DISCRETE CO-EVOLUTION OF MULTI-RESOLUTION PARTITION HYPERGRAPHS FOR MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Multi-view clustering (MVC) aims to exploit complementary information across different views to uncover a reliable consensus partition. However, existing methods often treat view-specific structures as fixed inputs to consensus learning, resulting in one-way information flow in which local structures guide the consensus but cannot be refined by it. Moreover, intrinsically categorical assignments are frequently optimized through continuous relaxation followed by post-hoc discretization, creating a mismatch between optimization and the final solution. To address these limitations, we propose **DCoE**, a discrete co-evolution framework for multi-view clustering. DCoE treats resolution-specific partitions as learnable discrete latent structures and jointly refines them with a common discrete consensus. These partitions induce multi-resolution hypergraphs that capture complementary group-wise structures, with adaptive weighting integrating their contributions. All partition variables are optimized directly in the discrete domain through exact reassignment gains, avoiding relaxation and post-hoc discretization while guaranteeing monotonic objective improvement. We further analyze a faithful row-simplex relaxation and show that, under a verifiable structural–spectral condition, every local maximizer is one-hot. A resolution–balance guarantee further connects this categorical geometry to the learned multi-resolution structures. Sparse anchor relations and group-level sufficient statistics enable matrix-free optimization with linear storage in the number of samples. Extensive experiments demonstrate the effectiveness and efficiency of DCoE.

1 INTRODUCTION

Different views of the same data often reveal complementary structures at different levels of granularity (Liu et al., 2025; Yang et al., 2025). Multi-view clustering (MVC) seeks to reconcile these heterogeneous structures into a unified partition while preserving informative view-specific relationships (Zhao et al., 2026; Zhou et al., 2026; Wang et al., 2025d). Graph-based methods explicitly encode structural relations among samples (Guo et al., 2026), while anchor-based representations improve scalability through sparse sample-to-anchor connections (Hu et al., 2026; He et al., 2025). However, computational compression alone does not resolve the structural heterogeneity across views.

A single structural resolution is often insufficient for complex data. Coarse partitions capture broad grouping tendencies, whereas finer partitions reveal local substructures (Meng et al., 2023; Wei et al., 2026). Multi-resolution modeling can therefore provide complementary structural evidence across granularities (Ju et al., 2025; Wang et al., 2025c; Yang et al., 2026; Su et al., 2025). Yet existing multi-resolution structures are commonly constructed before or separately from consensus learning and then treated as fixed inputs. Once fixed, they induce a one-way information flow in which view-specific structures guide consensus formation, while the refined cross-view consensus cannot feed back to correct structures that are noisy, initialization-dependent, or no longer compatible with the emerging global partition. This suggests that the view-specific structures themselves should be treated as learnable variables rather than static representations.

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.

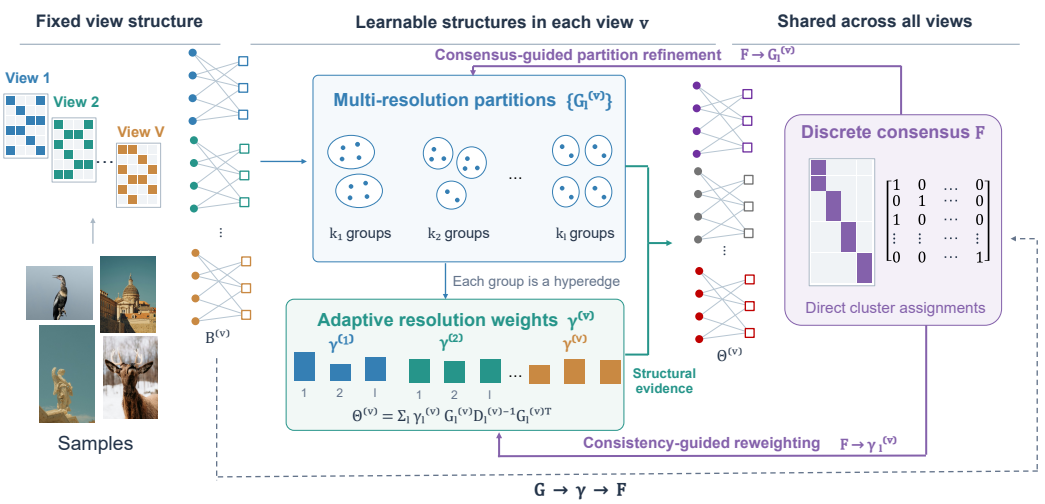


Figure 1: Overview of the DCoE framework.

A second issue concerns the categorical nature of clustering. Although the final output is a discrete partition, many clustering formulations optimize relaxed continuous assignments and recover labels afterward through rounding, argmax, or an additional clustering step. This creates a mismatch between the variables optimized during learning and the categorical solution ultimately returned. More fundamentally, it leaves open whether continuous solutions are actually favored by the objective or whether the objective geometry itself may already prefer categorical extreme points. These observations motivate a formulation in which discrete structures and consensus assignments are optimized directly, while the relationship between discreteness and the underlying objective geometry is explicitly characterized.

Motivated by these issues, we develop **DCoE**, a discrete co-evolution framework for multi-view clustering. Instead of treating multi-resolution structures as preconstructed inputs, DCoE models the resolution-specific partitions themselves as learnable discrete latent structures. Their groups induce partition hypergraphs at different resolutions, and the resulting view-specific structures are jointly refined with a common discrete consensus throughout optimization. The consensus therefore not only receives structural evidence from individual views but also feeds back to revise their resolution-specific partitions, forming a bidirectional co-evolution process. Adaptive resolution weighting integrates complementary structural evidence across granularities, while sparse anchor-induced affinities preserve view-specific neighborhood information.

All partition variables in DCoE remain discrete throughout optimization. We derive exact objective gains for single-sample reassignments and directly optimize the original discrete formulation without continuous relaxation or post-hoc discretization. Beyond this algorithmic design, we analyze a faithful row-simplex relaxation of the consensus block and establish conditions under which its local maximizers are necessarily one-hot. A structural–spectral certificate makes this property verifiable, while a resolution–balance guarantee further connects the categorical geometry to the resolution and balance of the learned partitions. By exploiting sparse anchor relations and group-level sufficient statistics, DCoE avoids materializing sample-wise affinity and hypergraph propagation matrices, resulting in linear storage with respect to the number of samples. The main contributions of this work are summarized as follows:

- We introduce learnable discrete multi-resolution partitions that induce partition hypergraphs and co-evolve bidirectionally with the cross-view consensus, rather than remaining fixed during consensus learning.
- We develop a fully discrete optimization scheme with exact reassignment gains for both the resolution-specific partitions and the consensus, avoiding continuous relaxation and guaranteeing monotonic objective improvement.

- We analyze a faithful row-simplex relaxation and derive a verifiable structural–spectral certificate and a resolution–balance guarantee for one-hot local maximizers. A matrix-free implementation further provides linear storage in the number of samples.

2 Related Work

Anchor-based multi-view clustering has received considerable attention due to its ability to model structural relationships with substantially lower computational cost than sample-wise graph construction (Peng et al., 2025; Chen et al., 2025c). Earlier methods mainly construct bipartite graphs between samples and a compact set of anchors and integrate the resulting view-specific structures for consensus learning (Li et al., 2015; Li & He, 2022). Li et al. (2022) adaptively fused view-specific sample–anchor bipartite graphs into a structured joint graph whose connected components directly indicate the consensus clusters. Beyond a single anchor configuration, recent studies have explored multiple anchor graphs, heterogeneous anchor configurations, and structural information at different scales (Zhang et al., 2023; Yu et al., 2024; Wang et al., 2024; Yang et al., 2023). Yu et al. (2025) further introduced a group of hybrid prototype quantities for each view to generate multi-scale view-specific graphs and more comprehensively characterize the underlying similarity structure.

Partition-based representations provide another clustering-oriented way to exploit structural information. Cluster ensemble methods combine multiple base partitions to obtain a more reliable consensus, with representative approaches operating directly in the partition space or aligning different clustering assignments before fusion (Strehl & Ghosh, 2002; Topchy et al., 2005). In multi-view clustering, partition-based methods further exploit view-specific clustering structures and integrate them to derive a common partition (Wan et al., 2022; Qin et al., 2026; Wang et al., 2025a). Closely related to partition representations, hypergraphs have been extensively used to characterize high-order relationships among groups of samples (Li et al., 2026; Feng et al., 2025). Hypergraph clustering methods typically encode vertex–hyperedge memberships through incidence matrices and perform learning based on normalized hypergraph operators (Wang et al., 2025b; Gao et al., 2022). Hypergraph modeling has also been introduced into multi-view clustering to exploit complementary high-order structures across views (Chen et al., 2025a; Gu et al., 2025).

3 METHODOLOGY

3.1 Multi-Resolution Partition Hypergraph Co-Evolution

Given a multi-view dataset $\{\mathbf{X}^{(v)}\}_{v=1}^{V}$ with n samples and c clusters, we aim to directly learn a discrete consensus partition $\mathbf{F} \in \{0,1\}^{n\times c}$ with $\mathbf{F}\mathbf{1}_c = \mathbf{1}_n$. For each view, we first construct a sparse row-stochastic sample-to-anchor relation $\mathbf{Z}^{(v)} \in \mathbb{R}^{n\times m_v}$ using adaptive-neighbor weighting (Nie et al., 2016), such that $\mathbf{Z}^{(v)}\mathbf{1}_{m_v} = \mathbf{1}_n$. Define

$$\mathbf{\Lambda}^{(v)} = \mathrm{Diag}\Big(\mathbf{Z}^{(v)\top}\mathbf{1}_n\Big), \qquad \mathbf{B}^{(v)} = \mathbf{Z}^{(v)}\mathbf{\Lambda}^{(v)-1/2}. \tag{1}$$

The induced sample-wise affinity is conceptually written as $\mathbf{K}^{(v)} = \mathbf{B}^{(v)}\mathbf{B}^{(v)\top} = \mathbf{Z}^{(v)}\mathbf{\Lambda}^{(v)-1}\mathbf{Z}^{(v)\top}$. Hence $\mathbf{K}^{(v)}$ is symmetric positive semidefinite and satisfies $\mathbf{K}^{(v)}\mathbf{1}_n = \mathbf{1}_n$. The matrix $\mathbf{K}^{(v)}$ is used only for formulation and is never explicitly constructed in implementation. To characterize structural information at different granularities, we associate the v-th view with L discrete partitions $\{\mathbf{G}_l^{(v)}\}_{l=1}^{L}$, where $\mathbf{G}_l^{(v)} \in \{0,1\}^{n\times k_l}$, $\mathbf{G}_l^{(v)}\mathbf{1}_{k_l} = \mathbf{1}_n$, and each partition group is nonempty. By regarding each partition group as a hyperedge, $\mathbf{G}_l^{(v)}$ naturally serves as the incidence matrix of a partition-induced hypergraph. Let $\mathbf{D}_l^{(v)} = \mathrm{Diag}(\mathbf{G}_l^{(v)\top}\mathbf{1}_n)$ denote the hyperedge-cardinality matrix. The corresponding normalized partition projection is defined as $\mathbf{P}_l^{(v)} = \mathbf{G}_l^{(v)}\mathbf{D}_l^{(v)-1}\mathbf{G}_l^{(v)\top}$, which captures group-wise structural relations while normalizing the contribution of each group by its cardinality. Its consistency with the consensus partition is measured by $T_l^{(v)} = \mathrm{Tr}(\mathbf{F}^\top\mathbf{P}_l^{(v)}\mathbf{F})$, which favors consistent consensus assignments within each partition group.

Meanwhile, the resolution-specific partitions should remain faithful to the intrinsic structure of their own views rather than being determined solely by the current consensus. We therefore define $\bar{\mathbf{K}}^{(v)} =$

$\mathbf{K}^{(v)} - \mathrm{Diag}(\mathrm{diag}(\mathbf{K}^{(v)}))$ and $\bar{\mathbf{E}} = \mathbf{1}_n\mathbf{1}_n^\top - \mathbf{I}_n$, and evaluate each partition by

$$R_l^{(v)} = \frac{\mathrm{Tr}\left(\mathbf{G}_l^{(v)\top}\bar{\mathbf{K}}^{(v)}\mathbf{G}_l^{(v)}\right)}{\mathrm{Tr}\left(\mathbf{G}_l^{(v)\top}\bar{\mathbf{E}}\mathbf{G}_l^{(v)}\right)}. \tag{2}$$

The denominator equals $\sum_{q=1}^{k_l} n_q(n_q-1)$ for group sizes n_q, thereby normalizing the accumulated within-group affinity by the number of within-group sample pairs.

