ONE-STEP CONSISTENCY-AWARE GRAPH DICTIONARY LEARNING FOR INCOMPLETE MULTI-VIEW CLUSTERING

Anonymous authors
Paper under double-blind review

ABSTRACT

Missing samples may degrade the clustering performance of incomplete multi-view clustering (IMVC), where one-step graph-based methods show great potential. However, existing graph learning methods for IMVC typically fail to exploit consistent and complementary information simultaneously at the graph embedding level. Moreover, they often impose low-rank constraints on high-dimensional affinity tensors with high computational complexity. To this end, we propose One-Step Consistency-aware Graph Dictionary Learning (OCGDL), which employs a partially shared graph dictionary to capture both consistency and complementarity simultaneously from the similarity graph obtained from the incomplete multiview data. We first decompose each partially observed similarity graph into shared and view-specific atoms under self-orthogonality and orthogonal-complement constraints, while a missing-indicator mask confines reconstruction to observed samples. Subsequently, we relax the strict constraint on decomposed graph features by considering graph features as dictionary atoms for affinity reconstruction, and row-wise $\ell_{2,1}$ coefficients encourage differentiated atom selection across views. Furthermore, we employ the tensor nuclear norm on the concatenated multiview full graph dictionary tensor of size $k \times V \times n$, which can effectively reduce computational complexity and also encourage high-order cross-view correlations at the embedding level. Discrete cluster indicators are obtained directly in the same objective via spectral rotation with adaptive view weights, forming one-step clustering. Extensive experiments on six benchmark multiview datasets validate the effectiveness of OCGDL under various view missing ratios.

1 INTRODUCTION

In real-world applications, multi-view data collected from diverse modalities, sensors, or feature spaces offer rich, multi-faceted semantic descriptions of complex objects (Xu et al., 2013; Xu & Tian, 2015; Yang & Wang, 2018; Li et al., 2018). Multi-view clustering (MVC) aims to partition unlabeled samples into coherent groups by jointly exploiting the cross-view consistency that reflects cluster shared features and the complementarity that captures view-specific discriminative information (Cao et al., 2015; Gao et al., 2015; Luo et al., 2018; Wang et al., 2019). Among existing paradigms, graph-based approaches are particularly appealing because affinity graphs can effectively encode nonlinear neighborhood structures and support spectral analysis for clustering (Ng et al., 2002; Von Luxburg, 2007; Zhan et al., 2017; Berahmand et al., 2025).

In open scenarios, however, sensor failures, transmission loss, or privacy constraints frequently lead to missing instances in partial views, giving rise to incomplete multi-view clustering (IMVC) (Li et al., 2014; Hu & Chen, 2019; Wen et al., 2023c). Compared with complete MVC, IMVC must contend with damaged pairwise relations and unbalanced observations across views, which readily bias similarity construction and degrade the subsequent embedding fusion (Wen et al., 2018; Li et al., 2021; Wang et al., 2022). Recent IMVC studies can be broadly grouped into imputation-based and imputation-free strategies (Wen et al., 2023c; Liu et al., 2026). The former recovers missing features or similarities before exploring cluster structures (Wang et al., 2021; Lin et al., 2021; Li et al., 2023a; Wang et al., 2026), whereas the latter learns consensus representations only from observed samples to avoid completion overhead (Hu & Chen, 2019; Wen et al., 2018; Liu et al., 2022a). Graph-

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.

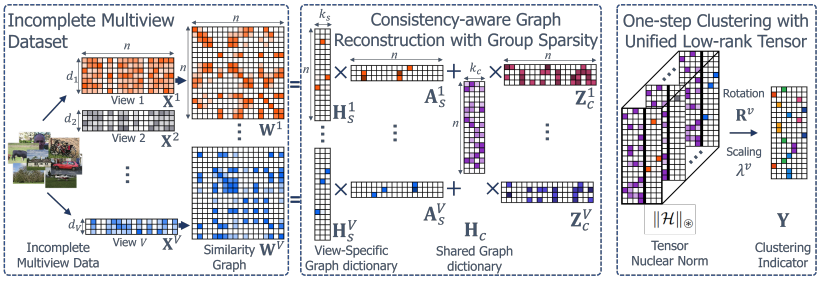


Figure 1: A graphical architecture of incomplete multi-view learning using the consistency-aware shared graph dictionary learning model with group sparsity from multi-view datasets.

oriented IMVC methods further transfer incompleteness into the similarity domain and seek a fused affinity or spectral embedding from partially observed graphs (Wen et al., 2018; 2023a; Li et al., 2021; Chen et al., 2024; 2025).

Despite notable progress, most existing graph-based IMVC pipelines still suffer from several limitations. First, consistency and complementarity are often modeled at different stages or on mismatched representations: some methods emphasize consensus affinity fusion while underusing view-specific structural cues (Wen et al., 2023d; Liu et al., 2025b; 2022b), whereas others preserve diversity without sufficiently coupling high-order cross-view correlations (Xu et al., 2025; Yu et al., 2025; Wang et al., 2026). Second, many approaches follow a two-step scheme that first learns graphs or embeddings and then applies spectral clustering or k-means as post-processing; the separation of representation learning and discrete partition readily causes error accumulation and suboptimal assignments (Tang et al., 2023; Chen et al., 2025; Xu et al., 2025; Wu et al., 2025). Third, low-rank modeling is commonly imposed on high-dimensional multi-view subspace coefficients or full affinity tensors, which is computationally expensive and may not directly regularize the spectral features that ultimately determine clustering decisions (Chen et al., 2023a; Xia et al., 2022; Chen et al., 2023b; 2025; Gu et al., 2026). These observations motivate a one-step framework that simultaneously captures shared and view-specific graph structures at the embedding level, while enforcing cross-view low-rank coupling with reduced complexity.

To this end, we propose One-Step Consistency-aware Graph Dictionary Learning (OCGDL) for incomplete multi-view clustering. Motivated by the spectral perturbation perspective that leading eigenvectors of multi-view affinities admit both aligned shared components and orthogonal view-specific residuals (Davis & Kahan, 1970; Yu et al., 2015), OCGDL decomposes each similarity graph into a partially shared graph dictionary with consensus and view-specific atoms (Liu et al., 2014; Zhao et al., 2022). Missing indicators are incorporated into the reconstruction objective so that unobserved blocks do not bias dictionary learning (Wen et al., 2018; Liu et al., 2024). We further relax classical low-rank graph factorization under orthogonality and orthogonal-complement constraints (Edelman et al., 1998), and encourage differentiated atom selection across views via row-wise $\ell_{2,1}$ sparsity on the reconstruction coefficients. Instead of imposing nuclear-norm constraints on high-dimensional subspace coefficients (Lu et al., 2019; Chen et al., 2023a; 2025), OCGDL stacks the concatenated full graph dictionaries into a third-order tensor and applies the tensor nuclear norm, which couples low-rank structures across views at the embedding level with smaller computational complexity. Finally, spectral rotation fusion with adaptive view weights is integrated into the same objective to produce discrete cluster indicators in one step (Huang et al., 2013; Tang et al., 2023), thereby avoiding the instability of post-hoc k-means.

The main contributions of this paper are summarized as follows:

- We present OCGDL, a one-step IMVC framework that performs consistency-aware graph dictionary learning and discrete partition jointly, bridging graph construction, cross-view fusion, and clustering assignment in a unified optimization.
- We devise a partially shared graph dictionary with orthogonal complement constraints and masked reconstruction, enabling simultaneous exploitation of consensus and complementary spectral structures from incomplete affinities, while $\ell_{2,1}$ coefficients promote view-adaptive atom selection.

- We impose tensor nuclear norm regularization on the stacked full graph-dictionary tensor rather than on high-dimensional subspace representations, which encourages high-order cross-view coupling and reduces the computational burden of low-rank modeling.

- Extensive experiments on six benchmark multi-view datasets under various missing ratios demonstrate the effectiveness and robustness of OCGDL against representative state-of-the-art IMVC methods.

2 Related Work

Incomplete multi-view clustering (IMVC) aims to uncover cluster structures when some samples are missing in partial views (Wen et al., 2023c; Liu et al., 2025a; Wen et al., 2023b). Existing methods are commonly divided into imputation-based and imputation-free paradigms (Liu et al., 2025a; Wen et al., 2023b). Early approaches extract compact representations from observed views via matrix factorization or subspace learning (Li et al., 2014; Hu & Chen, 2019), while generative and deep models further strengthen representation learning under missingness (Wang et al., 2021; Lin et al., 2021). Graph-based IMVC instead transfers incompleteness into the similarity domain, constructing partial graphs for spectral embedding or consensus graph learning (Wen et al., 2018; 2023a; Wang et al., 2022; Li et al., 2021). More recently, imputation-based methods jointly recover missing information and clustering representations via anchors, prototypes, or tensor regularization (Xu et al., 2025; Wang et al., 2026; Li et al., 2023a; Liu et al., 2022b). How to coordinate shared and view-specific structures while handling missing views remains a central challenge.

Beyond merely learning a consensus embedding, multi-view representation learning also needs to preserve distinctive cues in individual views. DiMSC encourages complementary subspace representations with HSIC (Cao et al., 2015), and dual contrastive prediction reinforces shared semantics while retaining view-dependent information (Lin et al., 2023). ARSL-IMVC further recovers each view with a shared latent factor and a view-specific estimator, then learns shared and specific subspace representations under a unified low-rank tensor (Wang et al., 2026). Partially shared dictionary models separate shared and specific representations with view-specific dictionaries (Liu et al., 2014; Zhao et al., 2022). However, most shared-private methods emphasize decomposition itself, while paying limited attention to regulating the expressive capacity between shared and specific components—especially for graph features under incompleteness.

Tensor learning provides another route to model high-order cross-view correlations by stacking multi-view representations into a third-order structure (Zhang et al., 2015; Xie et al., 2018; Lu et al., 2019). LTMSC imposes nuclear-norm constraints on unfolded self-representation tensors (Zhang et al., 2015), and t-SVD-based tensor nuclear norms better preserve multi-way consistency without conventional unfolding loss (Xie et al., 2018). For incomplete data, tensor completion and low-rank tensor regularization recover compact shared structures from partial observations (Xia et al., 2022; Chen et al., 2023a). Subsequent extensions include spectral block-diagonal alignment via t-SVD in OAGL (Li et al., 2023b; Chen et al., 2023c; 2025), shared and specific subspace block-diagonal alignment (Wang et al., 2026), and embedding-tensor regularization (Xu et al., 2025). These studies show the benefit of tensor low-rank coupling, yet they generally impose the low-rank tensor constraint on full affinity tensors, and seldom jointly constrain shared and view-specific embeddings within one unified tensor.

Finally, many graph-based IMVC pipelines still follow a two-stage scheme: fused affinities or consensus spectral embeddings are first obtained, and clustering labels are then produced by spectral clustering or k-means (Chen et al., 2025; 2024). Such separation can cause unstable assignments and suboptimal solutions (Chen et al., 2025). One-step methods therefore integrate representation learning and discrete partition into a unified objective, including UOMSC (Tang et al., 2023), one-step multi-view spectral clustering (Zhu et al., 2018), late-fusion one-stage IMVC (Zhang et al., 2021), imputed anchor-graph discretization (Wu et al., 2025), and spectral-rotation fusion at the indicator level (Chen et al., 2025; Huang et al., 2013). Motivated by the above gaps, this paper develops a one-step consistency-aware graph dictionary learning framework that jointly models shared and view-specific spectral dictionaries under missing indicators, couples them with a cross-view tensor nuclear norm, and obtains discrete assignments via spectral rotation.

