

KOALA: Kernel Coupling and Element Imputation Induced Multi-View Clustering

Tingting Wu^{*1,2}, Zhendong Li^{*1,2}, Zhibin Gu³, Jiazheng Yuan⁴, Songhe Feng^{†1,2}

¹ Key Laboratory of Big Data & Artificial Intelligence in Transportation (Beijing Jiaotong University), Ministry of Education

² School of Computer Science and Technology, Beijing Jiaotong University, Beijing, 100044, China

³ College of Computer and Cyber Security, Hebei Normal University, Hebei, China

⁴ College of Science and Technology, Beijing Open University, Beijing, China
{22120438, 23120371, guzhibin, shfeng}@bjtu.edu.cn, jzyuan@139.com

Abstract

Incomplete Multi-View Clustering (IMVC) has made significant progress by optimally merging multiple pre-specified incomplete views. Most existing IMVC algorithms operate under the assumption that view alignment is known, but in practice, the coupling information between views may be absent, thereby limiting the practical applicability of these methods. Being aware of this, we propose a novel IMVC method named **Kernel cOupling And eLement imputAtion induced Multi-View Clustering (KOALA)**, which sufficiently explores the nonlinear relationship among features and optimally processes a group of kernels with missing and unaligned elements to simultaneously resolve multi-view clustering problem under both uncoupled and incomplete scenarios. Specifically, we first introduce a cross-kernel alignment learning strategy to reconstruct the coupling relationships among multiple kernels, which effectively captures high-order nonlinear relationships among samples and enhances alignment accuracy. Additionally, a low-rank tensor constraint is imposed on the optimizable alignment kernel tensor, facilitating the effective imputation of missing kernel elements by leveraging consistency information across views. Subsequently, we develop an alternative optimization approach with promising convergence to solve the resultant optimization problem. Extensive experimental results on various multi-view datasets demonstrate that the KOALA method achieves remarkable clustering performance.

Introduction

Incomplete Multi-View Clustering (IMVC) has received considerable attention for exploring consistent and complementary knowledge from incomplete multi-view data (Hu and Chen 2019b; Liu et al. 2019b; Wen et al. 2019; Zhang et al. 2020; Liu et al. 2021; Wen et al. 2021; Lin et al. 2021, 2022b; Xu et al. 2022; Tang and Liu 2022; Zhao et al. 2023; Wen et al. 2023; Zhang et al. 2023b; Dong et al. 2023b,a; Jin et al. 2023). Generally speaking, existing IMVC methods can be categorized into three popular groups, including matrix factorization-based methods (Xu et al. 2018; Hu and Chen 2019a; Li et al. 2023; Zhang et al. 2023a; Xu et al.

2024a; Gu and Feng 2023; Gu, Li, and Feng 2024), graph-based methods (Wang et al. 2019; Wen et al. 2020; Li, Wan, and He 2021; Yan et al. 2022; Zhang et al. 2023c; Xu et al. 2024b; Gu et al. 2024) and kernel-based methods (Liu et al. 2019c,a, 2020; Zhang et al. 2021; Wan et al. 2022; Wang et al. 2024). i) Matrix factorization-based methods factorize each view data into two non-negative matrices and then attempt to find a unified view-consensus cluster partition. For example, partial multi-view clustering (Li, Jiang, and Zhou 2014) aims to identify a common latent representation for all views by assuming that samples are represented similarly across different views. ii) Graph-based methods first build undirected graphs using pairwise similarities between samples, followed by simultaneous spectral clustering and graph fusion. Typically, perturbation-oriented incomplete multi-view clustering produces comprehensive similarity matrices and alleviates the difficulties associated with spectral tension. iii) Kernel-based methods typically attempt to derive a unified representation or multiple latent representations corresponding to all views by employing kernels that have been reconstructed from these views. For instance, the work in (Liu et al. 2018) proposes a late fusion approach to simultaneously clustering and impute the incomplete matrices.

While the aforementioned IMVC methods effectively extract rich information from incomplete multi-view data, real-world scenarios present challenges beyond mere missing information in some views. The misalignment of samples across views poses significant hurdles for existing methods in exploring complementarity and achieving consensus representation in multi-view data (Xie et al. 2018). Unfortunately, research on uncoupled incomplete multi-view data has been relatively scarce in recent years. As far as we know, Lin et al. (Lin et al. 2022a) proposed a pioneering solution that investigates the relationship between the inferred self-representation matrix of misaligned data and the most reliable view. Despite advancement in handling uncoupled incomplete multi-view data, this method operates in Euclidean space, limiting its ability to capture only linear relationships between pairs of samples and between pairs of views. Consequently, high-order nonlinear relationships among samples and among views remain unexplored, leading to sub-optimal clustering results.

To effectively cluster uncoupled incomplete multi-view

*These authors contributed equally.

†Corresponding author

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

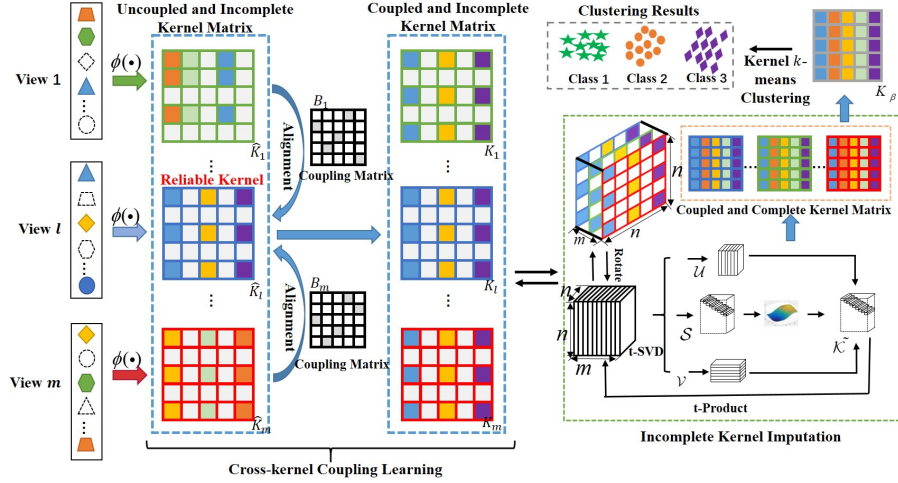


Figure 1: The Framework of KOALA: A coupling-imputation-clustering framework for uncoupled incomplete multi-view clustering.

