (\mathbf{X}^v - \mathbf{B}^v\mathbf{A}\mathbf{R}^v + \frac{\mathbf{\Lambda}^v}{\mu})\|_F^2. \tag{10}$$

Denoting $\widetilde{\mathbf{E}}^v = \mathbf{X}^v - \mathbf{B}^v\mathbf{A}\mathbf{R}^v + \mathbf{\Lambda}^v/\mu$ and $\widetilde{\mathbf{E}} = [\widetilde{\mathbf{E}}^1, \widetilde{\mathbf{E}}^2, \ldots, \widetilde{\mathbf{E}}^V]$, Eq. 10 can be rewritten as a standard $\ell_{2,1}$-norm proximal problem:

$$\min_{\mathbf{E}} \quad \lambda\|\mathbf{E}\|_{2,1} + \frac{\mu}{2}\left\|\mathbf{E} - \widetilde{\mathbf{E}}\right\|_F^2, \tag{11}$$

whose closed-form solution can be obtained by applying the $\ell_{2,1}$-norm shrinkage operator row-wise:

$$\mathbf{e}_i = \begin{cases} \left(1 - \dfrac{\lambda}{\mu\|\widetilde{\mathbf{e}}_i\|_2}\right)\widetilde{\mathbf{e}}_i, & \|\widetilde{\mathbf{e}}_i\|_2 > \dfrac{\lambda}{\mu}, \\ \mathbf{0}, & \text{otherwise}, \end{cases} \tag{12}$$

where $\mathbf{e}_i$ and $\widetilde{\mathbf{e}}_i$ denote the i-th rows of $\mathbf{E}$ and $\widetilde{\mathbf{E}}$, respectively. After updating $\mathbf{E}$, the reconstruction multipliers are updated according to the corresponding residuals:

$$\mathbf{\Lambda}^v = \mathbf{\Lambda}^v + \mu\left(\mathbf{X}^v - \mathbf{B}^v\mathbf{A}\mathbf{R}^v - \mathbf{E}^v\right), \quad v = 1, \ldots, V. \tag{13}$$

• **Updating $\mathcal{J}$ and $\mathcal{Q}$:** By fixing other variables, the subproblem w.r.t. $\mathcal{J}$ can be formulated as:

$$\min_{\mathcal{J}} \quad \Phi(\mathcal{J}) + \frac{\xi}{2}\|\mathcal{J} - (\mathcal{B} + \frac{\mathcal{Q}}{\xi})\|_F^2. \tag{14}$$

Denoting $\mathcal{T} = \mathcal{B} + \mathcal{Q}/\xi$ and $\widehat{\mathcal{T}} = \text{fft}(\mathscr{R}(\mathcal{T}), [], 3)$, Eq. (14) is solved slice-wise in the Fourier domain using the non-convex shrinkage and the updated frontal slices $\widehat{\mathcal{J}}^{(k)}$ are transformed as follows:

$$\mathcal{J} = \mathscr{R}^{-1}(\text{ifft}(\widehat{\mathcal{J}}, [], 3)). \tag{15}$$

The detailed ETR shrinkage procedure is provided in the Appendix C.1. After updating $\mathcal{J}$, the tensor-consistency multiplier is updated according to the corresponding residual:

$$\mathcal{Q} = \mathcal{Q} + \xi\left(\mathcal{B} - \mathcal{J}\right). \tag{16}$$

• **Updating $\mathbf{R}^v$:** By fixing other variables, the subproblem w.r.t. $\mathbf{R}^v$ can be formulated as follows:

$$\min_{\mathbf{R}^v\mathbf{R}^{vT}=\mathbf{I}_r} \quad \|\mathbf{X}^v - \mathbf{E}^v + \frac{\mathbf{\Lambda}^v}{\mu} - \mathbf{B}^v\mathbf{A}\mathbf{R}^v\|_F^2, \tag{17}$$

which can be equally transformed into:

$$\max_{\mathbf{R}^v\mathbf{R}^{vT}=\mathbf{I}_r} \quad \text{Tr}(\mathbf{R}^{vT}\mathbf{N}^v), \tag{18}$$

where $\mathbf{N}^v = \mathbf{A}^T\mathbf{B}^{vT}\widetilde{\mathbf{X}}^v$. By performing the compact SVD $\mathbf{N}^v = \mathbf{U}^v\mathbf{\Sigma}^v\mathbf{V}^{vT}$, the orthogonality constraint is preserved, yielding the optimal solution $\mathbf{R}^v = \mathbf{U}^v\mathbf{V}^{vT}$.

• **Updating $\mathbf{A}$:** By fixing other variables, the subproblem w.r.t. $\mathbf{A}$ can be formulated as:

$$\min_{\mathbf{A}\mathbf{A}^\top \preceq \mathbf{I}_m} \frac{1}{2}\text{Tr}(\mathbf{A}^T\mathbf{C}_A\mathbf{A}) - \text{Tr}(\mathbf{A}^T\mathbf{D}_A), \tag{19}$$

where $\mathbf{C}_A = \sum_{v=1}^{V} \mathbf{B}^{vT}\mathbf{B}^v$ and $\mathbf{D}_A = \sum_{v=1}^{V} \mathbf{B}^{vT}(\mathbf{X}^v - \mathbf{E}^v + \mathbf{\Lambda}^v/\mu)\mathbf{R}^{vT}$. Since $\mathbf{C}_A \succeq 0$ and the spectral-norm ball is convex, Eq. 19 is a convex problem and can be efficiently solved by an accelerated projected gradient method. The detailed optimization procedure is provided in Appendix C.2.

• **Updating G:** By fixing other variables, the subproblem w.r.t. $\mathbf{G}$ can be formulated as follows:

$$\min_{\mathbf{G}\mathbf{1}_m=\mathbf{1}_m,\ \mathbf{G}\geq 0,\ \mathrm{diag}(\mathbf{G})=\mathbf{0}} \quad \gamma \sum_{v=1}^{V} \mathrm{Tr}(\mathbf{B}^v \mathbf{L}_\mathbf{G} \mathbf{B}^{vT}) + \beta \|\mathbf{G}\|_F^2 + \eta\, \mathrm{Tr}(\mathbf{F}^T \mathbf{L}_\mathbf{G} \mathbf{F}), \tag{20}$$

which can be equivalently rewritten as:

$$\min_{\mathbf{g}_i \in \Delta_i} \quad \mathbf{d}_i \mathbf{g}_i^\top + \beta \|\mathbf{g}_i\|_2^2 \quad \Longleftrightarrow \quad \min_{\mathbf{g}_i \in \Delta_i} \quad \left\| \mathbf{g}_i + \frac{\mathbf{d}_i}{2\beta} \right\|_2^2, \tag{21}$$

where $\Delta_i = \{\mathbf{g}_i \mid \mathbf{g}_i \geq 0, \mathbf{g}_i \mathbf{1}_m = 1, g_{ii} = 0\}$ and $d_{ij} = \frac{\gamma}{2} \sum_{v=1}^{V} \|\mathbf{b}_{:,i}^v - \mathbf{b}_{:,j}^v\|_2^2 + \frac{\eta}{2} \|\mathbf{f}_i - \mathbf{f}_j\|_2^2$. Since Eq. 21 is separable across rows, each $\mathbf{g}_i$ is obtained by projecting $-\mathbf{d}_i/(2\beta)$ onto Δ_i.

