DIFUMVC: DISTRIBUTION FUSION MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Multi-view clustering (MVC) must decide how to represent each view, how to fuse information across views, and how to cluster. Most MVC pipelines operate on sample relations, point prototypes, or latent codes. Some additionally align or transport distributions in selected representation or fusion operations, while their final clustering remains sample-level. We introduce DIFUMVC, which adopts a distributional perspective through all three stages. *Distributional representation* retains k components in each view and encodes each empirical distribution by a kernel mean embedding (KME). *Distribution fusion* uses shared reference-object identities to compare source-view points with target-view KMEs while preserving view-specific geometry. *Distribution-driven clustering* assembles all ordered view pairs into an $n \times M (M \leq kV^2)$ point-distribution bipartite graph and applies transfer cut on its compact distribution side. We prove that the induced cross-view kernel is positive semidefinite, show that point–distribution fusion is an expected cross-view kernel affinity, and derive conditional finite-sample bounds that propagate both KME sampling error and random-reference approximation error through the fused graph to its compact spectral subspace under fixed-rule and split-reference conditions. Benchmark experiments demonstrate the efficiency and effectiveness of our proposed DIFUMVC.

1 INTRODUCTION

Multi-view clustering (MVC) partitions n unlabeled objects observed through V aligned but heterogeneous feature spaces into k clusters (Xu et al., 2013). We view this task as three coupled decisions: *representation* chooses the structural unit retained from each view, *fusion* combines those units, and *clustering* converts the fused information into partitions (as shown in Figure 1).

Existing methods expose different units at these stages. Spectral MVC fuses $n \times n$ sample relations (Kumar et al., 2011); scalable anchor methods instead use point prototypes (Li et al., 2015; Kang et al., 2020; Li et al., 2022); and deep methods learn sample codes or relationships (Trosten et al., 2021; Li et al., 2026; Peng et al., 2026). GMAE learns sample codes and adversarially aligns their cross-view distributions (Zou et al., 2026); CorreGen infers soft cross-view correspondence distributions during fusion (Zhou et al., 2026); and OPTION uses conditional flow matching for missing-view representation together with transport-inspired structural fusion (Zhou & Gu, 2026). Their reported final clustering acts on learned sample representations (points) rather than explicit distribution nodes. HAMC, by contrast, uses hyperbolic radius as a confidence signal and consensus pseudo-labels (Wang et al., 2026), rather than constructing an explicit probability measure. The crucial observation is that these methods differ in what they explicitly construct and manipulate at each stage, not in what their embeddings may latently carry. It reveals an open question: *can distributions remain explicit through representation, fusion, and clustering?*

A cluster is a population with points independently and identically sampled from the same underlying distribution, and the distribution can be summarized via a kernel mean embedding (KME) by its average feature map (Smola et al., 2007; Muandet et al., 2017), retaining population information while supporting simple inner products. The challenge is to compare such summaries across heterogeneous views and use them for the final partition without rebuilding a dense sample graph.

DIFUMVC answers this question with a three-stage pipeline. **(I) Distributional representation** discovers up to k components per view with density-adaptive isolation features (Ting et al., 2018)

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.

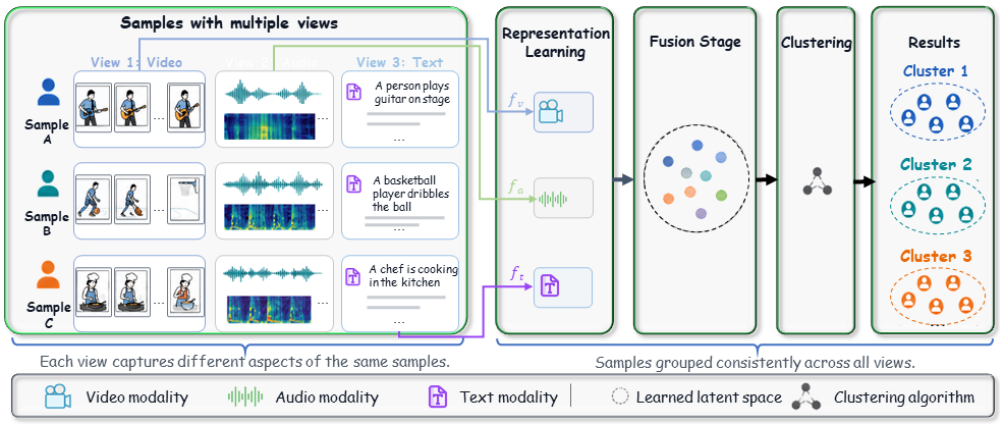


Figure 1: The MVC overview.

and represents each by an empirical KME. **(II) Distribution fusion** shares isolation reference *identities* across views but retains view-specific centers, radii, and metrics. Source-view points can therefore be compared with target-view KMEs, preserving all ordered within- and cross-view evidence. **(III) Distribution-driven clustering** concatenates these affinities into an $n \times M$ point–distribution graph, where $M \leq V^2 k$. Transfer cut Li et al. (2012) solves the spectral problem on its compact distribution-side nodes and transfers the embedding to samples. Distributions are thus explicit in the representatives, fusion edges, and compact graph that determines the labels.

Our contributions are:

- We formulate MVC as three distribution-explicit stages and realize them with per-view KMEs, shared-reference fusion, and compact-side distribution-driven clustering.
- We connect these stages through a KME fusion identity, conditional KME and random-reference concentration, affinity-ranking preservation, and explicit fused-graph and compact-subspace stability guarantees.
- We evaluate seven methods on 15 benchmarks using three clustering metrics, runtime, and cross-dataset significance tests to demonstrate the efficiency and effectiveness.

2 Related Work

We organize prior MVC methods by their representation, fusion, and final clustering objects.

Sample graphs and point anchors. Co-regularized spectral clustering couples view-specific sample graphs but retains dense eigensystem costs (Kumar et al., 2011). MVSC replaces sample graphs by sample–anchor bipartite graphs (Li et al., 2015); LMVSC learns compact subspace anchors (Kang et al., 2020); and SFMC fuses view-specific bipartite graphs without an extra scale parameter (Li et al., 2022). Their explicit graph nodes are individual anchors or point prototypes.

Kernel embeddings of distributions. A KME maps a probability law P to $\mu_P = \mathbb{E}_{X\sim P}[\phi(X)]$ in a reproducing-kernel Hilbert space (Smola et al., 2007; Sriperumbudur et al., 2010; Muandet et al., 2017). For characteristic kernels the population map can identify probability laws, while empirical means converge dimension-independently for bounded features. Isolation kernels construct data-dependent similarities through random partitions (Ting et al., 2018; Qin et al., 2019), and their distributional extension compares sets through empirical feature means (Ting et al., 2020). DIFUMVC uses these means explicitly across all three stages. We do *not* claim that one finite Monte Carlo feature realization is characteristic for every possible distribution; our guarantees concern its finite mean representation and explicitly separated target distributions.

Deep representation and correspondence models. CoMVC studies selective alignment of learned view representations (Trosten et al., 2021). The evaluated recent methods balance view optimization

Table 1: Three-stage comparison of representative MVC methods.

Criterion	CoregSC	MVSC	LMVSC	SFMC	CoMVC	BMvC	RAV	CorreGen	GMAE	HAMC	OPTION	**DiFuMVC**
Representation	1	2	2	2	1	1	1	1	1	1	3	3
Fusion	a	a	a	a	b	b	b	c	c	b	c	c
Scalable design	–	✓	✓	✓	B	B	B	B	✓	B	B	✓
Distribution stages	–	–	–	–	–	–	–	II	II	–	I II	I II III

Codes. 1/2/3: sample relation or code / point anchor / distribution-valued representation or flow; a/b/c: graph consensus / representation alignment / distribution alignment, correspondence, transport, or graph. C/I/U: complete/incomplete/unaligned; * denotes noisy correspondence. ✓/B/–: subquadratic-scale design / mini-batch / dense computation. I/II/III identify the stages that explicitly construct, align, transport, or cluster an empirical or probabilistic measure.

(BMvC) (Li et al., 2026), align sample relationships (RAV) (Peng et al., 2026), infer soft noisy-correspondence distributions (CorreGen) (Zhou et al., 2026), disentangle sample representations and adversarially align their cross-view distributions (GMAE) (Zou et al., 2026), use hyperbolic radius to model confidence (HAMC) (Wang et al., 2026), or combine conditional-flow imputation with transport-inspired latent-structure alignment (OPTION) (Zhou & Gu, 2026). Table 1 summarizes the representation, fusion, and clustering routes, generalized data settings, and scalability mechanisms.

3 Distribution Fusion Multi-View Clustering

Figure 2 details the three stages of DiFuMVC: per-view KMEs, ordered point–distribution fusion, and compact-side transfer cut on the distributional bipartite graph.

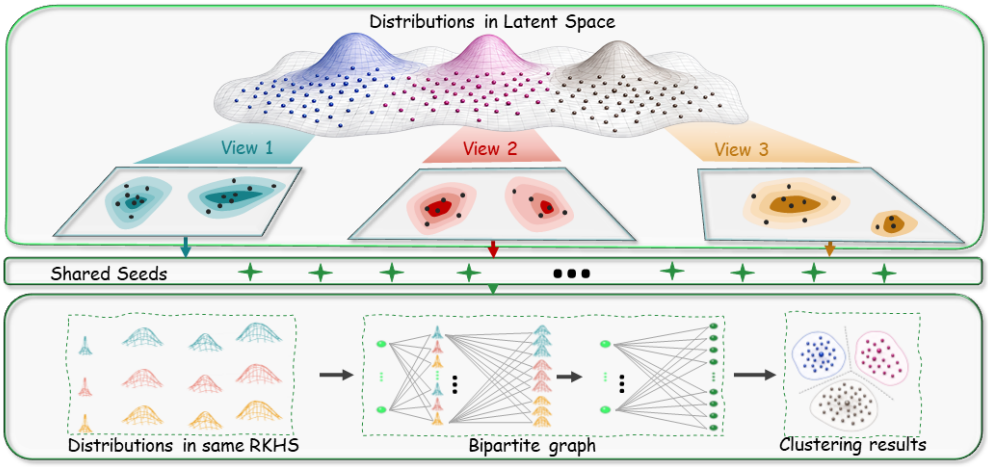


Figure 2: Overview of DiFuMVC.

3.1 Problem Setting and Three-Stage Objective

Let $\mathcal{X} = \{\boldsymbol{X}^{(v)}\}_{v=1}^{V}$ contain V complete, row-aligned views of n objects, with $\boldsymbol{X}^{(v)} \in \mathbb{R}^{n\times d_v}$. The ith rows $\{x_i^{(v)}\}_{v=1}^{V}$ describe the same object. Given the number of clusters k, MVC estimates one common partition without labels. In our formulation, Stage I produces the distribution set $\widehat{\mathcal{M}} = \{\widehat{\mu}_{vc} : v = 1,\dots,V,\ c = 1,\dots,k_v\}$, where $k_v = \min(c_v, k)$ and c_v is the number of connected components discovered in view v, and $A = \sum_{v=1}^{V} k_v \le Vk$. Stage II produces a

nonnegative bipartite matrix $\boldsymbol{B} \in \mathbb{R}_+^{n \times M}$, $M = VA \leq V^2 k$ (typically $M \ll n$), whose left nodes are objects and whose right nodes are view-conditioned distributions. We require $M \geq k$ so that the compact side supports a k-dimensional spectral embedding; a threshold violating this global condition is infeasible. Stage III maps this graph to labels $\boldsymbol{y} \in \{1, \ldots, k\}^n$ via transfer cut.

3.2 Stage I: Distributional representation

3.2.1 Shared-reference isolation coordinates

Let $R \geq 1$ and $2 \leq \psi \leq n$. For each random partition $r \in \{1, \ldots, R\}$, sample without replacement ψ object indices $I_r = \{i_{r1}, \ldots, i_{r\psi}\}$. Denote the realized partition by Ω_r and the full reference bank by $\mathcal{R} = (\Omega_1, \ldots, \Omega_R)$. The same index set I_r is used in every view. In view v, reference j is $z_{rj}^{(v)} = x_{i_{rj}}^{(v)}$, with nearest-reference radius $\rho_{rj}^{(v)} = \min_{\ell \neq j} d_v\left(z_{rj}^{(v)}, z_{r\ell}^{(v)}\right)$. The isolation feature activates reference j only when it is closest and the point lies inside its radius:

$$\phi_{rj}^{(v)}(x) = \mathbb{I}\left[j = \arg\min_{\ell} d_v(x, z_{r\ell}^{(v)}) \ \wedge \ d_v(x, z_{rj}^{(v)}) \leq \rho_{rj}^{(v)}\right]. \tag{1}$$

Ties use a deterministic rule. For an augmented point $\zeta = (v, x)$ and a realized partition Ω_r, write the corresponding one-partition feature vector as $\phi_{\Omega_r,\psi}(\zeta) = [\phi_{r1}^{(v)}(x), \ldots, \phi_{r\psi}^{(v)}(x)]^\top \in \{0,1\}^\psi$. Concatenate the R blocks and normalize, $\Psi_{\mathcal{R}}(v, x) = R^{-1/2}[\phi_{\Omega_1,\psi}((v,x)); \ldots; \phi_{\Omega_R,\psi}((v,x))] \in \mathbb{R}^{R\psi}$. Every partition activates at most one coordinate, so $\|\Psi_{\mathcal{R}}(v,x)\|_2 \leq 1$. Only coordinate identities are shared; d_v, $z_{rj}^{(v)}$, and $\rho_{rj}^{(v)}$ remain view-specific.

Consider the disjoint union $\mathcal{Z} = \bigsqcup_v \{v\} \times \mathcal{X}_v$ and define $\kappa_{\mathcal{R}}((p,x),(q,z)) = \langle \Psi_{\mathcal{R}}(p,x), \Psi_{\mathcal{R}}(q,z) \rangle$.

Theorem 1 (Valid cross-view kernel)**.** *Conditioned on the sampled references $\mathcal{R}$, $\kappa_{\mathcal{R}}$ is a symmetric positive-semidefinite kernel on $\mathcal{Z}$, with $0 \leq \kappa_{\mathcal{R}}(\zeta, \zeta') \leq 1$ for every $\zeta, \zeta' \in \mathcal{Z}$.*

Proof. For arbitrary $\zeta_1, \ldots, \zeta_N \in \mathcal{Z}$ and $a_1, \ldots, a_N \in \mathbb{R}$, $\sum_{ij} a_i a_j \kappa_{\mathcal{R}}(\zeta_i, \zeta_j) = \|\sum_i a_i \Psi_{\mathcal{R}}(\zeta_i)\|_2^2 \geq 0$. The range bound follows from nonnegative feature coordinates, Cauchy–Schwarz, and $\|\Psi_{\mathcal{R}}(\zeta)\|_2 \leq 1$. □

This validates the kernel interface between Stages I and II. Appendix D.3 also quantifies its finite-pair Monte Carlo approximation error as the number of independent reference partitions grows. Each fixed value of ψ defines a different population kernel: increasing R estimates that fixed-ψ kernel more accurately, whereas changing ψ changes the target kernel itself.

3.2.2 Estimating k distributions in each view

Distribution discovery remains view-specific, preserving heterogeneous geometry. Choose one shared subset $\mathcal{S}$ of s object identities. Within view v, connect two subset objects when their normalized isolation similarity exceeds a relative threshold:

$$A_{ij}^{(v)} = \mathbb{I}\left[\langle \Psi_{\mathcal{R}}(v, x_i), \Psi_{\mathcal{R}}(v, x_j) \rangle \geq \tau \, \kappa_{\max}^{(v)}\right], \quad i, j \in \mathcal{S}.$$

Let c_v be the number of connected components found in view v and $k_v = \min(c_v, k)$. We retain the k_v largest components and assign each remaining object to the corresponding component to obtain the distributions via kernel mean embedding.

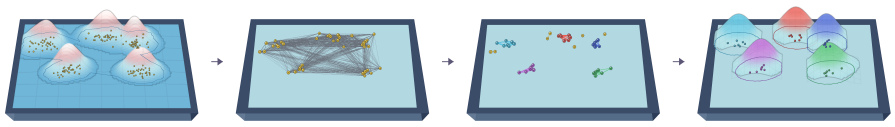


Figure 3: Estimating k_v distributions in each view. In each view, a thresholded similarity graph first exposes coherent components; the largest $k_v = \min(c_v, k)$ components are represented by empirical KMEs that average all assigned member features rather than selecting point anchors.