data, this paper introduces a new multi-view clustering algorithm termed Kernel Coupling and Element Imputation Induced Multi-View Clustering. The framework of our method is illustrated in Fig. 1. Specifically, KOALA first learns multiple non-coupled and incomplete kernel matrices from the provided uncoupled incomplete multi-view data. Subsequently, we adopt a cross-kernel alignment learning strategy by introducing a pair of coupling matrices in the reproducing kernel Hilbert space. This approach facilitates exploration of high-order nonlinear relationships among samples while realigning disordered kernel elements. Additionally, we introduce a strategy for recovering missing kernel elements using low-rank tensor constraints to explore high-order correlations among all kernel matrices, effectively imputing absent kernel elements. Overall, the main contributions of this paper are as follows:

1) To the best of our knowledge, we are the first work to utilize a kernel-based method, which integrates information from multiple views and flexibly handles complex and nonlinear data relationships, to solve the uncoupled and incomplete multi-view clustering problem.

2) KOALA introduces a cross-kernel alignment learning strategy to reconstruct the coupling relationship among kernels. Furthermore, a low-rank tensor constraint is used on the optimizable alignment kernel tensor to capture the consistency information to impute missing kernel elements.

3) We employ a curvilinear search algorithm to solve the difficult optimization problem with orthogonality constraint, followed by a six-step alternative optimization method for KOALA. Comprehensive experimental results clearly demonstrate the effectiveness and efficiency of our proposed method.

Preliminary

A. Multiple Kernel k -Means with Incomplete Kernels (MKKM- IK)

Given $\{\mathbf{x}_i\}_{i=1}^n \subseteq \mathcal{X}$ with d dimensions and n samples, and $\phi_p(\cdot) : \mathbf{x} \in \mathcal{X} \mapsto \mathcal{H}_p$ denotes the p th feature mapping that maps \mathbf{x} onto a reproducing kernel Hilbert space \mathcal{H}_p ($1 \leq p \leq m$). The kernel matrix \mathbf{K}_p can be expressed as $K_{i,j} = \phi_p(\mathbf{x}_i)^\top \phi_p(\mathbf{x}_j) = \kappa_p(\mathbf{x}_i, \mathbf{x}_j)$. In a multiple kernel setting, each sample is depicted as $\phi_\beta(\mathbf{x}) = [\beta_1 \phi_1(\mathbf{x})^\top, \dots, \beta_m \phi_m(\mathbf{x})^\top]^\top$ where $\beta = [\beta_1, \dots, \beta_m]^\top$ consists of the coefficients of the m base kernels $\{\kappa_p(\cdot, \cdot)\}_{p=1}^m$. Based on the definition of $\phi_\beta(\mathbf{x})$, the kernel function is denoted as $\kappa_\beta(\mathbf{x}_i, \mathbf{x}_j) = \phi_\beta(\mathbf{x}_i)^\top \phi_\beta(\mathbf{x}_j) = \sum_{p=1}^m \beta_p^2 \kappa_p(\mathbf{x}_i, \mathbf{x}_j)$. Let $\{\mathbf{K}_p\}_{p=1}^m$ denote a group of kernel matrices pre-calculated from coupled and incomplete multi-view data with m views. As noted in (Yu et al. 2011; Gönen and Margolin 2014), the fused kernel matrix \mathbf{K}_β can be described as $\mathbf{K}_\beta = \sum_{p=1}^m \beta_p^2 \mathbf{K}_p$, so the general formulation of MKKM- IK (Liu et al. 2019c) is

$$\begin{aligned} & \min_{\{\mathbf{H}, \beta, \{\mathbf{K}_p\}_{p=1}^m\}} \text{Tr}(\mathbf{K}_\beta (\mathbf{I}_n - \mathbf{H}\mathbf{H}^\top)), \\ & \text{s.t. } \mathbf{H} \in \mathbb{R}^{n \times k}, \mathbf{H}^\top \mathbf{H} = \mathbf{I}_k, \beta^\top \mathbf{1}_m = 1, \beta_p \geq 0, \end{aligned} \quad (1)$$

$$\mathbf{K}_p(s_p, s_p) = \mathbf{K}_p^{(cc)}, \mathbf{K}_p \succeq 0, \forall p,$$

where $s_p (1 \leq p \leq m)$ denotes the index of the sample presents in the p -th kernel and $\mathbf{K}_p^{(cc)}$ represents the kernel sub-matrix computed with the corresponding samples. The constraint $\mathbf{K}_p(s_p, s_p) = \mathbf{K}_p^{(cc)}$ is imposed to ensure that \mathbf{K}_p preserves the known entries throughout the procedure. Moreover, The matrix \mathbf{H} denotes the cluster partition matrix.

B. Tensor Nuclear Norm Theory

Let $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be a tensor, its block circulant matrix $\text{bcirc}(\mathcal{Y}) \in \mathbb{R}^{n_1 n_3 \times n_2 n_3}$ as

$$\text{bcirc}(\mathcal{Y}) = \begin{pmatrix} Y^{(1)} & Y^{(n_3)} & \dots & Y^{(2)} \\ Y^{(2)} & Y^{(1)} & \dots & Y^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ Y^{(n_3)} & Y^{(n_3-1)} & \dots & Y^{(1)} \end{pmatrix}. \quad (2)$$

Also, the block circular vectorization and its inverse operation are

$$bvec(\mathcal{Y}) = \begin{pmatrix} Y^{(1)} \\ Y^{(2)} \\ \vdots \\ Y^{(n_3)} \end{pmatrix}, \quad bvfod(bvec(\mathcal{Y})) = \mathcal{Y}. \quad (3)$$

Then, some related definitions are introduced below (Xie et al. 2018).

Definition 1 (t-product) Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$, the tensor $\mathcal{S} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$ is obtained by t-product of \mathcal{A} and \mathcal{B}

$$\mathcal{S} = \mathcal{A} * \mathcal{B} = bvfod(bcirc(\mathcal{A})bvec(\mathcal{B})),$$

Definition 2 (t-SVD) Given a tensor $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, it can be factorized by t-SVD as $\mathcal{Y} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$, where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal, and \mathcal{S} is an f-diagonal tensor whose size is $n_1 \times n_2 \times n_3$.