• **Updating F:** By fixing other variables, the optimization problem w.r.t. $\mathbf{F}$ becomes:

$$\min_{\mathbf{F}^T\mathbf{F}=\mathbf{I}_c} \mathrm{Tr}(\mathbf{F}^T \mathbf{L}_\mathbf{G} \mathbf{F}), \tag{22}$$

which can be efficiently solved by eigendecomposition of $\mathbf{L}_G$, with $\mathbf{F}$ formed by the eigenvectors corresponding to its c smallest eigenvalues.

• **Adaptive update of η:** To regulate the connectivity of the learned anchor graph, η is adaptively adjusted based on the eigenvalue distribution of $\mathbf{L}_\mathbf{G}$ as follows:

$$\eta^{(t+1)} = \begin{cases} \kappa \eta^{(t)}, & \text{if } \sum_{i=1}^{c} \sigma_i > \epsilon, \\ \eta^{(t)}/\kappa, & \text{if } \sum_{i=1}^{c+1} \sigma_i < \epsilon, \\ \eta^{(t)}, & \text{otherwise,} \end{cases} \tag{23}$$

where $\kappa > 1$ denotes the adjustment factor and ϵ is a small threshold. In this way, the learned graph is progressively guided toward c connected components.

• **Updating penalty parameters:** After updating the primal variables, the penalty parameters are updated as follows:

$$\mu^{(t+1)} = \min(\tau \mu^{(t)}, \mu_{\max}), \qquad \xi^{(t+1)} = \min(\tau \xi^{(t)}, \xi_{\max}), \tag{24}$$

where $\tau > 1$ denotes the increasing factor, while $\mu_{\max}$ and $\xi_{\max}$ are upper bounds.

After convergence, the sample-anchor representations are concatenated into $\mathbf{H} = [\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_V]$, from which a low-dimensional embedding is extracted by SVD and subsequently for k-means clustering. The overall optimization procedure is provided in Appendix C.3 as Algorithm 1.

Complexity Analysis: **Time complexity:** CORE is optimized by alternately updating all variables until convergence in an iterative optimization process. In each iteration, updating $\{\mathbf{B}^v\}_{v=1}^V$ requires $O(nmd + Vnm^2 + m^3)$, while updating $\mathbf{E}$ costs $O(nd)$. Updating $\mathcal{J}$ mainly requires $O(nmV \log n + Vnm^2)$, and the updates of $\{\mathbf{R}^v\}_{v=1}^V$ and $\mathbf{A}$ cost $O(nmd + Vnm^2)$. Moreover, updating $\mathbf{G}$ requires $O(Vnm^2)$, while updating $\mathbf{F}$ costs $O(m^3)$. Therefore, the overall time complexity over T iterations is $O\big(T(nmd + Vnm^2 + nmV \log n + m^3)\big)$, which remains comparable to that of existing tensor-based clustering methods.
Space complexity. The major storage costs mainly include the multi-view data and residual-related variables with $O(nd)$ space, and the sample-anchor representations and auxiliary tensors with $O(Vnm)$ space during the optimization process. The anchor topology and its Laplacian require an additional $O(m^2)$ storage. Therefore, the overall space complexity is $O(nd + Vnm + m^2)$, which scales linearly with the number of samples in practical large-scale applications.

4 Experiments

4.1 Experimental Setup

Datasets: To evaluate the effectiveness and scalability of CORE, ten multi-view datasets with varying scales and characteristics are adopted, including COIL20, NUS, BDGP, Scene, COIL100, Hdigit,

Table 1: Clustering performance comparison, where the best and second-best available results are highlighted in **bold** and underlined, respectively. "OM" denotes out-of-memory cases.