Figure 3 illustrates this component-selection and KME-estimation process. Let $C_c^{(v)}$ be the final member set of component c in view v. Its empirical distribution representative is

$$\widehat{\mu}_{vc} = \frac{1}{|C_c^{(v)}|} \sum_{i \in C_c^{(v)}} \Psi_{\mathcal{R}}(v, x_i^{(v)}), \qquad c = 1, \ldots, k_v.$$

Thus view v contributes k_v KMEs and the method estimates $A = \sum_v k_v \leq Vk$ view-specific distributions in total. Unlike a point anchor, $\widehat{\mu}_{vc}$ uses all assigned members and records their average occupancy of locally adaptive isolation cells. These KMEs are the output of Stage I.

3.3 Stage II: Distribution fusion

Stage II uses the Stage-I KMEs directly as fusion targets. Under a fixed component rule, let P_{qc} denote target component (q, c) and $\mu_{qc} = \mathbb{E}_{Y \sim P_{qc}} \Psi_{\mathcal{R}}(q, Y)$. For source view p and target view q, define the population point–distribution affinity and its empirical version as

$$a^*_{p \to (q,c)}(x) = \langle \Psi_{\mathcal{R}}(p, x), \mu_{qc} \rangle, \quad \widehat{a}_{p \to (q,c)}(x) = \langle \Psi_{\mathcal{R}}(p, x), \widehat{\mu}_{qc} \rangle.$$

Proposition 1 (KME fusion identity)**.** *Condition on the realized reference map $\mathcal{R}$. If $Y \sim P_{qc}$, then*

$$a^*_{p \to (q,c)}(x) = \mathbb{E}_Y \, \kappa_{\mathcal{R}}((p, x), (q, Y)), \quad \widehat{a}_{p \to (q,c)}(x) = \frac{1}{|C_c^{(q)}|} \sum_{j \in C_c^{(q)}} \kappa_{\mathcal{R}}((p, x), (q, x_j^{(q)})). \quad (2)$$

More explicitly, let $\mathbf{\Phi}_v$ contain $\Psi_{\mathcal{R}}(v, x_i^{(v)})^\top$ in row i, and let $(\boldsymbol{G}_q)_{jc} = \mathbb{I}[j \in C_c^{(q)}]/|C_c^{(q)}|$ for nonempty components. The complete raw block satisfies: $\boldsymbol{S}^{p \to q} = \mathbf{\Phi}_p \mathbf{\Phi}_q^\top \boldsymbol{G}_q = \boldsymbol{K}^{pq} \boldsymbol{G}_q$, $\boldsymbol{K}^{pq} = \mathbf{\Phi}_p \mathbf{\Phi}_q^\top$. Thus each fusion edge is the empirical expectation of the cross view kernel over a target distribution, and the factorization avoids materializing the dense $n \times n$ matrix $\boldsymbol{K}^{pq}$.

Proof. Boundedness permits exchanging the finite-dimensional inner product and expectation, so $\langle \Psi_{\mathcal{R}}(p, x), \mu_{qc} \rangle = \mathbb{E}_Y \langle \Psi_{\mathcal{R}}(p, x), \Psi_{\mathcal{R}}(q, Y) \rangle$. Substituting the empirical mean proves Equation 2. The cth column of $\mathbf{\Phi}_q^\top \boldsymbol{G}_q$ is exactly $\widehat{\mu}_{qc}$, proving the matrix identity. □

The resulting k_q empirical values for all n samples form $\boldsymbol{S}^{p \to q} \in \mathbb{R}_+^{n \times k_q}$. The implementation stores the unnormalized binary feature mean, so its matrix product is $R\boldsymbol{S}^{p \to q}$. Its normalized block is exactly the following expression after defining $\epsilon = \epsilon_{\text{code}}/R$; the common factor therefore changes only the numerical regularizer:

$$\widehat{\boldsymbol{S}}^{p \to q}_{i:} = \alpha_q \frac{\boldsymbol{S}^{p \to q}_{i:}}{\|\boldsymbol{S}^{p \to q}_{i:}\|_2 + \epsilon}. \quad (3)$$

The default uses equal nonnegative target-view weights α_q. Finally,

$$\boldsymbol{B} = \underset{p=1,\ldots,V;\ q=1,\ldots,V}{\text{concat}} \widehat{\boldsymbol{S}}^{p \to q} \in \mathbb{R}_+^{n \times M}, \qquad M = V \sum_{q=1}^{V} k_q = VA \leq V^2 k.$$

The right node (p, q, c) is distribution c estimated in target view q, observed through the sample coordinates of source view p. Keeping ordered pairs preserves directional cross-view evidence; their concatenation $\boldsymbol{B}$ is the output of Stage II.

3.4 Stage III: Distribution-driven clustering

Stage III applies transfer cut (Li et al., 2012) to the distribution nodes, avoiding a sample-side $n \times n$ diagonalization. Let

$$\boldsymbol{D}_x = \text{diag}(\boldsymbol{B}\mathbf{1}), \quad \boldsymbol{W}_y = \boldsymbol{B}^\top \boldsymbol{D}_x^{-1} \boldsymbol{B}, \quad \boldsymbol{D}_y = \text{diag}(\boldsymbol{W}_y \mathbf{1}). \quad (4)$$

Let $\boldsymbol{P} = \boldsymbol{D}_x^{-1/2} \boldsymbol{B} \boldsymbol{D}_y^{-1/2}$ denote the normalized point–distribution operator. Under this positive-degree assumption, diagonalize only $\widetilde{\boldsymbol{W}}_y = \boldsymbol{D}_y^{-1/2} \boldsymbol{W}_y \boldsymbol{D}_y^{-1/2} \in \mathbb{R}^{M \times M}$. If $\text{rank}(\boldsymbol{P}) \geq k$ and

$\boldsymbol{U} \in \mathbb{R}^{M\times k}$ contains top eigenvectors associated with positive eigenvalues, transfer them to samples by

$$\boldsymbol{H} = \boldsymbol{D}_x^{-1}\boldsymbol{B}\boldsymbol{D}_y^{-1/2}\boldsymbol{U}. \tag{5}$$

Row normalization and k-means (Lloyd, 1982) produce the final labels after the distribution-side subspace is transferred to samples. Appendix G proves that the small-side eigensystem constructs a basis spanning the corresponding top-k positive-eigenvalue sample-side normalized spectral subspace without an $n \times n$ eigendecomposition. It proves that the compact operator is a positive-semidefinite contraction with unit top eigenvalue, and that, under cutoff eigengap or equivariant tied-subspace selection, distribution-node, view, and component reindexing preserve transferred geometry.

3.5 Guarantees across the three stages

The guarantees mirror the pipeline: KME concentration controls an idealized Stage-I estimator, edge stability Stage II, and graph perturbation Stage III. They analyze an idealized conditional estimator in which the reference map and component rule are fixed independently of the KME-estimation samples, and observations within each nonempty component are conditionally i.i.d. This is exact for an independent pilot split or for new i.i.d. observations assigned by a fixed rule. For the finite fused graph we additionally fix the n source-query objects. Merely conditioning on references and components learned from those same observations does not preserve these sampling assumptions; that setting needs a data-dependent stability analysis.

Theorem 2 (Conditional empirical-KME concentration)**.** *Let $X_1, \dots, X_m$ be conditionally i.i.d. from a fixed cluster distribution P, let $\mu_P = \mathbb{E}_P \Psi_{\mathcal{R}}(v, X)$, and let $\widehat{\mu}_P = m^{-1}\sum_{\ell=1}^{m} \Psi_{\mathcal{R}}(v, X_\ell)$. For any $\delta \in (0,1)$, with conditional probability at least $1-\delta$,*

$$\|\widehat{\mu}_P - \mu_P\|_2 \le \frac{1+\sqrt{2\log(1/\delta)}}{\sqrt{m}}. \tag{6}$$

The proof, given in Appendix D, combines the second-moment bound with bounded differences (McDiarmid, 1989). A union bound replaces δ by δ/A to hold simultaneously for all view–cluster KMEs. It therefore controls the empirical distribution of the pilot-split or fixed-rule Stage-I estimator under the stated conditional-i.i.d. regime, not the same-data component-discovery procedure.

Corollary 1 (Edge, separation, and affinity-ranking stability)**.** *On the event in Theorem 2, for every source view p and every $x \in \mathcal{X}_p$,*

$$|\widehat{a}_{p\to(q,c)}(x) - a^*_{p\to(q,c)}(x)| \le \|\widehat{\mu}_{qc} - \mu_{qc}\|_2. \tag{7}$$

Moreover, for two target distributions a, b estimated with errors ϵ_a, ϵ_b, $\big|\|\widehat{\mu}_a - \widehat{\mu}_b\|_2 - \|\mu_a - \mu_b\|_2\big| \le \epsilon_a + \epsilon_b$. If, for a fixed (p, q, x), component c^* satisfies $a^*_{p\to(q,c^*)}(x) - a^*_{p\to(q,c)}(x) > \epsilon_{c^*} + \epsilon_c$ for every $c \neq c^*$, then c^* also uniquely maximizes the empirical affinities. Multiplication by the positive common scalar in Equation 3 preserves this within-block ordering when $\alpha_q > 0$.

Hence a finite family with positive population KME separation $\Delta = \min_{a\neq b} \|\mu_a - \mu_b\|_2 > 0$ remains separated whenever $\max_a \epsilon_a < \Delta/4$. The separation statement concerns Stage-I representations, while Equation 7 transfers the same estimation error to every Stage-II fusion edge. The ranking statement is local to one ordered block; neither statement claims that a component is a ground truth.

For the remaining propagation, Appendix C states the complete graph and spectral results; here we retain their rates. Let $m_{\min}$ be the smallest retained component size and assume the block-normalization denominators are uniformly positive. Simultaneous KME concentration gives, up to the displayed logarithmic factor and fixed block constants,

$$\|\widehat{\boldsymbol{B}} - \boldsymbol{B}^*\|_2 = O\left(V\sqrt{\frac{nk}{m_{\min}}}[1+\sqrt{\log(A/\delta)}]\right).$$

If the component rule is fixed independently and a fresh bank of R evaluation partitions is used, the joint reference–KME result instead gives

$$\|\widehat{\boldsymbol{B}}_{R,m,\psi} - \boldsymbol{B}_{\infty,\psi}\|_2 = O\left(V\sqrt{nk}\left[\frac{1+\sqrt{\log(A/\delta)}}{\sqrt{m_{\min}}} + \sqrt{\frac{\log(nVA/\delta)}{R}}\right]\right).$$

Thus, at a fixed resolution ψ, reference Monte Carlo error decays as $R^{-1/2}$ while KME sampling error decays as $m_{\min}^{-1/2}$. If the infinite-reference affinity margin is $\Delta_\psi > 0$, the same event preserves each finite-query affinity ranking once the combined error is below $\Delta_\psi/2$ (Corollary 2). Increasing R therefore improves the approximation guarantee, but need not monotonically improve a realized clustering score. Changing ψ changes the limiting kernel, graph degrees, separation margin, and eigengap, so no monotone ordering across resolutions follows.

Finally, if the population distribution-side operator has positive degrees and eigengap γ, its top-k subspace obeys the perturbation form: $\|\sin\Theta(\widehat{\boldsymbol{U}}, \boldsymbol{U}^*)\|_F \leq \frac{2\sqrt{k}\, L_B}{\gamma}\|\widehat{\boldsymbol{B}} - \boldsymbol{B}^*\|_2$, where L_B is the explicit degree-conditioning factor in Corollary 3. This closes the chain from Stage-I KME estimation through Stage-II graph fusion to the Stage-III compact spectral subspace. The guarantee stops before row normalization and final k-means clustering.

Complexity. Let $D = \sum_v d_v$, $A = \sum_v k_v$, and $M = VA \leq V^2 k$. Stage-I reference-feature construction and capped component discovery cost $O(R\psi^2 D + nR\psi D + Vs^2 R)$. Stage-II sparse point–distribution fusion costs $O(nV^2 kR)$. Stage-III transfer cut costs $O(nM^2 + M^3)$ because its eigensystem is formed on the $M(\ll n)$ distribution-side nodes rather than the n samples. Memory is $O(nM + M^2 + Vs^2 + nVR)$ when sparse features are retained. For fixed V, k, R, ψ, s, and feature dimensions, dependence on n is linear. Because $s = \min(n, 10{,}000)$, however, component discovery is $O(Vn^2 R)$ when $n \leq 10{,}000$ and becomes capped only for larger datasets [1].

4 Experiments

Datasets and metrics. We use 15 benchmark datasets with 400–86,796 samples, 2–7 views, and 7–101 clusters. They cover faces, objects, handwritten digits, generic images, and food. Appendix I reports sizes and source provenance. We measure normalized mutual information (NMI), adjusted Rand index (ARI) (Hubert & Arabie, 1985), and clustering F1, all in percent.

Competitors. We compare with BMvC (Li et al., 2026), RAV (Peng et al., 2026), CorreGen (Zhou et al., 2026), GMAE (Zou et al., 2026), HAMC (Wang et al., 2026), and OPTION (Zhou & Gu, 2026). GMAE has no valid CIFAR10 result; its score average uses 14 datasets.

All experiments were conducted on a server running Ubuntu 20.04.6 LTS. The server was equipped with a total of 72 physical cores and 144 logical threads, and 1 TB of system memory; The GPU is an NVIDIA RTX 5880 with 48 GB of memory.

4.1 Main Results in terms of NMI and Runtime

Table 2 shows the clustering results. DiFuMVC obtains 73.08 NMI on average, improving over the strongest score-average baseline, BMvC, by 3.85 points, and has the best common-block average rank of 1.57. The gains are not universal: BMvC leads on ORL, Flower17, both digit-feature datasets, and ALOI; HAMC leads CIFAR10, and CorreGen leads Food101. The lower block reports paired runtime ratios and places every absolute-time bar to the right of its number. DiFuMVC is fastest on all 15 datasets. Its geometric-mean paired speedups range from $6.53\times$ over GMAE to $240.25\times$ over RAV. These results demonstrate that our proposed method DiFuMVC outperforms recent methods in terms of both effectiveness and efficiency in multi-view clustering. Figure 4 reports Friedman–Nemenyi test results. Nemenyi separates DiFuMVC from RAV, GMAE, HAMC, and OPTION.

4.2 Spatially Resolved Transcriptomics

Following the BMvC spatial-transcriptomics setting (Li et al., 2026), we additionally report human dorsolateral prefrontal cortex (DLPFC) slice 151669, human breast cancer (HBC), and mouse brain anterior (MBA) sections. We reproduce the released BMvC-ST preprocessing: 3,000 highly variable genes define a 15-nearest-neighbor transcriptomic graph, spot coordinates define a spatial-radius graph, and deterministic row-normalized 128-dimensional truncated-SVD embeddings of the two graphs. Table 3 shows that DiFuMVC is competitive on all three datasets. On DLPFC, it remains

[1] DiFuMVC is efficient when $n \leq 10,000$ and only takes 16.5 seconds on a dataset with $n = 10,000$.

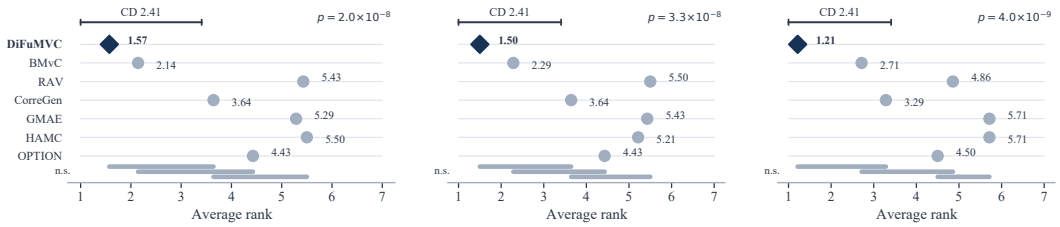
Table 2: NMI and runtime comparison. The upper block reports NMI, only the best mean is bold and only the second-best mean is underlined. The lower block reports $t_{\text{method}}/t_{\text{DiFuMVC}}$; bars to the right of the numbers encode absolute wall-clock seconds on a shared $\log_{10}$ scale.