Definition 3 (t-TNN) The t-SVD based tensor nuclear norm (t-TNN) is given by

$$\|\mathcal{Y}\|_* := \sum_{i=1}^{\min(n_1, n_2)} \sum_{k=1}^{n_3} |\mathcal{S}_f(i, i, k)|,$$

where $\mathcal{S}_f = fft(\mathcal{S}, [], 3)$ is the Fourier transform along the third dimension.

Proposed Method

This section introduces our proposed approach, which addresses the simultaneous tasks of imputing missing elements and aligning uncoupled values. Subsequently, we will provide detailed explanations of two sub-modules within the KOALA method: the cross-kernel alignment learning strategy and the missing kernel element recovery strategy.

Cross-kernel alignment learning strategy. As mentioned before, one of the main challenges is the unknown arrangement information among kernel matrices, which hinders view fusion. Thereafter, a key issue lies in *how to effectively reestablish the cross-kernel correspondence relationship under the incomplete scenario?*

Existing coupling methods generally function in Euclidean space and neglect high-order nonlinear relationships among samples and among views, resulting in suboptimal clustering performance. To solve this issue, KOALA conducts elements alignment in reproducing kernel Hilbert space. Precisely, let $\{\hat{\mathbf{K}}_p\}_{p=1}^m$ represent a collection of non-coupled and incomplete kernel matrices obtained from uncoupled incomplete multi-view data. Inspired by (Lin et al. 2022a), we explicitly introduce a set of coupling matrices $\mathbf{B}_p \in \mathbb{R}^{n \times n}$ ($1 \leq p \leq m$) to realign the uncoupled kernel matrices. This strategy randomly selects $\hat{\mathbf{K}}_q$ ($1 \leq q \leq m$) as the most reliable kernel and uses the following constraint to recouple the kernel matrix of each view with the q -th kernel:

$$\left\| \mathbf{B}_p^\top \hat{\mathbf{K}}_p \mathbf{B}_p - \hat{\mathbf{K}}_q \right\|_F^2, \quad (4)$$

By reconstructing the coupling relationship among kernels, each paired kernel matrix $\mathbf{B}_p^\top \hat{\mathbf{K}}_p \mathbf{B}_p$ can be obtained,

and the fused kernel matrix $\hat{\mathbf{K}}_\beta$ can be derived as $\hat{\mathbf{K}}_\beta = \sum_{p=1}^m \beta_p^2 \mathbf{B}_p^\top \hat{\mathbf{K}}_p \mathbf{B}_p$. This strategy maps complex nonlinear relationships in the original data into kernel space, making multi-view data more linearly separable, which captures the high-order nonlinear relationships sufficiently to reconstruct the alignment correlations among kernel matrices.

Missing kernel elements recovery strategy. For the incompleteness of kernel matrices, we propose a novel kernel tensor-based Missing Kernel Elements Recovery Strategy (MLR). This strategy focuses on *how to efficiently find significant information to fill in missing kernel elements*, which sufficiently considers intra-kernel correlations and high-order correlations among different kernels. Unlike traditional imputation strategies that operate at the kernel matrix level, MLR constructs a creative optimizable tensor representation and imposes a global low-rank constraint on the tensor. This approach captures high-order correlations among multiple kernels, thereby enhancing the accuracy of imputing missing kernel elements. Specifically, we construct an optimizable alignment kernel tensor structure defined as follows:

Definition 4 (Optimizable alignment kernel tensor) Given a set of coupled kernel matrices $\{\mathbf{B}_p^\top \hat{\mathbf{K}}_p \mathbf{B}_p\}_{p=1}^m$, the optimizable alignment kernel tensor is assembled by stacking the kernel matrices of different views, that is

$$\mathcal{K} = \Phi(\mathbf{B}_1^\top \hat{\mathbf{K}}_1 \mathbf{B}_1, \mathbf{B}_2^\top \hat{\mathbf{K}}_2 \mathbf{B}_2, \dots, \mathbf{B}_m^\top \hat{\mathbf{K}}_m \mathbf{B}_m), \quad (5)$$

where $\{\mathbf{B}_p\}_{p=1}^m$ are optimizable coupling matrices and $\Phi(\cdot)$ represents a mapping function that merges the kernel matrices $\mathbf{B}_p^\top \hat{\mathbf{K}}_p \mathbf{B}_p$ into a 3-dimension tensor of size $n \times n \times m$.

Based on Eq. (5), the incomplete optimizable alignment kernel tensor \mathcal{K} is formed, and $\|\mathcal{K}\|_*$ (definition 3) is applied to its rotated form $\mathcal{K} \in \mathbb{R}^{n \times m \times n}$, aiming to capture high-order correlations. The use of t-TNN facilitates the imputation of missing elements by capturing the global structure and latent patterns within $\hat{\mathbf{K}}_p$. Introducing the low-rank approximation on \mathcal{K} reduces model complexity and allows for the rational inference and completion of missing values based on the global trends and structure of the observed data. In short, this strategy implicitly fills in missing elements in the kernel matrix by effectively exploiting high-order correlations among incomplete multi-view data.

By integrating Eq. (4) and the low-rank tensor constraint $\|\mathcal{K}\|_*$ with Eq. (1), the objective formulation of our proposed KOALA is as follows:

$$\begin{aligned} \min_{\{\hat{\mathbf{K}}_p, \mathbf{B}_p\}_{p=1}^m, \beta, \mathbf{H}} & \quad \text{Tr}(\hat{\mathbf{K}}_\beta (\mathbf{I}_n - \mathbf{H}\mathbf{H}^\top)) + \lambda \|\mathcal{K}\|_* \\ & \quad + \theta \sum_{p=1}^m \left\| \mathbf{B}_p^\top \hat{\mathbf{K}}_p \mathbf{B}_p - \hat{\mathbf{K}}_q \right\|_F^2, \end{aligned}$$

$$\text{s.t. } \mathcal{K} = \Phi(\mathbf{B}_1^\top \hat{\mathbf{K}}_1 \mathbf{B}_1, \mathbf{B}_2^\top \hat{\mathbf{K}}_2 \mathbf{B}_2, \dots, \mathbf{B}_m^\top \hat{\mathbf{K}}_m \mathbf{B}_m), \quad (6)$$