Dataset	MCMF	HRFAL	PGAL	BGDR	DSTL	ESTMC-C2	TMC	**CORE**
					ACC (%)			
COIL20	$59.58_{\pm 1.12}$	$76.19_{\pm 4.15}$	$68.89_{\pm 0.85}$	$78.89_{\pm 0.32}$	$77.22_{\pm 0.98}$	$\underline{90.31}_{\pm 1.68}$	$73.17_{\pm 1.18}$	$\mathbf{96.01}_{\pm 0.85}$
NUS	$18.04_{\pm 1.58}$	$24.54_{\pm 0.78}$	$29.50_{\pm 1.02}$	$28.63_{\pm 0.47}$	$79.52_{\pm 1.80}$	$\mathbf{97.93}_{\pm 0.03}$	$27.26_{\pm 0.28}$	$\underline{97.20}_{\pm 0.06}$
BDGP	$30.92_{\pm 1.28}$	$50.04_{\pm 0.02}$	$48.24_{\pm 0.91}$	$35.40_{\pm 2.27}$	$\underline{97.88}_{\pm 0.00}$	$96.16_{\pm 0.00}$	$55.03_{\pm 0.03}$	$\mathbf{98.28}_{\pm 0.00}$
Scene	$31.55_{\pm 1.44}$	$42.39_{\pm 0.21}$	$43.08_{\pm 0.87}$	$34.90_{\pm 0.39}$	$67.91_{\pm 2.55}$	$\underline{68.59}_{\pm 0.17}$	$44.03_{\pm 0.08}$	$\mathbf{77.41}_{\pm 0.07}$
COIL100	$51.64_{\pm 1.65}$	$54.89_{\pm 2.04}$	$55.74_{\pm 1.11}$	$79.06_{\pm 0.52}$	$83.17_{\pm 2.51}$	$\underline{90.19}_{\pm 1.06}$	$59.93_{\pm 0.29}$	$\mathbf{92.08}_{\pm 1.01}$
Hdigit	$25.90_{\pm 1.18}$	$77.71_{\pm 0.01}$	$78.14_{\pm 0.79}$	$99.33_{\pm 0.29}$	$91.47_{\pm 7.76}$	$\mathbf{99.73}_{\pm 0.00}$	$89.84_{\pm 1.34}$	$\underline{99.54}_{\pm 0.00}$
ALOI100	$50.03_{\pm 1.35}$	$20.07_{\pm 5.67}$	$56.51_{\pm 0.97}$	$78.68_{\pm 0.43}$	$74.80_{\pm 0.91}$	$\underline{84.64}_{\pm 1.23}$	$66.85_{\pm 0.30}$	$\mathbf{87.22}_{\pm 1.38}$
Animal	OM	$18.11_{\pm 0.36}$	$16.25_{\pm 0.82}$	$19.64_{\pm 0.36}$	$79.53_{\pm 0.59}$	$\underline{90.08}_{\pm 1.49}$	$14.86_{\pm 0.10}$	$\mathbf{91.51}_{\pm 1.17}$
MNIST	OM	$33.75_{\pm 5.36}$	$43.02_{\pm 0.93}$	$49.41_{\pm 0.46}$	$63.76_{\pm 0.72}$	$\underline{86.33}_{\pm 0.00}$	$53.45_{\pm 1.29}$	$\mathbf{93.11}_{\pm 0.02}$
VoxCeleb	OM	$10.72_{\pm 2.05}$	$76.14_{\pm 1.05}$	OM	$84.04_{\pm 0.72}$	$\underline{89.60}_{\pm 1.34}$	$62.82_{\pm 0.08}$	$\mathbf{92.79}_{\pm 1.26}$
Ave. Score	-	40.84	51.55	-	79.93	<u>89.36</u>	54.72	**92.51**
Ave. Rank	7.85	6.30	5.60	4.85	3.20	<u>1.90</u>	5.10	**1.20**
					NMI (%)			
COIL20	$77.69_{\pm 1.29}$	$85.99_{\pm 2.27}$	$79.82_{\pm 0.78}$	$85.21_{\pm 0.34}$	$84.67_{\pm 0.10}$	$\underline{92.62}_{\pm 0.78}$	$82.57_{\pm 0.64}$	$\mathbf{95.01}_{\pm 0.85}$
NUS	$8.10_{\pm 1.47}$	$13.38_{\pm 0.47}$	$15.77_{\pm 0.90}$	$13.92_{\pm 0.41}$	$82.55_{\pm 0.89}$	$\mathbf{96.16}_{\pm 0.04}$	$14.02_{\pm 0.48}$	$\underline{94.69}_{\pm 0.07}$
BDGP	$8.21_{\pm 1.14}$	$31.31_{\pm 0.03}$	$26.79_{\pm 0.83}$	$9.85_{\pm 0.50}$	$\underline{93.46}_{\pm 0.00}$	$90.71_{\pm 0.00}$	$31.71_{\pm 0.04}$	$\mathbf{94.93}_{\pm 0.00}$
Scene	$23.02_{\pm 1.53}$	$30.07_{\pm 0.19}$	$31.97_{\pm 0.95}$	$20.75_{\pm 0.38}$	$\underline{72.03}_{\pm 2.53}$	$71.43_{\pm 0.10}$	$32.38_{\pm 0.06}$	$\mathbf{75.98}_{\pm 0.08}$
COIL100	$78.47_{\pm 1.37}$	$78.21_{\pm 0.82}$	$80.56_{\pm 0.86}$	$90.39_{\pm 0.45}$	$96.06_{\pm 0.63}$	$\underline{96.45}_{\pm 0.38}$	$82.75_{\pm 0.15}$	$\mathbf{97.47}_{\pm 0.21}$
Hdigit	$53.88_{\pm 1.19}$	$77.60_{\pm 0.04}$	$71.51_{\pm 0.74}$	$97.99_{\pm 0.30}$	$92.14_{\pm 2.90}$	$\mathbf{99.24}_{\pm 0.00}$	$88.61_{\pm 0.36}$	$\underline{98.79}_{\pm 0.00}$
ALOI100	$69.44_{\pm 1.42}$	$39.93_{\pm 6.77}$	$74.48_{\pm 0.92}$	$84.53_{\pm 0.40}$	$89.90_{\pm 0.03}$	$\underline{93.59}_{\pm 0.54}$	$79.87_{\pm 0.18}$	$\mathbf{94.19}_{\pm 0.51}$
Animal	OM	$14.59_{\pm 0.11}$	$13.12_{\pm 0.76}$	$13.48_{\pm 0.35}$	$84.58_{\pm 2.33}$	$\mathbf{93.10}_{\pm 0.51}$	$10.81_{\pm 0.09}$	$\underline{92.97}_{\pm 0.72}$
MNIST	OM	$21.44_{\pm 2.66}$	$27.94_{\pm 0.89}$	$36.39_{\pm 0.43}$	$61.33_{\pm 0.40}$	$\underline{77.68}_{\pm 0.00}$	$46.86_{\pm 0.84}$	$\mathbf{89.11}_{\pm 0.02}$
VoxCeleb	OM	$23.10_{\pm 2.65}$	$79.55_{\pm 0.97}$	OM	$86.82_{\pm 0.17}$	$\underline{96.91}_{\pm 0.28}$	$67.80_{\pm 0.05}$	$\mathbf{97.97}_{\pm 0.24}$
Ave. Score	-	41.56	50.15	-	84.35	<u>90.79</u>	53.74	**93.11**
Ave. Rank	7.65	6.00	5.70	5.35	3.10	<u>1.90</u>	5.00	**1.30**
					PUR (%)			
COIL20	$62.64_{\pm 1.25}$	$77.33_{\pm 3.54}$	$71.74_{\pm 0.81}$	$79.44_{\pm 0.34}$	$77.50_{\pm 1.37}$	$\underline{90.85}_{\pm 1.07}$	$75.19_{\pm 0.95}$	$\mathbf{96.01}_{\pm 0.85}$
NUS	$18.75_{\pm 1.53}$	$26.15_{\pm 0.56}$	$31.25_{\pm 0.94}$	$28.75_{\pm 0.44}$	$81.37_{\pm 2.12}$	$\mathbf{97.93}_{\pm 0.03}$	$28.53_{\pm 0.14}$	$\underline{97.20}_{\pm 0.06}$
BDGP	$36.80_{\pm 1.19}$	$54.31_{\pm 0.03}$	$48.24_{\pm 0.85}$	$35.40_{\pm 2.27}$	$\underline{97.88}_{\pm 0.00}$	$96.16_{\pm 0.00}$	$57.01_{\pm 0.04}$	$\mathbf{98.28}_{\pm 0.00}$
Scene	$35.75_{\pm 1.45}$	$43.29_{\pm 0.17}$	$46.65_{\pm 0.90}$	$35.57_{\pm 0.38}$	$\underline{71.11}_{\pm 3.29}$	$69.87_{\pm 0.05}$	$46.95_{\pm 0.08}$	$\mathbf{77.41}_{\pm 0.07}$
COIL100	$53.96_{\pm 1.33}$	$60.30_{\pm 1.48}$	$60.43_{\pm 0.88}$	$81.35_{\pm 0.47}$	$87.13_{\pm 1.72}$	$\underline{92.00}_{\pm 0.94}$	$64.37_{\pm 0.35}$	$\mathbf{93.52}_{\pm 0.66}$
Hdigit	$52.75_{\pm 1.27}$	$82.27_{\pm 0.01}$	$78.14_{\pm 0.80}$	$99.33_{\pm 0.30}$	$92.00_{\pm 7.01}$	$\mathbf{99.73}_{\pm 0.00}$	$91.62_{\pm 0.89}$	$\underline{99.54}_{\pm 0.00}$
ALOI100	$51.88_{\pm 1.40}$	$24.23_{\pm 7.08}$	$59.44_{\pm 0.93}$	$79.25_{\pm 0.42}$	$78.12_{\pm 0.88}$	$\underline{86.98}_{\pm 1.06}$	$69.25_{\pm 0.41}$	$\mathbf{88.86}_{\pm 1.10}$
Animal	OM	$21.47_{\pm 0.25}$	$19.88_{\pm 0.84}$	$20.59_{\pm 0.37}$	$82.25_{\pm 2.26}$	$\mathbf{94.08}_{\pm 0.67}$	$18.00_{\pm 0.17}$	$\underline{93.75}_{\pm 1.16}$
MNIST	OM	$38.88_{\pm 3.52}$	$43.61_{\pm 0.95}$	$50.53_{\pm 0.46}$	$64.99_{\pm 0.46}$	$\underline{86.33}_{\pm 0.00}$	$57.61_{\pm 1.02}$	$\mathbf{93.11}_{\pm 0.02}$
VoxCeleb	OM	$12.20_{\pm 2.10}$	$77.48_{\pm 1.02}$	OM	$85.52_{\pm 0.31}$	$\underline{91.96}_{\pm 0.98}$	$63.59_{\pm 0.06}$	$\mathbf{94.14}_{\pm 0.99}$
Ave. Score	-	44.04	53.69	-	81.79	<u>90.59</u>	57.21	**93.18**
Ave. Rank	7.65	6.10	5.70	5.15	3.10	<u>1.90</u>	5.10	**1.30**