Dataset	**BMvC** (PAMI26)	**RAV** (ICLR26)	**CorreGen** (ICLR26)	**GMAE** (PAMI26)	**HAMC** (ICML26)	**OPTION** (ICML26)	**DiFuMVC** (Ours)
Normalized mutual information (%)							
ORL	**94.95**$_{\pm 0.76}$	63.06$_{\pm 1.31}$	85.65$_{\pm 0.90}$	56.41$_{\pm 4.12}$	48.60$_{\pm 4.89}$	81.98$_{\pm 3.05}$	93.65$_{\pm 0.44}$
YaleB	29.86$_{\pm 3.34}$	11.24$_{\pm 0.82}$	26.84$_{\pm 3.15}$	21.62$_{\pm 3.02}$	3.55$_{\pm 1.23}$	7.22$_{\pm 5.22}$	**59.61**$_{\pm 1.43}$
Flw17	**63.30**$_{\pm 0.78}$	29.12$_{\pm 15.95}$	41.61$_{\pm 0.78}$	30.05$_{\pm 14.43}$	29.71$_{\pm 3.06}$	45.61$_{\pm 2.11}$	58.81$_{\pm 0.36}$
COIL20	85.10$_{\pm 0.78}$	85.50$_{\pm 1.93}$	83.92$_{\pm 0.66}$	84.47$_{\pm 1.73}$	70.24$_{\pm 4.80}$	79.88$_{\pm 1.42}$	**99.21**$_{\pm 0.09}$
Cal7	60.38$_{\pm 1.08}$	52.35$_{\pm 2.03}$	40.43$_{\pm 2.20}$	60.25$_{\pm 2.91}$	50.85$_{\pm 11.64}$	60.03$_{\pm 2.39}$	**66.51**$_{\pm 1.52}$
Leaves	95.87$_{\pm 0.22}$	80.66$_{\pm 0.69}$	92.23$_{\pm 0.25}$	82.73$_{\pm 0.86}$	84.81$_{\pm 1.26}$	86.21$_{\pm 1.95}$	**98.29**$_{\pm 0.11}$
Mfeat	**95.60**$_{\pm 0.55}$	85.69$_{\pm 4.97}$	86.98$_{\pm 1.44}$	91.81$_{\pm 3.09}$	82.35$_{\pm 2.37}$	91.94$_{\pm 1.73}$	94.59$_{\pm 0.30}$
Digits	**96.19**$_{\pm 0.41}$	81.49$_{\pm 5.83}$	86.67$_{\pm 1.36}$	92.88$_{\pm 2.19}$	82.32$_{\pm 1.70}$	91.34$_{\pm 2.62}$	94.59$_{\pm 0.30}$
NTU12	84.29$_{\pm 0.35}$	76.51$_{\pm 0.74}$	86.75$_{\pm 0.31}$	76.25$_{\pm 0.66}$	81.46$_{\pm 1.83}$	78.65$_{\pm 0.64}$	**87.51**$_{\pm 0.29}$
MNIST	70.13$_{\pm 1.88}$	74.53$_{\pm 0.19}$	72.46$_{\pm 0.11}$	66.57$_{\pm 1.65}$	75.55$_{\pm 0.15}$	74.54$_{\pm 0.41}$	**77.00**$_{\pm 0.61}$
Animal	76.24$_{\pm 0.31}$	51.49$_{\pm 0.40}$	69.38$_{\pm 0.34}$	63.24$_{\pm 2.45}$	65.58$_{\pm 0.67}$	50.19$_{\pm 4.24}$	**76.44**$_{\pm 0.38}$
ALOI	**92.70**$_{\pm 0.71}$	88.17$_{\pm 0.74}$	89.98$_{\pm 0.23}$	70.64$_{\pm 2.03}$	82.70$_{\pm 0.79}$	89.71$_{\pm 0.42}$	89.37$_{\pm 0.14}$
VGG50	16.02$_{\pm 0.45}$	13.70$_{\pm 0.23}$	14.60$_{\pm 0.28}$	7.10$_{\pm 1.03}$	10.70$_{\pm 1.04}$	10.26$_{\pm 0.65}$	**21.35**$_{\pm 0.27}$
CIFAR10	25.73$_{\pm 0.46}$	24.32$_{\pm 1.38}$	13.47$_{\pm 0.56}$	–	**26.96**$_{\pm 0.59}$	23.20$_{\pm 1.47}$	26.48$_{\pm 0.43}$
Food101	52.15$_{\pm 0.37}$	4.88$_{\pm 0.49}$	**53.71**$_{\pm 0.15}$	10.79$_{\pm 4.83}$	13.31$_{\pm 1.62}$	26.08$_{\pm 18.26}$	52.84$_{\pm 0.19}$
Avg. score ↑	69.23	54.85	62.98	58.20	53.91	59.79	**73.08**
Avg. rank ↓	2.14	5.43	3.64	5.29	5.50	4.43	**1.57**
Runtime and speedup							
ORL	7.2	31.0	78.3	2.9	6.6	13.4	**1.0**
YaleB	4.7	23.5	45.2	1.3	5.2	5.9	**1.0**
Flw17	17.2	197.0	103.6	8.1	34.0	25.2	**1.0**
COIL20	6.8	37.4	53.7	1.3	7.5	6.5	**1.0**
Cal7	119.1	751.4	373.9	27.5	133.5	88.5	**1.0**
Leaves	34.2	678.2	494.4	11.1	72.5	53.6	**1.0**
Mfeat	27.4	361.8	210.1	11.6	67.8	44.6	**1.0**
Digits	42.7	688.4	336.7	17.4	110.8	74.3	**1.0**
NTU12	38.2	747.1	1063.5	16.6	117.5	103.7	**1.0**
MNIST	126.4	489.7	234.7	9.9	91.6	51.6	**1.0**
Animal	58.3	182.1	181.5	3.6	41.9	28.6	**1.0**
ALOI	20.2	245.3	48.2	2.2	16.2	14.8	**1.0**
VGG50	28.1	153.7	79.8	3.1	28.2	20.7	**1.0**
CIFAR10	46.4	163.4	32.4	–	31.4	19.5	**1.0**
Food101	147.7	1830.5	467.4	25.5	178.2	148.2	**1.0**
GM (×)	**30.7**	**240.2**	**154.5**	**6.5**$^{(14)}$	**39.0**	**31.3**	**1.0**

Table 3: NMI comparison on spatially resolved transcriptomics datasets.

Dataset	**BMvC** (PAMI26)	**RAV** (ICLR26)	**CorreGen** (ICLR26)	**GMAE** (PAMI26)	**HAMC** (ICML26)	**OPTION** (ICML26)	**DiFuMVC** (Ours)
DLPFC	27.49$_{\pm 4.10}$	43.88$_{\pm 2.38}$	37.74$_{\pm 1.63}$	34.83$_{\pm 10.52}$	9.69$_{\pm 9.09}$	**52.65**$_{\pm 0.73}$	48.07$_{\pm 1.80}$
HBC	67.17$_{\pm 1.10}$	63.23$_{\pm 1.33}$	62.05$_{\pm 1.02}$	59.25$_{\pm 1.08}$	49.89$_{\pm 17.28}$	64.40$_{\pm 0.40}$	**67.87**$_{\pm 1.22}$
MBA	**70.88**$_{\pm 0.31}$	69.15$_{\pm 0.50}$	69.88$_{\pm 0.37}$	59.57$_{\pm 1.57}$	65.91$_{\pm 5.35}$	68.29$_{\pm 0.51}$	**70.88**$_{\pm 0.43}$

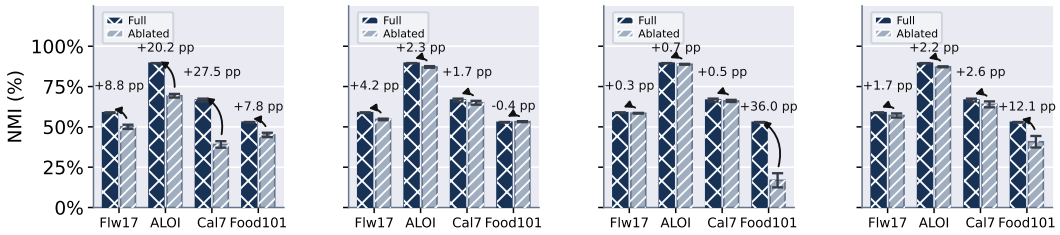
second behind OPTION. On HBC, it obtains the highest mean NMI, 67.87, exceeding BMvC by 0.70 points. On MBA, it reaches 70.88 NMI, the same as BMvC. Complete ARI and Hungarian-aligned macro-F1 results are reported in Appendix Table 7. Appendix J gives the spatial maps.



(a) NMI (b) ARI (c) F1

Figure 4: Friedman–Nemenyi analysis. (a–c) Mean ranks on 14 complete datasets; lower is better. Each panel gives the Friedman p-value and Nemenyi CD of 2.41 at $\alpha = 0.05$ (Demšar, 2006).

4.3 Ablation Study



Stage I (a) Point prototype
Stage II (b) Self-view only
Stage II (c) Cross-view only
Stage III (d) Direct k-means

Figure 5: Paired ablations in the three-stage pipeline. Each panel reports NMI (mean $\pm$ standard deviation over five seeds) for Full DiFuMVC and one ablated variant: **(a)** replace Stage-I distribution KMEs by point prototypes; **(b)** retain only self-view blocks in Stage II; **(c)** retain only cross-view blocks in Stage II; and **(d)** replace Stage-III transfer cut by direct k-means.

Figure 5 shows that replacing distribution KMEs by single-point prototypes decreases NMI by 7.8–27.5 points on all four datasets. Thus, Stage I benefits from representing a component as an empirical distribution rather than one representative sample. Keeping only self-view blocks decreases NMI by 1.7–4.2 points on Flower17, ALOI, and Caltech7; keeping only cross-view blocks decreases it by 0.3–0.7 points. Their joint use is therefore complementary on these datasets. Food101 is intentionally different: self-view only is 0.38 NMI and 0.66 ARI points above Full, while Full is 0.26 F1 points higher. This is because only one view contains useful information, while the small differences between the samples in the other two views introduce noise. Self-view only removes those transports and lets the informative within-view structure dominate. This small, metric-dependent reversal is therefore reasonable under severe view-resolution imbalance, rather than a contradiction of the fusion mechanism. The 36.0-point loss of cross-view only also shows that those low-resolution cross-view blocks are insufficient by themselves. Finally, replacing transfer cut by direct k-means lowers NMI by 1.7–12.1 points. Appendix L reports the paired ARI and F1 results.

5 Conclusion

DiFuMVC keeps distributions explicit through the complete pipeline: Stage I estimates per-view distributions, Stage II builds ordered point–distribution blocks, and Stage III clusters their compact-side graph. Shared identities enable cross-view inner products without equating view metrics, and the analysis links KME error to fusion-edge and spectral stability under its stated conditions. The experiments compared with recent SOTA methods on the benchmark datasets demonstrate the efficiency and effectiveness of DiFuMVC. Future work will sparsify view pairs, handle missing or unaligned views, and analyze data-dependent component discovery.

AI USE STATEMENT

Generative AI assisted with language editing, literature and reference checks, and the checking of theoretical statements and proofs. This introduced no new experimental measurements. The authors retain responsibility for the final manuscript, including verification of all AI-assisted content.

ETHICS STATEMENT

This work concerns unsupervised clustering of benchmark feature matrices. No new human-subject data were collected. Clustering can amplify representation bias, so deployment in consequential settings requires dataset-specific fairness, privacy, and error analysis.

REFERENCES

Qiong Cao, Li Shen, Weidi Xie, Omkar M. Parkhi, and Andrew Zisserman. VGGFace2: A dataset for recognising faces across pose and age. In *13th IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 67–74, 2018. doi: 10.1109/FG.2018.00020.

Ding-Yun Chen, Xiao-Pei Tian, Yu-Te Shen, and Ming Ouhyoung. On visual similarity based 3d model retrieval. *Computer Graphics Forum*, 22(3):223–232, 2003. doi: 10.1111/1467-8659.00669.

James S. Cope, Thibaut Beghin, Paolo Remagnino, and Sarah Barman. One-hundred plant species leaves data set. UCI Machine Learning Repository, 2013. URL https://archive.ics.uci.edu/dataset/241/one+hundred+plant+species+leaves.

Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006. URL https://www.jmlr.org/papers/v7/demsar06a.html.

Robert P. W. Duin. Multiple features data set. UCI Machine Learning Repository, 1998. URL https://archive.ics.uci.edu/dataset/72/multiple+features.

Li Fei-Fei, Rob Fergus, and Pietro Perona. One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):594–611, 2006. doi: 10.1109/TPAMI.2006.79.

Yifan Feng, Zizhao Zhang, Xibin Zhao, Rongrong Ji, and Yue Gao. GVCNN: Group-view convolutional neural networks for 3d shape recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 264–272, 2018. doi: 10.1109/CVPR.2018.00035.

Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001. doi: 10.1109/34.927464.

Jan-Mark Geusebroek, Gertjan J. Burghouts, and Arnold W. M. Smeulders. The amsterdam library of object images. *International Journal of Computer Vision*, 61(1):103–112, 2005. doi: 10.1023/B:VISI.0000042993.50813.60.

Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of Classification*, 2(1):193–218, 1985. doi: 10.1007/BF01908075.

Zhao Kang, Wangtao Zhou, Zhitong Zhao, Junming Shao, Meng Han, and Zenglin Xu. Large-scale multi-view subspace clustering in linear time. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 4412–4419, 2020. doi: 10.1609/aaai.v34i04.5867.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009. URL https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

Abhishek Kumar, Piyush Rai, and Hal Daumé, III. Co-regularized multi-view spectral clustering. In *Advances in Neural Information Processing Systems*, volume 24, pp. 1413–1421, 2011.

Christoph H. Lampert, Hannes Nickisch, and Stefan Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 951–958, 2009. doi: 10.1109/CVPR.2009.5206594.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791.

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. doi: 10.1109/TPAMI.2020.3011148.

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, pp. 2750–2756, 2015. doi: 10.1609/aaai.v29i1.9598.

Zhenglai Li, Jun Wang, Chang Tang, Xinzhong Zhu, Wei Zhang, and Xinwang Liu. Balanced multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026. doi: 10.1109/TPAMI.2026.3688728. Early Access.

Zhenguo Li, Xiao-Ming Wu, and Shih-Fu Chang. Segmentation using superpixels: A bipartite graph partitioning approach. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 789–796, 2012. doi: 10.1109/CVPR.2012.6247750.

Stuart P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28 (2):129–137, 1982. doi: 10.1109/TIT.1982.1056489.

Colin McDiarmid. On the method of bounded differences. In *Surveys in Combinatorics, 1989*, volume 141 of *London Mathematical Society Lecture Note Series*, pp. 148–188. Cambridge University Press, 1989. doi: 10.1017/CBO9781107359949.008.

Krikamol Muandet, Kenji Fukumizu, Bharath K. Sriperumbudur, and Bernhard Schölkopf. Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends in Machine Learning*, 10(1–2):1–141, 2017. doi: 10.1561/2200000060.

Sameer A. Nene, Shree K. Nayar, and Hiroshi Murase. Columbia object image library (COIL-20). Technical Report CUCS-005-96, Department of Computer Science, Columbia University, 1996. URL `https://cave.cs.columbia.edu/repository/COIL-20`.

Maria-Elena Nilsback and Andrew Zisserman. A visual vocabulary for flower classification. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pp. 1447–1454, 2006. doi: 10.1109/CVPR.2006.42.

Shuangmei Peng, Zhe Chen, Tianyang Xu, and Xiao-Jun Wu. Relationship alignment for view-aware multi-view clustering. In *International Conference on Learning Representations*, 2026. URL `https://openreview.net/forum?id=uRA9cT4MK6`.

Xiaoyu Qin, Kai Ming Ting, Ye Zhu, and Vincent C. S. Lee. Nearest-neighbour-induced isolation similarity and its impact on density-based clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pp. 4755–4762, 2019. doi: 10.1609/aaai.v33i01.33014755.

Ferdinando Samaria and Andy Harter. Parameterisation of a stochastic model for human face identification. In *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, pp. 138–142, 1994. doi: 10.1109/ACV.1994.341300.

Erich Schubert and Arthur Zimek. ELKI multi-view clustering data sets based on the amsterdam library of object images (ALOI). Zenodo, 2010.

Alex Smola, Arthur Gretton, Le Song, and Bernhard Schölkopf. A hilbert space embedding for distributions. In *Algorithmic Learning Theory*, volume 4754 of *Lecture Notes in Computer Science*, pp. 13–31, 2007. doi: 10.1007/978-3-540-75225-7_5.

Bharath K. Sriperumbudur, Arthur Gretton, Kenji Fukumizu, Bernhard Schölkopf, and Gert R. G. Lanckriet. Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research*, 11:1517–1561, 2010. URL `https://www.jmlr.org/papers/v11/sriperumbudur10a.html`.

Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *IEEE International Conference on Computer Vision*, pp. 945–953, 2015. doi: 10.1109/ICCV.2015.114.

Kai Ming Ting, Yue Zhu, and Zhi-Hua Zhou. Isolation kernel and its effect on SVM. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2329–2337, 2018. doi: 10.1145/3219819.3219990.

Kai Ming Ting, Bi-Cun Xu, Takashi Washio, and Zhi-Hua Zhou. Isolation distributional kernel: A new tool for kernel based anomaly detection. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 198–206, 2020. doi: 10.1145/3394486.3403062.