The partition groups across resolutions form a weighted multi-resolution hypergraph. Assigning all hyperedges at resolution l the shared weight $\gamma_l^{(v)}$, with $\boldsymbol{\gamma}^{(v)} \geq \mathbf{0}$ and $\mathbf{1}^\top\boldsymbol{\gamma}^{(v)} = 1$, yields the normalized propagation operator $\boldsymbol{\Theta}^{(v)} = \sum_{l=1}^{L}\gamma_l^{(v)}\mathbf{P}_l^{(v)}$. Its equivalence to the standard normalized hypergraph operator is established in Lemma 2 of Appendix A.2.

Combining the multi-resolution high-order structure with the anchor-induced similarity, the proposed model is formulated as

$$\begin{aligned}\max_{\mathbf{F},\{\mathbf{G}_l^{(v)}\},\{\boldsymbol{\gamma}^{(v)}\}} \quad & \frac{\sum_{v=1}^{V}\mathrm{Tr}\big(\mathbf{F}^\top[\alpha\boldsymbol{\Theta}^{(v)} + (1-\alpha)\mathbf{K}^{(v)}]\mathbf{F}\big)}{\mathrm{Tr}(\mathbf{F}^\top\mathbf{E}\mathbf{F})} + \eta\sum_{v=1}^{V}\sum_{l=1}^{L}R_l^{(v)} - \lambda_\gamma\sum_{v=1}^{V}\|\boldsymbol{\gamma}^{(v)}\|_2^2, \\ \text{s.t.} \quad & \mathbf{F} \in \{0,1\}^{n\times c},\ \mathbf{F}\mathbf{1}_c = \mathbf{1}_n,\ \mathbf{F}^\top\mathbf{1}_n > \mathbf{0},\ \boldsymbol{\gamma}^{(v)} \geq \mathbf{0},\ \mathbf{1}^\top\boldsymbol{\gamma}^{(v)} = 1, \\ & \mathbf{G}_l^{(v)} \in \{0,1\}^{n\times k_l},\ \mathbf{G}_l^{(v)}\mathbf{1}_{k_l} = \mathbf{1}_n,\ \mathbf{G}_l^{(v)\top}\mathbf{1}_n > \mathbf{0}.\end{aligned} \tag{3}$$

where $\mathbf{E} = \mathbf{1}_n\mathbf{1}_n^\top$, with $\alpha \in [0,1]$ and $\eta, \lambda_\gamma \geq 0$ as trade-off parameters. The denominator provides size-aware normalization for the discrete consensus.

The key property of Eq. (3) lies in the reciprocal interaction between $\mathbf{F}$ and $\{\mathbf{G}_l^{(v)}\}$. Given the current consensus, each resolution-specific partition is refined according to both its view-specific structural quality and its agreement with $\mathbf{F}$. The updated partition hypergraphs subsequently provide new high-order structural evidence for refining the consensus. In this way, multi-resolution view-specific structures and the discrete cross-view consensus co-evolve throughout the learning process.

For any feasible discrete partition, $\mathbf{P}_l^{(v)}$ is the orthogonal projector onto $\mathrm{col}(\mathbf{G}_l^{(v)})$. Hence maximizing $\mathrm{Tr}(\mathbf{F}^\top\mathbf{P}_l^{(v)}\mathbf{F})$ is equivalent to minimizing the projection residual $\|\mathbf{F} - \mathbf{P}_l^{(v)}\mathbf{F}\|_F^2$, providing a direct projection-consistency interpretation of the hypergraph term.

Moreover, since $\mathbf{E} = \mathbf{1}_n\mathbf{1}_n^\top$ and $\mathbf{F}$ is a one-hot partition matrix,

$$\mathrm{Tr}(\mathbf{F}^\top\mathbf{E}\mathbf{F}) = \frac{n^2}{c} + \left\|\mathbf{F}^\top\mathbf{1}_n - \frac{n}{c}\mathbf{1}_c\right\|_2^2. \tag{4}$$

Therefore, the denominator of the consensus trace ratio implicitly penalizes deviations from balanced cluster sizes. The projection interpretation is formally established in Proposition 1 of Appendix A.2.

3.2 Extreme-Point Geometry of the Consensus Relaxation

To examine whether the categorical consensus is favored by the objective geometry, fix $\{\mathbf{G}_l^{(v)}\}$ and $\{\boldsymbol{\gamma}^{(v)}\}$ and define

$$\mathbf{A} = \sum_{v=1}^{V}\left[\alpha\boldsymbol{\Theta}^{(v)} + (1-\alpha)\mathbf{K}^{(v)}\right]. \tag{5}$$

Replacing only the one-hot constraint on $\mathbf{F}$ by the row-simplex constraint gives the faithful relaxation

$$\max_{\mathbf{F}\geq\mathbf{0}} \quad \rho(\mathbf{F}) = \frac{\mathrm{Tr}(\mathbf{F}^\top\mathbf{A}\mathbf{F})}{\mathrm{Tr}(\mathbf{F}^\top\mathbf{E}\mathbf{F})}, \qquad \mathbf{F}\mathbf{1}_c = \mathbf{1}_n, \quad \mathbf{F}^\top\mathbf{1}_n > \mathbf{0}. \tag{6}$$

We first derive a local extreme-point condition, then an $\mathbf{F}$-independent certificate, and finally a structural guarantee in terms of partition resolution and balance.

Theorem 1 (Extreme-Point Property of the Relaxed Consensus)**.** *Let $\mathbf{F}^\star$ be a local maximizer of Eq. (6) and let $\rho^\star = \rho(\mathbf{F}^\star)$. If*

$$\min_i A_{ii} > \rho^\star, \tag{7}$$

then every row of $\mathbf{F}^\star$ is a probability-simplex vertex. Therefore, $\mathbf{F}^\star$ is one-hot.

With all other rows fixed, the corresponding difference function has Hessian $2(A_{ii} - \rho^\star)\mathbf{I}_c$. Under Eq. (7), it is strictly convex on the simplex, so a non-vertex point cannot be a local maximum.

To remove the dependence on the unknown $\rho^\star$, note that $\mathbf{A}\mathbf{1}_n = V\mathbf{1}_n$ and define the largest spectral response orthogonal to the constant direction as

$$\lambda_\perp(\mathbf{A}) = \max_{\substack{\mathbf{x}^\top \mathbf{1}_n = 0 \\ \|\mathbf{x}\|_2 = 1}} \mathbf{x}^\top \mathbf{A}\mathbf{x}. \tag{8}$$

Lemma 1 (Uniform Spectral Upper Bound)**.** *For every feasible row-simplex $\mathbf{F}$ in Eq. (6),*

$$\rho(\mathbf{F}) \le \frac{V + (c-1)\lambda_\perp(\mathbf{A})}{n}. \tag{9}$$

Combining Theorem 1 and Lemma 1 gives the following solution-independent certificate.

Theorem 2 (Structural–Spectral Certificate)**.** *If*

$$\min_i A_{ii} > \frac{V + (c-1)\lambda_\perp(\mathbf{A})}{n}, \tag{10}$$

then every local maximizer of Eq. (6) is one-hot.

To relate this certificate to the multi-resolution construction, define

$$\beta = \max_{v,l,q} \frac{k_l n_{lq}^{(v)}}{n}, \qquad r = \min_v \frac{\sum_{l=1}^{L} \gamma_l^{(v)} k_l}{c}, \tag{11}$$

where $n_{lq}^{(v)}$ is the size of the q-th group in $\mathbf{G}_l^{(v)}$. Here, $\beta \ge 1$ measures worst-case group imbalance, while $r > 0$ measures the minimum effective resolution relative to the target cluster number.

Theorem 3 (Resolution–Balance Guarantee)**.** *If*

$$\alpha r > \beta, \tag{12}$$

then Eq. (10) necessarily holds. Consequently, every local maximizer of Eq. (6) is one-hot.

Together, these results establish progressively verifiable sufficient conditions for one-hot local maximizers. Complete proofs are given in Appendix A.1.

4 OPTIMIZATION

We optimize Eq. (3) directly by block-coordinate ascent while keeping all partition assignments discrete. At each iteration, we sequentially update $\{\mathbf{G}_l^{(v)}\}$, $\{\boldsymbol{\gamma}^{(v)}\}$, and $\mathbf{F}$. For a generic ratio $\rho = N/D$, if a discrete reassignment changes its numerator and denominator by ΔN and ΔD, respectively, its exact variation is

$$\Delta\rho = \frac{\Delta N - \rho \Delta D}{D + \Delta D}. \tag{13}$$

This identity allows both discrete blocks to be updated by directly evaluating their exact objective gains.

Updating $\mathbf{G}_l^{(v)}$. With $\mathbf{F}$ and $\{\boldsymbol{\gamma}^{(v)}\}$ fixed, each resolution-specific partition can be updated independently. For a candidate move of sample i from group a to b, the exact changes $\Delta R_l^{(v)}$ and $\Delta T_l^{(v)}$ are evaluated from maintained group sizes and group-level anchor/consensus sums. Since $D_F = \mathrm{Tr}(\mathbf{F}^\top \mathbf{E}\mathbf{F})$ is unchanged, the overall gain is

$$\Delta\mathcal{J}_G = \frac{\alpha\gamma_l^{(v)}}{D_F}\Delta T_l^{(v)} + \eta \Delta R_l^{(v)}. \tag{14}$$

The feasible destination with the largest positive gain is accepted, while moves that empty the source group are excluded. Closed-form expressions for $\Delta R_l^{(v)}$ and $\Delta T_l^{(v)}$ are derived in Appendix A.3.

Updating $\boldsymbol{\gamma}^{(v)}$. With $\mathbf{F}$ and $\{\mathbf{G}_l^{(v)}\}$ fixed, the resolution weights of each view are independent. Let $\mathbf{t}^{(v)} = [T_1^{(v)}, \ldots, T_L^{(v)}]^\top$. The corresponding subproblem becomes

$$\max_{\boldsymbol{\gamma}^{(v)} \in \Delta_L} \quad \frac{\alpha}{D_F} \mathbf{t}^{(v)\top} \boldsymbol{\gamma}^{(v)} - \lambda_\gamma \|\boldsymbol{\gamma}^{(v)}\|_2^2, \tag{15}$$

where $\Delta_L = \{\boldsymbol{\gamma} \geq \mathbf{0} \mid \mathbf{1}^\top \boldsymbol{\gamma} = 1\}$. For $\lambda_\gamma > 0$, Eq. (15) is a strictly concave quadratic program with the unique solution

$$\boldsymbol{\gamma}^{(v)} = \Pi_{\Delta_L} \left(\frac{\alpha}{2\lambda_\gamma D_F} \mathbf{t}^{(v)} \right), \tag{16}$$

where $\Pi_{\Delta_L}(\cdot)$ denotes Euclidean projection onto the probability simplex. When $\lambda_\gamma = 0$, the linear subproblem attains its optimum at a simplex vertex associated with a maximal $T_l^{(v)}$.

Updating F. With $\{\mathbf{G}_l^{(v)}\}$ and $\{\boldsymbol{\gamma}^{(v)}\}$ fixed, let N_F and D_F denote the numerator and denominator of the consensus trace ratio and set $\theta_F = N_F / D_F$. For a candidate reassignment of sample i, the maintained anchor-cluster and partition-consensus statistics give exact changes ΔN_F and ΔD_F, and therefore

$$\Delta \mathcal{J}_F = \frac{\Delta N_F - \theta_F \Delta D_F}{D_F + \Delta D_F}. \tag{17}$$

The feasible destination cluster with the largest positive gain is accepted, with empty-cluster moves excluded. The closed-form sufficient-statistic updates for ΔN_F and ΔD_F are provided in Appendix A.3.