$$\mathbf{H} \in \mathbb{R}^{n \times k}, \quad \mathbf{H}^\top \mathbf{H} = \mathbf{I}_k, \quad \beta^\top \mathbf{1}_m = 1, \quad \beta_p \geq 0,$$

$$\hat{\mathbf{K}}_p(\mathbf{s}_p, \mathbf{s}_p) = \hat{\mathbf{K}}_p^{(cc)}, \quad \hat{\mathbf{K}}_\beta = \sum_{p=1}^m \beta_p^2 \mathbf{B}_p^\top \hat{\mathbf{K}}_p \mathbf{B}_p,$$

$$\hat{\mathbf{K}}_p \geq 0, \quad \mathbf{B}_p \geq 0, \quad \mathbf{B}_p^\top \mathbf{B}_p = \mathbf{I}_n, \quad \forall p,$$

where θ and λ are hyperparameters, $\|\mathcal{K}\|_*$ denotes the nuclear norm of the tensor. In this paper, we simply set $q = 1$

by assigning the first kernel as the most reliable kernel.

As seen from Eq. (6), the missing kernel elements are first filled with zero values, and then the cross-kernel alignment learning strategy is used to rebuild the coupling relationship among kernels. Subsequently, we create the optimizable alignment kernel tensor \mathcal{K} and employ the missing kernel elements recovery strategy to impute the missing kernel elements. In conclusion, we present a joint coupling-imputation-clustering framework where the kernel matrices $\{\hat{\mathbf{K}}_p\}_{p=1}^m$, the coupling matrices $\{\mathbf{B}_p\}_{p=1}^m$ and the clustering partition matrix \mathbf{H} can be jointly optimized to enhance the final clustering performance.

Optimization

Inspired by the alternating direction method of multipliers (ADMM) (Lin, Liu, and Su 2011), we introduce two auxiliary variables, $\mathcal{G} = \mathcal{K}$ and $\{\mathbf{M}\}_{p=1}^m = \{\mathbf{B}\}_{p=1}^m$ to make the variables separable. The Lagrange function of Eq. (6) can be expressed as the following unconstrained problem.

$$\begin{aligned} & \mathcal{L}(\mathbf{H}; \{\hat{\mathbf{K}}\}_{p=1}^m; \{\mathbf{B}\}_{p=1}^m; \mathcal{G}; \{\mathbf{M}\}_{p=1}^m; \beta) \\ &= \text{Tr}\left(\sum_{p=1}^m \beta_p^2 \mathbf{B}_p^\top \hat{\mathbf{K}}_p \mathbf{B}_p (\mathbf{I}_n - \mathbf{H}\mathbf{H}^\top)\right) + \lambda \|\mathcal{G}\|_* + \langle \mathcal{W}, \mathcal{K} - \mathcal{G} \rangle \\ &+ \frac{\rho}{2} \|\mathcal{K} - \mathcal{G}\|_F^2 + \theta \sum_{p=1}^m \left\| \mathbf{B}_p^\top \hat{\mathbf{K}}_p \mathbf{B}_p - \hat{\mathbf{K}}_q \right\|_F^2 \\ &+ \sum_{p=1}^m \left(\langle \mathbf{J}_p, \mathbf{B}_p - \mathbf{M}_p \rangle + \frac{\mu}{2} \|\mathbf{B}_p - \mathbf{M}_p\|_F^2 \right). \end{aligned} \quad (7)$$

Here, the tensor \mathcal{W} and the matrix $\{\mathbf{J}_p\}_{p=1}^m$ serve as Lagrange multipliers, while ρ and μ are penalty parameters controlling convergence. Then Eq. (6) is transformed into solving the following subproblems in a distinct manner.

Optimization H

Optimizing \mathbf{H} is equivalent to solving the following problem if other variables are fixed:

$$\min_{\mathbf{H}} \text{Tr}(\hat{\mathbf{K}}_\beta (\mathbf{I}_n - \mathbf{H}\mathbf{H}^\top)), \quad s.t. \mathbf{H} \in \mathbb{R}^{n \times k}, \mathbf{H}^\top \mathbf{H} = \mathbf{I}_k. \quad (8)$$

The Eq. (8) in optimizing \mathbf{H} is a traditional k -means clustering problem that can be solved by existing packages.

Optimization $\hat{\mathbf{K}}_p$

Fixing the other variables derives the subproblem for $\hat{\mathbf{K}}_p$,

$$\begin{aligned} & \min_{\hat{\mathbf{K}}_p} \text{Tr}(\beta_p^2 \mathbf{B}_p^\top \hat{\mathbf{K}}_p \mathbf{B}_p (\mathbf{I}_n - \mathbf{H}\mathbf{H}^\top)) + \langle \mathbf{W}_p, \hat{\mathbf{K}}_p - \mathbf{G}_p \rangle \\ &+ \frac{\rho}{2} \left\| \hat{\mathbf{K}}_p - \mathbf{G}_p \right\|_F^2 + \theta \sum_{p=1}^m \left\| \mathbf{B}_p^\top \hat{\mathbf{K}}_p \mathbf{B}_p - \hat{\mathbf{K}}_q \right\|_F^2, \quad (9) \\ &s.t. \hat{\mathbf{K}}_p(\mathbf{s}_p, \mathbf{s}_p) = \hat{\mathbf{K}}_p^{(cc)}, \hat{\mathbf{K}}_p \succeq 0, \forall p. \end{aligned}$$

By setting the derivative of Eq. (9), the closed form of $\hat{\mathbf{K}}_p$ can be obtained by

$$\hat{\mathbf{K}}_p = \frac{\rho \mathbf{G}_p + 2\theta \mathbf{B}_p \hat{\mathbf{K}}_q \mathbf{B}_p^\top - \mathbf{W}_p - \beta_p^2 \mathbf{B}_p^\top (\mathbf{I}_n - \mathbf{H}\mathbf{H}^\top) \mathbf{B}_p}{\rho + 2\theta}. \quad (10)$$

Optimization \mathcal{G}

If $\hat{\mathbf{K}}_p$ and $\mathbf{B}_p (p = 1, 2, \dots, m)$ are fixed, for updating the tensor \mathcal{G} , the problem can be written as

$$\mathcal{G}^* = \arg \min_{\mathcal{G}} \lambda \|\mathcal{G}\|_* + \frac{\rho}{2} \left\| \mathcal{G} - \left(\mathcal{K} + \frac{1}{\rho} \mathcal{W} \right) \right\|_F^2. \quad (11)$$

The following theorem can be used to solve the optimization problem Eq. (11).