3 METHODOLOGY

3.1 ONE-STEP CONSISTENCY-AWARE GRAPH DICTIONARY LEARNING

Given n samples with V views incomplete multi-view signals, i.e., $\{\mathbf{X}^v \in \mathbb{R}^{n\times d_v}\}_{v=1}^V$. The diagonal indicator matrix $\mathbf{P}^v \in \{0,1\}^{n\times n}$ identify existing samples that $\mathbf{P}^v_{ii} = 1$ when i-th sample in view v exist and 0 otherwise. The target of incomplete multi-view clustering is to obtain the discrete binary indicator matrix $\mathbf{Y} \in \{0,1\}^{n\times k}$ with one nonzero entry per row, where k denotes the number of clusters.

For a typical spectral clustering, we define the weighted undirected graph $\mathcal{G}^v(\mathcal{V}^v, \mathcal{E}^v)$ for view v, where node $v_i^v \in \mathcal{V}^v$ corresponds to the i-th sample in $\mathbf{X}^v$ and edge e_{ij}^v in $\mathcal{E}^v$ represents a connection between node i and node j. The relationship is described by the similarity graph $\mathbf{W}^v \in \mathbb{R}^{n\times n}$, where $\mathbf{W}^v_{ij}$ denotes the similarity between $\mathbf{x}_i^v$ and $\mathbf{x}_j^v$. The similarity graph can be obtained by various methods, such as self-representation with sparsity, low rank, and other constraints (Chen et al., 2023a;b; 2025), pairwise construction using a kernel or distance metric (Tang et al., 2023; Berahmand et al., 2025). With the constructed similarity graph, we can obtain the spectral embedding representation $\mathbf{H}^v \in \mathbb{R}^{n\times k}$ for view v in a low-rank structure by the orthogonality condition in (1).

$$\min_{\mathbf{H}^v} \|\mathbf{W}^v - \mathbf{H}^v \mathbf{H}^{v\top}\|_F^2 \quad \text{s.t.} \quad \mathbf{H}^{v\top}\mathbf{H}^v = \mathbf{I}_k. \tag{1}$$

The eigenvalue decomposition (EVD) is the closed-form solution for this orthogonality-constrained spectral approximation problem. For multi-view clustering, motivated by the spectral perturbation perspective that leading eigenvectors of multi-view affinities admit both aligned shared components and orthogonal view-specific residuals (Davis & Kahan, 1970; Yu et al., 2015), Corollary 1 suggests that spectral features across views share consistent principal directions while also carrying complementary information induced by view discrepancies. The Davis–Kahan formulation and the relation with the resulting corollary are stated in Appendix A.2.

Therefore, we can decompose the v-th view similarity graph as follows,

$$\mathbf{W}^v \approx \mathbf{H}_c \mathbf{\Lambda}_c^v \mathbf{H}_c^\top + \mathbf{H}_s^v \mathbf{\Lambda}_s^v (\mathbf{H}_s^v)^\top, \tag{2}$$

where $\mathbf{\Lambda}_c^v$ and $\mathbf{\Lambda}_s^v$ denote the view-shared and view-specific eigenvalue spectrum respectively. However, in incomplete multi-view clustering, view v is observed only on a subset, and we can only obtain a partially observable similarity graph $\hat{\mathbf{W}}^v = \mathbf{P}^v \mathbf{W}^v \mathbf{P}^v$. Missing samples may pull the reconstruction toward the unobserved block and bias the spectrum. Therefore, we relax the reconstruction in a dictionary learning form with the visible indicator $\mathbf{P}^v$, forming the graph reconstruction error $f(\mathbf{H}_c, \mathbf{H}_s^v, \mathbf{A}_c^v, \mathbf{A}_s^v)$ as follows,

$$\begin{aligned} \min f(\mathbf{H}_c, \mathbf{H}_s^v, \mathbf{A}_c^v, \mathbf{A}_s^v) &= \sum_{v=1}^V \frac{1}{2} \|\mathbf{P}^v \big(\hat{\mathbf{W}}^v - \mathbf{H}_c \mathbf{A}_c^v - \mathbf{H}_s^v \mathbf{A}_s^v\big) \mathbf{P}^v\|_F^2, \\ \text{s.t. } \mathbf{H}_c{}^\top \mathbf{H}_c = \mathbf{I}_{k_c}, \mathbf{H}_s^{v\top} \mathbf{H}_s^v = \mathbf{I}_{k_s}, \mathbf{H}_c^\top \mathbf{H}_s^v &= \mathbf{0}, \forall v, \end{aligned} \tag{3}$$

where $\mathbf{A}_c^v \approx \mathbf{\Lambda}_c^v \mathbf{H}_c^\top \in \mathbb{R}^{k_c\times n}$ and $\mathbf{A}_s^v \approx \mathbf{\Lambda}_s^v \mathbf{H}_s^{v\top} \in \mathbb{R}^{k_s\times n}$ allow each observed sample to use a sparse combination of shared and view-specific atoms, making $\mathbf{H}_c$ and $\mathbf{H}_s^v$ form the dictionary analogue for similarity graph reconstruction. The mask $\mathbf{P}^v$ removes missing samples from reconstruction in the objective function, confining the loss to visible entries. Moreover, we restrict the view-specific factors in the orthogonal complement of the consensus subspace with the constraint $\mathbf{H}_c^\top \mathbf{H}_s^v = \mathbf{0}$, making the split consistent with a genuine EVD decomposition rather than an unidentified linear factorization. Furthermore, we induce the sparsity constraint on the coefficients $\mathbf{A}_c^v$ and $\mathbf{A}_s^v$ as follows,

$$\Omega(\mathbf{A}_c^v, \mathbf{A}_s^v) = \sum_{v=1}^V \left(\|\mathbf{A}_c^v\|_{2,1} + \|\mathbf{A}_s^v\|_{2,1}\right), \quad \|\mathbf{A}\|_{2,1} = \sum_{i=1}^k \|\mathbf{a}_{i:}\|_2 = \sum_{i=1}^k \sqrt{\sum_{j=1}^n a_{ij}^2}, \tag{4}$$

where $\mathbf{a}_{i:}$ is i-th row in $\mathbf{A}$, and the group sparsity in the row encourages the coefficients in different views to select different graph dictionary atoms to reconstruct the similarity graph, which can mitigate the information loss because of the incompleteness.

Due to noise and incompleteness of the multiview data, simply fusing the graph dictionary learned from the inaccurate similarity graph may degrade the clustering performance. Thus, to ensure the consistency and complementarity across views at the graph dictionary level, we concatenate the shared and view-specific graph dictionary, i.e., $\mathbf{H}^v = [\mathbf{H}_c, \mathbf{H}_s^v] \in \mathbb{R}^{n\times k}$, and stack the full graph dictionary to form a third-order unified low-rank tensor $\mathcal{H} = \{\{\mathbf{h}_i^v\}_{v=1}^V\}_{i=1}^n \in \mathbb{R}^{k\times V\times n}$. In contrast to previous multi-view methods that impose low-rank tensor constraints on high-dimensional subspace coefficients or full affinity tensors (Zhang et al., 2015; Xie et al., 2018; Chen et al., 2023a; 2025), which are computationally expensive with a $n \times V \times n$ tensor as the sample size grows, we apply the tensor nuclear norm (TNN) on the spectral features (graph dictionaries) directly, i.e., $\|\mathcal{H}\|_{\circledast}$. Since $k \ll n$, the resulting $k \times V \times n$ tensor substantially reduces the cost of t-SVD-based TNN computations, while still encouraging a similar cross-view low-rank coupling effect at the embedding level that ultimately determines clustering.

Furthermore, to implement the one-step multiview clustering, we employ graph dictionary fusion with an orthogonal per-view rotation matrix $\mathbf{R}^v \in \mathbb{R}^{k\times k}$ to yield the binary indicator matrix $\mathbf{Y}$, i.e.,

$$\min \sum_{v=1}^{V} \left\| \mathbf{Y}(\mathbf{Y}^\top\mathbf{Y})^{-1/2} - \lambda_v \mathbf{H}^v\mathbf{R}^v \right\|_F^2, \text{ s.t. } \lambda_v \geq 0, \|\boldsymbol{\lambda}\|_2^2 = 1, \mathbf{R}^{v\top}\mathbf{R}^v = \mathbf{I}_k, \mathbf{y}_{i:}\mathbf{1}_k = 1. \tag{5}$$

where $\boldsymbol{\lambda} \in \mathbb{R}^V$ is the non-negative view weight vector which can balance the importance of different views.

Finally, denoting $\tilde{\mathbf{Y}} = \mathbf{Y}(\mathbf{Y}^\top\mathbf{Y})^{-1/2}$, we can summarize the objective function of the proposed model by unifying the partially shared graph dictionary learning term, the sparsity constraint, the full graph dictionary tensor constraint, and the one-step clustering term, which can be formulated as follows,

$$\begin{aligned}\min_{\substack{\mathbf{H}_c,\,\mathbf{H}_s^v,\,\mathbf{A}_c^v,\,\mathbf{A}_s^v,\\ \mathbf{Y},\,\mathbf{R}^v,\,\boldsymbol{\lambda}}} &\|\mathcal{H}\|_{\circledast} + \sum_{v=1}^{V}\frac{1}{2}\|\mathbf{P}^v\big(\hat{\mathbf{W}}^v - \mathbf{H}_c\mathbf{A}_c^v - \mathbf{H}_s^v\mathbf{A}_s^v\big)\mathbf{P}^v\|_F^2 + \eta\Omega(\mathbf{A}_c^v, \mathbf{A}_s^v) + \frac{\beta}{2}\sum_{v=1}^{V}\left\|\tilde{\mathbf{Y}} - \lambda_v\mathbf{H}^v\mathbf{R}^v\right\|_F^2 \\ &\text{s.t. } \mathbf{H}_c{}^\top\mathbf{H}_c = \mathbf{I}_{k_c}, \mathbf{H}_s^{v\top}\mathbf{H}_s^v = \mathbf{I}_{k_s}, \mathbf{H}_c^\top\mathbf{H}_s^v = \mathbf{0}, \lambda_v \geq 0, \|\boldsymbol{\lambda}\|_2^2 = 1, \mathbf{R}^{v\top}\mathbf{R}^v = \mathbf{I}_k, \mathbf{y}_{i:}\mathbf{1}_k = 1., \forall i, \forall v,\end{aligned} \tag{6}$$

where η and β are the balancing hyperparameters.