Matrix-free implementation. None of the $n \times n$ matrices $\mathbf{K}^{(v)}$, $\mathbf{P}_l^{(v)}$, or $\mathbf{E}$ is explicitly formed. In particular,

$$\mathrm{Tr}(\mathbf{F}^\top \mathbf{K}^{(v)} \mathbf{F}) = \|\mathbf{F}^\top \mathbf{B}^{(v)}\|_F^2, \qquad T_l^{(v)} = \|\mathbf{D}_l^{(v)-1/2} \mathbf{G}_l^{(v)\top} \mathbf{F}\|_F^2. \tag{18}$$

Analogous sufficient-statistic forms for the partition-quality terms are given in Appendix A.4.

5 CONVERGENCE AND COMPLEXITY

Theorem 4 (Monotonic Convergence)**.** *Assume that the trace ratios in Eq. (3) are well-defined. If each discrete reassignment is accepted only when its exact objective gain is positive and each $\boldsymbol{\gamma}^{(v)}$-subproblem is solved exactly, then the objective sequence is monotonically nondecreasing and converges to a finite limit.*

Complexity. Let $\nu_v = \mathrm{nnz}(\mathbf{B}^{(v)})/n$ denote the average number of nonzero anchor connections per sample and p_l the number of candidate groups evaluated at resolution l. One alternating sweep has time complexity $\mathcal{O}(n[\sum_v \nu_v(\sum_l p_l + c) + cVL] + VL \log L)$, where $p_l = k_l - 1$ under exhaustive reassignment. According to Eq. (18), no $n \times n$ matrix is explicitly formed, and the storage complexity is $\mathcal{O}(n[\sum_v \nu_v + VL] + \sum_v [(c + \sum_l k_l) m_v + c \sum_l k_l])$. Hence, for fixed model sizes and sparse degrees, both costs scale linearly with n while avoiding quadratic sample-wise storage.

6 EXPERIMENTS

6.1 EXPERIMENTAL SETUP

Datasets and metrics. We evaluate DCoE on eleven widely used multi-view benchmark datasets, whose statistics are summarized in Table 1. These datasets cover a wide range of sample sizes, view numbers, feature dimensions, and cluster numbers. We report ACC, NMI, and Purity, averaged over five independent random trials. Detailed preprocessing, parameter settings, and initialization procedures are provided in Appendix A.5. All experiments are conducted on a workstation equipped with an Intel Core i7-13700KF CPU, 32 GB of RAM, and an NVIDIA GeForce RTX 4090 GPU.

Table 1: Benchmark dataset statistics.

Dataset	#Samples	#Views	#Clusters	Dimensions
MSRC	210	5	7	24/576/512/256/254
NGs	500	3	5	2000/2000/2000
Handwritten	2,000	6	10	240/76/216/47/64/6
100Leaves	1,600	3	100	64/64/64
Scene	4,485	3	15	1800/1180/1240
OutScene	2,688	4	8	512/432/256/48
UCI	2,000	3	10	240/76/6
ALOI	10,800	4	100	77/13/64/125
MNIST-10k	10,000	3	10	30/9/30
COIL20	1,440	3	20	1024/944/4096
YouTubeFace	101,499	5	31	64/512/64/647/838

Compared methods. We compare against thirteen representative methods: AEVC (Liu et al., 2024), BONE (Xu et al., 2026), IB2MC (Guo et al., 2026), ALPC (Chen et al., 2025b), FMVACC (Wang et al., 2022), DMFC-SAGL (Cui et al., 2025), BGDR and CGDR (Ma et al., 2026), SparseMVC (Liu et al., 2026), RAV (Peng et al., 2026), AECoDDC (Trosten et al., 2023), WAH (Wang et al., 2026), and SEM (Xu et al., 2023).

6.2 Comparison with Competing Methods

Table 2: Clustering performance on eleven benchmark datasets. *OOM* denotes out-of-memory.

Dataset	AEVC CVPR'24	BONE AAAI'26	IB2MC TPAMI'26	ALPC AAAI'25	FMVACC NeurIPS'22	DMFC-SAGL MM'25	BGDR TKDE'26	CGDR TKDE'26	SparseMVC NeurIPS'25	RAV ICLR'26	AECoDDC CVPR'23	WAH AAAI'26	SEM NeurIPS'23	DCoE
ACC														
MSRC	0.7689	0.7762	0.8257	0.7930	0.6676	0.8190	0.8238	0.7810	0.6190	0.8285	0.4952	0.7238	0.7048	**0.9523**
NGs	0.8780	0.9440	0.6655	0.9299	0.9069	0.8040	0.7080	0.6880	0.6080	0.8820	0.3924	0.6080	0.5060	**0.9676**
Handwritten	0.7352	0.9700	0.9546	0.8001	0.9240	0.9235	0.8020	0.7075	0.8605	0.8910	0.5555	0.8300	0.8500	**0.9844**
100Leaves	0.7674	0.4788	0.9335	0.7537	0.4043	0.2243	0.7619	0.7712	0.5937	0.6550	0.5988	0.6062	0.4756	**0.9906**
Scene	0.4966	0.5987	0.5922	0.4550	0.5644	0.5173	0.4707	0.4149	0.3580	0.3157	0.4600	0.4618	0.4756	**0.6561**
OutScene	0.5088	0.7496	0.6484	0.6791	0.7054	0.7671	0.4297	0.3341	0.6867	0.5814	0.5740	0.6280	0.7210	**0.7754**
UCI	0.6659	0.8510	0.9495	0.7717	0.8912	0.9505	0.6430	0.5970	0.7785	0.8470	0.5190	0.7380	0.6640	**0.9517**
ALOI	0.6706	0.8111	0.7707	0.7334	0.3996	0.2452	0.6198	0.6363	0.5793	0.7688	0.6809	0.8284	0.7966	**0.9142**
MNIST-10k	0.6418	0.8855	0.6995	0.6605	0.8381	0.8493	0.7781	0.7652	0.8888	0.8961	0.7844	0.8998	0.7114	**0.9133**
COIL20	0.6564	0.7903	0.7875	0.6666	0.6772	0.6743	0.7569	0.7681	0.6201	0.7722	0.6173	0.7181	0.7389	**0.8050**
YouTubeFace	0.2215	0.1192	0.2961	0.2869	0.1149	0.1774	0.1609	0.2483	*OOM*	0.2732	0.2585	0.2924	0.3185	**0.3598**
NMI														
MSRC	0.6731	0.7402	0.7241	0.7209	0.5638	0.7770	0.7259	0.6715	0.5697	0.7386	0.3637	0.6721	0.6281	**0.9025**
NGs	0.7238	0.8918	0.4835	0.8368	0.7970	0.7221	0.4993	0.4786	0.4936	0.7504	0.1729	0.3861	0.3696	**0.8976**
Handwritten	0.7094	0.9301	0.8997	0.7789	0.8890	0.9193	0.7567	0.6387	0.8610	0.8754	0.5445	0.8140	0.8314	**0.9620**
100Leaves	0.9145	0.7452	0.9606	0.9008	0.4317	0.5936	0.8260	0.8568	0.8253	0.8374	0.8372	0.8200	0.7481	**0.9925**
Scene	0.4584	0.6127	0.5888	0.4660	0.5332	0.6046	0.4523	0.3872	0.4064	0.3907	0.4525	0.5143	**0.7481**	0.6105
OutScene	0.4311	0.6141	0.5072	0.5638	0.5629	**0.8479**	0.3277	0.2268	0.5805	0.5412	0.4770	0.5434	0.5954	0.6342
UCI	0.6213	0.8315	0.8964	0.7316	0.8386	**0.9509**	0.6496	0.5728	0.7536	0.8328	0.5355	0.6912	0.6963	0.9269
ALOI	0.8031	0.8753	0.8368	0.8736	0.6400	0.4608	0.7293	0.7478	0.7891	0.8922	0.8964	0.9072	0.9069	**0.9309**
MNIST-10k	0.5520	0.7740	0.6194	0.6248	0.7109	**0.9084**	0.6938	0.6817	0.7850	0.7920	0.6764	0.7954	0.6451	0.8323
COIL20	0.7572	0.8725	0.8464	0.7980	0.7796	0.7804	0.8277	0.8299	0.7467	0.8458	0.7159	0.7812	0.8194	**0.8758**
YouTubeFace	0.2030	0.0737	0.2806	0.2611	0.0803	0.1788	0.1184	0.0608	*OOM*	0.2529	0.2595	0.2638	**0.3511**	0.3417
Purity														
MSRC	0.7765	0.8048	0.8269	0.8015	0.6714	0.8190	0.8238	0.7810	0.6476	0.8285	0.5076	0.7333	0.7190	**0.9523**
NGs	0.8780	0.9440	0.6697	0.9334	0.9097	0.8040	0.7080	0.6880	0.6080	0.8820	0.4064	0.6080	0.5380	**0.9676**
Handwritten	0.7493	0.9700	0.9546	0.8146	0.9264	0.9235	0.8020	0.7075	0.8605	0.8910	0.5680	0.8300	0.8500	**0.9844**
100Leaves	0.8040	0.5000	0.9360	0.7855	0.4317	0.2568	0.7750	0.7844	0.6275	0.6725	0.6053	0.6462	0.5056	**0.9906**
Scene	0.5263	0.6586	0.6187	0.4946	0.6090	0.5644	0.4807	0.4234	0.3881	0.3580	0.4980	0.5113	0.5056	**0.6815**
OutScene	0.5628	0.7496	0.6484	0.6986	0.7082	**0.8203**	0.4438	0.3519	0.6867	0.5859	0.5924	0.6291	0.7210	0.7754
UCI	0.6957	0.8515	0.9495	0.7903	0.8964	0.9505	0.6710	0.5970	0.7785	0.8470	0.5414	0.7380	0.6840	**0.9520**
ALOI	0.6939	0.8179	0.7756	0.7589	0.4154	0.2597	0.6395	0.6464	0.6093	0.7887	0.6956	0.8334	0.8077	**0.9151**
MNIST-10k	0.6521	0.8855	0.7373	0.7072	0.8381	0.8890	0.7781	0.7652	0.8888	0.8961	0.7928	0.8998	0.7114	**0.9133**
COIL20	0.6793	0.7958	0.8014	0.6964	0.7072	0.7048	0.7743	0.7868	0.6437	0.7750	0.6298	0.7438	0.7389	**0.8112**
YouTubeFace	0.3263	0.2708	0.3970	0.3639	0.2722	0.2795	0.2829	0.2756	*OOM*	0.3792	0.3486	0.3875	0.3936	**0.4240**

Table 2 shows that DCoE achieves the best ACC on all eleven datasets, improving over the strongest competitor by 4.01 percentage points on average. It also ranks first in Purity on ten datasets and first/second in NMI on six/four datasets, respectively.

6.3 Ablation Study

We compare DCoE with five variants. *Initialization* directly uses the initial partition. *Relaxed-F* applies the row-simplex relaxation only to the consensus block $\mathbf{F}$, while retaining discrete co-evolving resolution-specific partitions and the same initialization as DCoE. The relaxed consensus is optimized by projected-gradient ascent with monotonic backtracking, and labels are obtained by row-wise argmax. *Relaxed-F, G* additionally relaxes $\{\mathbf{G}_l^{(v)}\}$ to row-simplex variables. Both relaxed variants use at most 80 outer iterations with early stopping. *w/o Partition Hypergraph* removes the

partition-hypergraph contribution, while *Fixed-G* keeps the initialized resolution-specific partitions fixed.

Table 3: Ablation study of DCoE in terms of ACC on eleven benchmark datasets.