Theorem 1 (Xie et al. 2018): Let $\tau > 0$, and $\mathcal{G} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\mathcal{F} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the best possible solution on a global scale for $\min_{\mathcal{G}} \tau \|\mathcal{G}\|_* + \frac{1}{2} \|\mathcal{G} - \mathcal{F}\|_F^2$ is given by the tensor tubal-shrinkage operator $\mathcal{G} = \mathcal{C}_{n,3\tau}(\mathcal{F}) = \mathcal{U} * \mathcal{C}_{n,3\tau}(\mathcal{S}) * \mathcal{V}^\top$. Noticed that $\mathcal{F} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$ and $\mathcal{C}_{n,3\tau} = \mathcal{S} * \mathcal{J}$, where $\mathcal{J} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ represents an f-diagonal tensor, whose diagonal element in the Fourier domain being $\mathcal{J}_f(i, i, j) = (1 - \frac{n,3\tau}{S_j^{(j)}}(i, i))_+$.

Optimization \mathbf{M}_p

By fixing all variables to expect \mathbf{M}_p , the problem in Eq. (7) can be written as

$$\min_{\mathbf{M}_p} \langle \mathbf{J}_p, \mathbf{B}_p - \mathbf{M}_p \rangle + \frac{\mu}{2} \|\mathbf{B}_p - \mathbf{M}_p\|_F^2. \quad (12)$$

By differentiating Eq. (12) and setting it to zero, the optimal solution \mathbf{M}_p can be obtained:

$$\mathbf{M}_p = \frac{\mathbf{J}_p + \mu \mathbf{B}_p}{\mu}, \quad (13)$$

Optimization \mathbf{B}_p

When other variables are fixed, the optimization for \mathbf{B}_p is formulated as

$$\begin{aligned} & \min_{\mathbf{B}_p} \beta_p^2 \text{Tr}(\mathbf{B}_p^\top \hat{\mathbf{K}}_p \mathbf{B}_p (\mathbf{I}_n - \mathbf{H}\mathbf{H}^\top)) + \theta \left\| \mathbf{B}_p^\top \hat{\mathbf{K}}_p \mathbf{B}_p - \hat{\mathbf{K}}_q \right\|_F^2 \\ &+ \langle \mathbf{J}_p, \mathbf{B}_p - \mathbf{M}_p \rangle + \frac{\mu}{2} \|\mathbf{B}_p - \mathbf{M}_p\|_F^2, \\ &s.t. \mathbf{B}_p \geq 0, \mathbf{B}_p^\top \mathbf{B}_p = \mathbf{I}_n. \end{aligned} \quad (14)$$

Let $\mathbf{Z}_p = \mathbf{J}_p - \mu \mathbf{M}_p$, $\mathbf{O}_p = \beta_p^2 (\mathbf{I}_n - \mathbf{H}\mathbf{H}^\top) - 2\theta \hat{\mathbf{K}}_q$. Due to the orthogonality constraint of \mathbf{B}_p , a curvilinear search algorithm is applied to optimize \mathbf{B}_p according to (Wen and Yin 2013). We first define $\mathcal{F}(\mathbf{B}_p) = \text{Tr}(\mathbf{O}_p \mathbf{B}_p^\top \hat{\mathbf{K}}_p \mathbf{B}_p + \mathbf{B}_p^\top \mathbf{Z}_p)$ and rewrite Eq. (14) as

$$\min_{\mathbf{B}_p} \mathcal{F}(\mathbf{B}_p), \quad s.t. \mathbf{B}_p^\top \mathbf{B}_p = \mathbf{I}_n. \quad (15)$$

We define $\mathbf{G} = \frac{\partial \mathcal{F}}{\partial \mathbf{B}_p}$ and $\mathbf{A} = \mathbf{G} \mathbf{B}_p^\top - \mathbf{B}_p \mathbf{G}^\top$. To preserve the orthogonality constraint, the next iteration of \mathbf{B}_p can be obtained according to (Wen and Yin 2013) as follows,

$$\hat{\mathbf{B}}_p(\tau) = \mathbf{B}_p - \frac{\tau}{2} \mathbf{A} (\mathbf{B}_p + \hat{\mathbf{B}}_p(\tau)). \quad (16)$$

Then we can easily do the calculation

$$\hat{\mathbf{B}}_p(\tau) = (\mathbf{I}_n + \frac{\tau}{2} \mathbf{A})^\dagger (\mathbf{I}_n - \frac{\tau}{2} \mathbf{A}) \mathbf{B}_p, \quad (17)$$

where $\tau \in \mathbb{R}$ is a step size, $\hat{\mathbf{B}}_p(\tau)$ can be solved since orthogonality is preserved by any skew-symmetric matrix \mathbf{A} .

Algorithm 1: The proposed KOALA

Input: $\{\hat{\mathbf{K}}_p^{(cc)}\}_{p=1}^m$, $\{\mathbf{s}_p\}_{p=1}^m$, λ , θ and ϵ_0 , $\rho = \mu = 10^{-5}$, $\rho_{max} = \mu_{max} = 10^{10}$;

Initialize: Initialize $\boldsymbol{\beta}^{(0)} = \mathbf{1}_m/m$, $\hat{\mathbf{K}}_p^{(0)}$ and $t = 1$, $\mathcal{G} = \mathcal{W} = 0$, $\{\mathbf{B}_p\}_{p=1}^m = \mathbf{I}_n$, $\{\mathbf{M}_p\}_{p=1}^m = \{\mathbf{J}_p\}_{p=1}^m = \mathbf{0}$;

- 1: **while** not converge **do**
- 2: $\hat{\mathbf{K}}_\beta^{(t)} = \sum_{p=1}^m (\beta_p^{(t-1)})^2 \mathbf{B}_p^{\top(t-1)} \hat{\mathbf{K}}_p^{(t-1)} \mathbf{B}_p^{(t-1)}$;
- 3: Update \mathbf{H}^t by solving a kernel k-means clustering optimization problem Eq. (8)
- 4: Update $\hat{\mathbf{K}}_p^{(t)}$ by Eq. (10);
- 5: Update $\mathcal{G}^{(t)}$ via subproblem Eq. (11) ;
- 6: Update $\mathcal{W}^{(t)}$ by Eq. (20);
- 7: Update $\mathbf{M}_p^{(t)}$ by Eq. (13);
- 8: Update $\mathbf{J}_p^{(t)}$ according to Eq. (21);
- 9: Update $\mathbf{B}_p^{(t)}$ by Eq. (18);
- 10: Update $\boldsymbol{\beta}^{(t)}$ by Eq. (19);
- 11: Check the convergence conditions: $obj^{(t)} - obj^{(t-1)} \leq \epsilon_0$;
- 12: **end while**
- 13: **Output:** \mathbf{H} .