3.2 Optimization

We solve Eq. (6) by the alternating direction method of multipliers (ADMM). Introducing an auxiliary tensor $\mathcal{T}$ and a Lagrange multiplier $\mathcal{M}$ for $\mathcal{H} = \mathcal{T}$ yields the augmented Lagrangian as follows,

$$\begin{aligned}\mathcal{L} = &\|\mathcal{T}\|_{\circledast} + \sum_{v=1}^{V}\frac{1}{2}\|\mathbf{P}^v\big(\hat{\mathbf{W}}^v - \mathbf{H}_c\mathbf{A}_c^v - \mathbf{H}_s^v\mathbf{A}_s^v\big)\mathbf{P}^v\|_F^2 \\ &+ \eta\Omega(\mathbf{A}_c^v, \mathbf{A}_s^v) + \frac{\beta}{2}\sum_{v=1}^{V}\left\|\tilde{\mathbf{Y}} - \lambda_v\mathbf{H}^v\mathbf{R}^v\right\|_F^2 + \langle\mathcal{M}, \mathcal{H} - \mathcal{T}\rangle + \frac{\mu}{2}\|\mathcal{H} - \mathcal{T}\|_F^2,\end{aligned} \tag{7}$$

where $\mu > 0$ is a penalty parameter. Let $\mathcal{G} = \mathcal{T} - \mathcal{M}/\mu$ and $\tilde{\mathbf{H}}^v = \lambda_v\tilde{\mathbf{Y}}\mathbf{R}^{v\top}$. The variables in Eq. (7) can be updated alternatively by fixing the others until convergence to obtain the final output as follows:

Update $\mathbf{Y}$ and $\mathbf{R}^v$. Matching $\tilde{\mathbf{Y}}$ with the fused embedding $\mathbf{Q} = \sum_{v=1}^V \lambda_v\mathbf{H}^v\mathbf{R}^v$ yields the discrete assignment as follows,

$$j_i = \arg\max_j Q_{ij}, \quad Y_{ij_i} = 1, \quad Y_{ij} = 0\,\forall\, j \neq j_i. \tag{8}$$

Each rotation has the SVD solution (Huang et al., 2013)

$$\mathbf{R}^v = \mathbf{U}_k^v\mathbf{V}_k^{v\top}, \quad \mathbf{H}^{v\top}\tilde{\mathbf{Y}} = \mathbf{U}_k^v\boldsymbol{\Sigma}_k^v\mathbf{V}_k^{v\top}. \tag{9}$$

Update $\mathbf{H}_c$ and $\mathbf{H}_s^v$. Linearizing the masked reconstruction together with the clustering and tensor terms gives

$$\max_{\mathbf{H}_c} tr(\mathbf{H}_c^\top\mathbf{R}_c) \quad \text{s.t. } \mathbf{H}_c{}^\top\mathbf{H}_c = \mathbf{I}_{k_c}. \tag{10}$$

where $\mathcal{G}^v_{k_c}$ is the first k_c components of view v from the tensor $\mathcal{G}$, and $\tilde{\mathbf{H}}^v_{k_c}$ is the first k_c column of $\tilde{\mathbf{H}}^v$, we have the stacked the linearized graph residual as follow,

$$\mathbf{R}_c = \sum_{v=1}^{V} \Big[\mathbf{P}^v \big(\hat{\mathbf{W}}^v - \mathbf{H}_c \mathbf{A}_c^v - \mathbf{H}_s^v \mathbf{A}_s^v \big) \mathbf{P}^v (\mathbf{A}_c^v)^\top + \beta \tilde{\mathbf{H}}^v_{k_c} + \mu \mathcal{G}^v_{k_c} \Big]. \tag{11}$$

We can then obtain $\mathbf{H}_c$ by the polar factor of $\mathbf{R}_c$, which returns the nearest Stiefel point with the orthogonal constraint ${\mathbf{H}_c}^\top \mathbf{H}_c = \mathbf{I}_{k_c}$ (Edelman et al., 1998). The view-specific update is analogous,

$$\max_{\mathbf{H}_s^v} tr(\mathbf{H}_s^{v\top} \mathbf{R}_s^v) \quad \text{s.t. } \mathbf{H}_s^{v\top} \mathbf{H}_s^v = \mathbf{I}_{k_s}, \mathbf{H}_c^\top \mathbf{H}_s^v = \mathbf{0}. \tag{12}$$

with $\mathbf{R}_s^v = \mathbf{P}^v \big(\hat{\mathbf{W}}^v - \mathbf{H}_c \mathbf{A}_c^v - \mathbf{H}_s^v \mathbf{A}_s^v \big) \mathbf{P}^v (\mathbf{A}_s^v)^\top + \beta \tilde{\mathbf{H}}^v_{k_s} + \mu \mathcal{G}^v_{k_s}$. We first project onto the orthogonal complement, $\tilde{\mathbf{R}}_s^v = (\mathbf{I}_n - \mathbf{H}_c \mathbf{H}_c^\top) \mathbf{R}_s^v$, and then take the polar factor of $\tilde{\mathbf{R}}_s^v$.

Update $\mathcal{T}$ and $\boldsymbol{\lambda}$. When fixing the other variables except for the tensor $\mathcal{T}$, the tensor subproblem

$$\min_{\mathcal{T}} \|\mathcal{T}\|_{\circledast} + \frac{\mu}{2} \big\| \mathcal{H} - \mathcal{T} + \mathcal{M}/\mu \big\|_F^2 \tag{13}$$

can be solved by tensor singular-value thresholding in the Fourier domain (Lu et al., 2019). The view weights admit a closed form on the non-negative unit sphere:

$$\lambda_v = l_v / \|\mathbf{l}\|_2, \quad l_v = \max\big(\mathrm{tr}(\mathbf{H}^{v\top} \tilde{\mathbf{Y}} \mathbf{R}^{v\top}), 0\big). \tag{14}$$

Update $\mathbf{A}_c^v$ and $\mathbf{A}_s^v$. The coefficient subproblem is a group sparse coding problem (Zhao et al., 2022). With a Lipschitz constant $\alpha_c^v \geq \|\mathbf{H}_c^\top \mathbf{H}_c\|_2$,

$$\mathbf{A}_c^v = \mathrm{Prox}_{2,1}\big(\mathbf{U}_c^v, \eta/\alpha_c^v\big), \quad \mathbf{U}_c^v = \mathbf{A}_c^v - \frac{1}{\alpha_c^v} \mathbf{H}_c^\top \mathbf{P}^v (\mathbf{H}_c \mathbf{A}_c^v + \mathbf{H}_s^v \mathbf{A}_s^v - \hat{\mathbf{W}}^v) \mathbf{P}^v, \tag{15}$$

where the row-wise proximal map is

$$\mathrm{Prox}_{2,1}(\mathbf{U}, \eta) = \max\{\|\mathbf{u}_{i:}\|_2 - \eta, 0\} \frac{\mathbf{u}_{i:}}{\|\mathbf{u}_{i:}\|_2}. \tag{16}$$

The update of $\mathbf{A}_s^v$ has the same form with $(\mathbf{H}_s^v, \alpha_s^v)$ in place of $(\mathbf{H}_c, \alpha_c^v)$:

$$\mathbf{A}_s^v = \mathrm{Prox}_{2,1}\big(\mathbf{U}_s^v, \eta/\alpha_s^v\big), \quad \mathbf{U}_s^v = \mathbf{A}_s^v - \frac{1}{\alpha_s^v} \big(\mathbf{H}_s^{v\top} \mathbf{P}^v (\mathbf{H}_c \mathbf{A}_c^v + \mathbf{H}_s^v \mathbf{A}_s^v - \hat{\mathbf{W}}^v) \mathbf{P}^v \big). \tag{17}$$

Finally, $\mathcal{M} \leftarrow \mathcal{M} + \mu(\mathcal{H} - \mathcal{T})$ and $\mu \leftarrow \min(\rho\mu, \mu_{\max})$ with $\rho > 1$. The summarized algorithm and computational complexity analysis are given in Appendix A.3.

4 Experiments

4.1 Datasets and Baselines

Datasets: We conduct a series of experiments on six benchmark public multiview datasets that cover object, digit, face, and text data with diverse scales and cluster numbers, including MSRCv1[1], Handwritten (Nie et al., 2018), Caltech101-20 (Li et al., 2020), ORL40 (Luo et al., 2018), the first 100 classes in ALOI (Schubert & Zimek, 2010), and BBCSport (Greene & Cunningham, 2006). The dataset details are presented in Table 4. The incomplete multiview data is constructed by removing random samples with a missing rate $1 - \|\mathbf{P}\|_0/nV \in \{0.1, 0.3, 0.5, 0.7\}$. Following the protocol of (Liu et al., 2026), every sample is observed in at least one view. Consequently, we omit the 0.7 missing setting on BBCSport and ORL40, for which $(V-1)/V < 0.7$.

Baselines. We compare against representative IMVC methods from two groups: two-step methods, including SCSL (Liu et al., 2024), SIMVC-SA (Wen et al., 2023d), LSIMVC (Liu et al., 2022a), and CAL (Liu et al., 2025b); and one-step methods, including OAGL (Chen et al., 2025), OIMC-IAGD (Wu et al., 2025), and OS-LF-IMVC (Zhang et al., 2021). All compared methods are implemented with their public source codes and recommended parameter settings. All experiments are conducted on an Intel Core i5-12600KF CPU with 32GB RAM. We repeat each experiment 10 times and report average values and standard deviations in terms of clustering accuracy (ACC), normalized mutual information (NMI), and Purity (PUR).

[1] https://www.microsoft.com/en-us/research/project/image-understanding/

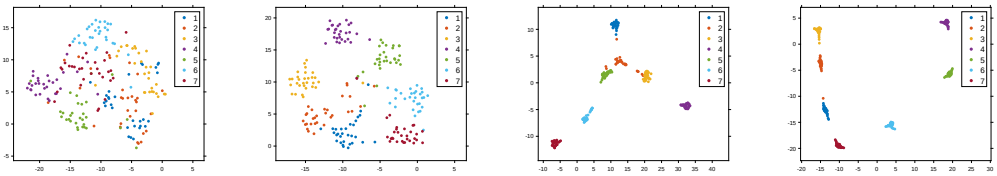
Table 1: Comparison of clustering results under different missing rates. The best and second-best results are highlighted in **bold** and underline, respectively.