	MSRC	NGs	Handwritten	100Leaves	Scene	OutScene	UCI	ALOI	MNIST-10k	COIL20	YouTubeFace
Initialization	0.7876	0.7708	0.8066	0.8643	0.5173	0.5798	0.6966	0.7486	0.5664	0.6949	0.2284
Relaxed-F	0.9000	0.9420	0.9810	0.9513	0.6441	0.7623	0.8350	0.8553	0.7589	0.7812	0.3090
Relaxed-F, G	0.8761	0.8828	0.9240	0.9420	0.5960	0.6995	0.8579	0.8474	0.7046	0.7789	0.2937
w/o Partition Hypergraph	0.9124	0.9340	0.9785	0.9515	0.5831	0.7424	0.8900	0.8451	0.8259	0.7656	0.3106
Fixed-G	0.9238	0.9480	0.9839	0.9618	0.6443	0.7712	0.9279	0.8908	0.8526	0.7966	0.3461
DCoE	**0.9523**	**0.9676**	**0.9844**	**0.9906**	**0.6561**	**0.7754**	**0.9517**	**0.9142**	**0.9133**	**0.8050**	**0.3598**

Table 3 shows that DCoE achieves the highest ACC on all eleven datasets. Under matched initialization, DCoE improves average ACC by 5.00 percentage points over *Relaxed-F*. *Relaxed-F* also exceeds *Relaxed-F, G* on ten of the eleven datasets, with an average improvement of 2.88 points. Removing the partition hypergraph reduces ACC by 4.83 points on average, while co-evolving the resolution-specific partitions improves ACC by 2.03 points over *Fixed-G*. Detailed runtime comparisons with the relaxed variants are provided in Appendix A.6.

6.4 Efficiency

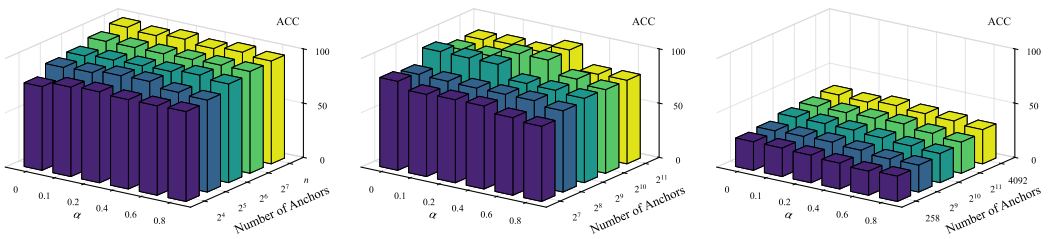
Table 4: Running time comparison (in seconds) on eleven benchmark datasets. *OOM* denotes out-of-memory failure.

Method	MSRC	NGs	Handwritten	100Leaves	Scene	OutScene	UCI	ALOI	MNIST-10k	COIL20	YouTubeFace
AEVC (CVPR'24)	0.060	0.102	0.460	3.630	1.250	0.620	0.930	24.730	2.320	0.460	85.490
BONE (AAAI'26)	2.6365	4.0280	34.6399	29.9833	45.8588	80.5379	19.4715	262.120	99.2123	31.9132	3925.710
IB2MC (TPAMI'26)	0.140	0.630	0.410	1.400	7.600	1.220	0.260	4.230	0.600	3.380	200.110
ALPC (AAAI'25)	0.090	0.300	0.120	0.700	0.530	0.150	0.140	7.600	0.160	0.610	25.660
FMVACC (NeurIPS'22)	0.361	0.389	0.389	1.975	13.104	1.233	0.866	21.120	5.208	2.220	532.350
DMFC-SAGL (MM'25)	1.330	2.740	8.490	6.800	43.900	6.360	2.210	88.700	14.210	14.790	3552.020
BGDR (TKDE'26)	0.410	0.540	0.340	0.500	5.050	0.762	0.190	4.590	0.620	2.290	141.660
CGDR (TKDE'26)	0.110	0.540	0.290	0.360	4.840	0.690	0.150	1.890	0.460	2.220	128.970
SparseMVC (NeurIPS'25)	124.040	110.080	256.630	389.400	410.820	178.200	145.500	969.390	282.980	218.470	*OOM*
RAV (ICLR'26)	111.920	65.620	1173.140	378.420	1102.670	899.320	549.330	4978.950	2497.430	351.730	49380.950
AECoDDC (CVPR'23)	13.200	19.100	95.500	35.100	185.400	117.800	45.800	326.500	221.400	61.800	6957.500
WAH (AAAI'26)	25.180	23.380	249.510	93.680	212.520	179.170	94.630	924.950	432.640	66.420	10803.860
SEM (NeurIPS'23)	44.700	31.300	310.100	254.900	276.700	242.800	127.300	1933.100	490.700	105.700	11715.000
DCoE	0.030	0.030	0.460	0.780	1.200	0.510	0.240	7.920	1.330	0.220	83.910

Table 4 shows that DCoE is highly efficient, achieving the lowest runtime on MSRC, NGs, and COIL20. It also remains practical on YouTubeFace (83.91 s), consistent with its matrix-free implementation.

6.5 Parameter Sensitivity and Convergence

We report representative sensitivity and convergence results on MSRC, MNIST-10k, and YouTube-Face, which span substantially different data scales. Additional parameter sensitivity results are provided in Appendix A.8.

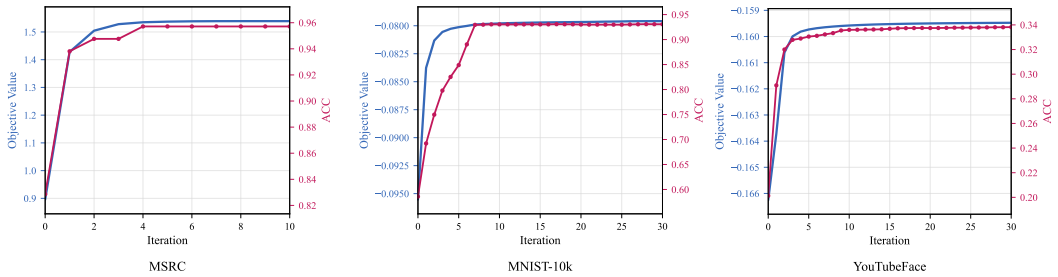


(a) MSRC (b) MNIST-10k (c) YouTubeFace

Figure 2: Sensitivity of DCoE to the number of anchors and α.

Sensitivity to the number of anchors and α. We vary $\alpha \in \{0, 0.1, 0.2, 0.4, 0.6, 0.8\}$ and five dataset-specific anchor configurations while fixing the remaining parameters. As shown in Fig. 2, DCoE remains robust over a broad range of configurations.

Convergence behavior. Figure 3 shows that the objective increases monotonically and rapidly stabilizes, consistent with Theorem 4. ACC also stabilizes rapidly, and DCoE typically terminates well before the maximum iteration budget.



(a) MSRC (b) MNIST-10k (c) YouTubeFace

Figure 3: Convergence behavior of DCoE on representative datasets.

6.6 Empirical Analysis of the Relaxed Consensus Geometry

We measure the discreteness of a relaxed consensus by $\mathrm{Disc}(\mathbf{F}) = n^{-1} \sum_i \max_j F_{ij}$, where a value of 1 denotes an exactly one-hot assignment.

Table 5: Consensus discreteness of the relaxed variants.

	MSRC	NGs	Handwritten	100Leaves	Scene	OutScene	UCI	ALOI	MNIST-10k	COIL20	YouTubeFace
Relaxed-F	1.0000	1.0000	0.9959	0.9981	0.9982	0.9977	0.9999	0.9950	0.9953	1.0000	0.9924
Relaxed-F, G	0.9998	0.9989	0.9975	0.9971	0.9973	0.9924	0.9991	0.9933	0.9965	0.9993	0.9821

As shown in Table 5, both relaxed variants produce highly categorical consensus assignments, with average $\mathrm{Disc}(\mathbf{F})$ values of 0.9975 and 0.9958 for *Relaxed-F* and *Relaxed-F, G*, respectively. In particular, the nearly categorical behavior of *Relaxed-F* is consistent with the extreme-point geometry of the faithful consensus relaxation analyzed in Section 3.2. Despite this near-categorical behavior, DCoE achieves higher ACC than both relaxed variants, as reported in Table 3, indicating that near-discrete assignments alone do not necessarily yield comparable clustering performance.

For the learned structures, we further evaluate $U_{\mathrm{cert}} = [V + (c-1)\lambda_{\perp}(\mathbf{A})]/n$ and $\Delta_{\mathrm{cert}} = \min_i A_{ii} - U_{\mathrm{cert}}$. All eleven datasets yield $\Delta_{\mathrm{cert}} > 0$, satisfying the sufficient condition in Theorem 2. Detailed certificate values are reported in Appendix A.7.

7 Conclusion

We proposed DCoE, a discrete co-evolution framework that jointly refines multi-resolution view-specific partitions and a common consensus without continuous relaxation. Theoretical analysis characterizes conditions under which the faithful consensus relaxation admits one-hot local maximizers, while the matrix-free optimization enables efficient large-scale learning. Experiments consistently demonstrate the effectiveness and efficiency of DCoE.

AI Use Statement

LLMs were used to assist with language editing and mathematical checking. All AI-assisted content was carefully reviewed and verified by the authors.

Reproducibility Statement

The optimization procedure and complexity analysis are summarized in the main paper, while complete proofs, exact reassignment-gain derivations, matrix-free identities, implementation details, parameter settings, and dataset preprocessing procedures are provided in the appendix and supplementary material. Anonymous source code is also provided with the submission to facilitate reproduction of the reported results.

References

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 *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10295–10304. IEEE, 2025a.

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.

Yawei Chen, Huibing Wang, Mingze Yao, Jinjia Peng, Guangqi Jiang, and Jiqing Zhang. Scalable multi-view clustering based on tight anchor distribution. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 7739–7747, 2025c.

Luyan Cui, Huibing Wang, Yawei Chen, Mingze Yao, Xianping Fu, and Jiqing Zhang. Dual-constraint multi-view fuzzy clustering with scalable anchor graph learning. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 7748–7756, 2025.

Yifan Feng, Yifan Zhang, Shihui Ying, Shaoyi Du, and Yue Gao. Kernelized hypergraph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(10):8938–8954, 2025.

Yue Gao, Zizhao Zhang, Haojie Lin, Xibin Zhao, Shaoyi Du, and Changqing Zou. Hypergraph learning: Methods and practices. *IEEE transactions on pattern analysis and machine intelligence*, 44(5):2548–2566, 2022.

Zhibin Gu et al. Hypergraph-enhanced contrastive learning for multi-view clustering with hyper-laplacian regularization. *Advances in Neural Information Processing Systems*, 38:54600–54620, 2025.

Shengzhao Guo, Jingyu Wang, Letian Zhang, Feiping Nie, and Xuelong Li. Ib2mc: Information bottleneck inspired balanced multiview clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026.

Wen-Jue He, Zheng Zhang, and Xiaofeng Zhu. Dual-correlation-guided anchor learning for scalable incomplete multi-view clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 36(9):17336–17349, 2025.

Xingchen Hu, Miao Jia, Jiyuan Liu, Siwei Wang, Ke Liang, and Wenjing Yang. Hierarchical anchor graph learning for multi-view clustering. In *Forty-third International Conference on Machine Learning*, 2026.

Hengrong Ju, Yang Lu, Weiping Ding, Wei Zhang, and Xibei Yang. Multigranularity information fused contrastive learning with multiview clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 36(10):18915–18929, 2025.

Fan Li, Xiaoyang Wang, Wenjie Zhang, Ying Zhang, and Xuemin Lin. Dhg-bench: A comprehensive benchmark for deep hypergraph learning. In *International Conference on Learning Representations*, volume 2026, pp. 47693–47727, 2026.

Lusi Li and Haibo He. Bipartite graph based multi-view clustering. *IEEE transactions on knowledge and data engineering*, 34(7):3111–3125, 2022.

Xuelong Li, Han Zhang, Rong Wang, and Feiping Nie. Multiview clustering: A scalable and parameter-free bipartite graph fusion method. *IEEE transactions on pattern analysis and machine intelligence*, 44(1):330–344, 2022.

Yeqing Li, Feiping Nie, Heng Huang, and Junzhou Huang. Large-scale multi-view spectral clustering via bipartite graph. In *Proceedings of the AAAI conference on artificial intelligence*, volume 29, 2015.

Ruimeng Liu, Xin Zou, Chang Tang, Xiao Zheng, Xingchen Hu, Kun Sun, and Xinwang Liu. Sparsemvc: probing cross-view sparsity variations for multi-view clustering. *Advances in Neural Information Processing Systems*, 38:154594–154622, 2026.

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

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*, 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*, 2026.