To reduce the complexity of the matrix inversion, we further improve Eq. (17) according to (Zhang et al. 2022). Let $\mathbf{U} = [\mathbf{G}, \mathbf{B}_p]$ and $\mathbf{V} = [\mathbf{B}_p, \mathbf{G}]$, then the SMW formula is applied to $\mathbf{I}_n + \frac{\tau}{2} \mathbf{A} = \mathbf{I}_n + \frac{\tau}{2} \mathbf{U} \mathbf{V}^\top$ and $(\mathbf{I}_n + \frac{\tau}{2} \mathbf{A})^\dagger = \mathbf{I}_n - \frac{\tau}{2} \mathbf{U} (\mathbf{I}_n + \frac{\tau}{2} \mathbf{V}^\top \mathbf{U})^\dagger \mathbf{V}^\top$. After that, Eq. (17) can be rewritten as

$$\hat{\mathbf{B}}_p(\tau) = \mathbf{B}_p - \tau \mathbf{U} (\mathbf{I}_n + \frac{\tau}{2} \mathbf{V}^\top \mathbf{U})^\dagger \mathbf{V}^\top \mathbf{B}_p, \quad (18)$$

where τ is selected using a one-dimensional line searching strategy.

Optimization β

Fixing the other variables, the optimization in Eq. (6) regarding β is transformed to

$$\min_{\beta} \sum_{p=1}^m \beta_p^2 \text{Tr}(\mathbf{B}_p^\top \hat{\mathbf{K}}_p \mathbf{B}_p (\mathbf{I}_n - \mathbf{H} \mathbf{H}^\top)), \quad (19)$$

s.t. $\boldsymbol{\beta}^\top \mathbf{1}_m = 1$, $\beta_p \geq 0$, $\forall p$.

We summarize the solution of Eq. (6) in Algorithm 1. In addition, the Lagrange multipliers \mathcal{W} and \mathbf{G}_p can be optimized as follows,

$$\mathcal{W}^* = \mathcal{W} + \rho(\mathcal{K} - \mathcal{G}), \quad (20)$$

$$\mathbf{J}_p = \mathbf{J}_p + \mu(\mathbf{B}_p - \mathbf{M}_p). \quad (21)$$

Experiment

Experimental Settings

Datasets: We assessed the performance of our method using seven benchmark multi-view datasets, covering various sizes and numbers of clusters: **ORL**, **proteinFold**, **Cora**, **UCI Digits**, **Caltech101-7**, **Caltech101-20**, and **Caltech101-all**. The detailed information for each dataset is presented in Table 1.

Datasets	Samples	Views	Classes
ORL	400	3	40
proteinFold	694	12	27
Caltech101-7	1474	6	7
UCI Digits	2000	6	10
Caltech101-20	2386	6	20
Cora	2708	2	7
Caltech101-all	9144	6	102

Table 1: Datasets used in our experiments.

Comparison methods: There are several commonly used imputation approaches including zero filling (ZF), mean filling (MF), and k -nearest-neighbor filling (KNN). The widely used Multiple Kernel k -Means (MKKM) (Gönen and Margolin 2014) is combined with the above imputation methods to form "two-stage" kernel-based methods, and KOALA is compared with these two-stage methods, including MKKM+ZF and MKKM+MF. Similarly, we compare our method with "one-stage" kernel-based approaches that optimize both imputation and clustering simultaneously. These are LRKT-IMVC (Wu, Feng, and Yuan 2024) and the variants of MKKM-IK (Liu et al. 2019c) include MKKM-IK+MF, MKKM-IK+ZF, MKKM-IK+KNN, and MKKM-IK-MKC. In addition to the kernel-based IMKC methods, UMIC (Lin et al. 2022a), a technique specifically designed for uncoupled incomplete multi-view data, is also used for comparison with our approach. For a fair comparison, we directly employ the source codes provided by the respective literature.

Uncoupled incomplete data construction: In this article, the underlying assumption is that there is at least one view available for each sample. The proportion of samples with missing views is controlled by the missing ratio parameter ϵ . Specifically, we set ϵ to $[0.1 : 0.2 : 0.5]$ for all datasets. As the mapping relationship of uncoupled data is unknown, and not all instances of uncoupled data are completely independent in the real world, we randomly shuffle the data arrangement for each view to create the uncoupled incomplete multi-view data in our experiments. Additionally, we randomly generate the 'incomplete and uncoupled' patterns ten times and present the statistical results.

Evaluation metrics: To evaluate clustering performance, five commonly used criteria are employed: clustering Accuracy (ACC), Purity, Adjusted Rand Index (ARI), Fscore, and Precision. Due to space limitations, only partial results are presented in Table 2. More experimental results are presented in the supplementary materials.

Experimental Results

Overall Clustering Performances. We evaluate the clustering performance of all the comparison methods mentioned above, along with the proposed KOALA, considering different missing ratios. The partial results are listed in Table 2. The best results are highlighted in bold, and the second best results are underlined. From Table 2, three observations can be made.