Data	Methods	Missing rate											
		0.1			0.3			0.5			0.7		
		ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
MSRCv1	SCSL	82.67±1.39	72.44±1.73	82.67±1.39	77.29±4.43	67.34±2.35	78.67±2.96	64.81±16.43	55.39±17.51	66.71±16.62	31.81±4.72	20.28±7.65	34.00±5.92
	SIMVC-SA	75.87±4.83	65.21±3.88	76.23±4.25	70.17±4.03	57.11±4.11	70.81±3.38	59.51±4.72	45.85±3.27	60.80±4.18	49.60±4.54	35.08±4.18	51.28±5.10
	LSIMVC	66.29±3.04	58.69±2.06	69.48±1.45	55.05±6.49	46.32±6.62	58.14±6.71	41.62±6.00	31.29±4.98	43.19±6.23	24.57±2.70	11.62±3.15	25.38±2.88
	OIMC-IAGD	83.76±2.44	74.32±2.05	83.76±2.44	77.38±6.32	67.97±4.03	78.43±4.68	61.05±6.86	52.32±4.37	63.95±5.46	45.19±4.36	35.09±4.51	47.86±4.13
	OS-LF-IMVC	60.67±3.51	48.81±3.91	61.19±3.07	53.33±5.06	41.94±3.00	54.57±4.73	42.29±3.29	30.89±2.11	44.33±2.19	35.52±2.13	23.77±1.92	37.95±2.13
	CAL	99.33±0.56	98.65±0.96	99.33±0.56	91.65±11.16	86.88±15.17	91.68±11.08	**98.92±0.80**	**97.62±1.72**	**98.92±0.80**	**94.62±1.83**	89.29±3.16	**94.62±1.83**
	OAGL	99.10±1.56	98.43±2.08	99.10±1.56	96.48±6.66	96.17±4.51	97.10±4.78	97.19±2.64	94.98±3.81	97.19±2.64	89.71±6.33	85.59±5.61	89.95±5.95
	OCGDL (ours)	**99.67±0.45**	**99.28±0.93**	**99.67±0.45**	**99.33±0.68**	**98.63±1.40**	**99.33±0.68**	97.62±2.02	95.59±3.23	97.62±2.02	93.24±6.37	**91.09±5.29**	93.29±6.28
Caltech101-20	SCSL	56.41±1.84	62.69±1.56	75.46±2.19	41.27±8.15	55.13±10.76	73.78±7.93	43.08±2.28	53.08±1.05	72.80±1.04	35.30±9.12	37.48±13.89	60.47±11.16
	SIMVC-SA	42.42±0.88	58.46±0.61	76.27±0.54	40.25±1.06	54.46±0.78	73.36±0.60	36.93±0.97	47.17±0.71	67.74±0.73	29.07±0.92	34.92±1.24	58.25±1.26
	LSIMVC	37.64±1.91	41.94±0.44	65.52±0.31	39.17±1.99	40.98±1.07	63.40±1.08	25.36±3.12	29.07±3.05	55.53±3.51	36.65±5.86	30.16±2.33	54.85±3.09
	OIMC-IAGD	58.06±2.40	61.17±1.85	73.56±2.15	53.03±4.10	58.58±1.94	74.48±1.92	38.01±3.05	48.52±1.61	67.82±1.46	31.46±3.88	40.28±1.91	63.38±1.29
	OS-LF-IMVC	40.91±1.42	49.30±1.58	69.14±1.68	33.92±1.22	39.95±1.15	62.17±1.69	25.41±0.95	30.57±1.15	54.79±1.40	20.35±1.36	24.14±1.31	48.81±1.72
	CAL	58.97±1.50	72.57±0.57	**88.35±0.40**	61.97±1.38	73.43±0.59	88.41±0.56	61.84±2.18	73.27±0.89	**89.08±0.58**	52.28±3.17	62.27±1.88	83.26±1.46
	OAGL	55.52±5.09	66.48±3.03	74.72±2.18	53.09±2.50	63.74±2.55	74.00±1.81	44.19±5.85	52.03±2.62	67.43±1.95	51.12±7.54	68.06±4.29	76.99±3.71
	OCGDL (ours)	**66.30±3.66**	**77.26±1.17**	85.56±1.54	**66.58±1.57**	**82.32±1.41**	**88.65±0.93**	**65.61±2.58**	**79.23±1.74**	86.53±1.53	**62.23±1.75**	**77.36±1.49**	**84.89±1.27**
ALOI	SCSL[1]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	SIMVC-SA	50.67±0.55	68.78±0.24	53.36±0.54	37.59±0.62	60.55±0.35	40.16±0.60	29.68±0.28	55.13±0.28	32.66±0.33	21.80±0.44	48.85±0.30	24.77±0.39
	LSIMVC	41.05±0.34	62.06±0.96	44.69±2.23	32.57±4.93	43.98±5.68	37.54±5.99	41.27±1.07	51.96±1.62	45.96±1.16	25.99±2.96	34.85±3.57	27.87±3.25
	OIMC-IAGD	60.96±1.74	75.81±0.32	62.85±1.41	55.77±1.32	70.96±0.49	57.51±1.31	43.72±1.32	62.52±0.59	46.27±1.03	33.09±0.64	54.39±0.52	35.81±0.91
	OS-LF-IMVC	41.58±0.87	66.55±0.46	45.35±0.80	33.31±0.33	59.33±0.25	37.18±0.31	26.51±0.28	52.36±0.31	30.14±0.22	20.85±0.46	45.03±0.27	23.79±0.48
	CAL[2]	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM	OOM
	OAGL	84.09±2.22	91.71±0.45	85.44±1.64	83.92±1.90	91.77±0.30	85.33±1.33	81.86±1.70	91.58±0.40	83.96±1.19	81.23±2.58	91.95±0.34	83.90±1.81
	OCGDL (ours)	**85.44±1.08**	**93.46±0.31**	**87.15±1.01**	**91.39±1.09**	**94.87±0.18**	**91.88±1.12**	**91.11±2.32**	**95.06±0.37**	**91.76±2.05**	**86.24±2.34**	**94.28±0.71**	**88.21±1.84**
Handwritten	SCSL[1]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	SIMVC-SA	80.55±8.04	73.27±4.42	81.61±6.42	76.24±7.16	68.60±4.04	77.82±5.52	69.90±3.22	59.51±2.51	70.36±2.87	55.89±2.08	45.17±1.38	56.61±1.83
	LSIMVC	85.33±6.47	83.65±3.32	86.14±5.70	85.14±7.67	81.61±4.00	85.91±6.21	82.37±4.29	75.83±2.38	83.04±3.26	74.12±2.45	68.92±1.16	75.27±2.26
	OIMC-IAGD	86.20±6.76	85.07±2.37	88.25±5.00	89.71±4.80	83.46±1.93	90.11±3.94	77.86±4.67	72.26±2.34	78.67±3.60	66.11±5.68	62.65±3.00	69.84±4.09
	OS-LF-IMVC	72.26±5.02	66.73±1.97	73.47±3.74	62.23±2.65	58.34±1.69	64.50±1.96	55.34±4.04	47.88±1.98	57.16±3.65	35.89±2.66	29.69±2.12	37.79±2.91
	CAL	**100.00±0.02**	**99.99±0.04**	**100.00±0.02**	99.97±0.04	99.94±0.11	99.97±0.04	99.94±0.06	99.84±0.16	99.94±0.06	**99.77±0.13**	**99.39±0.34**	**99.77±0.13**
	OAGL	98.52±4.68	99.54±1.45	99.00±3.16	99.96±0.03	99.91±0.09	99.96±0.03	98.50±4.17	99.14±1.30	98.83±3.11	99.45±0.19	98.65±0.42	99.45±0.19
	OCGDL (ours)	99.76±0.13	99.35±0.37	99.76±0.13	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	98.90±3.15	99.42±0.37	98.90±3.15
ORL40[3]	SCSL	76.35±2.06	87.40±1.03	79.10±1.80	64.58±5.20	81.77±4.60	69.73±6.12	16.12±2.41	30.56±4.75	18.65±2.13	–	–	–
	SIMVC-SA	78.52±1.52	89.84±0.83	81.46±1.26	74.45±2.01	86.85±1.40	77.46±2.09	61.00±3.36	75.97±2.11	63.40±2.91	–	–	–
	LSIMVC	71.03±2.25	82.37±1.24	73.90±1.65	61.88±3.08	75.85±2.05	65.35±2.35	48.00±2.54	66.43±1.60	52.50±1.67	–	–	–
	OIMC-IAGD	60.23±3.85	78.12±2.35	63.48±3.47	57.33±1.86	76.39±1.66	61.72±1.41	48.62±1.92	68.63±1.80	52.57±1.82	–	–	–
	OS-LF-IMVC	74.38±3.59	86.77±1.68	76.95±2.97	63.28±3.80	78.72±1.90	65.77±3.56	47.08±1.50	67.35±1.34	49.70±1.62	–	–	–
	CAL	95.40±0.48	98.42±0.20	96.43±0.45	96.57±0.60	98.67±0.31	97.26±0.48	92.43±2.42	95.98±1.08	93.58±2.10	–	–	–
	OAGL	90.80±4.15	97.39±0.98	93.03±2.95	93.30±3.85	97.34±1.03	94.62±2.70	88.80±4.42	95.82±1.27	91.38±3.22	–	–	–
	OCGDL (ours)	**99.78±0.18**	**99.80±0.17**	**99.78±0.18**	**97.42±2.02**	**98.94±0.64**	**97.60±1.70**	**96.75±2.37**	**98.64±0.55**	**96.92±1.73**	–	–	–
BBCSport[3]	SCSL	79.54±6.33	66.80±5.15	80.09±5.49	72.96±14.62	57.16±21.18	74.78±14.46	45.92±10.16	26.86±10.01	52.90±9.15	–	–	–
	SIMVC-SA	80.08±3.52	65.52±1.87	81.66±0.79	75.88±4.11	59.52±3.76	78.99±1.56	45.55±5.18	27.16±6.35	54.47±4.76	–	–	–
	LSIMVC	86.91±2.12	72.97±2.07	86.91±2.12	83.53±5.03	69.24±2.94	84.63±2.39	43.44±4.56	24.15±6.54	51.38±4.55	–	–	–
	OIMC-IAGD	74.23±8.69	58.71±8.57	76.86±5.98	71.93±9.66	51.87±7.92	73.51±7.70	38.62±5.50	20.06±7.07	48.53±6.06	–	–	–
	OS-LF-IMVC	78.22±8.03	63.91±5.95	81.47±4.73	72.63±7.99	58.43±5.38	77.83±4.21	41.47±3.66	21.44±3.30	52.34±3.22	–	–	–
	CAL	99.14±0.17	97.11±0.53	99.14±0.17	94.72±1.20	84.04±2.87	94.72±1.20	74.56±5.04	68.29±4.62	78.84±3.62	–	–	–
	OAGL	**99.93±0.10**	**99.71±0.38**	**99.93±0.10**	97.32±5.23	96.58±5.56	97.56±4.68	90.92±8.77	88.18±6.46	92.94±5.44	–	–	–
	OCGDL (ours)	99.87±0.15	99.52±0.56	99.87±0.15	**99.15±0.26**	**97.17±0.74**	**99.15±0.26**	**97.79±5.06**	**96.65±4.50**	**98.25±3.67**	–	–	–

[1] SCSL is N/A in Handwritten and ALOI because it require $d_v > k$.

[2] CAL reports OOM on ALOI due to insufficient memory.

[3] "–" denotes an inapplicable missing rate, as discussed in Section 4.1.



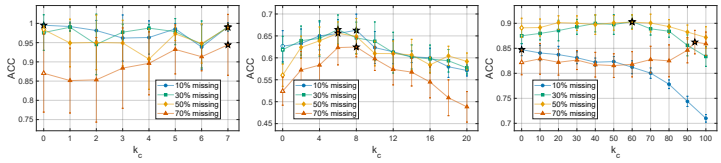
(a) OS-LF-IMVC (b) CAL (c) OAGL (d) OCGDL

Figure 2: t-SNE visualization of the learned representations obtained by different methods on MSRCv1 under a 30% missing rate. Colors denote the ground-truth classes.

4.2 Performance Comparison and Analysis

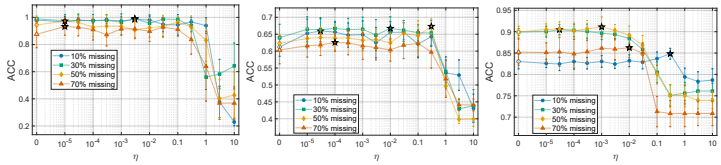
Table 1 reports the performance of different algorithms under different missing rates in the six datasets. Overall, OCGDL achieves highly competitive results across datasets and missing levels, and is particularly strong on challenging large-scale or multi-class settings. On Caltech101-20 and ALOI, OCGDL ranks first on nearly all reported metrics under all missing rates, with clear margins over both two-stage graph methods and one-step competitors. On MSRCv1 and Handwritten, OCGDL remains on par with the strongest baselines, i.e., CAL and OAGL, and often attains the best or second-best scores. These results suggest that jointly modeling shared and view-specific graph dictionaries, together with compact tensor coupling and one-step spectral rotation, yields robust embeddings when view observations become sparse.