Qianwen Meng, Hangwei Qian, Yong Liu, Lizhen Cui, Yonghui Xu, and Zhiqi Shen. Mhccl: masked hierarchical cluster-wise contrastive learning for multivariate time series. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 9153–9161, 2023.

Feiping Nie, Xiaoqian Wang, Michael Jordan, and Heng Huang. The constrained laplacian rank algorithm for graph-based clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016.

Jinjia Peng, Tianhang Cheng, Guangqi Jiang, and Huibing Wang. Prior-oriented anchor learning with coalesced semantics for multi-view clustering. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 1141–1150, 2025.

Shuangmei Peng, Zhe Chen, Tianyang Xu, and Xiaojun Wu. Relationship alignment for view-aware multi-view clustering. In *International Conference on Learning Representations*, volume 2026, pp. 147335–147347, 2026.

Yalan Qin, Xinpeng Zhang, and Guorui Feng. Explainable k-means neural networks for multi-view clustering. In *The Fourteenth International Conference on Learning Representations*, 2026.

Alexander Strehl and Joydeep Ghosh. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of machine learning research*, 3(Dec):583–617, 2002.

Peng Su, Shudong Huang, Weihong Ma, Deng Xiong, and Jiancheng Lv. Multi-view granular-ball contrastive clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 20637–20645, 2025.

Alexander Topchy, Anil K Jain, and William Punch. Clustering ensembles: Models of consensus and weak partitions. *IEEE transactions on pattern analysis and machine intelligence*, 27(12): 1866–1881, 2005.

Daniel J Trosten, Sigurd Løkse, Robert Jenssen, and Michael C Kampffmeyer. On the effects of self-supervision and contrastive alignment in deep multi-view clustering. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23976–23985. IEEE, 2023.

Xinhang Wan, Jiyuan Liu, Weixuan Liang, Xinwang Liu, Yi Wen, and En Zhu. Continual multi-view clustering. In *Proceedings of the 30th ACM International Conference on Multimedia*, pp. 3676–3684, 2022.

Jun Wang, Zhenglai Li, Chang Tang, Suyuan Liu, Hao Yu, Chuan Tang, Miaomiao Li, and Xinwang Liu. Scalable cross-view sample alignment for multi-view clustering with view structure similarity. In *Advances in Neural Information Processing Systems*, volume 38, 2025a.

Qianqian Wang, Bowen Zhao, Zhengming Ding, Xiangdong Zhang, and Quanxue Gao. A simple yet effective hypergraph clustering network. In *IJCAI*, pp. 6352–6360, 2025b.

Rui Wang, Yuting Jiang, Xiaoqing Luo, Xiao-Jun Wu, Nicu Sebe, and Ziheng Chen. Wasserstein-aligned hyperbolic multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 26444–26452, 2026.

Siwei Wang, Xinwang Liu, Suyuan Liu, Jiaqi Jin, Wenxuan Tu, Xinzhong Zhu, and En Zhu. Align then fusion: Generalized large-scale multi-view clustering with anchor matching correspondences. *Advances in Neural Information Processing Systems*, 35:5882–5895, 2022.

Siwei Wang, Xinwang Liu, Suyuan Liu, Wenxuan Tu, and En Zhu. Scalable and structural multi-view graph clustering with adaptive anchor fusion. *IEEE Transactions on Image Processing*, 33: 4627–4639, 2024.

Suixue Wang, Shilin Zhang, Qingchen Zhang, Peng Li, and Weiliang Huo. Master: A multi-granularity invariant structure clustering scheme for multi-view clustering. In *IJCAI*, pp. 6415–6423, 2025c.

Xinxin Wang, Yongshan Zhang, and Yicong Zhou. Bidirectional probabilistic multi-graph learning and decomposition for multi-view clustering. *IEEE Transactions on Image Processing*, 34:3609–3621, 2025d.

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.

Huilang Xu, Xiang-Xiang Su, Simin Chen, Guang-Yong Chen, and Xing Chen. Bridging optimization and neural networks for efficient multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 16066–16074, 2026.

Jie Xu, Shuo Chen, Yazhou Ren, Xiaoshuang Shi, Hengtao Shen, Gang Niu, and Xiaofeng Zhu. Self-weighted contrastive learning among multiple views for mitigating representation degeneration. *Advances in neural information processing systems*, 36:1119–1131, 2023.

Ben Yang, Xuetao Zhang, Zhongheng Li, Feiping Nie, and Fei Wang. Efficient multi-view k-means clustering with multiple anchor graphs. *IEEE Transactions on Knowledge and Data Engineering*, 35(7):6887–6900, 2023.

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*, pp. 6794–6802, 2025.

Jie Yang, Cheng-You Lu, Zhongli Wang, Hsiang-Ting Chen, Guang-Kui Xu, Chenglong Zhang, Shuting Dong, Xinyan Liang, and Bingbing Jiang. Multi-view clustering with granularity-aware pseudo supervision. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 27538–27546, 2026.

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.

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

Pei Zhang, Siwei Wang, Liang Li, Changwang Zhang, Xinwang Liu, En Zhu, Zhe Liu, Lu Zhou, and Lei Luo. Let the data choose: Flexible and diverse anchor graph fusion for scalable multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 11262–11269, 2023.

Tongzheng Zhao, Yangyang Wen, Yukai Shi, Xinyan Liang, Feijiang Li, Peng Zhou, and Liang Du. Contractive anchor resolvent diffusion for incomplete multi-view clustering. In *Forty-third International Conference on Machine Learning*, 2026.

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

A Appendix

Additional theoretical derivations, proofs, implementation details, and supplementary experimental results are provided in this appendix.

A.1 Proofs of the Main Theorems

This subsection first provides the proofs of the formal results stated in the main paper. We begin with the extreme-point theorem, then establish the uniform spectral upper bound, the structural–spectral certificate, and the resolution–balance guarantee, followed by the monotonic convergence result.

Proof of Theorem 1. Let $\mathbf{f}_j = \mathbf{F}_{j:}^\top$ and fix all rows of $\mathbf{F}$ except the i-th row, denoted by $\mathbf{x} = \mathbf{F}_{i:}^\top \in \Delta_c$. Define

$$\mathbf{b}_i = \sum_{j \neq i} A_{ij}\mathbf{f}_j, \qquad \mathbf{h}_{-i} = \sum_{j \neq i} \mathbf{f}_j. \tag{19}$$

Up to terms independent of $\mathbf{x}$, the numerator and denominator of Eq. (6) can be written as

$$N_i(\mathbf{x}) = A_{ii}\|\mathbf{x}\|_2^2 + 2\mathbf{b}_i^\top \mathbf{x} + C_N, \qquad D_i(\mathbf{x}) = \|\mathbf{x}\|_2^2 + 2\mathbf{h}_{-i}^\top \mathbf{x} + C_D, \tag{20}$$

where C_N and C_D are constants with respect to $\mathbf{x}$.

Let $\mathbf{x}^\star = \mathbf{F}_{i:}^{\star\top}$ and $\rho^\star = \rho(\mathbf{F}^\star)$. Since $\mathbf{F}^\star$ is a local maximizer and the denominator is positive, every sufficiently close feasible $\mathbf{x}$ satisfies

$$N_i(\mathbf{x}) - \rho^\star D_i(\mathbf{x}) \leq 0, \tag{21}$$

with equality at $\mathbf{x}^\star$. Hence $\mathbf{x}^\star$ is a local maximizer of

$$\phi_i(\mathbf{x}) = N_i(\mathbf{x}) - \rho^\star D_i(\mathbf{x}). \tag{22}$$

Its Hessian is

$$\nabla^2 \phi_i(\mathbf{x}) = 2\,(A_{ii} - \rho^\star)\,\mathbf{I}_c. \tag{23}$$

Under Eq. (7), $\nabla^2\phi_i(\mathbf{x}) \succ 0$, so ϕ_i is strictly convex on the feasible simplex. A strictly convex function cannot have a local maximum at a non-extreme point of a convex set. Hence $\mathbf{x}^\star$ must be a simplex vertex. Since i is arbitrary, every row of $\mathbf{F}^\star$ is a simplex vertex, and therefore $\mathbf{F}^\star$ is one-hot.

Proof of Lemma 1. For every hard partition $\mathbf{G}_l^{(v)}$,

$$\mathbf{P}_l^{(v)}\mathbf{1}_n = \mathbf{G}_l^{(v)}\mathbf{D}_l^{(v)-1}\mathbf{G}_l^{(v)\top}\mathbf{1}_n = \mathbf{1}_n. \tag{24}$$

Moreover, using Eq. (1) and $\mathbf{Z}^{(v)}\mathbf{1}_{m_v} = \mathbf{1}_n$,

$$\mathbf{K}^{(v)}\mathbf{1}_n = \mathbf{Z}^{(v)}\mathbf{\Lambda}^{(v)-1}\mathbf{Z}^{(v)\top}\mathbf{1}_n = \mathbf{Z}^{(v)}\mathbf{1}_{m_v} = \mathbf{1}_n. \tag{25}$$

Both matrices are symmetric positive semidefinite. Since $\sum_l \gamma_l^{(v)} = 1$, Eq. (5) therefore gives

$$\mathbf{A}\mathbf{1}_n = V\mathbf{1}_n. \tag{26}$$

Let $\mathbf{f}_s$ denote the s-th column of a feasible row-simplex $\mathbf{F}$, let $h_s = \mathbf{1}_n^\top \mathbf{f}_s$, and decompose

$$\mathbf{f}_s = \frac{h_s}{n}\mathbf{1}_n + \mathbf{u}_s, \qquad \mathbf{u}_s^\top \mathbf{1}_n = 0. \tag{27}$$

By Eq. (26) and the definition of $\lambda_\perp(\mathbf{A})$,

$$\mathbf{f}_s^\top \mathbf{A}\mathbf{f}_s \le \frac{V}{n}h_s^2 + \lambda_\perp(\mathbf{A})\|\mathbf{u}_s\|_2^2. \tag{28}$$

Summing over s and defining $\mathbf{h} = \mathbf{F}^\top \mathbf{1}_n$ yields

$$\mathrm{Tr}(\mathbf{F}^\top \mathbf{A}\mathbf{F}) \le \frac{V}{n}\|\mathbf{h}\|_2^2 + \lambda_\perp(\mathbf{A})\left(\|\mathbf{F}\|_F^2 - \frac{1}{n}\|\mathbf{h}\|_2^2\right). \tag{29}$$

For every row-simplex $\mathbf{F}$,

$$\|\mathbf{F}\|_F^2 \le n, \qquad \|\mathbf{h}\|_2^2 \ge \frac{n^2}{c}, \tag{30}$$

where the second inequality follows from $\mathbf{1}_c^\top \mathbf{h} = n$ and Cauchy–Schwarz. Furthermore, each $\mathbf{P}_l^{(v)}$ is an orthogonal projector and each $\mathbf{K}^{(v)}$ is symmetric positive semidefinite and row-stochastic. Hence their eigenvalues lie in $[0, 1]$, which implies $0 \le \lambda_\perp(\mathbf{A}) \le V$.

Since $\mathrm{Tr}(\mathbf{F}^\top \mathbf{E}\mathbf{F}) = \|\mathbf{h}\|_2^2$, Eq. (29) gives

$$\begin{aligned}\rho(\mathbf{F}) &\le \frac{V - \lambda_\perp(\mathbf{A})}{n} + \lambda_\perp(\mathbf{A})\frac{\|\mathbf{F}\|_F^2}{\|\mathbf{h}\|_2^2} \\ &\le \frac{V - \lambda_\perp(\mathbf{A})}{n} + \frac{c\,\lambda_\perp(\mathbf{A})}{n} \\ &= \frac{V + (c-1)\lambda_\perp(\mathbf{A})}{n},\end{aligned} \tag{31}$$

which proves Lemma 1.

Proof of Theorem 2. By Lemma 1, every feasible row-simplex $\mathbf{F}$ satisfies

$$\rho(\mathbf{F}) \le \frac{V + (c-1)\lambda_\perp(\mathbf{A})}{n}. \tag{32}$$

Therefore, under Eq. (10),

$$\min_i A_{ii} > \frac{V + (c-1)\lambda_\perp(\mathbf{A})}{n} \ge \rho(\mathbf{F}) \tag{33}$$

for every feasible $\mathbf{F}$. In particular, $\min_i A_{ii} > \rho(\mathbf{F}^\star)$ at every local maximizer $\mathbf{F}^\star$. Theorem 1 then implies that every such $\mathbf{F}^\star$ is one-hot.