Dataset	ϵ	Metric(%)	*ZF	*MF	LRKT-IMVC	*IK-ZF	*IK-MF	*IK-KNN	*IK-MKC	UIMC	Ours
ORL	0.1	ACC	16.18 (0.60)	16.15 (0.70)	16.63 (0.60)	16.82 (0.92)	16.65 (0.80)	16.57 (0.74)	16.43 (0.69)	31.32 (13.36)	23.15 (1.13)
		Purity	16.77 (0.69)	16.98 (0.81)	17.28 (0.70)	17.25 (0.89)	17.17 (0.71)	17.15 (0.78)	16.95 (0.76)	32.95 (14.13)	28.40 (0.83)
		Fscore	2.60 (0.20)	2.56 (0.26)	2.65 (0.24)	2.72 (0.37)	2.77 (0.34)	2.62 (0.35)	2.52 (0.14)	19.06 (14.71)	42.01 (2.72)
	0.3	ACC	16.22 (1.04)	16.40 (0.65)	16.45 (0.59)	16.30 (0.63)	16.40 (0.82)	16.30 (0.42)	16.45 (0.95)	16.27 (0.10)	21.52 (0.79)
		Purity	16.72 (1.09)	16.98 (0.59)	17.05 (0.66)	16.83 (0.76)	16.95 (0.92)	17.05 (0.43)	17.18 (1.11)	16.96 (0.09)	25.95 (0.70)
		Fscore	2.81 (0.32)	2.79 (0.30)	2.51 (0.30)	2.53 (0.29)	2.66 (0.38)	2.54 (0.20)	2.64 (0.55)	2.66 (0.06)	35.13 (3.20)
0.5	ACC	15.85 (0.58)	15.98 (0.75)	16.50 (0.69)	16.32 (0.78)	15.98 (0.41)	16.23 (0.91)	16.53 (0.86)	16.43 (0.17)	21.85 (1.30)	
	Purity	16.32 (0.70)	16.40 (0.79)	17.10 (0.74)	16.92 (0.92)	16.55 (0.48)	17.03 (0.88)	17.10 (1.04)	17.04 (0.28)	25.65 (0.72)	
	Fscore	2.98 (0.52)	3.09 (0.26)	2.71 (0.47)	2.64 (0.35)	2.55 (0.18)	2.55 (0.38)	2.68 (0.41)	2.84 (0.08)	31.84 (1.56)	
proteinFold	0.1	ACC	11.79 (0.40)	11.69 (0.45)	11.95 (0.55)	11.70 (0.31)	11.73 (0.49)	11.97 (0.36)	12.00 (0.44)	12.88 (1.03)	15.98 (0.43)
		Purity	15.94 (0.77)	16.04 (0.55)	16.17 (0.47)	16.02 (0.59)	15.73 (0.52)	16.05 (0.81)	16.40 (0.52)	17.26 (1.20)	20.85 (0.34)
		Fscore	4.06 (0.23)	4.02 (0.18)	3.86 (0.14)	3.99 (0.14)	3.96 (0.18)	3.96 (0.16)	3.98 (0.20)	4.33 (0.49)	42.65 (1.28)
	0.3	ACC	12.07 (0.47)	12.13 (0.64)	11.86 (0.32)	11.77 (0.38)	11.67 (0.28)	12.07 (0.20)	11.59 (0.37)	11.90 (0.09)	15.71 (0.49)
		Purity	16.73 (0.79)	15.98 (0.88)	16.04 (0.31)	16.07 (0.67)	15.91 (0.61)	16.04 (0.48)	15.79 (0.49)	16.06 (0.07)	20.26 (0.55)
		Fscore	4.15 (0.22)	5.48 (0.34)	3.84 (0.14)	3.93 (0.18)	3.94 (0.16)	3.95 (0.20)	3.73 (0.19)	4.36 (0.33)	36.78 (1.30)
0.5	ACC	12.23 (0.56)	12.38 (0.46)	11.71 (0.38)	11.84 (0.35)	11.82 (0.51)	12.02 (0.38)	11.86 (0.34)	11.81 (0.05)	15.40 (0.48)	
	Purity	16.28 (0.75)	16.08 (0.74)	15.94 (0.60)	15.85 (0.45)	15.98 (0.57)	16.05 (0.49)	16.15 (0.54)	15.90 (0.09)	19.38 (0.40)	
	Fscore	4.31 (0.29)	7.10 (1.36)	3.84 (0.14)	3.93 (0.16)	3.94 (0.21)	3.97 (0.15)	3.81 (0.15)	5.55 (0.30)	29.42 (1.33)	
Cora	0.1	ACC	36.05 (1.16)	36.34 (1.37)	48.58 (0.47)	36.35 (0.79)	36.66 (0.77)	36.55 (0.70)	48.53 (0.40)	14.07 (3.66)	55.27 (1.12)
		Purity	45.54 (1.20)	45.89 (1.16)	54.30(0.78)	45.72 (0.91)	46.11 (0.95)	45.93 (0.92)	54.21 (0.67)	21.35 (7.74)	58.08 (0.96)
		Fscore	27.09 (1.29)	27.59 (1.41)	32.16(0.53)	27.72 (0.46)	28.04 (0.49)	27.82 (0.43)	31.97 (0.43)	14.58 (3.71)	35.01 (0.51)
	0.3	ACC	31.66 (1.90)	29.67 (2.43)	43.14 (0.88)	32.69 (1.55)	33.21 (1.46)	32.61 (1.54)	44.17 (1.17)	8.53 (5.55)	50.87 (1.47)
		Purity	40.16 (2.22)	39.00 (2.55)	50.16 (1.08)	41.81 (1.78)	42.49 (1.74)	41.75 (1.82)	51.29 (1.12)	11.40 (9.58)	53.74 (1.23)
		Fscore	23.22 (1.68)	22.73 (1.80)	29.14 (0.67)	25.28 (0.93)	26.59 (0.86)	25.21 (0.99)	28.84 (0.62)	8.25 (5.65)	31.28 (1.01)
0.5	ACC	27.53 (2.98)	27.96 (2.32)	39.79 (1.37)	32.01 (1.70)	32.47 (1.49)	32.28 (1.45)	40.29 (1.69)	5.96 (5.28)	46.18 (1.44)	
	Purity	35.97 (2.23)	36.77 (2.33)	48.38 (1.26)	40.14 (1.29)	41.35 (1.45)	40.48 (0.97)	46.85 (1.71)	9.03 (8.24)	49.78 (1.20)	
	Fscore	21.59 (0.94)	22.27 (1.54)	27.19 (0.79)	23.15 (0.70)	26.44 (0.65)	23.21 (0.64)	25.87 (0.84)	5.79 (5.44)	27.78 (1.02)	