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

Jinjun Wang, Jianchao Yang, Kai Yu, Fengjun Lv, Thomas Huang, and Yihong Gong. Locality-constrained linear coding for image classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3360–3367, 2010. doi: 10.1109/CVPR.2010.5540018.

Xin Wang, Devinder Kumar, Nicolas Thome, Matthieu Cord, and Frederic Precioso. Recipe recognition with large multimodal food dataset. In *IEEE International Conference on Multimedia & Expo Workshops*, pp. 1–6, 2015. doi: 10.1109/ICMEW.2015.7169757.

Yiming Wang, Qun Li, Dongxia Chang, Jie Wen, Hua Dai, and Fu Xiao. Asymmetric multi-view clustering with hyperbolic uncertainty modeling. In *International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=OTKdKRLXf7. Spotlight.

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

Yi Yu, Tengyao Wang, and Richard J. Samworth. A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015. doi: 10.1093/biomet/asv008.

Haochen Zhou, Guofeng Ding, Mouxing Yang, Peng Hu, Yijie Lin, and Xi Peng. Uncover underlying correspondence for robust multi-view clustering. In *International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=a4S1nQay3b.

Siyuan Zhou and Zhibin Gu. OPTION: Optimal transport–guided flow matching for incomplete and unaligned multi-view clustering. In *International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=1jBsENo5ii.

Xin Zou, Ruimeng Liu, Chang Tang, Zhenglai Li, Xinwang Liu, Kunlun He, and Wanqing Li. Learning disentangled representations for generalized multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026. doi: 10.1109/TPAMI.2026.3687339. Early Access.

A LIMITATIONS

DIFUMVC assumes complete alignment and known k; its compact side has at most V^2k nodes and grows quadratically with V. When there are many views, the time spent fusing information from the views will increase significantly, and the constructed bipartite graph will also be large. Although the current data views are small, how to reduce the complexity of the V is a task to be explored when facing more data views in the future.

B ALGORITHMIC DETAILS

Algorithm 1: Distribution Fusion Multi-View Clustering

Input: $\{\boldsymbol{X}^{(v)}\}_{v=1}^{V}$, cluster count k, isolation partitions R, resolution ψ, threshold τ, subset cap $s_{\max}$.
Output: labels $\boldsymbol{y} \in \{1, \dots, k\}^n$.

1. **Stage I: distributional representation.** Obatain the kernel features of points via the R shared reference-index sets of size ψ, and form k_v empirical KMEs for each view.
2. **Stage II: distribution fusion.** Compute point-to-distribution affinities, row-normalize the block, and concatenate it into $\boldsymbol{B}$.
3. **Stage III: distribution-driven clustering.** Form the normalized distribution graph with $M = V\sum_v k_v \le V^2k$ nodes, take its top k eigenvectors, transfer them to samples, and return k-means labels after row normalization.

The pseudo code of DIFUMVC is shown in Algorithm 1[2]. If fewer than k connected components are produced in a view, the adaptive rule retains all available components; if more are produced, it retains the largest k.

C COMPLETE GRAPH AND SPECTRAL GUARANTEES

The following full statements support the compact rate summary in Section 3.5.

To propagate these simultaneous edge errors, define raw population and empirical block rows

$$\boldsymbol{s}_{i,*}^{p\to q} = \big(a^*_{p\to(q,c)}(x_i^{(p)})\big)_{c=1}^{k_q}, \qquad \widehat{\boldsymbol{s}}_i^{p\to q} = \big(\widehat{a}_{p\to(q,c)}(x_i^{(p)})\big)_{c=1}^{k_q},$$

and let, for $\epsilon > 0$ and $\alpha_q \ge 0$, $g_q(\boldsymbol{s}) = \alpha_q \boldsymbol{s}/(\|\boldsymbol{s}\|_2 + \epsilon)$. Concatenating $g_q(\boldsymbol{s}_{i,*}^{p\to q})$ produces $\boldsymbol{B}^*$; concatenating $g_q(\widehat{\boldsymbol{s}}_i^{p\to q})$ produces $\widehat{\boldsymbol{B}}$, which is the graph $\boldsymbol{B}$ used by the algorithm.

Theorem 3 (Uniform concentration of the fused graph)**.** *Under the conditional sampling assumptions of Theorem 2, fix the source-query design, assume all A target components are nonempty, and let $m_{qc} \ge 1$ be the size of target component (q, c). Define*

$$b_{qc}(\delta) = \frac{1 + \sqrt{2\log(A/\delta)}}{\sqrt{m_{qc}}}, \qquad e_q(\delta) = \left(\sum_{c=1}^{k_q} b_{qc}^2(\delta)\right)^{1/2},$$

$$a_q = \min_{i,p} \|\boldsymbol{s}_{i,*}^{p\to q}\|_2 > 0, \qquad \ell_q = \frac{2\alpha_q}{a_q + \epsilon}.$$

Then, with conditional probability at least $1 - \delta$,

$$\|\widehat{\boldsymbol{B}} - \boldsymbol{B}^*\|_2 \le \|\widehat{\boldsymbol{B}} - \boldsymbol{B}^*\|_F \le C_B(\delta), \qquad C_B(\delta) = \left[nV\sum_{q=1}^{V} \ell_q^2 e_q^2(\delta)\right]^{1/2}.$$

If $m_{qc} \ge m_{\min}$ and $\ell_q \le \ell_{\max}$ for all q, c, then

$$C_B(\delta) \le \ell_{\max} V\sqrt{nk}\, \frac{1 + \sqrt{2\log(A/\delta)}}{\sqrt{m_{\min}}}. \tag{8}$$

[2]The source code is avaliable in `anonymous.4open.science/r/DiFuMVC-BAB7`

The event is uniform over all nV source-view queries because one KME error bound controls its inner product with every norm-bounded source feature; it therefore needs no additional union factor in n.

Theorem 4 (Joint random-reference and KME concentration)**.** *Fix an isolation resolution ψ and condition on a fixed observed reference pool and a fixed source-view query family. Draw independent evaluation partitions $\Omega_1, \dots, \Omega_R \sim \Pi_\psi$ from that pool. Let $\{\xi_u\}_{u=1}^{nV}$ denote the queries, and let the $A = \sum_{q=1}^{V} k_q$ nonempty target distributions P_{qc} and their component rule be fixed independently of this evaluation reference bank. For $Y \sim P_{qc}$, define*

$$h_{\Omega,\psi}(\xi_u,(q,Y)) = \langle \phi_{\Omega,\psi}(\xi_u), \phi_{\Omega,\psi}((q,Y))\rangle, \quad \kappa_{R,\psi}(\zeta,\zeta') = \frac{1}{R}\sum_{r=1}^{R} h_{\Omega_r,\psi}(\zeta,\zeta'),$$
$$a_{\infty,\psi;u,qc} = \mathbb{E}_{\Omega,Y} h_{\Omega,\psi}(\xi_u,(q,Y)), \quad a_{R,\psi;u,qc} = \mathbb{E}_Y \kappa_{R,\psi}(\xi_u,(q,Y)),$$
$$\widehat{a}_{R,m,\psi;u,qc} = \frac{1}{m_{qc}}\sum_{j=1}^{m_{qc}} \kappa_{R,\psi}(\xi_u,(q,Y_{qcj})),$$

where the Y_{qcj} are conditionally i.i.d. from P_{qc} and independent of the evaluation partitions. For each fixed edge,

$$\mathbb{E}\widehat{a}_{R,m,\psi;u,qc} = a_{\infty,\psi;u,qc}, \qquad \mathrm{Var}(\widehat{a}_{R,m,\psi;u,qc}) \le \frac{1}{4R} + \frac{1}{4m_{qc}}. \tag{9}$$

Let $L_{\mathrm{aff}} = nVA$ and define

$$r_R(\delta) = \sqrt{\frac{\log(4L_{\mathrm{aff}}/\delta)}{2R}}, \qquad b_{qc}^{(m)}(\delta) = \frac{1+\sqrt{2\log(2A/\delta)}}{\sqrt{m_{qc}}},$$
$$d_{qc}^{R,m}(\delta) = b_{qc}^{(m)}(\delta) + r_R(\delta), \qquad e_{q,R,m}(\delta) = \left(\sum_{c=1}^{k_q}[d_{qc}^{R,m}(\delta)]^2\right)^{1/2}.$$

Then, with probability at least $1-\delta$ over both the evaluation partitions and KME samples, simultaneously for all u, q, c,

$$|\widehat{a}_{R,m,\psi;u,qc} - a_{\infty,\psi;u,qc}| \le d_{qc}^{R,m}(\delta). \tag{10}$$

Construct $\boldsymbol{B}_{\infty,\psi}$ from the infinite-reference affinity rows $\boldsymbol{s}_{i,\infty,\psi}^{p\to q}$ and $\widehat{\boldsymbol{B}}_{R,m,\psi}$ from their empirical counterparts, using the same fixed normalized-coordinate regularizer $\epsilon > 0$. Define

$$a_{q,\psi} = \min_{i,p} \|\boldsymbol{s}_{i,\infty,\psi}^{p\to q}\|_2 > 0, \qquad \ell_{q,\psi} = \frac{2\alpha_q}{a_{q,\psi}+\epsilon}.$$

On the same event,

$$\|\widehat{\boldsymbol{B}}_{R,m,\psi} - \boldsymbol{B}_{\infty,\psi}\|_2 \le \|\widehat{\boldsymbol{B}}_{R,m,\psi} - \boldsymbol{B}_{\infty,\psi}\|_F \le C_{R,m,\psi}(\delta), \tag{11}$$

where

$$C_{R,m,\psi}(\delta) = \left[nV\sum_{q=1}^{V} \ell_{q,\psi}^2 e_{q,R,m}^2(\delta)\right]^{1/2}. \tag{12}$$

If $m_{qc} \ge m_{\min}$ and $\ell_{q,\psi} \le \ell_{\max,\psi}$ for all q, c, then

$$C_{R,m,\psi}(\delta) \le \ell_{\max,\psi} V\sqrt{nk}\left[\frac{1+\sqrt{2\log(2A/\delta)}}{\sqrt{m_{\min}}} + \sqrt{\frac{\log(4nVA/\delta)}{2R}}\right]. \tag{13}$$

Theorem 4 is a split-reference/fixed-rule idealization: a pilot bank may determine the component rule, but the R partitions analyzed above must be a fresh independent evaluation bank. It does not justify conditioning on components learned from those same R partitions. Appendix D.4 gives the proof.

Corollary 2 (Reference budget for affinity-ranking preservation)**.** *For the fixed query family in Theorem 4, let c^*_{uq} uniquely maximize the infinite-reference population affinity and define*

$$\Delta_\psi = \min_{u,q} \min_{c \neq c^*_{uq}} \left[a_{\infty,\psi;u,qc^*_{uq}} - a_{\infty,\psi;u,qc} \right] > 0.$$

If only reference Monte Carlo error is present, the population ranking is preserved with probability at least $1-\delta$ whenever

$$R > \frac{2\log(2L_{\mathrm{aff}}/\delta)}{\Delta_\psi^2}; \quad \textit{in particular,} \quad R^{\mathrm{ref}}_{\min} = \left\lfloor \frac{2\log(2L_{\mathrm{aff}}/\delta)}{\Delta_\psi^2} \right\rfloor + 1 \tag{14}$$

is sufficient. With finite KME samples, let $b^{(m)}_{\max}(\delta) = \max_{q,c} b^{(m)}_{qc}(\delta)$. If $\Delta_\psi > 2b^{(m)}_{\max}(\delta)$, the joint empirical ranking is preserved with probability at least $1-\delta$ whenever

$$R > \frac{2\log(4L_{\mathrm{aff}}/\delta)}{[\Delta_\psi - 2b^{(m)}_{\max}(\delta)]^2}. \tag{15}$$

Corollary 3 (Conditional KME-to-spectral-subspace stability)**.** *Condition on the fixed query design in Theorem 3 and a population graph $\boldsymbol{B}^*$ satisfying the following deterministic conditions. Let $d^*_x = \min_i(\boldsymbol{B}^*\mathbf{1})_i > 0$, $d^*_y = \min_j((\boldsymbol{B}^*)^\top\mathbf{1})_j > 0$, $M = VA$, and*

$$\gamma = \lambda_k(T(\boldsymbol{B}^*)) - \lambda_{k+1}(T(\boldsymbol{B}^*)) > 0, \qquad T(\boldsymbol{B}) = \widetilde{\boldsymbol{W}}_y.$$

Write $C_\delta = C_B(\delta)$ and suppose

$$\sqrt{M}\, C_\delta \le d^*_x/2, \qquad \sqrt{n}\, C_\delta \le d^*_y/2. \tag{16}$$

Set $d_x = d^*_x/2$, $d_y = d^*_y/2$, $\beta_\delta = \|\boldsymbol{B}^*\|_2 + C_\delta$, and

$$L_{B,\delta} = \frac{2\beta_\delta}{d_x d_y} + \frac{\beta_\delta^2\sqrt{M}}{d_x^2 d_y} + \frac{\beta_\delta^2\sqrt{n}}{d_x d_y^2}.$$

If $\boldsymbol{U}^*$ and $\widehat{\boldsymbol{U}}$ span the top-k eigenspaces of $T(\boldsymbol{B}^*)$ and $T(\widehat{\boldsymbol{B}})$, respectively, then, with conditional probability at least $1-\delta$,

$$\|T(\widehat{\boldsymbol{B}}) - T(\boldsymbol{B}^*)\|_2 \le L_{B,\delta} C_B(\delta), \tag{17}$$

$$\|\sin\Theta(\widehat{\boldsymbol{U}}, \boldsymbol{U}^*)\|_F \le \frac{2\sqrt{k}\, L_{B,\delta}}{\gamma} C_B(\delta). \tag{18}$$

Corollary 4 (Joint reference-to-spectral stability)**.** *Under Theorem 4, let $d_{x,\psi} = \min_i(\boldsymbol{B}_{\infty,\psi}\mathbf{1})_i > 0$, $d_{y,\psi} = \min_j(\boldsymbol{B}^\top_{\infty,\psi}\mathbf{1})_j > 0$, and*

$$\gamma_\psi = \lambda_k(T(\boldsymbol{B}_{\infty,\psi})) - \lambda_{k+1}(T(\boldsymbol{B}_{\infty,\psi})) > 0.$$

Write $C_\psi = C_{R,m,\psi}(\delta)$ and suppose $\sqrt{M}C_\psi \le d_{x,\psi}/2$ and $\sqrt{n}C_\psi \le d_{y,\psi}/2$. Define $\underline{d}_{x,\psi} = d_{x,\psi}/2$, $\underline{d}_{y,\psi} = d_{y,\psi}/2$, $\beta_\psi = \|\boldsymbol{B}_{\infty,\psi}\|_2 + C_\psi$, and

$$L_{B,\psi} = \frac{2\beta_\psi}{\underline{d}_{x,\psi}\underline{d}_{y,\psi}} + \frac{\beta_\psi^2\sqrt{M}}{\underline{d}^2_{x,\psi}\underline{d}_{y,\psi}} + \frac{\beta_\psi^2\sqrt{n}}{\underline{d}_{x,\psi}\underline{d}^2_{y,\psi}}.$$

If $\boldsymbol{U}_{\infty,\psi}$ and $\widehat{\boldsymbol{U}}_{R,m,\psi}$ span the corresponding top-k compact-side eigenspaces, then, with probability at least $1-\delta$,

$$\|T(\widehat{\boldsymbol{B}}_{R,m,\psi}) - T(\boldsymbol{B}_{\infty,\psi})\|_2 \le L_{B,\psi} C_{R,m,\psi}(\delta),$$

$$\|\sin\Theta(\widehat{\boldsymbol{U}}_{R,m,\psi}, \boldsymbol{U}_{\infty,\psi})\|_F \le \frac{2\sqrt{k}\, L_{B,\psi}}{\gamma_\psi} C_{R,m,\psi}(\delta).$$

Corollary 4 exposes two statistically distinct errors: KME sampling contributes $O(m_{\min}^{-1/2})$, whereas finite reference averaging contributes $O(R^{-1/2})$. The resolution ψ changes the limiting graph itself

and therefore enters through $a_{q,\psi}$, its degrees, $\|\boldsymbol{B}_{\infty,\psi}\|_2$, and γ_ψ. The corollary imposes conditions on these quantities at each fixed ψ but does not order them across resolutions.

Appendix E proves the graph bound, and Appendix F proves the spectral corollary with all degree perturbations exposed. These equations display their dependence on n, V, k, block conditioning, degrees, and eigengap; no dimension-free rate is claimed. The result controls the compact spectral subspace, not subsequent row normalization or final k-means labels.