Proof of Theorem 3. Fix any sample i. At resolution l of view v, let $q_l^{(v)}(i)$ denote the group containing sample i. Since $\mathbf{G}_l^{(v)}$ is one-hot,

$$P_{l,ii}^{(v)} = \frac{1}{n_{l,q_l^{(v)}(i)}^{(v)}}. \tag{34}$$

By the definition of β in Eq. (11),

$$n_{lq}^{(v)} \le \beta\frac{n}{k_l}, \tag{35}$$

and therefore

$$P_{l,ii}^{(v)} \ge \frac{k_l}{\beta n}. \tag{36}$$

Moreover, $\mathbf{K}^{(v)}$ is positive semidefinite, so $K_{ii}^{(v)} \geq 0$. Therefore, using Eq. (5),

$$\begin{aligned} A_{ii} &= \sum_{v=1}^{V} \left[\alpha \sum_{l=1}^{L} \gamma_l^{(v)} P_{l,ii}^{(v)} + (1-\alpha) K_{ii}^{(v)} \right] \\ &\geq \frac{\alpha}{\beta n} \sum_{v=1}^{V} \sum_{l=1}^{L} \gamma_l^{(v)} k_l. \end{aligned} \tag{37}$$

By the definition of r in Eq. (11),

$$\sum_{l=1}^{L} \gamma_l^{(v)} k_l \geq rc \tag{38}$$

for every view v. Hence

$$A_{ii} \geq \frac{\alpha V r c}{\beta n}, \tag{39}$$

and consequently

$$\min_i A_{ii} \geq \frac{\alpha V r c}{\beta n}. \tag{40}$$

Each $\mathbf{P}_l^{(v)}$ is an orthogonal projector, and each $\mathbf{K}^{(v)}$ is symmetric positive semidefinite and row-stochastic. Thus their eigenvalues lie in $[0, 1]$. Since $\mathbf{\Theta}^{(v)}$ is a convex combination of $\{\mathbf{P}_l^{(v)}\}_{l=1}^{L}$, $\alpha\mathbf{\Theta}^{(v)} + (1-\alpha)\mathbf{K}^{(v)}$ has spectral norm at most one. Consequently,

$$\lambda_\perp(\mathbf{A}) \leq V. \tag{41}$$

It follows that

$$\frac{V + (c-1)\lambda_\perp(\mathbf{A})}{n} \leq \frac{Vc}{n}. \tag{42}$$

Finally, Eq. (12) gives $\alpha r/\beta > 1$. Combining Eqs. (40) and (42) yields

$$\min_i A_{ii} \geq \frac{\alpha V r c}{\beta n} > \frac{Vc}{n} \geq \frac{V + (c-1)\lambda_\perp(\mathbf{A})}{n}. \tag{43}$$

Therefore Eq. (10) holds, and Theorem 2 implies that every local maximizer of Eq. (6) is one-hot.

Proof of Theorem 4. Let $\mathcal{J}^{(t)}$ denote the objective value of Eq. (3) after the t-th complete alternating iteration.

During the update of each $\mathbf{G}_l^{(v)}$, only reassignments with positive exact gain $\Delta\mathcal{J}_G$ are accepted. Each $\boldsymbol{\gamma}^{(v)}$-subproblem is solved exactly over the probability simplex, and the consensus update accepts only reassignments with positive exact gain $\Delta\mathcal{J}_F$. Therefore, none of the three block updates decreases the objective, and hence

$$\mathcal{J}^{(t+1)} \geq \mathcal{J}^{(t)}. \tag{44}$$

It remains to show that the objective is upper bounded. Define $\mathbf{A}^{(v)} = \alpha\mathbf{\Theta}^{(v)} + (1-\alpha)\mathbf{K}^{(v)}$. Both $\mathbf{\Theta}^{(v)}$ and $\mathbf{K}^{(v)}$ are symmetric positive semidefinite and row-stochastic, so $\mathbf{A}^{(v)}$ is also symmetric positive semidefinite and row-stochastic, with $\|\mathbf{A}^{(v)}\|_2 = 1$. Since $\mathbf{F}$ is one-hot, $\|\mathbf{F}\|_F^2 = n$, and thus $\mathrm{Tr}(\mathbf{F}^\top \mathbf{A}^{(v)} \mathbf{F}) \leq n$.

Let $\mathbf{h} = \mathbf{F}^\top \mathbf{1}_n$ denote the cluster-size vector. Since $\mathbf{1}_c^\top \mathbf{h} = n$, $\|\mathbf{h}\|_2^2 \geq n^2/c$. Moreover, $\mathrm{Tr}(\mathbf{F}^\top \mathbf{E} \mathbf{F}) = \|\mathbf{h}\|_2^2$. Therefore,

$$\frac{\sum_{v=1}^{V} \mathrm{Tr}(\mathbf{F}^\top \mathbf{A}^{(v)} \mathbf{F})}{\mathrm{Tr}(\mathbf{F}^\top \mathbf{E} \mathbf{F})} \leq \frac{Vc}{n}. \tag{45}$$

For each view, let $\kappa_v = \max_{i \neq j} K_{ij}^{(v)}$. Since $R_l^{(v)}$ is an average of off-diagonal affinities within partition groups, $R_l^{(v)} \leq \kappa_v$ whenever it is well-defined. Hence $\eta \sum_{v,l} R_l^{(v)} \leq \eta L \sum_v \kappa_v$. Furthermore, $-\lambda_\gamma \sum_v \|\boldsymbol{\gamma}^{(v)}\|_2^2 \leq 0$. Combining these bounds gives

$$\mathcal{J}^{(t)} \leq \frac{Vc}{n} + \eta L \sum_{v=1}^{V} \kappa_v < +\infty. \tag{46}$$

Thus, $\{\mathcal{J}^{(t)}\}$ is monotonically nondecreasing by Eq. (44) and bounded above by Eq. (46). Therefore, $\mathcal{J}^{(t)}$ converges to a finite limit.

A.2 Supplementary Structural Results

This subsection collects supplementary structural results used to interpret the multi-resolution hypergraph construction, projection consistency, and size-aware consensus normalization. Each statement is followed immediately by its proof.

Lemma 2 (Normalized Hypergraph Equivalence)**.** *Let* $\mathbf{H}^{(v)} = [\mathbf{G}_1^{(v)}, \ldots, \mathbf{G}_L^{(v)}]$ *be the incidence matrix of the multi-resolution hypergraph, whose* l*-th resolution hyperedges have weight* $\gamma_l^{(v)}$ *and degree matrix* $\mathbf{D}_l^{(v)}$*. Then* $\mathbf{D}_v^{(v)} = \mathbf{I}_n$*, and*

$$\mathbf{D}_v^{(v)-1/2}\mathbf{H}^{(v)}\mathbf{W}^{(v)}\mathbf{D}_e^{(v)-1}\mathbf{H}^{(v)\top}\mathbf{D}_v^{(v)-1/2} = \sum_{l=1}^{L}\gamma_l^{(v)}\mathbf{P}_l^{(v)} = \mathbf{\Theta}^{(v)}. \tag{47}$$

Proof of Lemma 2. For the v-th view, define the hyperedge-weight matrix as

$$\mathbf{W}^{(v)} = \mathrm{BlkDiag}\left(\gamma_1^{(v)}\mathbf{I}_{k_1}, \ldots, \gamma_L^{(v)}\mathbf{I}_{k_L}\right), \tag{48}$$

and the hyperedge-degree matrix as

$$\mathbf{D}_e^{(v)} = \mathrm{BlkDiag}\left(\mathbf{D}_1^{(v)}, \ldots, \mathbf{D}_L^{(v)}\right), \tag{49}$$

where $\mathrm{BlkDiag}(\cdot)$ denotes the block-diagonal concatenation of its matrix arguments.

Since $\mathbf{G}_l^{(v)}$ is a one-hot partition matrix, each sample belongs to exactly one hyperedge at the l-th resolution, i.e., $\sum_{q=1}^{k_l} G_{l,iq}^{(v)} = 1$ for every sample i. All hyperedges at the l-th resolution share the weight $\gamma_l^{(v)}$. Hence, the weighted degree of vertex i is

$$\begin{aligned} d_i^{(v)} &= \sum_{l=1}^{L}\gamma_l^{(v)}\sum_{q=1}^{k_l}G_{l,iq}^{(v)} \\ &= \sum_{l=1}^{L}\gamma_l^{(v)} = 1, \end{aligned} \tag{50}$$

where the last equality follows from $\mathbf{1}^\top\boldsymbol{\gamma}^{(v)} = 1$. Thus, all vertices have unit weighted degree, and the corresponding vertex-degree matrix is

$$\mathbf{D}_v^{(v)} = \mathrm{Diag}\left(d_1^{(v)}, \ldots, d_n^{(v)}\right) = \mathbf{I}_n. \tag{51}$$

Using $\mathbf{H}^{(v)} = [\mathbf{G}_1^{(v)}, \ldots, \mathbf{G}_L^{(v)}]$, together with Eqs. (48), (49), and (51), we obtain

$$\begin{aligned} &\mathbf{D}_v^{(v)-1/2}\mathbf{H}^{(v)}\mathbf{W}^{(v)}\mathbf{D}_e^{(v)-1}\mathbf{H}^{(v)\top}\mathbf{D}_v^{(v)-1/2} \\ &= \mathbf{H}^{(v)}\mathbf{W}^{(v)}\mathbf{D}_e^{(v)-1}\mathbf{H}^{(v)\top} \\ &= \sum_{l=1}^{L}\gamma_l^{(v)}\mathbf{G}_l^{(v)}\mathbf{D}_l^{(v)-1}\mathbf{G}_l^{(v)\top} \\ &= \sum_{l=1}^{L}\gamma_l^{(v)}\mathbf{P}_l^{(v)} = \mathbf{\Theta}^{(v)}. \end{aligned}$$

which proves Lemma 2.

Proposition 1 (Projection Interpretation)**.** *For any feasible resolution-specific partition* $\mathbf{G}_l^{(v)}$*, the matrix* $\mathbf{P}_l^{(v)}$ *is the orthogonal projection onto* $\mathrm{col}(\mathbf{G}_l^{(v)})$*. For fixed* $\{\mathbf{G}_l^{(v)}\}$ *and* $\{\boldsymbol{\gamma}^{(v)}\}$*, maximizing the multi-resolution hypergraph consistency term is equivalent to*

$$\max_{\mathbf{F}} \sum_{v=1}^{V} \mathrm{Tr}\left(\mathbf{F}^\top \boldsymbol{\Theta}^{(v)} \mathbf{F}\right) \Longleftrightarrow \min_{\mathbf{F}} \sum_{v=1}^{V} \sum_{l=1}^{L} \gamma_l^{(v)} \left\| \mathbf{F} - \mathbf{P}_l^{(v)} \mathbf{F} \right\|_F^2, \tag{52}$$

subject to the discrete constraints on $\mathbf{F}$*.*

Proof of Proposition 1. For any feasible partition $\mathbf{G}_l^{(v)}$, each sample belongs to exactly one nonempty group. Therefore, the columns of $\mathbf{G}_l^{(v)}$ have disjoint supports, and

$$\mathbf{G}_l^{(v)\top} \mathbf{G}_l^{(v)} = \mathbf{D}_l^{(v)}. \tag{53}$$

By definition, $\mathbf{P}_l^{(v)} = \mathbf{G}_l^{(v)} \mathbf{D}_l^{(v)-1} \mathbf{G}_l^{(v)\top}$, which is symmetric. Moreover, using Eq. (53),

$$\begin{aligned} \mathbf{P}_l^{(v)2} &= \mathbf{G}_l^{(v)} \mathbf{D}_l^{(v)-1} \mathbf{G}_l^{(v)\top} \mathbf{G}_l^{(v)} \mathbf{D}_l^{(v)-1} \mathbf{G}_l^{(v)\top} \\ &= \mathbf{G}_l^{(v)} \mathbf{D}_l^{(v)-1} \mathbf{D}_l^{(v)} \mathbf{D}_l^{(v)-1} \mathbf{G}_l^{(v)\top} = \mathbf{P}_l^{(v)}. \end{aligned} \tag{54}$$

Thus, $\mathbf{P}_l^{(v)}$ is symmetric and idempotent, and hence is the orthogonal projection onto $\mathrm{col}(\mathbf{G}_l^{(v)})$. Accordingly,

$$\begin{aligned} \left\| \mathbf{F} - \mathbf{P}_l^{(v)} \mathbf{F} \right\|_F^2 &= \mathrm{Tr}\left[\mathbf{F}^\top (\mathbf{I} - \mathbf{P}_l^{(v)}) \mathbf{F}\right] \\ &= n - \mathrm{Tr}\left(\mathbf{F}^\top \mathbf{P}_l^{(v)} \mathbf{F}\right), \end{aligned} \tag{55}$$

where $\mathrm{Tr}(\mathbf{F}^\top \mathbf{F}) = n$ follows from the one-hot constraint on $\mathbf{F}$.