ALOI100, Animal, MNIST, and VoxCeleb. The detailed statistics are summarized in Appendix D.1.

Comparative Methods: To comprehensively evaluate the proposed CORE, several representative and state-of-the-art multi-view clustering methods are selected for comparison, including ***Representation-based methods*****:** (1) Multi-View Clustering via Multi-Stage Fusion (MCMF Gan et al. (2025)); ***Graph-based methods*****:** (2) Hypergraph Regularization-based Anchor Learning (HRFAL Zeng et al. (2026)), (3) Prior-Guided Anchor Learning (PGAL Wu et al. (2024b)), and (4) Bipartite Graph Discrete Reconstruction (BGDR Ma et al. (2026)); ***Tensor-based methods*****:** (5) Fast Disentangled Slim Tensor Learning (DSTL Xu et al. (2025)), (6) Efficient and Scalable Tensorial Multi-View Subspace Clustering (ESTMC-C2 Ji & Feng (2025)), and (7) Tensor Multi-Rank Constraint Guided Anchor-Wise Adaptive Alignment (TMC Wang et al. (2026)). For a fair comparison, the parameters of all competing methods are tuned according to the recommended settings in their original papers or released implementations. For CORE, the number of anchors is selected from $\{2c, 3c, 5c, 10c\}$, the latent dimension is set as $r = \min\{c, d_1, \ldots, d_V\}$, and the parameters λ, γ, and β are selected from $\{10^{-6}, \cdots, 10^{2}\}$. The spectral parameter η is initialized to 10^{-4} and adaptively updated during optimization according to the connectivity of the learned anchor graph.

Evaluation Protocol: To mitigate the influence of random initialization, k-means is independently repeated 10 times, with the mean and standard deviation reported over all runs. Clustering performance is evaluated in terms of ACC, NMI and Purity.

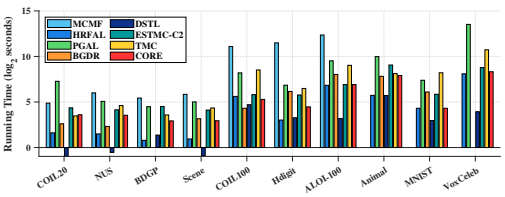


Figure 2: Running time comparison.

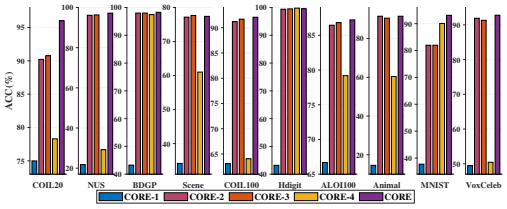
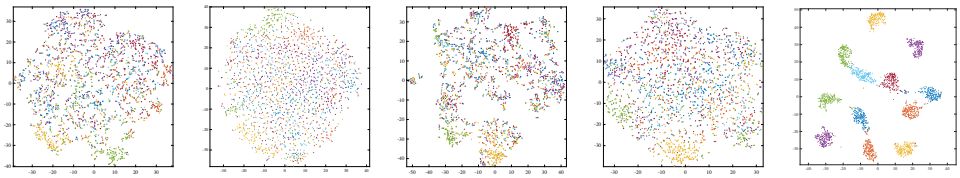


Figure 3: Ablation results of CORE.



(a) HRFAL (b) TMC (c) BGDR (d) ESTMC-C2 (e) CORE

Figure 4: t-SNE visualization of the clustering embeddings obtained by different clustering methods.

4.2 Comparison Results

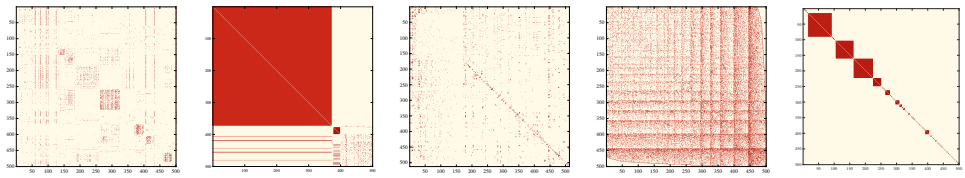
Clustering Performance: The clustering performance on ten multi-view datasets is reported in Table 1 with means and standard deviations, where the best and second-best results are highlighted in bold and underlined, respectively. Based on the results, the following observations can be drawn: (1) Across datasets with varying scales and characteristics, CORE achieves competitive performance, demonstrating its overall effectiveness and robustness. (2) Compared with representation-based and graph-based methods (i.e., HRFAL, PGAL and BGDR), CORE shows clear advantages, demonstrating the benefit of coupling representation learning with adaptive topology optimization. (3) The gains over tensor-based methods (i.e., DSTL, ESTMC and TMC) demonstrate that incorporating cluster-aware topology effectively enhances the discriminability beyond cross-view consistency.
Running Time: Fig. 2 compares the running time of different methods across all datasets. Specifically, CORE achieves competitive computational efficiency while maintaining clustering performance, highlighting the scalability advantage of topology learning in the compact anchor space.

4.3 Ablation Study

To investigate the effectiveness of different modules, four variants of CORE are considered: (1) **CORE-1** excludes tensor regularization, removing high-order consistency modeling; (2) **CORE-2** fixes the anchor topology $\mathbf{G}$ after initialization, removing iterative representation-topology coupling; (3) **CORE-3** removes the spectral connectivity constraint while retaining adaptive anchor-topology learning; (4) **CORE-4** removes anchor-topology learning entirely, such that the sample-anchor representations are learned without graph-based structural guidance. As shown in Fig. 3, the performance degradation of the four variants verifies the effectiveness of tensor regularization (capturing high-order cross-view consistency), iterative representation-topology refinement (enabling mutual structural enhancement), spectral connectivity (promoting cluster-aware anchor structures), and anchor-topology learning (providing structural guidance for refining sample-anchor representations) across diverse real-world multi-view clustering scenarios. These results further validate the representation-topology coupling of CORE for mutual refinement.