The joint statements keep ϵ fixed in normalized-feature coordinates. The implementation instead stores affinities multiplied by R, so a fixed raw-code floor gives $\epsilon_R = \epsilon_{\mathrm{code}}/R$. Denote the corresponding empirical implementation graph by $\widehat{\boldsymbol{B}}^{(\epsilon_R)}_{R,m,\psi}$; this is distinct from the fixed-ϵ graph in Theorem 4. To compare with a zero-floor limit, let $\boldsymbol{B}^{(0)}_{\infty,\psi}$ denote the population graph whose blocks are normalized with $\epsilon = 0$. If $\|\boldsymbol{s}^{p\to q}_{i,\infty,\psi}\|_2 \geq a_{q,\psi}$, the change from zero-floor to code-floor normalization contributes at most

$$u_{q,R} = \frac{\alpha_q \epsilon_{\mathrm{code}}}{R(a_{q,\psi} + \epsilon_{\mathrm{code}}/R)} \tag{19}$$

per normalized block row. Hence, with $\ell_{q,\psi,R} = 2\alpha_q/(a_{q,\psi} + \epsilon_{\mathrm{code}}/R)$,

$$\|\widehat{\boldsymbol{B}}^{(\epsilon_R)}_{R,m,\psi} - \boldsymbol{B}^{(0)}_{\infty,\psi}\|_F \leq C^{(0)}_{R,m,\psi}(\delta) := \left[nV \sum_{q=1}^{V} \left(\ell_{q,\psi,R} e_{q,R,m}(\delta) + u_{q,R}\right)^2\right]^{1/2}. \tag{20}$$

Any spectral consequence of this comparison must recompute the degree lower bounds, eigengap, and L_B using $\boldsymbol{B}^{(0)}_{\infty,\psi}$ and $C^{(0)}_{R,m,\psi}$. Alternatively, scaling $\epsilon_{\mathrm{code}} = R\epsilon$ keeps the normalized-coordinate regularizer fixed across an R sweep and retains the original population baseline.

D Proofs for Stage-I Representation and Stage-II Fusion

D.1 KME concentration

Assume the reference map and cluster rule were fixed independently of the KME-estimation observations, and condition on these objects and m. All expectations and probabilities in this subsection are conditional on this sigma-field. The conditional-i.i.d. premise is an assumption; merely conditioning on a same-sample learned map or rule would not establish it. Write $Z_\ell = \Psi_{\mathcal{R}}(v, X_\ell)$, so $\|Z_\ell\| \leq 1$, $\mu = \mathbb{E}Z_\ell$, and $\widehat{\mu} = m^{-1}\sum_\ell Z_\ell$. Independence gives

$$\mathbb{E}\|\widehat{\mu} - \mu\|_2^2 = \frac{1}{m}\left(\mathbb{E}\|Z\|_2^2 - \|\mu\|_2^2\right) \leq \frac{1}{m}.$$

Jensen's inequality therefore yields $\mathbb{E}\|\widehat{\mu} - \mu\|_2 \leq m^{-1/2}$. Replacing one observation changes the norm by at most $2/m$. McDiarmid's bounded-difference inequality (McDiarmid, 1989) gives, for every $u > 0$,

$$\Pr[\|\widehat{\mu} - \mu\|_2 - \mathbb{E}\|\widehat{\mu} - \mu\|_2 \geq u] \leq \exp(-m\,u^2/2).$$

Taking $u = \sqrt{2\log(1/\delta)/m}$ proves Equation 6.

D.2 Point–distribution edges and finite-family separation

Cauchy–Schwarz and $\|\Psi_{\mathcal{R}}(p, x)\| \leq 1$ give

$$|\widehat{a}(x) - a^*(x)| = |\langle \Psi_{\mathcal{R}}(p, x), \widehat{\mu} - \mu\rangle| \leq \|\widehat{\mu} - \mu\|_2,$$

uniformly over source points x. For two distributions, the reverse triangle inequality gives

$$\begin{aligned} |\|\widehat{\mu}_a - \widehat{\mu}_b\|_2 - \|\mu_a - \mu_b\|_2| &\leq \|(\widehat{\mu}_a - \mu_a) - (\widehat{\mu}_b - \mu_b)\|_2 \\ &\leq \epsilon_a + \epsilon_b. \end{aligned}$$

If every error is below $\Delta/4$, every empirical pair remains at least $\Delta/2$ apart. For the affinity ranking, Equation 7 gives

$$\widehat{a}_{p\to(q,c^*)}(x) - \widehat{a}_{p\to(q,c)}(x) \geq a^*_{p\to(q,c^*)}(x) - a^*_{p\to(q,c)}(x) - \epsilon_{c^*} - \epsilon_c > 0.$$

The block normalization multiplies all k coordinates by the same positive scalar, so it cannot change this ordering.

D.3 FINITE-PAIR RANDOM-REFERENCE CONCENTRATION

Fix ψ. For one partition r, write

$$h_{r,\psi}(\zeta,\zeta') = \langle \phi_{\Omega_r,\psi}(\zeta), \phi_{\Omega_r,\psi}(\zeta')\rangle \in \{0,1\}, \qquad \kappa_{R,\psi}(\zeta,\zeta') = \frac{1}{R}\sum_{r=1}^{R} h_{r,\psi}(\zeta,\zeta'). \tag{21}$$

Theorem 5 (Finite-pair Monte Carlo concentration)**.** *Condition on a fixed observed reference pool. Suppose the R reference partitions are drawn independently from the same single-partition rule Π_ψ; sampling may be without replacement within each partition. Define*

$$\kappa_{\infty,\psi}(\zeta,\zeta') = \mathbb{E}_{\Omega\sim\Pi_\psi} h_{\Omega,\psi}(\zeta,\zeta').$$

Then $\kappa_{\infty,\psi}$ is positive semidefinite. Write $p_\psi(\zeta,\zeta') = \kappa_{\infty,\psi}(\zeta,\zeta')$. For each fixed pair,

$$\begin{aligned} \mathbb{E}\kappa_{R,\psi}(\zeta,\zeta') &= p_\psi(\zeta,\zeta'),\\ \mathrm{Var}(\kappa_{R,\psi}(\zeta,\zeta')) &= \mathbb{E}\big[(\kappa_{R,\psi}(\zeta,\zeta') - p_\psi(\zeta,\zeta'))^2\big] = \frac{p_\psi(\zeta,\zeta')[1-p_\psi(\zeta,\zeta')]}{R} \le \frac{1}{4R}. \end{aligned} \tag{22}$$

For any fixed collection $\mathcal{Q}$ of $L \ge 1$ augmented-point pairs chosen independently of the R partition draws, with probability at least $1-\delta$ over the reference partitions,

$$\max_{(\zeta,\zeta')\in\mathcal{Q}} |\kappa_{R,\psi}(\zeta,\zeta') - \kappa_{\infty,\psi}(\zeta,\zeta')| \le \sqrt{\frac{\log(2L/\delta)}{2R}}. \tag{23}$$

Moreover, if $v_\psi = \max_{(\zeta,\zeta')\in\mathcal{Q}} p_\psi(\zeta,\zeta')[1-p_\psi(\zeta,\zeta')]$, the variance-sensitive bound

$$\max_{(\zeta,\zeta')\in\mathcal{Q}} |\kappa_{R,\psi}(\zeta,\zeta') - \kappa_{\infty,\psi}(\zeta,\zeta')| \le \sqrt{\frac{2v_\psi \log(2L/\delta)}{R}} + \frac{2\log(2L/\delta)}{3R} \tag{24}$$

also holds with probability at least $1-\delta$.

Proof. Each $h_{\Omega,\psi}$ is an inner-product kernel, so its expectation is positive semidefinite. For a fixed pair, the R variables in Equation 21 are independent, identically distributed Bernoulli variables with mean $p_\psi(\zeta,\zeta')$, proving Equation 22. Hoeffding's inequality gives

$$\Pr(|\kappa_{R,\psi}(\zeta,\zeta') - \kappa_{\infty,\psi}(\zeta,\zeta')| \ge t) \le 2\exp(-2Rt^2).$$

A union bound over the L fixed pairs and $t = \sqrt{\log(2L/\delta)/(2R)}$ proves Equation 23. Applying Bernstein's inequality to each centered average and taking the same union bound gives Equation 24. □

This theorem controls only Monte Carlo randomness for a fixed finite query family and a fixed ψ. It does not give a uniform bound over the input domains, and it does not make the realized finite-dimensional kernel characteristic.

D.4 PROOF OF THE JOINT REFERENCE–KME BOUNDS

For a fixed edge (u,q,c), let

$$\eta_{r,u,qc} = \mathbb{E}_{Y\sim P_{qc}} h_{\Omega_r,\psi}(\xi_u,(q,Y)) \in [0,1].$$

The variables $\eta_{r,u,qc}$ are i.i.d. across r and $a_{R,\psi;u,qc} = R^{-1}\sum_r \eta_{r,u,qc}$. Thus $\mathbb{E}a_{R,\psi;u,qc} = a_{\infty,\psi;u,qc}$ and $\mathrm{Var}(a_{R,\psi;u,qc}) \le 1/(4R)$. Conditional on the evaluation reference bank, the m_{qc} summands defining $\widehat{a}_{R,m,\psi;u,qc}$ are i.i.d. in $[0,1]$. The law of total variance therefore gives

$$\begin{aligned} \mathrm{Var}(\widehat{a}_{R,m,\psi;u,qc}) &= \mathrm{Var}_\Omega(a_{R,\psi;u,qc}) + \mathbb{E}_\Omega \mathrm{Var}(\widehat{a}_{R,m,\psi;u,qc} \mid \Omega_{1:R})\\ &\le \frac{1}{4R} + \frac{1}{4m_{qc}}, \end{aligned}$$

which proves Equation 9.

For the high-probability result, Hoeffding's inequality applied to the $\eta_{r,u,qc}$ and a union bound over $L_{\text{aff}} = nVA$ fixed edges give, with probability at least $1 - \delta/2$,

$$\max_{u,q,c} |a_{R,\psi;u,qc} - a_{\infty,\psi;u,qc}| \le r_R(\delta). \tag{25}$$

Conditional on any realized evaluation bank, apply Theorem 2 to the A target distributions with failure probability $\delta/(2A)$. A union bound gives

$$\|\widehat{\mu}_{R,qc} - \mu_{R,qc}\|_2 \le b_{qc}^{(m)}(\delta) \qquad \text{for all } q, c \tag{26}$$

with conditional probability at least $1 - \delta/2$. Here $\widehat{\mu}_{R,qc}$ and $\mu_{R,qc}$ denote, respectively, the empirical and population KMEs in the realized finite reference feature space. Since every source feature has norm at most one, Equation 26 controls all nV source queries without a further union bound. Combining Equations 25 and 26 by the triangle inequality proves Equation 10.

Within target view q, the k_q coordinate errors in Equation 10 give

$$\|\widehat{\boldsymbol{s}}_{i,R,m,\psi}^{p\to q} - \boldsymbol{s}_{i,\infty,\psi}^{p\to q}\|_2 \le e_{q,R,m}(\delta).$$

The normalization bound in Equation 28 then gives a normalized block-row error at most $\ell_{q,\psi} e_{q,R,m}(\delta)$. Summing its square over n objects, V source views, and V target views proves Equations 11–12. Under the uniform assumptions, $e_{q,R,m}(\delta) \le \sqrt{k}[b_{\max}^{(m)}(\delta) + r_R(\delta)]$, which proves Equation 13. Applying Proposition 4 and the same Davis–Kahan argument as Corollary 3 proves Corollary 4.

For Corollary 2, a uniform edge error r can reduce an affinity gap by at most $2r$. Substituting $r = \sqrt{\log(2L_{\text{aff}}/\delta)/(2R)}$ and requiring $2r < \Delta_\psi$ proves Equation 14. With finite KME samples, the gap is at least $\Delta_\psi - 2[b_{\max}^{(m)}(\delta) + r_R(\delta)]$; requiring this quantity to be positive proves Equation 15. Finally, for $g_\epsilon(s) = \alpha s/(\|s\|_2 + \epsilon)$ and $\|s\|_2 \ge a$,

$$\|g_{\epsilon_{\text{code}}/R}(s) - g_0(s)\|_2 = \frac{\alpha\epsilon_{\text{code}}/R}{\|s\|_2 + \epsilon_{\text{code}}/R} \le \frac{\alpha\epsilon_{\text{code}}}{R(a + \epsilon_{\text{code}}/R)},$$

which proves the floor-drift term in Equation 19. Adding this drift to the empirical block error and summing squared bounds over all nV source-view rows and V target-view blocks proves Equation 20; the subsequent spectral constants must therefore be those of its zero-floor population baseline.

D.5 Why isolation resolution is not monotone

Proposition 2 (No distribution-free monotonicity in ψ)**.** *Under the isolation feature in Equation 1, increasing ψ can either increase or decrease a population pair affinity and its associated two-component affinity margin. Hence neither $\kappa_{\infty,\psi}$ nor Δ_ψ has a distribution-free monotonicity in ψ. This proposition makes no monotonicity claim about the derived compact eigengap γ_ψ. For any N fixed augmented points, define the realized Gram matrix $\boldsymbol{K}_{R,\psi}$ by $(\boldsymbol{K}_{R,\psi})_{ij} = \kappa_{R,\psi}(\zeta_i, \zeta_j)$. It also satisfies*

$$\operatorname{rank}(\boldsymbol{K}_{R,\psi}) \le \min\{N, R\psi\}.$$

Proof. Consider one view on the real line with Euclidean distance, and let Π_ψ sample uniformly among size-ψ subsets of a three-point reference pool. For the pool $\{0, 10, 20\}$ and query pair $(4, 6)$, the single-partition collision indicators for the three size-two subsets are $0, 1, 1$, whereas the size-three indicator is 0. Thus $\kappa_{\infty,2}(4, 6) = 2/3$ and $\kappa_{\infty,3}(4, 6) = 0$. Take the one-query family $\{4\}$ and two target distributions $P_1 = \delta_4$ and $P_2 = \delta_6$. Since the self-affinity of query 4 is one in both cases, the resulting Δ_ψ increases from $1/3$ to 1. Conversely, for the pool $\{0, 50, 100\}$ and query pair $(49, 51)$, the size-two indicators are again $0, 1, 1$, while the size-three indicator is 1. Hence $\kappa_{\infty,2}(49, 51) = 2/3$ and $\kappa_{\infty,3}(49, 51) = 1$. The same partition calculation gives $\kappa_{\infty,2}(49, 100) = 1/3$ and $\kappa_{\infty,3}(49, 100) = 0$. Take the query family $\{49\}$, $P_1 = \delta_{49}$, and $P_2 = \frac{3}{4}\delta_{51} + \frac{1}{4}\delta_{100}$. Its competing affinity increases from $7/12$ to $3/4$, while the unique-winner margin Δ_ψ decreases strictly from $5/12$ to $1/4$. All assignments obey the nearest-reference radius gate in Equation 1. This proves both possible directions. The rank bound follows because the realized feature map has dimension $R\psi$. □

For fixed ψ, Equation 22 gives the exact pairwise mean-square error $p_\psi(1-p_\psi)/R$ and a high-probability radius of order $R^{-1/2}$. This is not a sample-path monotonicity statement: for a Bernoulli pair affinity with mean $1/2$, the realized outcomes $(1, 0, 1)$ give absolute errors $1/2, 0, 1/6$ at $R = 1, 2, 3$. Downstream component discovery, spectral selection, row normalization, and k-means add further nonlinearities, so larger R does not imply monotonically larger NMI, ARI, or F1.

D.6 Why finite features are not claimed characteristic

Proposition 3. *Let the realized feature dimension be $d = R\psi$. If the augmented domain contains at least $d+2$ distinct points, the finite map $\Psi_{\mathcal{R}}$ cannot identify every probability distribution on that domain.*

Proof. For any $d+2$ points ζ_i, the augmented vectors $(1, \Psi_{\mathcal{R}}(\zeta_i)) \in \mathbb{R}^{d+1}$ are linearly dependent. Hence nonzero coefficients a_i exist with $\sum_i a_i = 0$ and $\sum_i a_i \Psi_{\mathcal{R}}(\zeta_i) = 0$. Normalizing the positive and negative parts of a produces two distinct discrete probability distributions with the same finite mean feature. □

For a fixed realized feature map, the empirical KME remains a consistent estimator of the finite feature mean; injectivity is needed only for universal distribution identification, not for the concentration and edge bounds above.