Using $\boldsymbol{\Theta}^{(v)} = \sum_{l=1}^{L} \gamma_l^{(v)} \mathbf{P}_l^{(v)}$ and $\sum_{l=1}^{L} \gamma_l^{(v)} = 1$, we obtain

$$\sum_{v=1}^{V} \mathrm{Tr}\left(\mathbf{F}^\top \boldsymbol{\Theta}^{(v)} \mathbf{F}\right) = Vn - \sum_{v=1}^{V} \sum_{l=1}^{L} \gamma_l^{(v)} \left\| \mathbf{F} - \mathbf{P}_l^{(v)} \mathbf{F} \right\|_F^2. \tag{56}$$

Thus, maximizing the multi-resolution hypergraph consistency term is equivalent to minimizing the weighted projection residual, which proves Proposition 1.

A.3 Derivation of Exact Reassignment Gains

We provide the derivations of the sufficient-statistic updates used for the exact reassignment gains in Section 4.

Reassignment of $\mathbf{G}_l^{(v)}$. For clarity, we omit the superscripts (v, l) whenever no ambiguity arises. Recall that

$$N_R = \mathrm{Tr}\left(\mathbf{G}^\top \bar{\mathbf{K}}^{(v)} \mathbf{G}\right), \qquad D_R = \mathrm{Tr}\left(\mathbf{G}^\top \bar{\mathbf{E}} \mathbf{G}\right). \tag{57}$$

Let $\mathbf{s}_q = \sum_{j: G_{jq}=1} \mathbf{b}_j^{(v)}$ and $n_q = \sum_j G_{jq}$. Using $\mathbf{K}^{(v)} = \mathbf{B}^{(v)} \mathbf{B}^{(v)\top}$, the numerator can be written as

$$N_R = \sum_q \|\mathbf{s}_q\|_2^2 - \left\| \mathbf{B}^{(v)} \right\|_F^2. \tag{58}$$

Suppose sample i is moved from group a to group b. Then

$$\mathbf{s}_a' = \mathbf{s}_a - \mathbf{b}_i^{(v)}, \qquad \mathbf{s}_b' = \mathbf{s}_b + \mathbf{b}_i^{(v)}. \tag{59}$$

Since all other group statistics remain unchanged,

$$\begin{aligned}\Delta N_R &= \left\|\mathbf{s}_a - \mathbf{b}_i^{(v)}\right\|_2^2 + \left\|\mathbf{s}_b + \mathbf{b}_i^{(v)}\right\|_2^2 - \|\mathbf{s}_a\|_2^2 - \|\mathbf{s}_b\|_2^2 \\ &= 2\left(\mathbf{b}_i^{(v)\top}\mathbf{s}_b - \mathbf{b}_i^{(v)\top}\mathbf{s}_a + \|\mathbf{b}_i^{(v)}\|_2^2\right).\end{aligned} \tag{60}$$

Similarly, because $\bar{\mathbf{E}} = \mathbf{1}_n\mathbf{1}_n^\top - \mathbf{I}_n$,

$$D_R = \sum_q n_q(n_q - 1). \tag{61}$$

After the reassignment, $n_a' = n_a - 1$ and $n_b' = n_b + 1$, and hence

$$\begin{aligned}\Delta D_R &= (n_a - 1)(n_a - 2) + (n_b + 1)n_b \\ &\quad - n_a(n_a - 1) - n_b(n_b - 1) \\ &= 2(n_b - n_a + 1).\end{aligned} \tag{62}$$

For the hypergraph consistency term, define $\mathbf{z}_q = \sum_{j:G_{jq}=1} \mathbf{f}_j$. Then

$$T_l^{(v)} = \sum_q \frac{\|\mathbf{z}_q\|_2^2}{n_q}. \tag{63}$$

The reassignment changes only groups a and b, with

$$\mathbf{z}_a' = \mathbf{z}_a - \mathbf{f}_i, \qquad \mathbf{z}_b' = \mathbf{z}_b + \mathbf{f}_i. \tag{64}$$

Therefore,

$$\begin{aligned}\Delta T_l^{(v)} = &\frac{\|\mathbf{z}_a - \mathbf{f}_i\|_2^2}{n_a - 1} + \frac{\|\mathbf{z}_b + \mathbf{f}_i\|_2^2}{n_b + 1} \\ &- \frac{\|\mathbf{z}_a\|_2^2}{n_a} - \frac{\|\mathbf{z}_b\|_2^2}{n_b}.\end{aligned} \tag{65}$$

Substituting Eqs. (60), (62), and (65) into the generic ratio variation in Eq. (13) yields the exact gain used for updating $\mathbf{G}_l^{(v)}$.

Reassignment of F. For the anchor-induced term of view v, let

$$A^{(v)} = \mathrm{Tr}\left(\mathbf{F}^\top\mathbf{K}^{(v)}\mathbf{F}\right) = \sum_{s=1}^{c}\left\|\mathbf{u}_s^{(v)}\right\|_2^2, \tag{66}$$

where $\mathbf{u}_s^{(v)} = \sum_{j:F_{js}=1}\mathbf{b}_j^{(v)}$. If sample i is moved from consensus cluster a to cluster b, then

$$\mathbf{u}_a^{(v)\prime} = \mathbf{u}_a^{(v)} - \mathbf{b}_i^{(v)}, \qquad \mathbf{u}_b^{(v)\prime} = \mathbf{u}_b^{(v)} + \mathbf{b}_i^{(v)}. \tag{67}$$

Thus,

$$\begin{aligned}\Delta A^{(v)} &= \left\|\mathbf{u}_a^{(v)} - \mathbf{b}_i^{(v)}\right\|_2^2 + \left\|\mathbf{u}_b^{(v)} + \mathbf{b}_i^{(v)}\right\|_2^2 \\ &\quad - \left\|\mathbf{u}_a^{(v)}\right\|_2^2 - \left\|\mathbf{u}_b^{(v)}\right\|_2^2 \\ &= 2\left(\mathbf{b}_i^{(v)\top}\mathbf{u}_b^{(v)} - \mathbf{b}_i^{(v)\top}\mathbf{u}_a^{(v)} + \|\mathbf{b}_i^{(v)}\|_2^2\right).\end{aligned} \tag{68}$$

For a fixed resolution-specific partition, let q denote the group containing sample i, and recall that

$$T_l^{(v)} = \sum_{q=1}^{k_l}\frac{\|\mathbf{C}_{l,q:}^{(v)}\|_2^2}{n_q}, \qquad \mathbf{C}_l^{(v)} = \mathbf{G}_l^{(v)\top}\mathbf{F}. \tag{69}$$

Only the a-th and b-th entries of $\mathbf{C}_{l,q:}^{(v)}$ change:

$$C_{l,qa}^{(v)\prime} = C_{l,qa}^{(v)} - 1, \qquad C_{l,qb}^{(v)\prime} = C_{l,qb}^{(v)} + 1. \tag{70}$$

Therefore,

$$\begin{aligned}\Delta T_l^{(v)} &= \frac{(C_{l,qa}^{(v)}-1)^2+(C_{l,qb}^{(v)}+1)^2-(C_{l,qa}^{(v)})^2-(C_{l,qb}^{(v)})^2}{n_q} \\ &= \frac{2\left(C_{l,qb}^{(v)}-C_{l,qa}^{(v)}+1\right)}{n_q}.\end{aligned} \tag{71}$$

Finally, let h_s denote the size of consensus cluster s. Since

$$D_F = \mathrm{Tr}\left(\mathbf{F}^\top \mathbf{E}\mathbf{F}\right) = \sum_{s=1}^{c} h_s^2, \tag{72}$$

the reassignment $a \to b$ gives

$$\begin{aligned}\Delta D_F &= (h_a-1)^2+(h_b+1)^2-h_a^2-h_b^2 \\ &= 2(h_b-h_a+1).\end{aligned} \tag{73}$$

Combining the view-specific changes gives

$$\Delta N_F = \sum_{v=1}^{V}\left[(1-\alpha)\Delta A^{(v)} + \alpha\sum_{l=1}^{L}\gamma_l^{(v)}\Delta T_l^{(v)}\right]. \tag{74}$$

Together with Eq. (73), substitution into Eq. (13) yields the exact consensus reassignment gain in Eq. (17).

A.4 Additional Matrix-Free Identities

Besides Eq. (18), the partition-quality terms can be evaluated without explicitly forming $\bar{\mathbf{K}}^{(v)}$ or $\bar{\mathbf{E}}$. For a hard partition $\mathbf{G}_l^{(v)}$,

$$\mathrm{Tr}\left(\mathbf{G}_l^{(v)\top}\bar{\mathbf{K}}^{(v)}\mathbf{G}_l^{(v)}\right) = \|\mathbf{G}_l^{(v)\top}\mathbf{B}^{(v)}\|_F^2 - \|\mathbf{B}^{(v)}\|_F^2, \tag{75}$$

and

$$\mathrm{Tr}\left(\mathbf{G}_l^{(v)\top}\bar{\mathbf{E}}\mathbf{G}_l^{(v)}\right) = \|\mathbf{G}_l^{(v)\top}\mathbf{1}_n\|_2^2 - n, \qquad D_F = \|\mathbf{F}^\top\mathbf{1}_n\|_2^2. \tag{76}$$

These identities, together with the group-level statistics in Appendix A.3, yield the matrix-free implementation used throughout DCoE.

A.5 Implementation Details

All features are normalized before constructing the sample-to-anchor relations. Specifically, ℓ_2 normalization is applied to NGs, while min-max normalization is used for all other datasets. All hyperparameters are selected through grid search. We search $\alpha \in \{0, 0.1, \ldots, 0.9\}$, $\eta \in \{0.5, 1, 2\}$, and $\lambda_\gamma \in \{10^{-3}, 10^{-2}, 10^{-1}\}$ for all datasets. The number of anchors is selected from dataset-specific candidate sets. For the multi-resolution partition construction, we use $\{3c, 5c, 7c\}$, $\{3c, 4c, 5c\}$, and $\{2c, 3c, 4c\}$ for small-, medium-, and large-scale datasets, respectively. The consensus partition $\mathbf{F}$ is initialized by applying MiniBatchKMeans to the concatenation of all preprocessed views. For each view and resolution, $\mathbf{G}_l^{(v)}$ is initialized by applying MiniBatchKMeans to the corresponding sparse anchor representation $\mathbf{B}^{(v)}$ with k_l clusters. The resolution weights are initialized uniformly as $\gamma_l^{(v)} = 1/L$.

A.6 Additional Ablation Results

To further examine the computational behavior of continuous relaxation, we compare the optimization time of *Relaxed-F* and *Relaxed-F, G*. Both variants use the same initialization and experimental protocol as in the ablation study.

As shown in Table 6, DCoE achieves lower optimization times than both relaxed variants across all eleven datasets. Moreover, *Relaxed-F, G* requires more optimization time than *Relaxed-F* across all datasets.

Table 6: Optimization time (seconds) of DCoE and its relaxed variants.