To provide an intuitive comparison of representation quality, we visualize the embeddings using t-SNE in Figure 2. OCGDL produces a well-organized embedding, where same-class samples form compact groups and different classes are well separated. OAGL and CAL also recover relatively distinct clusters, yet OCGDL exhibits particularly clean inter-class margins on this example, whereas OS-LF-IMVC still shows noticeable overlap. These qualitative observations further support that shared and specific graph-dictionary learning with cross-view tensor regularization helps retain cluster geometry under missing views. Complete t-SNE plots of all compared methods are provided in Appendix A.4.



(a) MSRCv1 (b) Caltech101-20 (c) ALOI

Figure 3: Sensitivity of OCGDL to the shared dimension k_c under different missing rates.



(a) MSRCv1 (b) Caltech101-20 (c) ALOI

Figure 4: Sensitivity of OCGDL to the sparsity weight η under different missing rates.

4.3 Parameter Sensitivity Study

In the following subsection, we only present results on 3 datasets due to page limitations. The complete results on all six datasets are provided in the Appendix A.4. We first present the effect of k_c, which allocates capacity between the view-shared and view-specific graph dictionaries, as shown in Figure 3. Overall, the preferred k_c range depends on the dataset. With a proper k_c, OCGDL outperforms both the shared-only extreme ($k_c = k$) and the private-only extreme ($k_c = 0$) on Caltech101-20 and in most ALOI settings, supporting the effectiveness of the shared-specific graph embedding decomposition. We also observe a trend that when the missing rate is low, smaller k_c is often preferable on MSRCv1 and ALOI, indicating that view-specific cues remain informative. As missing rates increase, larger k_c becomes more beneficial, suggesting that consensus across views is more reliable once view-specific neighborhoods are too incomplete to be trusted.

We also present how the sparsity weight η that controls row-wise $\ell_{2,1}$ selection of graph dictionary atoms affects the performance, as shown in Figure 4. Across datasets, a moderate η generally yields stable and better ACC than that of $\eta = 0$, especially with 30% and 50% missing rates, supporting the effectiveness of group sparse coding in graph dictionary learning.

4.4 Ablation Study

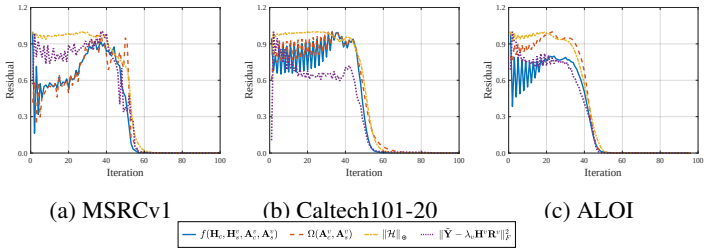
To isolate the contribution of each component, we compare OCGDL with five variants in Table 2 at a 50% missing rate. Overall, the full OCGDL model achieves the best or tied-best performance on nearly all metrics across the six datasets. V1, which removes the shared dictionary, consistently degrades performance, confirming that a consensus spectral subspace is necessary for cross-view alignment under incompleteness. V2 disables $\ell_{2,1}$ sparsity, which also hurts, especially on MSRCv1, indicating that capacity-aware atom selection stabilizes reconstruction when observations are noisy or partial. V3 replaces the group $\ell_{2,1}$ sparsity with an element-wise ℓ_1 norm also degrades performance, suggesting that unstructure sparsity yields a more biased dictionary than row-wise group selection. V4 replaces the $k\times V\times n$ embedding tensor with a affinity-style tensor $\hat{\mathcal{H}} = \{\mathbf{H}^v\mathbf{H}^{v\top}\}_{v=1}^{V} \in \mathbb{R}^{n\times V\times n}$, which is consistently inferior, suggesting that imposing TNN on compact graph-dictionary embeddings better couples cross-view structure than low-rankness on reconstructed affinities. V5 discards sparse coefficients relaxation and reconstructs $\mathbf{W}^v$ directly by Eq. (2), which degrades the clustering performance except for Caltech101-20, showing that flexible coefficient coding is important for OCGDL.

4.5 Convergence Analysis and Running Time Comparison

We further present the convergence diagram of module-wise residuals under a 50% missing rate in Figure 5. Generally, OCGDL can typically converge within 100 iterations in different datasets. Tensor-consistency and one-step clustering errors decrease smoothly from early iterations, while the similarity reconstruction error and the sparsity term tend to show a short plateau followed by

Table 2: The ablation comparison of OCGDL at 50% missing ratio. V1 removes the shared dictionary ($k_c = 0$); V2 removes sparsity regularization ($\eta = 0$); V3 replaces row-wise $\ell_{2,1}$ sparsity with element-wise ℓ_1; V4 applies TNN to the affinity-style tensor $\hat{\mathcal{H}} = \{\mathbf{H}^v\mathbf{H}^{v\top}\}_{v=1}^{V} \in \mathbb{R}^{n\times V\times n}$; V5 reconstructs $\mathbf{W}^v$ by Eq. (2) without sparse codes.

Variant	MSRCv1			Caltech101-20			ALOI			Handwritten			ORL40			BBCSport		
	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
Full	**97.62±2.02**	**95.59±3.23**	**97.62±2.02**	**65.61±2.58**	79.23±1.74	86.53±1.53	**91.11±2.32**	**95.06±0.37**	**91.76±2.05**	**100.00±0.00**	**100.00±0.00**	**100.00±0.00**	**96.75±2.37**	**98.64±0.55**	**96.92±1.73**	**97.79±5.06**	**96.65±4.50**	**98.25±3.67**
V1	93.33±7.32	92.52±6.12	93.38±7.21	56.13±2.68	76.62±1.25	86.40±1.12	89.02±0.78	94.21±0.15	89.02±0.78	99.97±0.03	99.92±0.10	99.97±0.03	91.33±3.67	96.13±1.48	92.53±2.97	92.35±10.49	91.49±9.90	93.82±8.28
V2	90.33±10.90	90.11±7.84	91.14±9.69	62.28±2.46	78.45±1.09	86.45±1.16	89.73±2.00	94.63±0.28	89.97±1.97	99.98±0.05	99.96±0.13	99.98±0.05	96.28±1.29	98.09±0.61	96.60±0.98	92.26±12.27	90.39±12.98	94.26±8.94
V3	96.74±2.38	94.28±3.47	96.74±2.38	62.82±5.50	78.59±1.95	**86.76±1.67**	88.41±1.87	94.36±0.31	89.82±1.42	99.99±0.02	99.98±0.04	99.99±0.02	96.18±2.62	98.41±0.96	96.75±2.25	96.91±6.08	94.62±7.03	97.68±3.80
V4	94.19±6.90	92.70±4.30	94.57±6.05	59.06±3.49	76.41±1.91	85.18±1.22	80.35±1.25	93.23±0.29	81.20±1.01	99.99±0.02	99.97±0.06	99.99±0.02	94.90±1.55	97.81±0.46	95.20±1.19	91.38±10.72	87.42±11.08	92.81±7.50
V5	91.86±6.95	87.99±6.37	91.86±6.95	62.51±2.94	**79.67±1.63**	85.75±1.43	87.20±5.67	87.24±1.25	74.18±3.62	99.61±0.22	99.15±0.40	99.61±0.22	84.25±6.49	91.18±2.39	84.38±6.57	92.06±10.54	90.26±5.82	94.82±4.89



(a) MSRCv1 (b) Caltech101-20 (c) ALOI

Figure 5: Module-wise convergence of OCGDL in different datasets under a 50% missing rate.

a clear drop after a certain number of iterations. These consistent patterns empirically support the numerical stability of the proposed ADMM-style alternating optimization.

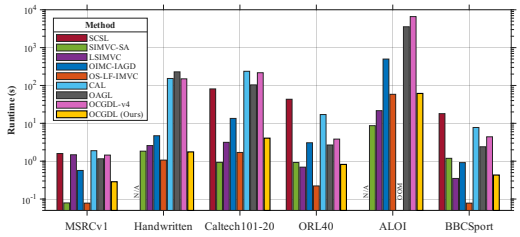


Figure 6: Running time comparison of the different methods on six benchmark datasets under a 50% missing rate. OCGDL-V4 denotes the ablation variant in Table 2.

Figure 6 compares the convergence running time costs. OCGDL maintains a practical runtime across all six datasets. We additionally report OCGDL-V4 that applies TNN to the $n\times V\times n$ affinity tensor, which requires substantially more time than OCGDL, suggesting that coupling views through the stacked graph-dictionary embeddings improves not only accuracy but also scalability.

5 Conclusion

In this paper, we presented OCGDL, a one-step consistency-aware graph dictionary learning framework for incomplete multi-view clustering. Starting from a spectral perturbation view of multi-view affinities, OCGDL decomposes each similarity graph into shared and view-specific graph dictionary atoms under orthogonality and orthogonal-complement constraints, while missing indicators prevent reconstruction bias on unobserved blocks. Row-wise $\ell_{2,1}$ sparsity regulates atom selection across views, and a tensor nuclear norm on the compact stacked graph-dictionary tensor couples high-order cross-view correlations at the embedding level with favorable computational complexity. Discrete cluster indicators are obtained jointly via spectral rotation with adaptive view weights, avoiding unstable post-hoc k-means. Extensive experiments on six benchmarks under diverse missing rates demonstrate that OCGDL achieves competitive or superior clustering performance against representative IMVC baselines in different metrics, produces clearer cluster geometry in t-SNE visualizations, and benefits from each of its key components according to ablation studies. Sensitivity analyses further show that moderate sparsity weights and carefully allocated shared dimensions are important for balancing consistency and complementarity. Runtime and convergence studies confirm that the compact tensor design improves efficiency over affinity-tensor alternatives and that the alternating optimization is empirically stable. Future work includes extending OCGDL to large-scale settings, exploring adaptive allocation of k_c and k_s without grid search, and investigating tighter theoretical guarantees for masked graph-dictionary factorization under missing views.

Reproducibility statement

To ensure the reproducibility of our work, the complete source code of the OCGDL will be provided in the supplementary materials. The public datasets used in the experiments are all cited in the paper and the construction of incomplete datasets can be carried out according to the specific methods mentioned in Section 4.1 of the main text.

Ethics statement

This paper does not involve any potential ethics issues.

AI use statement

We employ LLM-based AI to refine the language and wording for improved fluency and precision of the manuscript, with no further application beyond this purpose.

References

Kamal Berahmand, Farid Saberi-Movahed, Razieh Sheikhpour, Yuefeng Li, and Mahdi Jalili. A comprehensive survey on spectral clustering with graph structure learning. *arXiv preprint arXiv:2501.13597*, 2025.

Xiaochun Cao, Changqing Zhang, Huazhu Fu, Si Liu, and Hua Zhang. Diversity-induced multi-view subspace clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–594, 2015.

Jie Chen, Zhiqiang Wang, Hua Mao, and Xi Peng. Low-rank tensor learning for incomplete multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 35(11):11556–11569, 2023a. doi: 10.1109/TKDE.2022.3230964.

Jie Chen, Yingke Chen, Zhu Wang, Haixian Zhang, and Xi Peng. Spectral embedding fusion for incomplete multiview clustering. *IEEE Transactions on Image Processing*, 33:4116–4130, 2024.

Jie Chen, Hua Mao, Wai Lok Woo, Chuanbin Liu, Zhu Wang, and Xi Peng. One-step adaptive graph learning for incomplete multiview subspace clustering. *IEEE Transactions on Knowledge and Data Engineering*, 37(5):2771–2783, 2025.