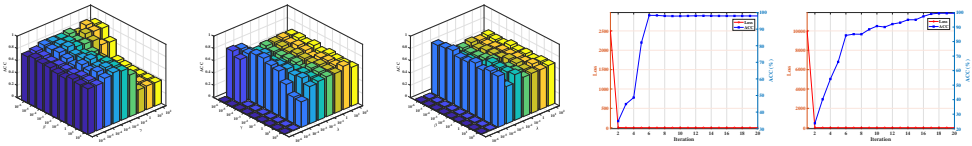
4.4 Visualization

Representation Visualization: To visualize the final clustering embeddings, t-SNE (van der Maaten & Hinton, 2008) is employed on the NUS dataset to project the high-dimensional representations into a two-dimensional space, as shown in Fig. 4. Compared with representative baselines, CORE exhibits more compact intra-cluster distributions and clearer inter-cluster separation in the projected space, indicating stronger clustering discriminability.



(a) HRFAL (b) TMC (c) BGDR (d) ESTMC-C2 (e) CORE

Figure 5: Visualization of the anchor topology structures learned by different clustering methods.



(a) β vs. γ (b) λ vs. γ (c) λ vs. β

Figure 6: Parameter sensitivity on ALIO100.

(a) BDGP (b) Hdigit

Figure 7: Convergence analysis.

Topology Visualization: To evaluate the learned anchor structure, the pairwise relationships among anchors on the ALOI100 dataset are visualized in Fig. 5. Specifically, while the baseline methods exhibit less distinct structural organization among anchors, CORE presents a clearer block-diagonal structure, indicating that anchors associated with the same cluster are more tightly organized within the learned topology and exhibit more coherent connectivity.

4.5 Parameter Sensitivity and Convergence

Parameter Sensitivity: To investigate the parameter sensitivity of CORE, one parameter is fixed while the other two are varied jointly, as shown in Fig. 6. Specifically, CORE maintains relatively stable performance over a broad range of parameter combinations, whereas the performance drops sharply when β or γ becomes excessively large. This indicates that appropriately balancing the graph regularization and representation-topology coupling is important for clustering.

Convergence Analysis: To evaluate the convergence of CORE, the objective value and ACC are tracked over iterations on BDGP and Hdigit. As shown in Fig. 7, the objective value decreases rapidly during the early iterations and gradually stabilizes, while ACC increases correspondingly and gradually reaches a stable level, demonstrating the stable convergence behavior of CORE.

5 Conclusion

In this paper, a Coupled Representation and Cluster-Aware Topology (CORE) framework is proposed for scalable multi-view clustering. Specifically, CORE jointly learns robust sample-anchor representations and adaptive anchor topology, enabling representation learning and structural refinement to mutually reinforce each other. Meanwhile, tensor regularization is introduced to capture high-order cross-view consistency, while spectral connectivity further guides the learned topology toward clustering-discriminative structures. By conducting structural learning in the compact anchor space, CORE avoids expensive sample-wise graph construction and maintains favorable computational efficiency. Extensive experiments on multi-view datasets with varying scales demonstrate the effectiveness, robustness, and scalability of CORE.

Although CORE achieves promising performance and favorable scalability, there remain several aspects for further improvement. First, the predefined anchor numbers reduce its adaptability to unknown data structures. Second, the focus on complete multi-view data limits its applicability to incomplete scenarios. Future work will investigate adaptive structure determination and extend CORE to incomplete scenarios to enhance its flexibility and practical applicability.

AI USE STATEMENT

Generative AI tools were used only to assist with language polishing and improve the grammar and readability of the manuscript; the authors take full responsibility for all content.

REFERENCES

Dimitri P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, New York, 1982. ISBN 9780120934805. doi: 10.1016/C2013-0-10366-2.

Guoqing Chao, Shiliang Sun, and Jinbo Bi. A survey on multi-view clustering. *IEEE Transactions on Artificial Intelligence*, 2(2):146–168, 2021.

Jingyi Chen, Zhibin Dong, Tiejun Li, and Yibo Han. Dc-span: A dual contrastive attention network for multi-view clustering. In *Proceedings of the Fortieth AAAI Conference on Artificial Intelligence*, volume 40, pp. 20127–20135, 2026a.

Zhe Chen, Fanhui Meng, Tianyang Xu, and Xiao-Jun Wu. Cluster-aware anchor learning for multi-view clustering. In *Proceedings of the Forty-Third IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17714–17723, June 2026b.

Guanghao Du, Peng Song, Yuanbo Cheng, Zhaowei Liu, Yanwei Yu, and Wenming Zheng. Weighted tensor-based consistent anchor graph learning for multi-view clustering. *Neurocomputing*, 620:129253, 2025.

Yu Gan, Yunning You, Junjie Huang, Sen Xiang, Chang Tang, Wei Hu, and Shan An. Multi-view clustering via multi-stage fusion. *IEEE Transactions on Multimedia*, 27:4571–4583, 2025.

Zhibin Gu, Zhendong Li, and Songhe Feng. Topology-driven multi-view clustering via tensorial refined sigmoid rank minimization. In *Proceedings of the Thirtieth ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 920–931, 2024.

Jipeng Guo, Xiang Xu, Yu Cao, Man Cao, Mengyuan Xin, Tianxiang Zhao, Ye Su, Junbin Gao, Mingliang Cui, and Youqing Wang. Anchor-to-graph structural co-regularization for scalable multi-view clustering. *Pattern Recognition*, 177, 2026.

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*, pp. 5336–5344, 2025.

Jintian Ji and Songhe Feng. Anchors crash tensor: Efficient and scalable tensorial multi-view subspace clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(4):2660–2675, 2025.

Jiyuan Liu, Xinwang Liu, Yuexiang Yang, Li Liu, Siqi Wang, Weixuan Liang, and Jiangyong Shi. One-pass multi-view clustering for large-scale data. In *Proceedings of the Eighteenth IEEE/CVF International Conference on Computer Vision*, pp. 12324–12333, 2021.

Jiyuan Liu, Xinwang Liu, Chuankun Li, Xinhang Wan, Hao Tan, Yi Zhang, Weixuan Liang, Qian Qu, Yu Feng, Renxiang Guan, and Ke Liang. Large-scale multi-view tensor clustering with implicit linear kernels. In *Proceedings of the Forty-Second IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20727–20736, 2025.

Xinwang Liu, Ke Liang, Jun Wang, Suyuan Liu, Xiangke Wang, and Huaimin Wang. Two decades of multi-view clustering: Taxonomy, application, and challenge. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(3):3744–3764, 2026.

Zhoumin Lu, Yongbo Yu, Linru Ma, Feiping Nie, and Rong Wang. Capturing individuality and commonality between anchor graphs for multi-view clustering. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 5860–5868, 2025.