E Proof of Fused-Graph Concentration

Apply Theorem 2 with failure probability δ/A to the A target KMEs. A union bound gives, simultaneously for every (q, c),

$$\|\widehat{\mu}_{qc} - \mu_{qc}\|_2 \le b_{qc}(\delta).$$

Because $\|\Psi_{\mathcal{R}}(p, x)\|_2 \le 1$, the same event controls every source view and every source point:

$$|\langle \Psi_{\mathcal{R}}(p, x), \widehat{\mu}_{qc} - \mu_{qc} \rangle| \le b_{qc}(\delta).$$

Consequently,

$$\|\widehat{\boldsymbol{s}}_i^{p\to q} - \boldsymbol{s}_{i,*}^{p\to q}\|_2 \le e_q(\delta).$$

For arbitrary vectors $\widehat{\boldsymbol{s}}, \boldsymbol{s}$, $\epsilon > 0$, and $\alpha_q \ge 0$, addition and subtraction using the population denominator give

$$\|g_q(\widehat{\boldsymbol{s}}) - g_q(\boldsymbol{s})\|_2 \le \frac{\alpha_q \|\widehat{\boldsymbol{s}} - \boldsymbol{s}\|_2}{\|\boldsymbol{s}\|_2 + \epsilon} + \frac{\alpha_q \|\widehat{\boldsymbol{s}}\|_2 |\|\widehat{\boldsymbol{s}}\|_2 - \|\boldsymbol{s}\|_2|}{(\|\widehat{\boldsymbol{s}}\|_2 + \epsilon)(\|\boldsymbol{s}\|_2 + \epsilon)} \tag{27}$$

$$\le \frac{2\alpha_q}{\|\boldsymbol{s}\|_2 + \epsilon} \|\widehat{\boldsymbol{s}} - \boldsymbol{s}\|_2. \tag{28}$$

Thus each normalized (p, q) block row has error at most $\ell_q e_q(\delta)$. Summing squared errors over n objects, V source views, and V target views yields

$$\|\widehat{\boldsymbol{B}} - \boldsymbol{B}^*\|_F^2 \le \sum_{i=1}^{n} \sum_{p=1}^{V} \sum_{q=1}^{V} \ell_q^2 e_q^2(\delta) = nV \sum_{q=1}^{V} \ell_q^2 e_q^2(\delta).$$

Finally $\|\cdot\|_2 \le \|\cdot\|_F$. Under the uniform assumptions, $e_q^2(\delta) \le k[1 + \sqrt{2\log(A/\delta)}]^2/m_{\min}$ and $\sum_q \ell_q^2 e_q^2(\delta)$ is at most $V\ell_{\max}^2$ times this quantity, proving Equation 8.

F Conditional Stage-III Spectral Stability

First consider block normalization $g_\epsilon(s) = \alpha s/(\|s\|_2 + \epsilon)$ with $\epsilon > 0$ and $\alpha \ge 0$. If a population block satisfies $\|s\|_2 \ge a_0 > 0$, direct addition and subtraction gives

$$\|g_\epsilon(\widehat{s}) - g_\epsilon(s)\|_2 \le \frac{2\alpha}{a_0 + \epsilon} \|\widehat{s} - s\|_2.$$

Without a_0 the numerical ϵ yields only the looser global constant $2\alpha/\epsilon$.

Let $B_0 = B^*$ and $B_1 = \widehat{B}$ be nonnegative $n \times M$ matrices, $\eta = \|B_1 - B_0\|_2$, and $\beta = \max(\|B_0\|_2, \|B_1\|_2)$. For $j \in \{0, 1\}$ define

$$D_{x,j} = \mathrm{diag}(B_j \mathbf{1}), \qquad D_{y,j} = \mathrm{diag}(B_j^\top \mathbf{1}), \qquad T(B_j) = D_{y,j}^{-1/2} B_j^\top D_{x,j}^{-1} B_j D_{y,j}^{-1/2}.$$

Assume all row and column degrees of both matrices are at least $d_x > 0$ and $d_y > 0$, respectively.

Proposition 4 (Small-side graph stability)**.**

$$\|T(B_1) - T(B_0)\|_2 \le L_B \eta,$$

where

$$L_B = \frac{2\beta}{d_x d_y} + \frac{\beta^2 \sqrt{M}}{d_x^2 d_y} + \frac{\beta^2 \sqrt{n}}{d_x d_y^2}. \tag{29}$$

Proof. Set $E = B_1 - B_0$, $R_j = D_{x,j}^{-1}$, and $C_j = D_{y,j}^{-1/2}$. Degree perturbations satisfy

$$\|R_1 - R_0\|_2 \le \frac{\sqrt{M}\,\eta}{d_x^2}, \qquad \|C_1 - C_0\|_2 \le \frac{\sqrt{n}\,\eta}{2 d_y^{3/2}}.$$

For $F_j = B_j^\top R_j B_j$, a three-term telescoping expansion gives

$$\|F_1 - F_0\|_2 \le \left(\frac{2\beta}{d_x} + \frac{\beta^2 \sqrt{M}}{d_x^2} \right) \eta.$$

Expanding $C_1 F_1 C_1 - C_0 F_0 C_0$, applying $\|C_j\|_2 \le d_y^{-1/2}$ and $\|F_j\|_2 \le \beta^2/d_x$, yields Equation 29. □

Let U_0, U_1 span the top-k eigenspaces of $T(B_0), T(B_1)$ and suppose $\gamma = \lambda_k(T(B_0)) - \lambda_{k+1}(T(B_0)) > 0$. The Davis–Kahan variant of Yu et al. (2015) gives

$$\|\sin\Theta(U_1, U_0)\|_F \le \frac{2\sqrt{k}\, L_B \eta}{\gamma},$$

and an orthogonal O exists such that

$$\|U_1 O - U_0\|_F \le \frac{2^{3/2}\sqrt{k}\, L_B \eta}{\gamma}.$$

The result is informative when $L_B \eta \ll \gamma$. It does not by itself ensure stable row normalization or final k-means labels.

F.1 Proof of Corollary 3

On the event of Theorem 3, set $\boldsymbol{E} = \widehat{\boldsymbol{B}} - \boldsymbol{B}^*$ and $\eta = \|\boldsymbol{E}\|_2 \le C_\delta$. The row- and column-degree perturbations satisfy

$$\|\boldsymbol{E}\mathbf{1}_M\|_\infty \le \sqrt{M}\,\eta, \qquad \|\boldsymbol{E}^\top \mathbf{1}_n\|_\infty \le \sqrt{n}\,\eta.$$

Equation 16 therefore ensures that every empirical row and column degree is at least $d_x = d_x^*/2$ and $d_y = d_y^*/2$; the population degrees obey the same lower bounds. Moreover,

$$\|\widehat{\boldsymbol{B}}\|_2 \le \|\boldsymbol{B}^*\|_2 + \eta \le \beta_\delta.$$

Proposition 4 applied with these deterministic bounds gives Equation 17. The Davis–Kahan variant of Yu et al. (2015), at the population eigengap γ, then gives Equation 18.

G WHY STAGE-III TRANSFER CUT SOLVES THE SMALL-SIDE PROBLEM

Let $\boldsymbol{B} \in \mathbb{R}_+^{n\times M}$ have positive row and column degrees. First,

$$\boldsymbol{W}_y\mathbf{1} = \boldsymbol{B}^\top \boldsymbol{D}_x^{-1}\boldsymbol{B}\mathbf{1} = \boldsymbol{B}^\top\mathbf{1},$$

so $\boldsymbol{D}_y$ in Equation 4 is exactly the column-degree matrix of $\boldsymbol{B}$. Set

$$\boldsymbol{P} = \boldsymbol{D}_x^{-1/2}\boldsymbol{B}\boldsymbol{D}_y^{-1/2}, \qquad \boldsymbol{T} = \boldsymbol{P}^\top\boldsymbol{P} = \widetilde{\boldsymbol{W}}_y. \tag{30}$$

Proposition 5 (Spectrum of the compact distribution graph)**.** *The matrix $\boldsymbol{T}$ is positive semidefinite, every eigenvalue belongs to $[0,1]$, and*

$$\boldsymbol{T}\boldsymbol{D}_y^{1/2}\mathbf{1} = \boldsymbol{D}_y^{1/2}\mathbf{1}.$$

The multiplicity of eigenvalue one equals the number of connected components of the weighted bipartite graph induced by the positive entries of $\boldsymbol{B}$. In particular, if this graph is connected, then $1 = \lambda_1(\boldsymbol{T}) > \lambda_2(\boldsymbol{T})$.

Proof. For any $\boldsymbol{z} \in \mathbb{R}^M$, weighted Cauchy–Schwarz gives

$$\|\boldsymbol{P}\boldsymbol{z}\|_2^2 = \sum_i \frac{1}{d_{x,i}}\left(\sum_j B_{ij}\frac{z_j}{\sqrt{d_{y,j}}}\right)^2 \le \sum_i\sum_j B_{ij}\frac{z_j^2}{d_{y,j}} = \sum_j z_j^2.$$

Hence $0 \preceq \boldsymbol{T} \preceq \boldsymbol{I}$. Moreover,

$$\boldsymbol{P}\boldsymbol{D}_y^{1/2}\mathbf{1} = \boldsymbol{D}_x^{1/2}\mathbf{1}, \qquad \boldsymbol{P}^\top\boldsymbol{D}_x^{1/2}\mathbf{1} = \boldsymbol{D}_y^{1/2}\mathbf{1},$$

which proves the unit eigenvalue. Equality in the weighted Cauchy–Schwarz step requires $z_j/\sqrt{d_{y,j}}$ to be constant among all distribution nodes adjacent to each sample node. This equality propagates along bipartite paths, leaving exactly one free constant per connected component. Conversely, these componentwise constants attain equality, so the stated multiplicity follows. □

Proposition 6 (Transfer-cut eigensystem)**.** *Let $\boldsymbol{U}$ have orthonormal columns and satisfy*

$$\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{\Lambda}, \qquad \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1,\dots,\lambda_r) \succ 0.$$

Then $\boldsymbol{Q} = \boldsymbol{P}\boldsymbol{U}\boldsymbol{\Lambda}^{-1/2}$ has orthonormal columns and $\boldsymbol{P}\boldsymbol{P}^\top\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{\Lambda}$. The implemented transfer embedding satisfies

$$\boldsymbol{H} = \boldsymbol{D}_x^{-1/2}\boldsymbol{P}\boldsymbol{U} = \boldsymbol{D}_x^{-1/2}\boldsymbol{Q}\boldsymbol{\Lambda}^{1/2}. \tag{31}$$

It is also a right-eigenvector matrix of the sample-side random-walk operator

$$\boldsymbol{R}_x = \boldsymbol{D}_x^{-1}\boldsymbol{B}\boldsymbol{D}_y^{-1}\boldsymbol{B}^\top, \qquad \boldsymbol{R}_x\boldsymbol{H} = \boldsymbol{H}\boldsymbol{\Lambda}.$$

For every nonzero row, row normalization removes the positive left scaling in Equation 31; hence the implementation uses the row directions of the eigenvalue-weighted basis $\boldsymbol{P}\boldsymbol{U} = \boldsymbol{Q}\boldsymbol{\Lambda}^{1/2}$. It is generally not the same as row-normalizing $\boldsymbol{Q}$. The latter directions would be recovered by row-normalizing $\boldsymbol{H}\boldsymbol{\Lambda}^{-1/2}$.

Proof. The identity $\boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I}$ and the symmetric sample-side eigen-equation follow directly from $\boldsymbol{P}^\top\boldsymbol{P}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{\Lambda}$. Equation 5 and Equation 30 give the first equality in Equation 31; the second follows from the definition of $\boldsymbol{Q}$. Finally, $\boldsymbol{R}_x = \boldsymbol{D}_x^{-1/2}(\boldsymbol{P}\boldsymbol{P}^\top)\boldsymbol{D}_x^{1/2}$, which gives $\boldsymbol{R}_x\boldsymbol{H} = \boldsymbol{H}\boldsymbol{\Lambda}$. □

G.1 INVARIANCE TO DISTRIBUTION-NODE REINDEXING

Proposition 7 (Reindexing invariance)**.** *After $\boldsymbol{B}$ is formed, let $\boldsymbol{\Pi}$ be any $M \times M$ permutation matrix and set $\boldsymbol{B}' = \boldsymbol{B}\boldsymbol{\Pi}$. Then*

$$\boldsymbol{D}_x' = \boldsymbol{D}_x, \quad \boldsymbol{D}_y' = \boldsymbol{\Pi}^\top\boldsymbol{D}_y\boldsymbol{\Pi}, \quad \boldsymbol{P}' = \boldsymbol{P}\boldsymbol{\Pi}, \quad \boldsymbol{T}' = \boldsymbol{\Pi}^\top\boldsymbol{T}\boldsymbol{\Pi}. \tag{32}$$

Assume the selected top-k subspace has the cutoff gap $\lambda_k(\boldsymbol{T}) > \lambda_{k+1}(\boldsymbol{T})$, and let $\boldsymbol{U}, \boldsymbol{U}'$ be orthonormal bases of the top-k subspaces of $\boldsymbol{T}, \boldsymbol{T}'$. Then an orthogonal $\boldsymbol{O}$ exists such that $\boldsymbol{U}' = \boldsymbol{\Pi}^\top\boldsymbol{U}\boldsymbol{O}$, and the transferred embedding obeys $\boldsymbol{H}' = \boldsymbol{H}\boldsymbol{O}$. With zero rows left zero, row normalization, all pairwise Euclidean distances, and the k-means objective of every fixed partition are unchanged. Consequently, the set of globally optimal partitions is unchanged, with cluster names arbitrary.

Proof. The first two identities follow from row sums and permuted column sums. Functional calculus for a permuted positive diagonal matrix gives $(\boldsymbol{D}'_y)^{-1/2} = \boldsymbol{\Pi}^\top \boldsymbol{D}_y^{-1/2} \boldsymbol{\Pi}$, proving the last two identities. Substitution in Equation 5 gives $\boldsymbol{H}' = \boldsymbol{H}\boldsymbol{O}$; uniqueness of the cutoff subspace gives the stated orthogonal change of basis. Right multiplication by an orthogonal matrix preserves row norms and all Euclidean distances, hence it preserves the global k-means objective. □

Changing source-view order, target-view order, or component labels within a target view induces such a column permutation when the view weights are reindexed consistently. Before block normalization, this statement applies to structured permutations that preserve the (p, q) block grouping; after $\boldsymbol{B}$ is formed, any column permutation satisfies the algebraic identities in Equation 32. At an exact cutoff tie, output invariance additionally needs the full tied eigenspace or an equivariant subspace selector. Finite-precision ties and randomized k-means initialization can also change an implementation's selected representative optimum.

H Parameters for each Methods

The hyperparameters of the algorithms used in the experiments are shown in Table 4.

Table 4: Parameters for the methods.

Algorithm	Parameter
DIFUMVC	$\psi \in \{64, 128, 256, 512, 1024\},\ s = 10000$ $\tau \in \{0.1, 0.2, 0.3, 0.4, 0.5\},\ R = 100$
RAV	$\lambda \in \{10^{-5}, 10^{-4}, \ldots, 10^{2}, 10^{3}\}$ $\beta \in \{10^{-5}, 10^{-4}, \ldots, 10^{-1}, 1\}$
BMvC	$\lambda \in \{10^{-5}, 10^{-4}, \ldots, 10^{4}, 10^{5}\}$
CorreGen	$\tau = 0.5,\ \beta = 0.5$ $\lambda = 3 \times 10^{-2},\ \rho = 0.2$
GMAE	$\alpha \in \{0.01, 0.02, \ldots, 0.06, 0.07\}$ $\beta \in \{0.01, 0.02, \ldots, 0.06, 0.07\}$
HAMC	$d \in \{64, 128, 256\}$ $\lambda_{\text{wc}} \in \{10^{-3}, 10^{-2}, \ldots, 10, 10^{2}\}$ $\lambda_1 \in \{10^{-3}, 10^{-2}, \ldots, 10, 10^{2}\}$ $\lambda_2 \in \{10^{-3}, 10^{-2}, \ldots, 10, 10^{2}\}$ $m \in \{0.9, 0.95, 0.98, 0.99\}$ $p = 0.5,\ \tau = 0.3,\ \tau_c = 0.7$
OPTION	$\lambda_{\text{GW}} \in \{0.05, 0.1, 0.2\},\ \lambda_{\text{clu}} \in \{1, 2, 3\}$ $\lambda_{\text{rec}} \in \{0.5, 1, 1.5\},\ \lambda_{\text{con}} = 0.3$

I Dataset Statistics and Provenance

Table 5: Dataset statistics. n, V, and k are the numbers of samples, views, and clusters.