Yongyong Chen, Xiaolin Wang, Chong Han, and Yigang Liu. Enhanced tensor low-rank and sparse representation recovery for incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 10696–10704, 2023b. doi: 10.1609/aaai.v37i9.26323.

Yongyong Chen, Xiaojia Zhao, Zheng Zhang, Youfa Liu, Jingyong Su, and Yicong Zhou. Tensor learning meets dynamic anchor learning: From complete to incomplete multiview clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2023c.

Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970. doi: 10.1137/0707001.

Alan Edelman, Tomás A Arias, and Steven T Smith. The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(2):303–353, 1998.

Hongchang Gao, Feiping Nie, Xuelong Li, and Heng Huang. Multi-view subspace clustering. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4238–4246, 2015.

Derek Greene and Pádraig Cunningham. Practical solutions to the problem of diagonal dominance in kernel document clustering. In *Proceedings of the 23rd international conference on Machine learning*, pp. 377–384, 2006.

Zhibin Gu, Zhenhao Zhong, Xi Zhang, and Bing Li. Constant degree matrix-driven incomplete multi-view clustering via connectivity-structure and embedding tensor learning. In *The Fourteenth International Conference on Learning Representations*, 2026.

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

Jin Huang, Feiping Nie, and Heng Huang. Spectral rotation versus k-means in spectral clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 27, pp. 431–437, 2013.

Haobin Li, Yunfan Li, Mouxing Yang, Peng Hu, Dezhong Peng, and Xi Peng. Incomplete multi-view clustering via prototype-based imputation. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pp. 3911–3919, 2023a.

Jing Li, Quanxue Gao, Qianqian Wang, Ming Yang, and Wei Xia. Orthogonal non-negative tensor factorization based multi-view clustering. In *Advances in Neural Information Processing Systems*, volume 36, pp. 18186–18202, 2023b.

Lusi Li, Zhiqiang Wan, and Haibo He. Incomplete multi-view clustering with joint partition and graph learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(1):589–602, 2021.

Shao-Yuan Li, Yuan Jiang, and Zhi-Hua Zhou. Partial multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28, 2014.

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, 2020.

Yingming Li, Ming Yang, and Zhongfei Zhang. A survey of multi-view representation learning. *IEEE Transactions on Knowledge and Data Engineering*, 31(10):1863–1883, 2018.

Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. Completer: Incomplete multi-view clustering via contrastive prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11174–11183, 2021.

Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4447–4461, 2023.

Chengliang Liu, Zhihao Wu, Jie Wen, Yong Xu, and Chao Huang. Localized sparse incomplete multi-view clustering. *IEEE Transactions on Multimedia*, 25:5539–5551, 2022a.

Chengliang Liu, Jie Wen, Yong Xu, Bob Zhang, Liqiang Nie, and Min Zhang. Reliable representation learning for incomplete multi-view missing multi-label classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6):4940–4956, 2025a.

Haolu Liu, Xiyue Wang, Xuanting Xie, Liangjian Wen, and Zhao Kang. Rethinking incompleteness: Formalizing protocol divergence and train-once learning for robust imvc. *arXiv preprint arXiv:2606.04857*, 2026.

Jing Liu, Yu Jiang, Zechao Li, Zhi-Hua Zhou, and Hanqing Lu. Partially shared latent factor learning with multiview data. *IEEE Transactions on Neural Networks and Learning Systems*, 26(6):1233–1246, 2014.

Qian Liu, Huibing Wang, Jinjia Peng, Yawei Chen, Mingze Yao, Xianping Fu, and Yang Wang. Consensus-guided incomplete multi-view clustering via cross-view affinities learning. In *IJCAI*, pp. 5761–5769, 2025b.

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

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

Canyi Lu, Jiashi Feng, Yudong Chen, Wei Liu, Zhouchen Lin, and Shuicheng Yan. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE transactions on pattern analysis and machine intelligence*, 42(4):925–938, 2019.

Shirui Luo, Changqing Zhang, Wei Zhang, and Xiaochun Cao. Consistent and specific multi-view subspace clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*, volume 14, pp. 849–856, 2002.

Feiping Nie, Lai Tian, and Xuelong Li. Multiview clustering via adaptively weighted procrustes. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2022–2030, 2018.

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

Chang Tang, Zhenglai Li, Jun Wang, Xinwang Liu, Wei Zhang, and En Zhu. Unified one-step multi-view spectral clustering. *IEEE Transactions on Knowledge and Data Engineering*, 35(6): 6449–6460, 2023. doi: 10.1109/TKDE.2022.3172687.

Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007. doi: 10.1007/s11222-007-9033-z.

Hao Wang, Yan Yang, and Bing Liu. Gmc: Graph-based multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 32(6):1116–1129, 2019.

Qianqian Wang, Zhengming Ding, Zhiqiang Tao, Quanxue Gao, and Yun Fu. Generative partial multi-view clustering with adaptive fusion and cycle consistency. *IEEE Transactions on Image Processing*, 30:1771–1783, 2021.

Yiming Wang, Dongxia Chang, Zhiqiang Fu, Jie Wen, and Yao Zhao. Incomplete multiview clustering via cross-view relation transfer. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(1):367–378, 2022.

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

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

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

Jie Wen, Gehui Xu, Zhanyan Tang, Wei Wang, Lunke Fei, and Yong Xu. Graph regularized and feature aware matrix factorization for robust incomplete multi-view clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(5):3728–3741, 2023b.

Jie Wen, Zheng Zhang, Lunke Fei, Bob Zhang, Yong Xu, Zhao Zhang, and Jinxing Li. A survey on incomplete multiview clustering. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(2):1136–1149, 2023c. doi: 10.1109/TSMC.2022.3192635.

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

Jinghan Wu, Rankun Chen, Haoran Wang, Xuetao Zhang, Ben Yang, and Badong Chen. One-step incomplete multi-view clustering via imputed anchor graph discretization. *Knowledge-Based Systems*, pp. 114935, 2025.

Wei Xia, Quanxue Gao, Qianqian Wang, and Xinbo Gao. Tensor completion-based incomplete multiview clustering. *IEEE Transactions on Cybernetics*, 52(12):13635–13644, 2022.

Yuan Xie, Dacheng Tao, Wensheng Zhang, Yan Liu, Lei Zhang, and Yanyun Qu. On unifying multi-view self-representations for clustering by tensor multi-rank minimization. *International Journal of Computer Vision*, 126(11):1157–1179, 2018.

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

Deng Xu, Chao Zhang, Cong Guo, Chunlin Chen, and Huaxiong Li. Fast incomplete multi-view clustering with adaptive similarity completion and reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 21734–21742, 2025.

Dongkuan Xu and Yingjie Tian. A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2(2):165–193, 2015.

Yan Yang and Hao Wang. Multi-view clustering: A survey. *Big Data Mining and Analytics*, 1(2): 83–107, 2018.

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

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.

Kun Zhan, Changqing Zhang, Junpeng Guan, and Junsheng Wang. Graph learning for multiview clustering. *IEEE Transactions on Cybernetics*, 48(10):2887–2895, 2017.

Changqing Zhang, Huazhu Fu, Si Liu, Guangcan Liu, and Xiaochun Cao. Low-rank tensor constrained multiview subspace clustering. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1582–1590, 2015.

Yi Zhang, Xinwang Liu, Siwei Wang, Jiyuan Liu, Sisi Dai, and En Zhu. One-stage incomplete multi-view clustering via late fusion. In *Proceedings of the 29th ACM international conference on multimedia*, pp. 2717–2725, 2021.

Haoli Zhao, Peng Zhong, Haiqin Chen, Zhenni Li, Wuhui Chen, and Zibin Zheng. Group non-convex sparsity regularized partially shared dictionary learning for multi-view learning. *Knowledge-Based Systems*, 242:108364, 2022.

Xiaofeng Zhu, Shichao Zhang, Wei He, Rongyao Hu, Cong Lei, and Pengfei Zhu. One-step multi-view spectral clustering. *IEEE Transactions on Knowledge and Data Engineering*, 31(10):2022–2034, 2018.

A APPENDIX

A.1 NOTATIONS

In this paper, the boldface uppercase letter, like $\mathbf{A}$, denotes a matrix, $\mathbf{a}_i$ denotes the ith column of the matrix, and $\mathbf{a}_{i:}$ denotes the ith row of the matrix. Boldface lowercase letter $\mathbf{a}$ denotes a vector, and a_i denotes the ith entry of $\mathbf{a}$. The upper right corner of $\mathbf{A}^v$ is the view number v. $\mathbf{A}^\top$ denotes the transpose of the matrix $\mathbf{A}$. Specific notations are listed in Table 3.

Table 3: Summary of the main notations used in OCGDL.

Notation	Description
V	Number of views.
n	Number of samples.
d_v	Feature dimension of the v-th view.
k	Number of clusters.
k_c	Dimension of the shared graph dictionary.
k_s	Dimension of the view-specific graph dictionary, where $k_s = k - k_c$.
$\mathbf{X}^v$	Data matrix of the v-th view, $\mathbf{X}^v \in \mathbb{R}^{n\times d_v}$.
$\mathbf{P}^v$	Observation-indicator matrix of the v-th view, $\mathbf{P}^v \in \{0,1\}^{n\times n}$.
$\mathbf{W}^v$	Similarity graph of the v-th view, $\mathbf{W}^v \in \mathbb{R}^{n\times n}$.
$\widehat{\mathbf{W}}^v$	Partially observed similarity graph of the v-th view.
$\mathbf{H}_c$	Shared graph dictionary, $\mathbf{H}_c \in \mathbb{R}^{n\times k_c}$.
$\mathbf{H}_s^v$	View-specific graph dictionary of the v-th view, $\mathbf{H}_s^v \in \mathbb{R}^{n\times k_s}$.
$\mathbf{H}^v$	Full graph dictionary of the v-th view, $\mathbf{H}^v = [\mathbf{H}_c, \mathbf{H}_s^v]$.
$\mathbf{A}_c^v$	Sparse coefficient matrix of the shared graph dictionary, $\mathbf{A}_c^v \in \mathbb{R}^{k_c\times n}$.
$\mathbf{A}_s^v$	Sparse coefficient matrix of the view-specific graph dictionary, $\mathbf{A}_s^v \in \mathbb{R}^{k_s\times n}$.
$\mathcal{H}$	Stacked full graph-dictionary tensor, $\mathcal{H} \in \mathbb{R}^{k\times V\times n}$.
$\mathbf{Y}$	Discrete binary clustering indicator matrix, $\mathbf{Y} \in \{0,1\}^{n\times k}$.
$\widetilde{\mathbf{Y}}$	Normalized clustering indicator, $\widetilde{\mathbf{Y}} = \mathbf{Y}(\mathbf{Y}^\top\mathbf{Y})^{-1/2}$.
$\mathbf{R}^v$	Orthogonal rotation matrix of the v-th view, $\mathbf{R}^v \in \mathbb{R}^{k\times k}$.
λ_v	Adaptive weight of the v-th view.
λ	Non-negative view-weight vector, $\lambda \in \mathbb{R}^V$.
$\mathcal{T}$	Auxiliary tensor used in tensor nuclear norm optimization.
$\mathcal{M}$	Lagrange multiplier tensor in the ADMM optimization.
η	Sparsity regularization parameter.
β	One-step clustering trade-off parameter.
μ	ADMM penalty parameter.
$\|\cdot\|_F$	Frobenius norm.
$\|\cdot\|_{2,1}$	Row-wise $\ell_{2,1}$ norm.
$\|\cdot\|_\circledast$	Tensor nuclear norm.