Zhenyu Ma, Shengzhao Guo, Jingyu Wang, Feiping Nie, and Xuelong Li. Scalable graph discrete reconstruction for efficient multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 38(7):4641–4657, 2026.

Bojan Mohar. The laplacian spectrum of graphs. In Y. Alavi, G. Chartrand, O. R. Oellermann, and A. J. Schwenk (eds.), *Graph Theory, Combinatorics, and Applications*, volume 2, pp. 871–898. Wiley, 1991.

Yinghui Sun, Xingfeng Li, Quansen Sun, Min-Ling Zhang, and Zhenwen Ren. Improved weighted tensor schatten p-norm for fast multi-view graph clustering. In *Proceedings of the Thirty-Second ACM International Conference on Multimedia*, pp. 1427–1436, 2024.

Chuan Tang, Minhui Wang, and Kun Sun. One-step multiview clustering via adaptive graph learning and spectral rotation. *IEEE Transactions on Neural Networks and Learning Systems*, 36(3):5442–5453, 2025.

Chuan Tang, Miaomiao Li, Jun Wang, Renxiang Guan, Siwei Wang, Chang Tang, En Zhu, and Xinwang Liu. Align then tensorize: Multi-level consistent anchor graph learning for scalable multi-view clustering. *IEEE Transactions on Image Processing*, 35:7049–7062, 2026.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008.

Xinhang Wan, Xinwang Liu, Jiyuan Liu, Siwei Wang, Yi Wen, Weixuan Liang, En Zhu, Zhe Liu, and Lu Zhou. Auto-weighted multi-view clustering for large-scale data. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence*, volume 37, pp. 10078–10086, 2023.

Jun Wang, Miaomiao Li, Zhenglai Li, Hao Yu, Suyuan Liu, Dayu Hu, Chang Tang, and Xinwang Liu. Tensor multi-rank constraint guided anchor-wise adaptive alignment for multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 38(3):2015–2027, 2026.

Xiaohui Wei, Yuting Chen, Feiping Nie, Haibo Liu, Qiya Song, and Lin Xiao. View-adaptive multi-granularity anchor learning for multi-view clustering. *IEEE Transactions on Image Processing*, 35:3113–3126, 2026.

Song Wu, Yan Zheng, Yazhou Ren, Jing He, Xiaorong Pu, Shudong Huang, Zhifeng Hao, and Lifang He. Self-weighted contrastive fusion for deep multi-view clustering. *IEEE Transactions on Multimedia*, 26:9150–9162, 2024a.

Xi Wu, Hanchen Wang, Shuxiao Li, Jian Dai, and Zhenwen Ren. Prior indicator guided anchor learning for multi-view subspace clustering. *IEEE Transactions on Consumer Electronics*, 70(1): 144–154, 2024b.

Deng Xu, Chao Zhang, Zechao Li, Chunlin Chen, and Huaxiong Li. Fast disentangled slim tensor learning for multi-view clustering. *IEEE Transactions on Multimedia*, 27:1254–1265, 2025.

Ben Yang, Jinghan Wu, Xuetao Zhang, Xinhu Zheng, Feiping Nie, and Badong Chen. Discrete correntropy-based multi-view anchor-graph clustering. *Information Fusion*, 103:102097, 2024.

Jie Yang, Wei Chen, Feng Liu, Peng Zhou, Zhongli Wang, Xinyan Liang, and Bingbing Jiang. Multi-view clustering via multi-granularity ensemble. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, 2025.

Yunpeng Zeng, Peng Song, Beihua Yang, Changjia Wang, Guanghao Du, Yanwei Yu, and Wenming Zheng. Hypergraph regularization-based anchor learning for multi-view clustering. *Pattern Recognition*, 172:112465, 2026.

Chao Zhang, Haoxing Chen, Huaxiong Li, and Chunlin Chen. Learning latent disentangled embeddings and graphs for multi-view clustering. *Pattern Recognition*, 156:110839, 2024a.

Chao Zhang, Xiuyi Jia, Zechao Li, Chunlin Chen, and Huaxiong Li. Learning cluster-wise anchors for multi-view clustering. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence*, volume 38, pp. 16696–16704, 2024b.

Yunhe Zhang, Jinyu Cai, Zhihao Wu, Pengyang Wang, and See-Kiong Ng. Mixture of experts as representation learner for deep multi-view clustering. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence*, volume 39, pp. 22704–22713, 2025.

A NOTATION SUMMARY

For clarity, the main notations used throughout this paper are summarized in Table 2.

Table 2: Description of key notations.

Notation	Description	Notation	Description
n	Number of samples	V	Number of views
c	Number of clusters	m	Number of anchors
r	Dimension of the shared latent space	d_v	Feature dimension of the v-th view
$\mathbf{X}^v \in \mathbb{R}^{n\times d_v}$	Data matrix of the v-th view	$\mathbf{B}^v \in \mathbb{R}^{n\times m}$	v-th view sample-anchor representation
$\mathbf{A} \in \mathbb{R}^{m\times r}$	Shared anchor basis	$\mathbf{R}^v \in \mathbb{R}^{r\times d_v}$	View-specific projection
$\mathbf{E}^v \in \mathbb{R}^{n\times d_v}$	Reconstruction residual	$\mathcal{B} \in \mathbb{R}^{n\times m\times V}$	Sample-anchor tensor
$\mathcal{J} \in \mathbb{R}^{n\times m\times V}$	Auxiliary tensor	$\mathbf{G} \in \mathbb{R}^{m\times m}$	Anchor topology matrix
$\mathbf{L}_G \in \mathbb{R}^{m\times m}$	Graph Laplacian	$\mathbf{D}_G \in \mathbb{R}^{m\times m}$	Diagonal degree matrix
$\mathbf{F} \in \mathbb{R}^{m\times c}$	Spectral embedding of the anchor topology	$\mathbf{\Lambda}^v \in \mathbb{R}^{n\times d_v}$	Reconstruction multiplier
$\mathcal{Q} \in \mathbb{R}^{n\times m\times V}$	Tensor-consistency multiplier	$\mathbf{H} \in \mathbb{R}^{n\times mV}$	Final clustering representation

B GRAPH CONNECTIVITY PROPERTY

Theorem 1 *The multiplicity of the zero eigenvalue of the graph Laplacian matrix $\mathbf{L}_S$ is equal to the number of connected components in the graph $\mathbf{S}$ Mohar (1991).*

According to Theorem 1, a graph with c connected components satisfies $\mathrm{rank}(\mathbf{L}_S) = m - c$. Therefore, the connectivity structure can be characterized through the eigenvalue distribution of the graph Laplacian, where the number of zero eigenvalues directly reflects the corresponding number of connected components in the learned graph structure. This property provides the theoretical basis for the spectral connectivity regularization adopted in CORE during topology learning.