Dataset	n	V	k	Dataset	n	V	k
ORL	400	3	40	NTU2012	2,012	2	67
YaleB	650	3	10	MNIST	10,000	3	10
Flower17	1,360	7	17	Animal	10,158	2	50
COIL20	1,440	3	20	ALOI	10,800	4	100
Caltech7	1,474	6	7	VGGFace2	34,027	4	50
100Leaves	1,600	3	100	CIFAR10	60,000	7	10
Mfeat	2,000	6	10	Food101	86,796	3	101
UCI Digits	2,000	6	10				

The feature packs are treated as complete and row-aligned. Sources include ORL (Samaria & Harter, 1994), YaleB (Georghiades et al., 2001), Flower17 (Nilsback & Zisserman, 2006), COIL20 (Nene et al., 1996), Caltech101 (Fei-Fei et al., 2006), 100Leaves (Cope et al., 2013), UCI Multiple Features (Duin, 1998), and MNIST (LeCun et al., 1998). NTU2012 uses 3D model data (Chen et al., 2003) with MVCNN/GVCNN features (Su et al., 2015; Feng et al., 2018); Animal derives from Animals with Attributes (Lampert et al., 2009); ALOI uses its multi-view release (Geusebroek et al., 2005; Schubert & Zimek, 2010); and VGGFace2-50 derives from VGGFace2 (Cao et al., 2018). The CIFAR10 pack combines image and LLC features (Krizhevsky, 2009; Wang et al., 2010), while Food101 matches the UPMC multimodal release size (Wang et al., 2015).

J Spatial Transcriptomics Details and Visualizations

The processed MAT files store both exact sparse graphs (`fadj` and `sadj`), the normalized 3,000-HVG expression matrix, spot barcodes and coordinates, manual annotations, histopathology paths, and the two 128-dimensional graph-SVD views in `X`. The data manifest records the BMvC-ST generator commit hashes and the DLPFC/HBC, MBA-expression, and MBA-annotation source URLs. The refined DLPFC 151669 run evaluates all 5,050 ψ–τ–seed candidates: 3,305 are feasible, 1,745 are recorded as infeasible, and none terminates with an unexpected error. Across the 15 selected DiFuMVC runs reported here, $\psi = 16, 32, 64, 128, 256$, and 512 are chosen 2, 5, 2, 2, 1, and 3 times, respectively; $\psi = 2, 4, 8$, and 1024 are searched but never selected. The selected ψ/τ pair for every dataset–seed run is retained in the released result audit.

Figures 6, 7, and 8 use the requested fixed 3×3 layout. Panels (a) and (b) show the histopathological image and manual annotation; panels (c)–(i) show BMvC, RAV, CorreGen, GMAE, HAMC, OPTION, and DiFuMVC. The manual-label legend is placed below the nine-panel grid. For every displayed method, we select the saved partition with the highest NMI among its five runs, recompute NMI/ARI/F1 from that partition, and Hungarian-align cluster identifiers only for colors and macro-F1.

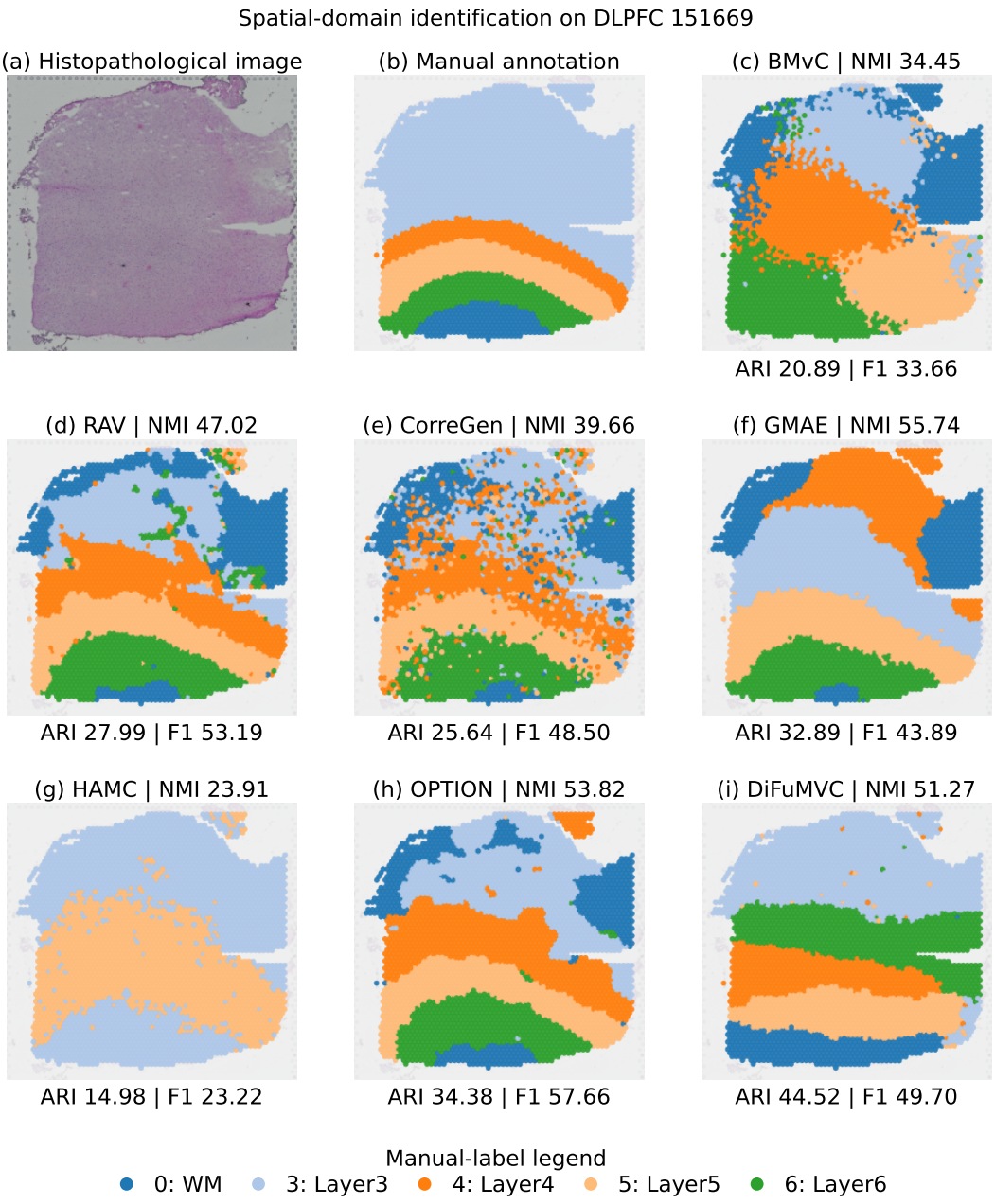


Figure 6: Spatial-domain identification on DLPFC slice 151669. Panel annotations report NMI, ARI, and Hungarian-aligned macro-F1 (percent) for the highest-NMI saved run. All clustering panels use the manual-label palette shown in the legend below the grid.

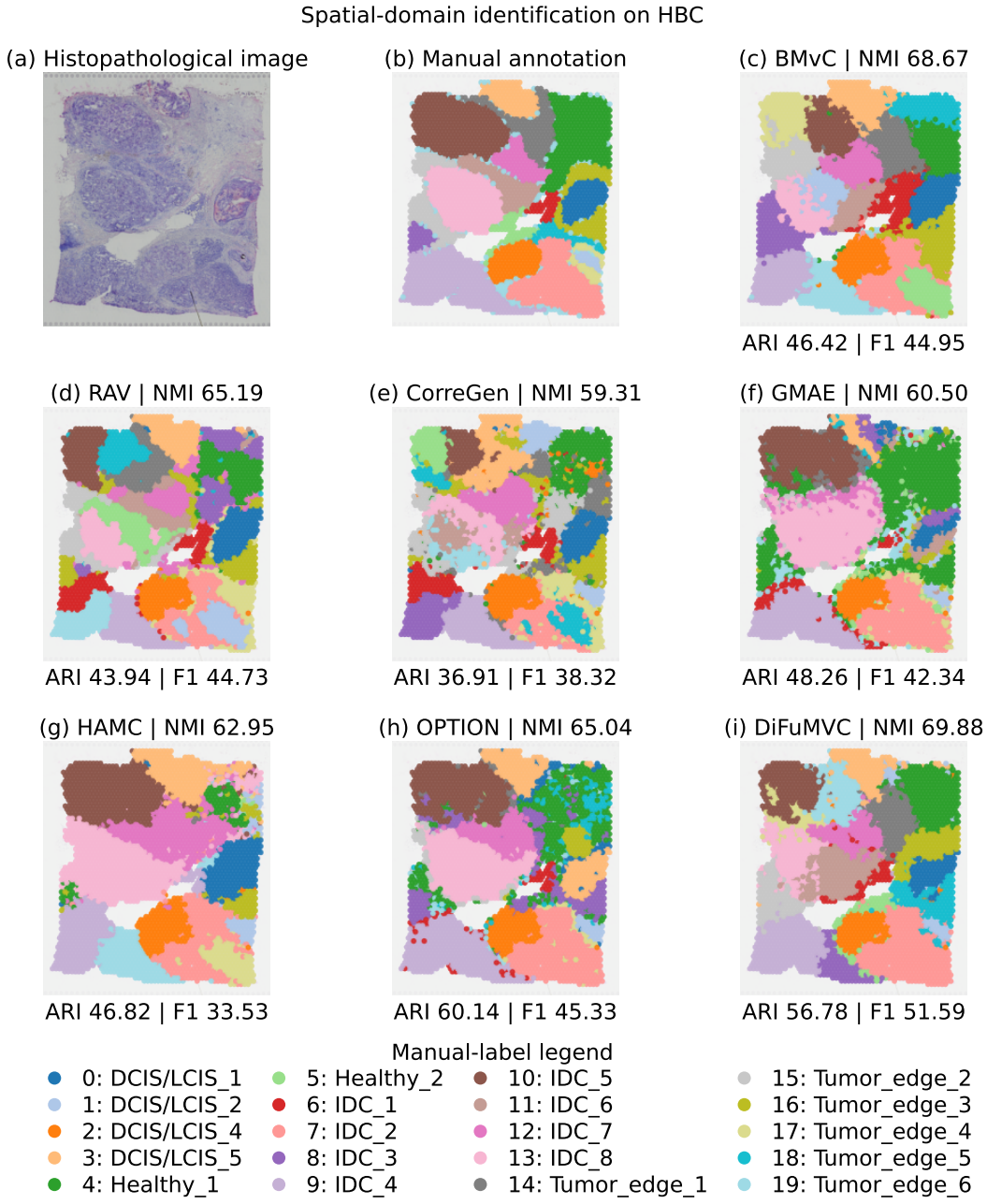


Figure 7: Spatial-domain identification on HBC. The layout, best-run rule, metric annotation, and color alignment match Figure 6.

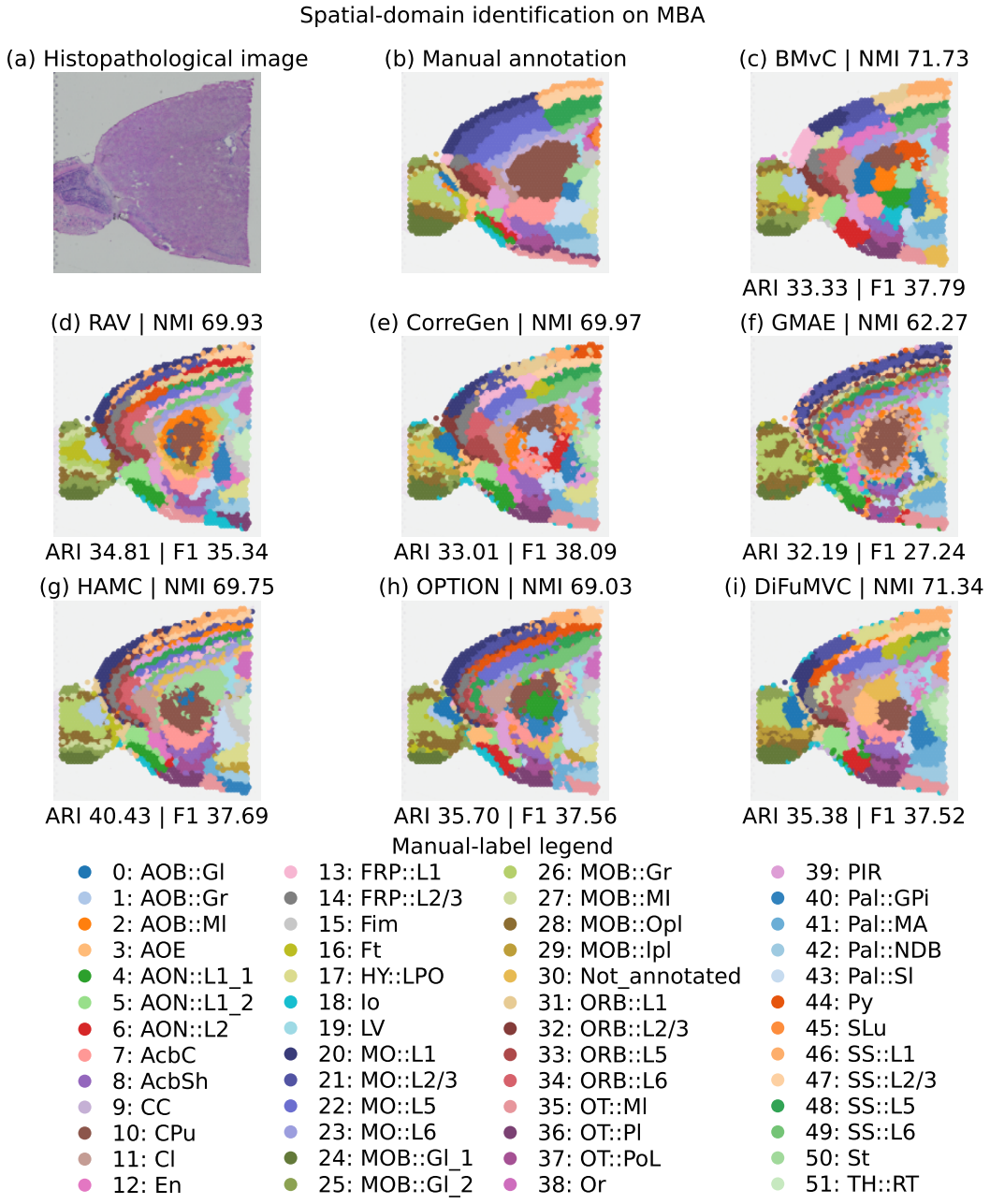


Figure 8: Spatial-domain identification on MBA. The layout, best-run rule, metric annotation, and color alignment match Figure 6.

K Complete ARI and F1 Results

The results of ARI and F1 are shown in Table 6.

Table 6: Complete ARI and F1 comparison. Methods are columns and datasets are rows. Entries are means (%) with standard deviations as small lower-right subscripts; only the best mean is bold and only the second-best mean is underlined. Average scores use all available entries, whereas average ranks use the 14 complete datasets shared by all seven methods.