A.2 DAVIS–KAHAN BACKGROUND FOR SHARED-SPECIFIC SPECTRAL DECOMPOSITION

Lemma 1 (Davis–Kahan $\sin\Theta$ theorem (Davis & Kahan, 1970; Yu et al., 2015)). *Let* $\mathbf{A}, \widehat{\mathbf{A}} \in \mathbb{R}^{n\times n}$ *be symmetric matrices with eigenvalue decompositions*

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T, \quad \widehat{\mathbf{A}} = \widehat{\mathbf{U}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{U}}^T, \tag{18}$$

where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, $\widehat{\mathbf{\Lambda}} = \mathrm{diag}(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_n)$, $\lambda_1 \geq \cdots \geq \lambda_n$, $\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_n$, *and both* $\mathbf{U}$ *and* $\widehat{\mathbf{U}}$ *are orthogonal. Fix integers* $1 \leq r \leq s \leq n$ *and set* $k = s - r + 1$. *Write*

$$\mathbf{U}_{r:s} = \mathbf{U}_{:,r:s} \in \mathbb{R}^{n \times k}, \quad \widehat{\mathbf{U}}_{r:s} = \widehat{\mathbf{U}}_{:,r:s} \in \mathbb{R}^{n \times k} \tag{19}$$

for the orthonormal eigenbases associated with $\{\lambda_r, \ldots, \lambda_s\}$ *and* $\{\widehat{\lambda}_r, \ldots, \widehat{\lambda}_s\}$, *respectively. Let* $\mathbf{\Theta}(\mathbf{U}_{r:s}, \widehat{\mathbf{U}}_{r:s})$ *be the diagonal matrix of principal angles between* $\mathrm{range}(\mathbf{U}_{r:s})$ *and* $\mathrm{range}(\widehat{\mathbf{U}}_{r:s})$, *and define the eigengap*

$$\gamma \;=\; \min\bigl\{\lambda_{r-1} - \lambda_r,\; \lambda_s - \lambda_{s+1}\bigr\} \;>\; 0, \tag{20}$$

with the conventions $\lambda_0 = +\infty$ *and* $\lambda_{n+1} = -\infty$. *Then*

$$\bigl\|\sin \mathbf{\Theta}(\mathbf{U}_{r:s}, \widehat{\mathbf{U}}_{r:s})\bigr\|_F \;\leq\; \frac{2\min\bigl\{\sqrt{k}\,\|\widehat{\mathbf{A}} - \mathbf{A}\|_2,\; \|\widehat{\mathbf{A}} - \mathbf{A}\|_F\bigr\}}{\gamma}. \tag{21}$$

Application to multi-view affinities. Setting $\mathbf{A} = \mathbf{W}^\star$ and $\widehat{\mathbf{A}} = \mathbf{W}^v = \mathbf{W}^\star + \mathbf{E}^v$ in Lemma 1 immediately yields the $\sin\Theta$ bound in Corollary 1. The subsequent shared and view-specific split $\mathbf{H}^v\mathbf{R}^v = [\mathbf{H}_c, \mathbf{H}_s^v]$ follows by orthogonally aligning the leading eigenspace of $\mathbf{W}^v$ and partitioning the aligned basis into a cross-view stable block of size k_c and a complementary view-specific block of size $k_s = k - k_c$. This provides the spectral motivation for the partially shared graph-dictionary model in the main text.

Corollary 1 (Shared and view-specific spectral factors)**.** *Let* $\mathbf{W}^\star \in \mathbb{R}^{n \times n}$ *be a latent consensus affinity whose leading* k *eigenvalues are separated by the gap* $\gamma^\star = \lambda_k^\star - \lambda_{k+1}^\star > 0$, *and let* $\mathbf{H}^\star \in \mathbb{R}^{n \times k}$ *collect the corresponding orthonormal eigenvectors. For each view* v, *write*

$$\mathbf{W}^v = \mathbf{W}^\star + \mathbf{N}^v, \tag{22}$$

where $\mathbf{N}^v$ *is a view-specific information and perturbation, and let* $\mathbf{H}^v \in \mathbb{R}^{n \times k}$ *collect the leading orthonormal eigenvectors of* $\mathbf{W}^v$. *Then the Davis–Kahan* $\sin\Theta$ *theorem yields*

$$\bigl\|\sin \mathbf{\Theta}\bigl(\mathbf{H}^\star, \mathbf{H}^v\bigr)\bigr\|_F \;\leq\; \frac{2\min\bigl\{\sqrt{k}\,\|\mathbf{N}^v\|_2,\; \|\mathbf{N}^v\|_F\bigr\}}{\gamma^\star}. \tag{23}$$

Consequently, there exist orthogonal alignments $\mathbf{R}^v \in \mathbb{R}^{k \times k}$ *such that*

$$\mathbf{H}^v\mathbf{R}^v \;=\; \bigl[\mathbf{H}_c,\; \mathbf{H}_s^v\bigr], \tag{24}$$

where $\mathbf{H}_c \in \mathbb{R}^{n \times k_c}$ *spans cross-view shared spectral features aligned with* $\mathbf{H}^\star$, *and* $\mathbf{H}_s^v \in \mathbb{R}^{n \times k_s}$ *spans view-specific spectral features induced by* $\mathbf{N}^v$, *with* $k_s = k - k_c$.

A.3 Algorithm and Computational Complexity Analysis

We summarize the algorithm for the overall objective in Eq. (6) in Algorithm 1. Let d denote the average number of nonzero entries per row of the sparse similarity graph $\hat{\mathbf{W}}^v$, which is independent of the feature dimension d_v. The computational complexity of OCGDL consists of the following updates. First, updating $\mathbf{Y}$ requires forming $\mathbf{Q} = \sum_{v=1}^{V} \lambda_v \mathbf{H}^v\mathbf{R}^v$ and thus costs $\mathcal{O}(Vnk^2)$. Second, updating $\mathbf{R}^v$ by the SVD of $\mathbf{H}^{v\top}\tilde{\mathbf{Y}} \in \mathbb{R}^{k \times k}$ costs $\mathcal{O}(Vnk^2)$ over all views, where the $k \times k$ SVD is absorbed by the matrix product. Next, updating $\mathbf{H}_c$ and $\mathbf{H}_s^v$ costs $\mathcal{O}(V(dnk + nk^2))$, because the masked residual is evaluated by sparse–dense products with $\hat{\mathbf{W}}^v$ together with $n \times k$ Gram multiplies, followed by polar factorizations of size $n \times k_c$ and $n \times k_s$; the orthogonal-complement projection is applied as $(\mathbf{I}_n - \mathbf{H}_c\mathbf{H}_c^\top)\mathbf{R}_s^v$ without materializing $\mathbf{I}_n$. Then, updating $\mathcal{T}$ by tensor singular-value thresholding on the compact tensor $\mathcal{H} \in \mathbb{R}^{k \times V \times n}$ costs $\mathcal{O}(nkV \log n + nkV \min(k, V))$, where the first term is the FFT along the sample mode and the second term is the economy SVD of the $k \times V$ frontal slices. Subsequently, updating $\boldsymbol{\lambda}$ costs $\mathcal{O}(Vnk^2)$. Finally, updating $\mathbf{A}_c^v$ and $\mathbf{A}_s^v$ uses the identity $\mathbf{H}^\top(\mathbf{H}\mathbf{A} - \hat{\mathbf{W}}^v) = (\mathbf{H}^\top\mathbf{H})\mathbf{A} - \mathbf{H}^\top\hat{\mathbf{W}}^v$: the sparse–dense products and Gram matrices are formed once at $\mathcal{O}(V(dnk + nk^2))$, after which each of the T_{in} inner proximal-gradient iterations costs $\mathcal{O}(Vnk^2)$, giving $\mathcal{O}(V(dnk + nk^2) + T_{\text{in}}Vnk^2)$ in total. The dual update of $\mathcal{M}$ is only $\mathcal{O}(nkV)$. Since $nkV \min(k, V) \leq Vnk^2$ and T_{in} is a small constant independent of n, the complexity of Algorithm 1 is $\mathcal{O}\bigl(t(Vnk^2 + Vdnk + nkV \log n)\bigr)$, where t is the number of outer iterations. The final computational complexity of Algorithm 1 is $\mathcal{O}(tn \log n)$ since $V \ll n$, $k \ll n$ and $d = \mathcal{O}(1)$.

Algorithm 1 The proposed OCGDL algorithm by solving the problem in Eq. (6)

Input: The partially observed similarity graph $\{\hat{\mathbf{W}}^v\}_{v=1}^V$, $\{\mathbf{P}^v\}_{v=1}^V$, k, k_c, η, β.

Initialize: $\rho = 1.2$, $\mu = 10^{-4}$, $\mu_{\max} = 10^6$, $\mathrm{tol} = 10^{-5}$, $t = 1$ and $\mathrm{maxIters} = 200$;
while not converged **do**
  Update $\mathbf{Y}$ via Eq. (8);
  Update $\{\mathbf{R}^v\}_{v=1}^V$ via Eq. (9);
  Update $\mathbf{H}_c$ via Eq. (10);
  Update $\{\mathbf{H}_s^v\}_{v=1}^V$ via Eq. (12);
  Update $\mathcal{T}$ via Eq. (13);
  Update $\boldsymbol{\lambda}$ via Eq. (14);
  Update $\{\mathbf{A}_c^v, \mathbf{A}_s^v\}_{v=1}^V$ via Eq. (15) and Eq. (17) respectively;
  Update the Lagrange multiplier $\mathcal{M}$ via $\mathcal{M} \leftarrow \mathcal{M} + \mu(\mathcal{H} - \mathcal{T})$;
  Update the penalty parameter μ via $\mu \leftarrow \min(\rho\mu, \mu_{\max})$;
  Check the convergence condition $\|\mathcal{H} - \mathcal{T}\|_{\max} < \mathrm{tol}$;
  if $t > \mathrm{maxIters}$ **or** converged **then**
    break;
  end if
  $t \leftarrow t + 1$;
end while

Output: $\mathbf{Y}$.

A.4 Experiments Details Supplementary results of OCGDL

This appendix provides dataset statistics omitted from the main paper, complete t-SNE comparisons, full sensitivity curves of η and k_c on all six datasets, and module-wise convergence plots.