C DETAILED OPTIMIZATION

C.1 OPTIMIZATION OF $\mathcal{J}$

By fixing the remaining variables, the subproblem of $\mathcal{J}$ is given in Eq. (14) as follows:

$$\min_{\mathcal{J}} \quad \Phi(\mathcal{J}) + \frac{\xi}{2}\|\mathcal{J} - \mathcal{T}\|_F^2, \qquad \mathcal{T} = \mathcal{B} + \frac{\mathcal{Q}}{\xi}.$$

Accordingly, $\mathcal{T} = \mathcal{B} + \frac{\mathcal{Q}}{\xi}$ and $\mathcal{J}$ are first rearranged and transformed into the Fourier domain, where $\widehat{\mathcal{T}} = \mathrm{fft}(\mathscr{R}(\mathcal{T}), [], 3)$ and $\widehat{\mathcal{J}} = \mathrm{fft}(\mathscr{R}(\mathcal{J}), [], 3)$, respectively. Then, the ETR regularization term can be expressed in the Fourier domain as:

$$\Phi(\mathcal{J}) = \frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{q}\phi_\delta\left(\sigma_j(\widehat{\mathcal{J}}^{(k)})\right), \quad \phi_\delta(s) = \frac{e^{\delta^2}s}{s+\delta},$$

where $q = \min(m, V)$. For each frontal slice, let $\widehat{\mathcal{T}}^{(k)} = \mathbf{U}_k\,\mathrm{diag}(\boldsymbol{\sigma}_k)\mathbf{V}_k^H$. The corresponding singular-value subproblem is solved using an iterative reweighted shrinkage scheme. Starting from $d_{k,j}^{(0)} = \sigma_{k,j}$, the weights and singular values are iteratively updated as follows:

$$w_{k,j}^{(t)} = \frac{\delta e^{\delta^2}}{(d_{k,j}^{(t)} + \delta)^2}, \quad d_{k,j}^{(t+1)} = \max\{\sigma_{k,j} - \frac{w_{k,j}^{(t)}}{\xi}, 0\}.$$

The corresponding frontal slice is reconstructed as $\widehat{\mathcal{J}}^{(k)} = \mathbf{U}_k\,\mathrm{diag}(\mathbf{d}_k)\mathbf{V}_k^H$. The inner iteration terminates when the change in the singular values satisfies $\sum_{j=1}^{q}(d_{k,j}^{(t+1)} - d_{k,j}^{(t)})^2 < 10^{-4}$, or when the maximum number of 100 iterations is reached. Finally, $\mathcal{J}$ is recovered by:

$$\mathcal{J} = \mathscr{R}^{-1}(\mathrm{ifft}(\widehat{\mathcal{J}}, [], 3)).$$

C.2 Optimization of $\mathbf{A}$

By fixing the remaining variables, the subproblem of $\mathbf{A}$ in Eq. (19) is given as follows:

$$\min_{\mathbf{A}\mathbf{A}^\top \preceq \mathbf{I}_m} \frac{1}{2}\operatorname{Tr}(\mathbf{A}^T\mathbf{C}_A\mathbf{A}) - \operatorname{Tr}(\mathbf{A}^T\mathbf{D}_A),$$

where $\mathbf{C}_A = \sum_{v=1}^{V}(\mathbf{B}^v)^\top\mathbf{B}^v$ and $\mathbf{D}_A = \sum_{v=1}^{V}(\mathbf{B}^v)^\top(\mathbf{X}^v - \mathbf{E}^v + \mathbf{\Lambda}^v/\mu)(\mathbf{R}^v)^\top$. Since $\mathbf{C}_A \succeq 0$ and $\mathbf{A}\mathbf{A}^\top \preceq \mathbf{I}_m$ (i.e., $\|\mathbf{A}\|_2 \leq 1$), this subproblem is convex, with the gradient given by $\nabla f(\mathbf{A}) = \mathbf{C}_A\mathbf{A} - \mathbf{D}_A$. Considering that the gradient is Lipschitz continuous with constant $L_A = \lambda_{\max}(\mathbf{C}_A)$, an accelerated projected gradient method is employed to solve this problem. For a matrix $\mathbf{M} = \mathbf{U}\operatorname{diag}(\boldsymbol{\sigma})\mathbf{V}^\top$, its projection onto the feasible spectral-norm ball is given by

$$\Pi(\mathbf{M}) = \mathbf{U}\operatorname{diag}(\min(\boldsymbol{\sigma}, 1))\,\mathbf{V}^\top,$$

where the minimum is applied elementwise. In the experiments, we set $q_0 = 1$ and $\mathbf{Y}^{(0)} = \mathbf{A}^{(0)}$. When $\mathbf{C}_A = \mathbf{0}$ and $\mathbf{D}_A = \mathbf{0}$, the objective is constant over the feasible set, and we select $\mathbf{A} = [\mathbf{I}_r; \mathbf{0}]$ as a canonical feasible solution. Otherwise, the following accelerated projected gradient procedure is applied as follows:

$$\mathbf{A}^{(t+1)} = \Pi\left(\mathbf{Y}^{(t)} - \frac{\mathbf{C}_A\mathbf{Y}^{(t)} - \mathbf{D}_A}{L_A}\right), q_{t+1} = \frac{1+\sqrt{1+4q_t^2}}{2},$$
$$\mathbf{Y}^{(t+1)} = \mathbf{A}^{(t+1)} + \frac{q_t - 1}{q_{t+1}}\left(\mathbf{A}^{(t+1)} - \mathbf{A}^{(t)}\right).$$

The convergence of the inner solver is evaluated using the Frank–Wolfe gap

$$g_A(\mathbf{A}) = \langle \nabla f(\mathbf{A}), \mathbf{A}\rangle + \|\nabla f(\mathbf{A})\|_*,$$

where $\|\cdot\|_*$ denotes the nuclear norm. The Frank–Wolfe gap measures the optimality of the current solution and approaches zero as $\mathbf{A}$ converges to the optimum. The inner iteration terminates when the normalized Frank–Wolfe gap is below 10^{-4} or when the maximum of 200 iterations is reached.

Algorithm 1 Optimization Procedures for CORE

Input: Multi-view data $\{\mathbf{X}^v\}_{v=1}^V$, cluster number c, anchor number m, latent dimension r, and parameters λ, γ, and β;

Initialize $\{\mathbf{B}^v, \mathbf{R}^v, \mathbf{E}^v, \mathbf{\Lambda}^v\}_{v=1}^V$, $\mathbf{A}$, $\mathcal{J}$, and $\mathcal{Q}$ as zero matrices/tensors; set $\mathbf{G} = (\mathbf{1}_m\mathbf{1}_m^\top - \mathbf{I}_m)/(m-1)$, initialize $\mathbf{F}$ using the c eigenvectors associated with the smallest eigenvalues of $\mathbf{L}_\mathbf{G}$, and initialize the penalty parameters $\mu = \xi = \eta = 10^{-4}$;
repeat
  Update $\mathbf{B}^v$ by Eq. (9) and Construct $\mathcal{B} = \operatorname{cat}(\mathbf{B}^1, \ldots, \mathbf{B}^V)$;
  Update $\mathbf{E}$ by Eq. (12) and Update $\mathbf{\Lambda}^v$ by Eq. (13);
  Update $\mathcal{J}$ by Eq. (15) and Update $\mathcal{Q}$ by Eq. (16);
  Update $\mathbf{R}^v$ by Eq. (18), $v = 1, \ldots, V$;
  Update $\mathbf{A}$ by solving Eq. (19);
  Update $\mathbf{G}$ by Eq. (21);
  Update $\mathbf{F}$ by Eq. (22);
  Update η according to Eq. (23);
  Update μ and ξ according to Eq. (24);
until convergence
Construct the final representation $\mathbf{H} = [\mathbf{B}^1, \mathbf{B}^2, \ldots, \mathbf{B}^V]$;

Output: Obtain the clustering labels by applying SVD and k-means to $\mathbf{H}$.