Method	MSRC	NGs	Handwritten	100Leaves	Scene	OutScene	UCI	ALOI	MNIST-10k	COIL20	YouTubeFace
Relaxed-F	0.48	0.19	3.49	8.95	4.47	1.98	1.55	80.84	6.53	2.69	406.37
Relaxed-F, G	0.55	0.40	5.60	45.06	16.10	5.05	2.77	448.80	14.57	5.47	562.63
DCoE	0.03	0.03	0.46	0.78	1.20	0.51	0.24	7.92	1.33	0.22	83.91

A.7 Detailed Verification of the Structural–Spectral Certificate

For the learned $\{\mathbf{G}_l^{(v)}\}$ and $\{\boldsymbol{\gamma}^{(v)}\}$, we evaluate

$$U_{\mathrm{cert}} = \frac{V + (c-1)\lambda_{\perp}(\mathbf{A})}{n}, \qquad \Delta_{\mathrm{cert}} = \min_i A_{ii} - U_{\mathrm{cert}}.$$

According to Theorem 2, $\Delta_{\mathrm{cert}} > 0$ certifies that every local maximizer of the faithful row-simplex consensus relaxation is one-hot.

Table 7: Empirical evaluation of the structural–spectral certificate. A positive Δ_{cert} certifies one-hot local maximizers of the faithful row-simplex relaxation.

	MSRC	NGs	Handwritten	100Leaves	Scene	OutScene	UCI	ALOI	MNIST-10k	COIL20	YouTubeFace
$\min_i A_{ii}$	0.3752	0.0581	0.0896	0.2295	0.0199	0.0376	0.0277	0.0728	0.0112	0.0748	0.0065
$\lambda_{\perp}(\mathbf{A})$	4.0793	1.6737	5.7534	2.5904	2.5026	3.2138	2.8469	3.7612	2.8681	2.9506	3.2117
U_{cert}	0.1404	0.0194	0.0289	0.1622	0.0085	0.0099	0.0143	0.0348	0.0029	0.0410	0.0010
Δ_{cert}	0.2348	0.0387	0.0607	0.0674	0.0114	0.0277	0.0134	0.0379	0.0084	0.0338	0.0055
Cert.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

As shown in Table 7, all evaluated datasets yield positive certificate margins, confirming that the sufficient condition in Theorem 2 is satisfied for the learned structures.

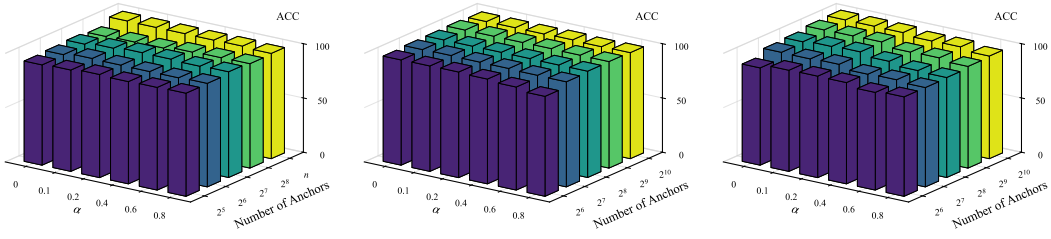
A.8 Additional Parameter Sensitivity and Convergence Results

This subsection provides supplementary sensitivity and convergence results. For anchor–α sensitivity and convergence, we report the eight datasets omitted from the main paper. Since the η–λ_γ plots are omitted from the main text, their results are reported here for all eleven datasets using the same experimental protocol.

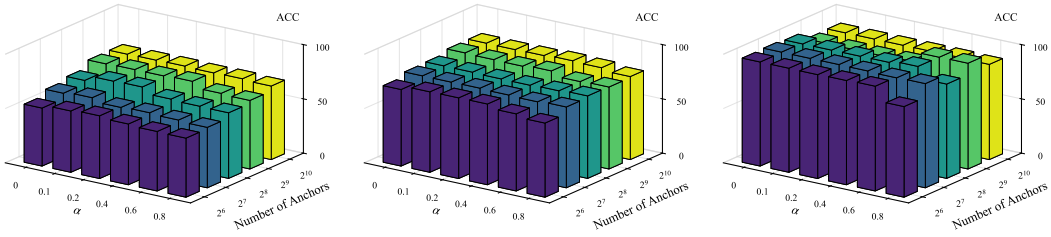
Sensitivity to the number of anchors and α. Figure 4 shows the results on the remaining eight datasets. The overall behavior is consistent with that in the main paper: DCoE remains effective across a broad range of anchor numbers and α values.

Sensitivity to η and λ_γ. Figure 5 shows that DCoE is generally stable with respect to the two regularization parameters across all eleven datasets.

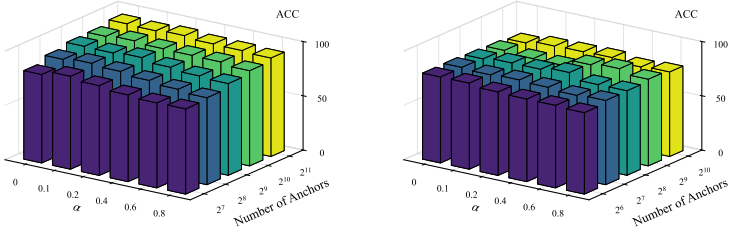
Additional convergence results. As shown in Fig. 6, the objective sequences on the remaining datasets exhibit rapid stabilization, consistent with the convergence behavior observed in the main paper.



(a) NGs (b) Handwritten (c) 100Leaves

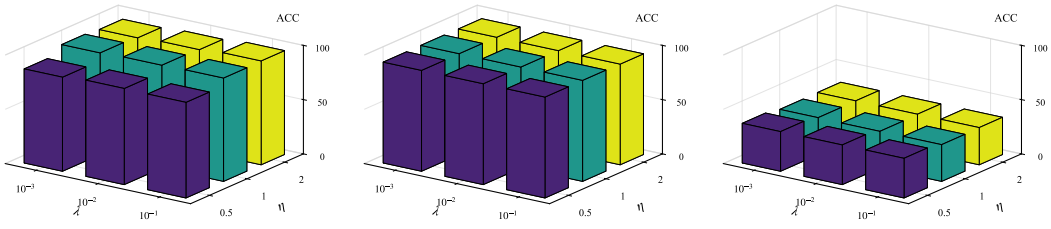


(d) Scene (e) OutScene (f) UCI

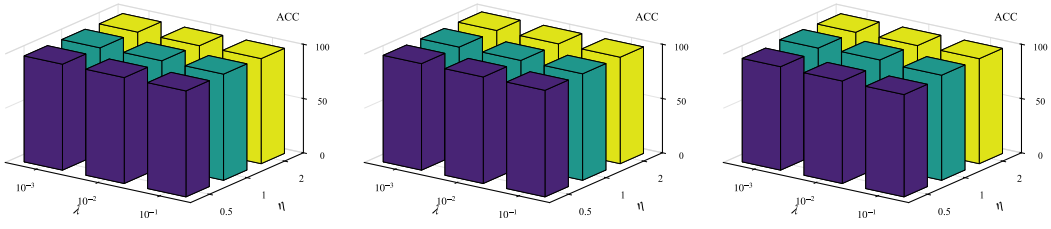


(g) ALOI (h) COIL20

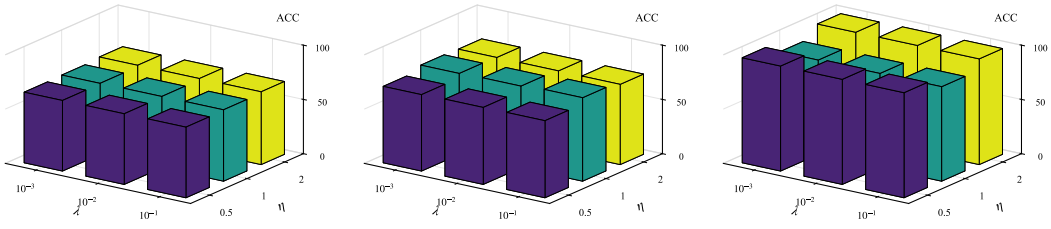
Figure 4: Additional sensitivity of DCoE to the number of anchors and α.



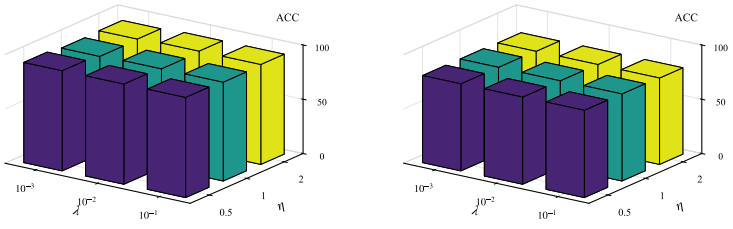
(a) MSRC (b) MNIST-10k (c) YouTubeFace



(d) NGs (e) Handwritten (f) 100Leaves

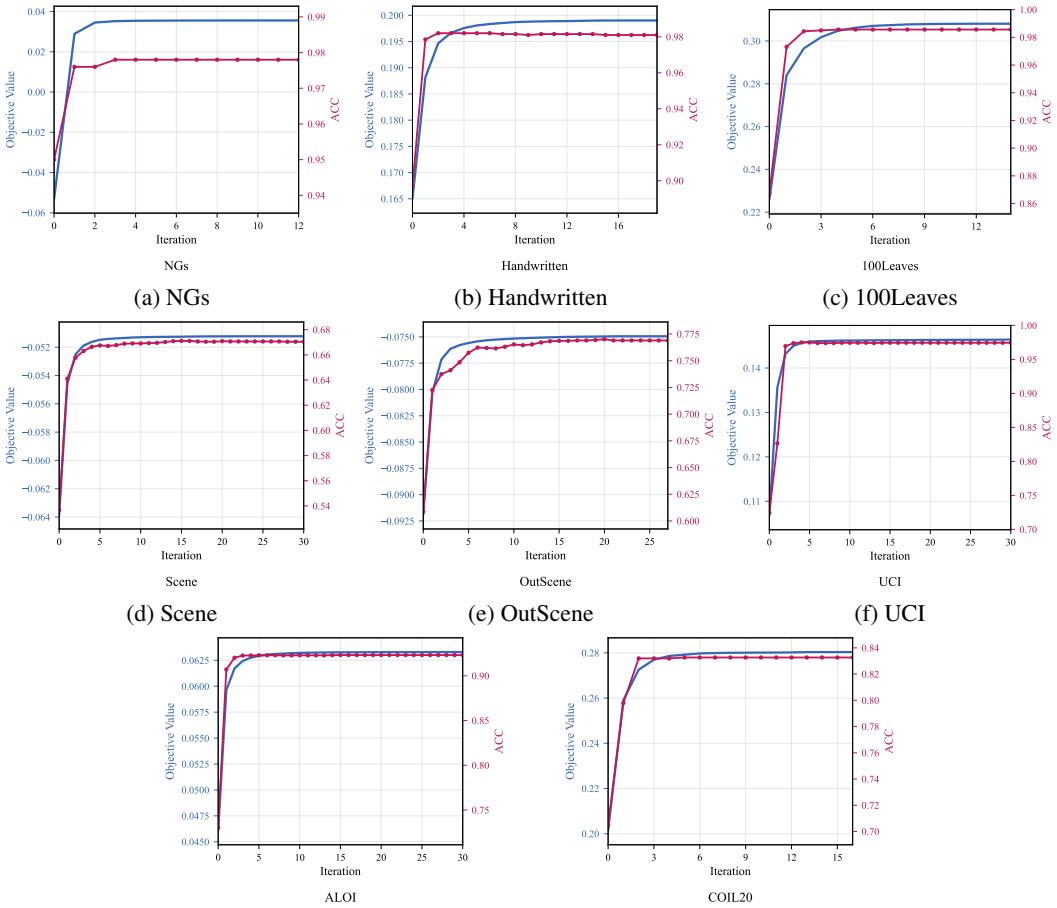


(g) Scene (h) OutScene (i) UCI



(j) ALOI (k) COIL20

Figure 5: Sensitivity of DCoE to η and λ_γ on all benchmark datasets.



(a) NGs (b) Handwritten (c) 100Leaves

(d) Scene (e) OutScene (f) UCI

(g) ALOI (h) COIL20

Figure 6: Additional convergence curves of DCoE on the remaining benchmark datasets.