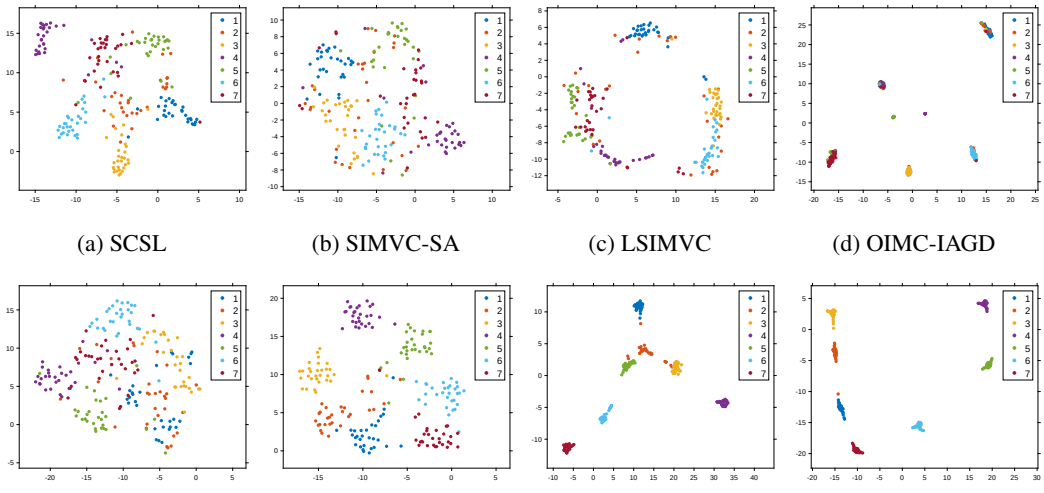
Datasets. Table 4 summarizes the six benchmarks. They span object images (MSRCv1, Caltech101-20, ALOI), handwritten digits (Handwritten), faces (ORL40), and text (BBCSport), with sample sizes from $n = 210$ to $n = 10{,}800$, view numbers from $V = 2$ to $V = 6$, and cluster numbers from $k = 5$ to $k = 100$. This diversity makes the suite a representative test of incomplete multi-view clustering under both small-scale and large-class regimes. Incomplete views are generated as described in Section 4.1: samples are removed independently at random, each instance remains visible in at least one view, and the 0.7 missing setting is omitted when $(V-1)/V < 0.7$.

Table 4: Details of the six datasets.

Dataset	n	V	d_v	k
MSRCv1	210	5	$(24, 576, 512, 256, 254)$	7
Handwritten	2,000	6	$(240, 76, 216, 47, 64, 6)$	10
Caltech101-20	2,386	6	$(48, 40, 254, 1984, 512, 928)$	20
ORL40	400	3	$(4096, 3304, 6750)$	40
ALOI	10,800	4	$(77, 13, 64, 64)$	100
BBCSport	544	2	$(3183, 3203)$	5

Additional t-SNE visualizations. Figure 7 compares all eight methods on MSRCv1 at a 30% missing rate. SCSL, SIMVC-SA, LSIMVC, and OIMC-IAGD exhibit substantial inter-class overlap, indicating that incomplete affinities are not sufficiently aligned into a discriminative embedding. OS-LF-IMVC improves compactness for some classes but still mixes several categories. CAL and OAGL recover much clearer clusters, yet a few class boundaries remain ambiguous. OCGDL yields the most compact intra-class groups and the cleanest inter-class margins, which is consistent with its leading clustering performance in this setting in Table 1.

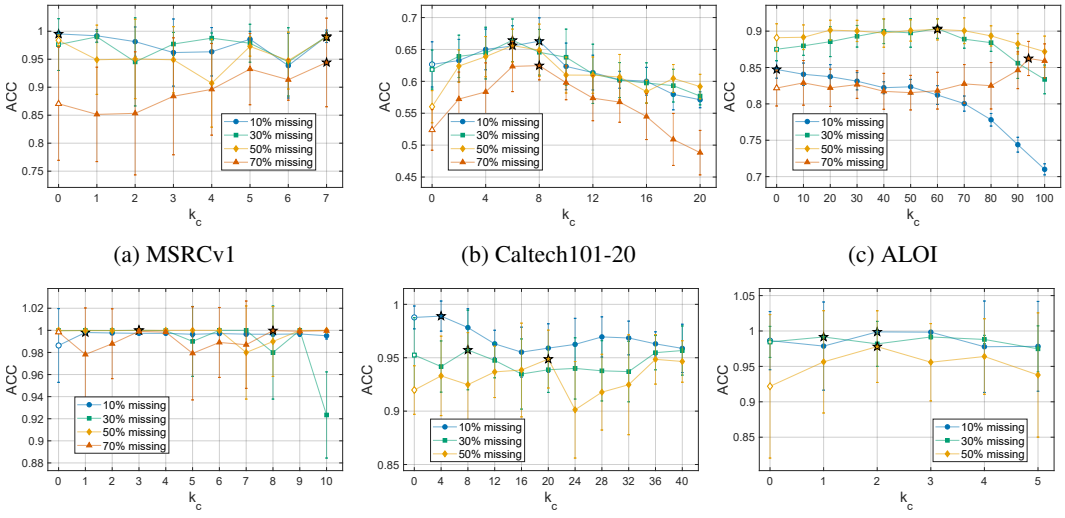
Sensitivity to k_c. Figure 8 reports ACC versus the shared dimension k_c on all six datasets. Neither extreme is uniformly optimal: $k_c = 0$ (view-specific only) and $k_c = k$ (shared only) are often inferior to an interior split. The trend with missingness is consistent with the main-text discussion. On MSRCv1, ORL40, and ALOI, a small k_c is preferable at a 10% missing rate, whereas larger k_c



(a) SCSL (b) SIMVC-SA (c) LSIMVC (d) OIMC-IAGD

(e) OS-LF-IMVC (f) CAL (g) OAGL (h) OCGDL

Figure 7: Complete t-SNE comparison on MSRCv1 under a 30% missing rate. Colors denote the ground-truth classes.



(a) MSRCv1 (b) Caltech101-20 (c) ALOI

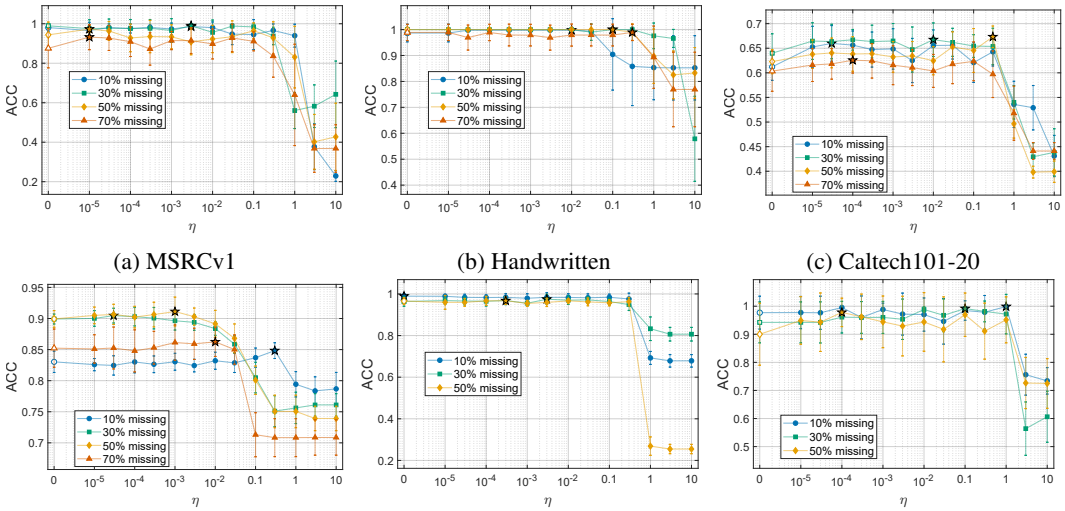
(d) Handwritten (e) ORL40 (f) BBCSport

Figure 8: Complete sensitivity of OCGDL to the shared dimension k_c under different missing rates.

becomes beneficial as missingness grows, indicating a shift from exploiting complementary neighborhoods toward enforcing cross-view consensus. Caltech101-20 prefers an intermediate fraction around $0.3k$–$0.4k$ at all rates; Handwritten and ORL40 likewise move from a small shared part at low missingness toward a larger shared subspace at higher missing rates. BBCSport, with only two views and $k = 5$, is less sensitive, but an interior k_c still matches or exceeds the extremes. All these results support that the shared-specific graph embedding split is not a redundant parameterization.

Sensitivity to η. Figure 9 reports ACC versus the sparsity weight η on all six datasets. A moderate η is consistently preferable to $\eta = 0$, confirming that row-wise $\ell_{2,1}$ coding improves atom selection under incomplete observations. The preferred magnitude is dataset-dependent: MSRCv1 favors relatively small values, Handwritten and BBCSport remain strong over a broader interval, and excessively large η eventually degrades ACC by overshrinking coefficients. The curves are generally smooth, indicating that OCGDL is not overly brittle to η once it lies in a moderate range.

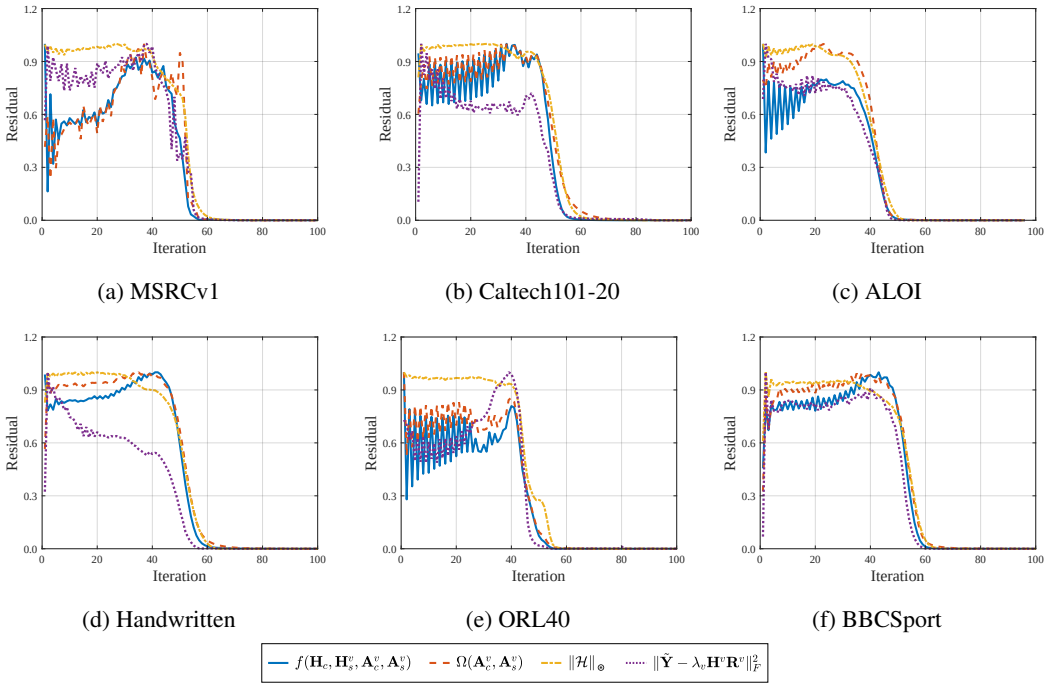
Complete convergence curves. Figure 10 plots module-wise residuals on all six datasets under a 50% missing rate. OCGDL converges within about 100 iterations in every case, including the



(a) MSRCv1 (b) Handwritten (c) Caltech101-20

(d) ALOI (e) ORL40 (f) BBCSport

Figure 9: Complete sensitivity of OCGDL to the sparsity weight η under different missing rates.



(a) MSRCv1 (b) Caltech101-20 (c) ALOI

(d) Handwritten (e) ORL40 (f) BBCSport

Figure 10: Complete module-wise convergence of OCGDL under a 50% missing rate.

largest setting ALOI. Tensor-consistency and one-step clustering residuals decrease smoothly from early iterations. The graph-reconstruction and sparsity terms typically stay on a short plateau and then drop from 30 to 50 iterations, after which all residuals settle at a low level. The same qualitative pattern appears on both small and large datasets, supporting the empirical stability of the ADMM updates.