C.3 Overall Optimization Procedure

The overall optimization procedure of CORE is summarized in Algorithm 1. All primal and dual variables are initialized to zero unless otherwise specified. During the first iteration, the SVD-based update provides feasible row-orthogonal factors $\mathbf{R}^v$. If the $\mathbf{A}$-subproblem becomes constant, i.e., $\mathbf{C}_A = \mathbf{0}$ and $\mathbf{D}_A = \mathbf{0}$, a canonical feasible basis $\mathbf{A} = [\mathbf{I}_r; \mathbf{0}]$ is adopted.

D ADDITIONAL EXPERIMENTS

D.1 DATASET DETAILS

To provide a comprehensive evaluation, ten widely used multi-view benchmark datasets with different scales and feature dimensions are adopted in the experiments. The detailed statistics of these datasets are summarized in Table 3.

Table 3: Statistics of the multi-view datasets.

Dataset	Classes	Samples	Features
COIL20	20	1440	2801 (512/420/1239/630)
NUS	12	2400	1134 (64/144/73/128/225/500)
BDGP	5	2500	1750 (1000/500/250)
Scene	8	2688	1248 (512/432/256/48)
COIL100	100	7200	90 (30/30/30)
Hdigit	10	10000	1040 (784/256)
ALOI100	100	10800	279 (77/13/64/125)
Animal	20	11673	8690 (2689/2000/2001/2000)
MNIST	10	20000	459 (256/144/59)
VoxCeleb	100	23000	1458 (192/512/160/80/514)

D.2 ADDITIONAL CLUSTERING RESULTS

Table 4: Clustering performance comparison on ARI (%), where the best and second-best available results are highlighted in **bold** and underlined, respectively. OM denotes out-of-memory cases.

Dataset	MCMF	HRFAL	PGAL	BGDR	DSTL	ESTMC-C2	TMC	**CORE**
COIL20	$53.76_{\pm 4.12}$	$72.58_{\pm 4.64}$	$63.23_{\pm 2.85}$	$72.88_{\pm 1.63}$	$63.20_{\pm 2.46}$	$\underline{86.59}_{\pm 1.98}$	$66.24_{\pm 1.24}$	$\mathbf{92.15}_{\pm 1.51}$
NUS	$2.67_{\pm 3.47}$	$7.23_{\pm 0.23}$	$8.84_{\pm 2.31}$	$8.55_{\pm 1.52}$	$70.06_{\pm 3.19}$	$\mathbf{95.55}_{\pm 0.06}$	$8.04_{\pm 1.08}$	$\underline{94.04}_{\pm 0.13}$
BDGP	$7.07_{\pm 3.86}$	$25.78_{\pm 0.03}$	$18.80_{\pm 2.47}$	$6.11_{\pm 0.00}$	$\underline{94.84}_{\pm 0.00}$	$90.89_{\pm 0.00}$	$29.30_{\pm 1.35}$	$\mathbf{95.74}_{\pm 0.00}$
Scene	$12.15_{\pm 3.24}$	$20.68_{\pm 0.17}$	$21.79_{\pm 2.18}$	$14.16_{\pm 1.47}$	$\underline{61.30}_{\pm 4.45}$	$56.17_{\pm 0.13}$	$23.08_{\pm 1.16}$	$\mathbf{62.59}_{\pm 0.13}$
COIL100	$44.02_{\pm 4.35}$	$45.53_{\pm 1.55}$	$44.11_{\pm 2.72}$	$71.87_{\pm 1.58}$	$81.23_{\pm 2.64}$	$\underline{88.75}_{\pm 1.36}$	$53.82_{\pm 1.29}$	$\mathbf{91.31}_{\pm 0.96}$
Hdigit	$36.79_{\pm 3.91}$	$71.74_{\pm 0.02}$	$66.40_{\pm 2.56}$	$98.52_{\pm 1.41}$	$88.51_{\pm 5.10}$	$\mathbf{99.40}_{\pm 0.00}$	$85.39_{\pm 1.22}$	$\underline{98.98}_{\pm 0.00}$
ALOI100	$37.27_{\pm 4.08}$	$9.59_{\pm 0.47}$	$44.02_{\pm 2.63}$	$58.92_{\pm 1.55}$	$69.27_{\pm 2.25}$	$\underline{81.47}_{\pm 1.30}$	$42.45_{\pm 1.19}$	$\mathbf{83.52}_{\pm 1.20}$
Animal	OM	$5.47_{\pm 0.12}$	$5.36_{\pm 2.24}$	$6.42_{\pm 1.38}$	$70.56_{\pm 6.56}$	$\underline{88.10}_{\pm 1.43}$	$3.83_{\pm 1.05}$	$\mathbf{88.46}_{\pm 1.87}$
MNIST	OM	$13.21_{\pm 0.38}$	$19.84_{\pm 2.49}$	$24.60_{\pm 1.61}$	$52.34_{\pm 3.83}$	$\underline{69.34}_{\pm 0.00}$	$38.64_{\pm 1.31}$	$\mathbf{85.74}_{\pm 0.04}$
VoxCeleb	OM	$3.37_{\pm 0.29}$	$65.42_{\pm 2.37}$	OM	$82.35_{\pm 2.40}$	$\underline{89.01}_{\pm 1.01}$	$51.43_{\pm 1.27}$	$\mathbf{92.59}_{\pm 1.08}$
Ave. Score	-	27.52	35.78	-	73.37	<u>84.53</u>	40.22	**88.51**
Ave. Rank	7.75	6.00	5.60	5.05	3.30	<u>2.00</u>	5.10	**1.20**

To provide a more comprehensive evaluation, Table 4 reports the clustering performance in terms of ARI on all datasets. According to Table 1 and Table 4, CORE consistently achieves superior or highly competitive results across all four evaluation metrics and diverse datasets. Specifically, CORE obtains average scores of 92.51%, 93.11%, 93.18%, and 88.51% on ACC, NMI, PUR, and ARI, respectively, outperforming the second-best method by 3.15, 2.32, 2.59, and 3.98 percentage points. Moreover, CORE achieves the best average ranks of 1.20, 1.30, 1.30, and 1.20 across the four metrics, further demonstrating its stable clustering performance across different datasets.