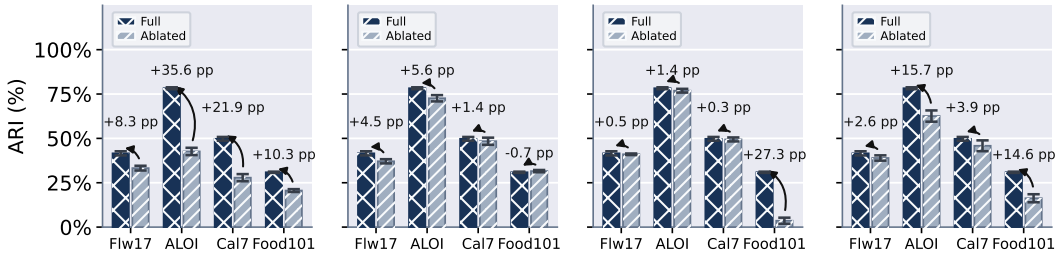
Dataset	**BMvC** (PAMI26)	**RAV** (ICLR26)	**CorreGen** (ICLR26)	**GMAE** (PAMI26)	**HAMC** (ICML26)	**OPTION** (ICML26)	**DiFuMVC** (Ours)
Adjusted Rand Index (%)							
ORL	**84.21**$_{\pm 1.67}$	21.98$_{\pm 2.16}$	63.15$_{\pm 1.77}$	15.09$_{\pm 4.49}$	11.15$_{\pm 3.80}$	54.39$_{\pm 7.22}$	<u>80.74</u>$_{\pm 1.04}$
YaleB	<u>16.66</u>$_{\pm 2.45}$	4.23$_{\pm 0.49}$	13.01$_{\pm 1.73}$	11.00$_{\pm 0.63}$	0.40$_{\pm 0.35}$	1.96$_{\pm 2.44}$	**44.52**$_{\pm 2.94}$
Flw17	**47.65**$_{\pm 1.43}$	13.52$_{\pm 10.13}$	23.46$_{\pm 0.59}$	14.80$_{\pm 9.86}$	10.93$_{\pm 3.11}$	26.03$_{\pm 1.94}$	<u>42.23</u>$_{\pm 0.77}$
COIL20	<u>73.92</u>$_{\pm 0.83}$	73.89$_{\pm 4.26}$	71.26$_{\pm 1.82}$	70.50$_{\pm 2.62}$	47.80$_{\pm 9.99}$	62.99$_{\pm 2.29}$	**98.37**$_{\pm 0.25}$
Cal7	38.23$_{\pm 2.14}$	28.84$_{\pm 0.89}$	22.18$_{\pm 1.82}$	45.47$_{\pm 2.02}$	**56.25**$_{\pm 9.89}$	49.33$_{\pm 3.71}$	<u>49.61</u>$_{\pm 1.33}$
Leaves	<u>84.91</u>$_{\pm 0.79}$	47.54$_{\pm 1.58}$	77.26$_{\pm 0.99}$	46.16$_{\pm 3.06}$	55.89$_{\pm 4.28}$	60.49$_{\pm 4.36}$	**94.94**$_{\pm 0.30}$
Mfeat	**95.78**$_{\pm 0.68}$	80.05$_{\pm 6.63}$	84.39$_{\pm 1.70}$	87.26$_{\pm 7.90}$	75.19$_{\pm 4.45}$	91.37$_{\pm 2.28}$	<u>94.61</u>$_{\pm 0.53}$
Digits	**96.46**$_{\pm 0.37}$	74.80$_{\pm 7.76}$	84.50$_{\pm 1.71}$	90.53$_{\pm 5.90}$	78.34$_{\pm 1.70}$	90.37$_{\pm 3.68}$	<u>94.61</u>$_{\pm 0.53}$
NTU12	65.57$_{\pm 1.32}$	45.19$_{\pm 1.37}$	<u>74.64</u>$_{\pm 0.77}$	47.44$_{\pm 1.91}$	63.37$_{\pm 2.86}$	55.90$_{\pm 2.48}$	**78.02**$_{\pm 0.92}$
MNIST	65.98$_{\pm 3.37}$	72.16$_{\pm 0.21}$	70.27$_{\pm 0.13}$	58.14$_{\pm 1.61}$	<u>73.35</u>$_{\pm 0.17}$	72.41$_{\pm 0.69}$	**74.46**$_{\pm 0.80}$
Animal	<u>60.18</u>$_{\pm 1.07}$	26.91$_{\pm 0.54}$	52.94$_{\pm 0.78}$	36.27$_{\pm 3.86}$	43.82$_{\pm 1.30}$	20.43$_{\pm 1.88}$	**62.45**$_{\pm 0.38}$
ALOI	**81.54**$_{\pm 2.02}$	69.66$_{\pm 2.23}$	76.47$_{\pm 0.48}$	36.16$_{\pm 4.02}$	56.54$_{\pm 2.62}$	70.09$_{\pm 2.53}$	<u>78.25</u>$_{\pm 0.50}$
VGG50	<u>4.44</u>$_{\pm 0.34}$	3.50$_{\pm 0.14}$	4.04$_{\pm 0.13}$	1.31$_{\pm 0.24}$	2.23$_{\pm 0.21}$	2.10$_{\pm 0.32}$	**7.69**$_{\pm 0.16}$
CIFAR10	17.12$_{\pm 0.34}$	15.35$_{\pm 1.22}$	7.22$_{\pm 0.53}$	–	**17.82**$_{\pm 0.77}$	14.75$_{\pm 1.43}$	<u>17.78</u>$_{\pm 0.51}$
Food101	29.04$_{\pm 0.33}$	0.22$_{\pm 0.08}$	**33.11**$_{\pm 0.26}$	1.33$_{\pm 0.74}$	1.51$_{\pm 0.86}$	8.29$_{\pm 9.50}$	<u>30.90</u>$_{\pm 0.35}$
Avg. score ↑	<u>57.45</u>	38.52	50.53	40.10	39.64	45.39	**63.28**
Avg. rank ↓	<u>2.29</u>	5.50	3.64	5.43	5.21	4.43	**1.50**
F1 (%)							
ORL	**88.41**$_{\pm 1.75}$	38.81$_{\pm 1.63}$	72.00$_{\pm 0.85}$	34.23$_{\pm 4.63}$	16.93$_{\pm 3.30}$	66.55$_{\pm 6.32}$	<u>84.98</u>$_{\pm 1.74}$
YaleB	26.88$_{\pm 2.53}$	18.43$_{\pm 0.19}$	<u>29.81</u>$_{\pm 2.68}$	21.44$_{\pm 4.52}$	5.90$_{\pm 2.02}$	17.22$_{\pm 4.51}$	**65.30**$_{\pm 1.84}$
Flw17	<u>51.99</u>$_{\pm 1.34}$	21.36$_{\pm 19.62}$	42.27$_{\pm 0.92}$	27.74$_{\pm 9.83}$	12.59$_{\pm 2.58}$	40.05$_{\pm 4.44}$	**58.42**$_{\pm 1.73}$
COIL20	75.86$_{\pm 0.79}$	<u>78.41</u>$_{\pm 3.46}$	75.67$_{\pm 1.08}$	72.43$_{\pm 4.58}$	48.51$_{\pm 14.09}$	70.28$_{\pm 2.22}$	**99.16**$_{\pm 0.15}$
Cal7	**58.14**$_{\pm 0.92}$	25.91$_{\pm 1.17}$	23.29$_{\pm 0.87}$	34.63$_{\pm 4.46}$	25.85$_{\pm 2.42}$	39.35$_{\pm 2.88}$	<u>47.17</u>$_{\pm 1.54}$
Leaves	<u>88.59</u>$_{\pm 0.70}$	58.46$_{\pm 1.66}$	82.77$_{\pm 1.28}$	54.01$_{\pm 3.27}$	62.21$_{\pm 4.86}$	67.70$_{\pm 3.83}$	**96.45**$_{\pm 0.33}$
Mfeat	<u>96.27</u>$_{\pm 0.59}$	86.81$_{\pm 4.39}$	92.39$_{\pm 0.89}$	87.76$_{\pm 10.10}$	82.04$_{\pm 1.96}$	95.95$_{\pm 1.19}$	**97.55**$_{\pm 0.26}$
Digits	<u>96.86</u>$_{\pm 0.33}$	83.43$_{\pm 6.22}$	92.49$_{\pm 0.88}$	92.30$_{\pm 7.33}$	82.85$_{\pm 0.84}$	95.39$_{\pm 1.95}$	**97.55**$_{\pm 0.26}$
NTU12	67.38$_{\pm 0.59}$	52.05$_{\pm 1.98}$	<u>71.63</u>$_{\pm 1.06}$	42.41$_{\pm 1.93}$	56.03$_{\pm 7.14}$	51.76$_{\pm 2.46}$	**72.22**$_{\pm 0.66}$
MNIST	69.52$_{\pm 3.00}$	85.70$_{\pm 0.11}$	84.60$_{\pm 0.07}$	68.94$_{\pm 2.64}$	<u>86.34</u>$_{\pm 0.12}$	85.80$_{\pm 0.35}$	**86.90**$_{\pm 0.44}$
Animal	<u>61.47</u>$_{\pm 0.68}$	36.70$_{\pm 0.76}$	59.28$_{\pm 0.79}$	31.08$_{\pm 4.87}$	48.63$_{\pm 1.44}$	26.85$_{\pm 4.85}$	**64.37**$_{\pm 0.21}$
ALOI	<u>84.14</u>$_{\pm 1.82}$	71.36$_{\pm 2.72}$	81.36$_{\pm 0.62}$	42.94$_{\pm 3.43}$	60.43$_{\pm 3.13}$	70.70$_{\pm 2.59}$	**87.03**$_{\pm 0.43}$
VGG50	6.76$_{\pm 0.22}$	11.26$_{\pm 0.20}$	<u>12.52</u>$_{\pm 0.30}$	5.26$_{\pm 0.59}$	8.47$_{\pm 0.77}$	7.51$_{\pm 0.31}$	**19.11**$_{\pm 0.39}$
CIFAR10	25.29$_{\pm 0.32}$	35.12$_{\pm 2.61}$	24.16$_{\pm 1.10}$	–	**39.20**$_{\pm 3.22}$	34.44$_{\pm 3.11}$	<u>38.26</u>$_{\pm 1.11}$
Food101	31.27$_{\pm 0.32}$	2.93$_{\pm 0.30}$	**48.36**$_{\pm 0.50}$	5.35$_{\pm 1.93}$	1.16$_{\pm 0.53}$	19.11$_{\pm 18.52}$	<u>44.91</u>$_{\pm 1.26}$
Avg. score ↑	<u>61.92</u>	47.12	59.51	44.32	42.48	52.58	**70.63**
Avg. rank ↓	<u>2.71</u>	4.86	3.29	5.71	5.71	4.50	**1.21**

L Complete Structural Ablations

Figures 9 and 10 report the ARI and F1 counterparts of the NMI ablation in Figure 5.

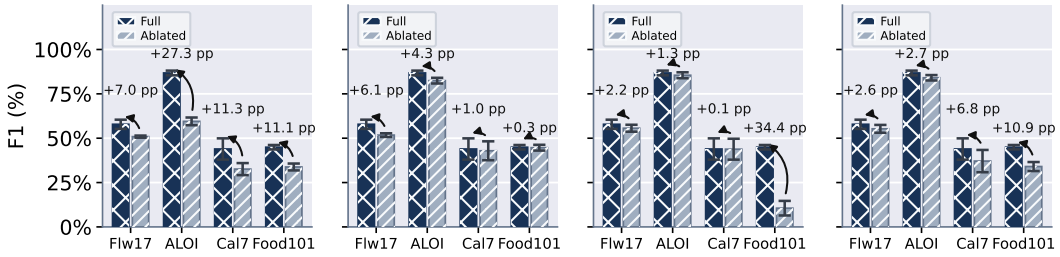
Table 7: Complete ARI and Hungarian-aligned macro-F1 comparison on spatially resolved transcriptomics datasets. Methods are columns and datasets are rows. Entries are means (%) with standard deviations as small lower-right subscripts over five runs; only the best mean is bold and only the second-best mean is underlined. DLPFC reports slice 151669.

Dataset	BMvC (PAMI26)	RAV (ICLR26)	CorreGen (ICLR26)	GMAE (PAMI26)	HAMC (ICML26)	OPTION (ICML26)	DiFuMVC (Ours)
Adjusted Rand Index (%)							
DLPFC 151669	$15.46_{\pm 3.15}$	$24.91_{\pm 2.52}$	$24.50_{\pm 1.50}$	$12.94_{\pm 9.98}$	$2.12_{\pm 6.94}$	$\mathbf{34.75}_{\pm 0.59}$	$\underline{33.07}_{\pm 10.30}$
HBC	$44.68_{\pm 1.34}$	$41.39_{\pm 1.51}$	$41.61_{\pm 1.73}$	$47.37_{\pm 1.48}$	$34.46_{\pm 14.81}$	$\mathbf{50.94}_{\pm 5.85}$	$\underline{49.57}_{\pm 3.67}$
MBA	$32.35_{\pm 0.80}$	$32.27_{\pm 1.65}$	$32.42_{\pm 0.92}$	$28.70_{\pm 3.39}$	$\mathbf{36.63}_{\pm 3.81}$	$32.79_{\pm 1.98}$	$\underline{35.17}_{\pm 0.35}$
F1 (%)							
DLPFC 151669	$43.14_{\pm 2.15}$	$\underline{54.65}_{\pm 1.17}$	$17.29_{\pm 0.55}$	$36.67_{\pm 4.08}$	$19.51_{\pm 5.34}$	$\mathbf{59.11}_{\pm 0.97}$	$41.87_{\pm 7.76}$
HBC	$\mathbf{51.30}_{\pm 1.27}$	$41.78_{\pm 2.11}$	$42.16_{\pm 1.61}$	$39.76_{\pm 2.71}$	$24.50_{\pm 10.69}$	$43.41_{\pm 3.14}$	$\underline{46.72}_{\pm 2.62}$
MBA	$\mathbf{42.62}_{\pm 0.77}$	$36.33_{\pm 1.35}$	$37.27_{\pm 0.51}$	$22.97_{\pm 2.23}$	$28.65_{\pm 8.90}$	$34.79_{\pm 2.46}$	$\underline{38.21}_{\pm 0.71}$



Stage I (a) Point prototype | Stage II (b) Self-view only | Stage II (c) Cross-view only | Stage III (d) Direct k-means

Figure 9: **Paired structural ablations measured by ARI.** Bars report mean $\pm$ population standard deviation. Curved annotations are Full-minus-ablation percentage-point changes.



Stage I (a) Point prototype | Stage II (b) Self-view only | Stage II (c) Cross-view only | Stage III (d) Direct k-means

Figure 10: **Paired structural ablations measured by F1.** The protocol and visual encoding are identical to Figure 9.

Across all four datasets, the Stage-I point-prototype replacement is again a large intervention in ARI and F1, and direct k-means is consistently weaker than transfer cut. The Stage-II ablations expose the same Food101 asymmetry as NMI: self-view only improves ARI by 0.66 points but lowers F1 by 0.26, whereas cross-view only loses 27.3 ARI and 34.4 F1 points. This metric-level split is consistent with low-resolution cross-view blocks perturbing global partition agreement while preserving some class-aligned assignments.

L.1 FOOD101 ADAPTIVE COMPONENT DIAGNOSTIC

Food101 exposes a large disparity in per-view graph resolution. Rather than discarding a view or fabricating empty components, the general Stage-I rule keeps the largest $k_v = \min(c_v, k)$ components in each view and constructs all ordered blocks from those available distributions. Table 8 shows the retained counts at each seed. The image view consistently supplies all 101 nodes, the class-text view supplies one, and the description-text view supplies 11–60. Thus the rerun is genuinely three-view, but its compact graph is strongly imbalanced across target views; this diagnostic explains why cross-view fusion can be slightly weaker than self-view-only fusion for NMI and ARI on this dataset.

Table 8: **Food101 adaptive distribution counts at $\psi = 128$.** For each seed, Stage I retains the largest $k_v = \min(c_v, k)$ components in view v.

Seed	seed1	seed2	seed3	seed4	seed5
Description-text	17	18	60	11	12

M PARAMETER SENSITIVITY

The DIFUMVC is not sensitive to parameter τ and ψ in some range, as shown in Figure 11.

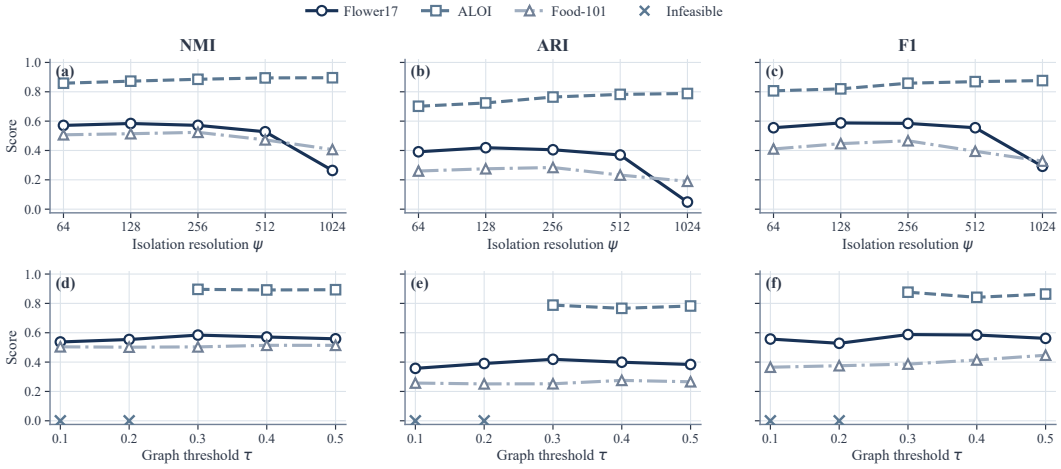


Figure 11: **Parameter sensitivity.** The top and bottom rows vary the isolation resolution ψ and graph threshold τ, respectively; columns report NMI, ARI, and F1 as fractions. Navy, blue-gray, and slate curves denote Flower17, ALOI, and Food101. Crosses at ALOI $\tau \in \{0.1, 0.2\}$ mark infeasible component constructions and are not zero-valued clustering observations.