UCLDigits	0.1	ACC	13.12 (0.24)	13.11 (0.22)	13.13 (0.26)	13.10 (0.24)	13.10 (0.24)	13.16 (0.35)	13.08 (0.26)	9.91 (5.94)	13.66 (0.49)
		Purity	20.00 (0.02)	20.00 (0.00)	20.01 (0.03)	20.00 (0.00)	20.01 (0.03)	20.00 (0.00)	20.00 (0.00)	10.23 (6.16)	20.38 (0.22)
		Fscore	10.26 (0.11)	10.23 (0.11)	10.28 (0.10)	10.27 (0.10)	10.22 (0.09)	10.20 (0.11)	10.20 (0.10)	7.66 (4.37)	20.58 (3.37)
	0.3	ACC	12.89 (0.39)	12.83 (0.29)	12.93 (0.39)	12.91 (0.40)	12.84 (0.28)	12.92 (0.38)	12.86 (0.39)	10.68 (6.59)	13.42 (0.30)
		Purity	20.00 (0.02)	20.01 (0.03)	20.02 (0.03)	20.00 (0.02)	20.01 (0.03)	20.00 (0.00)	20.00 (0.02)	11.34 (7.00)	20.27 (0.16)
		Fscore	10.33 (0.15)	10.47 (0.12)	10.31 (0.13)	10.33 (0.15)	10.38 (0.12)	10.31 (0.16)	10.28 (0.16)	8.93 (5.22)	14.58 (4.25)
0.5	ACC	13.11 (0.39)	13.09 (0.29)	13.15 (0.39)	13.07 (0.39)	13.13 (0.30)	13.13 (0.38)	13.16 (0.37)	11.26 (9.44)	13.47 (0.30)	
	Purity	20.06 (0.10)	20.06 (0.09)	20.05 (0.10)	20.06 (0.10)	20.04 (0.09)	20.06 (0.10)	20.06 (0.10)	12.15 (10.52)	20.34 (0.22)	
	Fscore	11.86 (0.14)	12.87 (0.24)	11.81 (0.14)	11.84 (0.14)	12.49 (0.24)	11.13 (0.65)	11.77 (0.13)	9.48 (8.03)	16.28 (2.20)	
Caltech101-7	0.1	ACC	17.90 (0.70)	17.94 (0.69)	17.02 (0.33)	17.59 (0.73)	17.69 (0.70)	17.59 (0.69)	17.37 (0.37)	23.74 (3.73)	53.11 (0.41)
		Purity	54.14 (0.00)	54.14 (0.00)	54.14 (0.00)	54.14 (0.00)	54.14 (0.00)	54.14 (0.00)	54.14 (0.00)	54.14 (0.00)	58.39 (3.82)
		Fscore	14.48 (0.12)	14.48 (0.13)	14.32 (0.07)	14.42 (0.12)	14.43 (0.13)	14.44 (0.11)	14.36 (0.06)	18.61 (2.86)	89.43 (0.74)
	0.3	ACC	17.54 (1.06)	17.95 (0.77)	17.33 (0.49)	17.28 (0.40)	17.63 (0.58)	17.24 (0.49)	17.16 (0.46)	19.70 (1.07)	50.61 (0.87)
		Purity	54.14 (0.00)	54.14 (0.00)	54.14 (0.00)	54.14 (0.00)	54.14 (0.00)	54.14 (0.00)	54.14 (0.00)	54.14 (0.00)	54.72 (0.10)
		Fscore	14.44 (0.17)	14.51 (0.14)	14.38 (0.09)	14.40 (0.10)	14.47 (0.12)	14.37 (0.08)	14.34 (0.08)	15.61 (0.09)	73.08 (1.11)
0.5	ACC	18.01 (0.93)	18.22 (1.12)	18.32 (0.91)	17.39 (0.49)	17.86 (1.49)	17.43 (0.40)	18.15 (1.19)	21.62 (1.79)	48.13 (0.82)	
	Purity	54.14 (0.00)	54.14 (0.00)	54.14 (0.00)	54.14 (0.00)	54.14 (0.00)	54.14 (0.00)	54.14 (0.00)	54.14 (0.00)	56.87 (2.44)	
	Fscore	14.58 (0.27)	14.71 (0.29)	14.66 (0.31)	14.44 (0.13)	14.61 (0.33)	14.44 (0.12)	14.64 (0.29)	16.59 (0.79)	61.22 (0.65)	
Caltech101-20	0.1	ACC	8.39 (0.31)	8.39 (0.35)	8.54 (0.20)	8.43 (0.39)	8.37 (0.42)	8.29 (0.31)	8.53 (0.42)	9.20 (4.65)	33.95 (0.25)
		Purity	33.46 (0.03)	33.47 (0.04)	33.47 (0.08)	33.45 (0.01)	33.45 (0.00)	33.46 (0.04)	33.45 (0.00)	23.88 (10.16)	36.83 (0.26)
		Fscore	5.05 (0.05)	5.05 (0.05)	5.12 (0.04)	5.04 (0.06)	5.04 (0.06)	5.04 (0.04)	5.11 (0.07)	4.67 (2.32)	88.29 (0.85)
	0.3	ACC	8.52 (0.32)	8.54 (0.28)	8.65 (0.34)	8.56 (0.29)	8.67 (0.34)	8.67 (0.30)	8.82 (0.30)	9.07 (4.85)	32.16 (0.58)
		Purity	33.48 (0.08)	33.47 (0.08)	33.60 (0.11)	33.49 (0.09)	33.45 (0.00)	33.47 (0.06)	33.52 (0.12)	28.96 (8.74)	36.48 (0.31)
		Fscore	5.15 (0.06)	5.12 (0.08)	5.16 (0.08)	5.15 (0.09)	5.16 (0.08)	5.19 (0.10)	5.18 (0.06)	4.93 (2.14)	70.41 (1.29)
0.5	ACC	8.53 (0.21)	8.66 (0.14)	8.42 (0.16)	8.76 (0.41)	8.81 (0.33)	8.76 (0.27)	8.40 (0.19)	6.86 (1.19)	30.42 (0.31)	
	Purity	33.45 (0.00)	33.45 (0.00)	33.45 (0.03)	33.45 (0.00)	33.46 (0.05)	33.47 (0.05)	33.45 (0.03)	26.45 (4.63)	36.00 (0.24)	
	Fscore	5.14 (0.05)	5.13 (0.05)	5.09 (0.04)	5.23 (0.11)	5.23 (0.11)	5.22 (0.08)	5.09 (0.04)	4.08 (0.71)	58.37 (0.81)	
Caltech101-all	0.1	ACC	4.04 (0.04)	4.02 (0.08)	4.01 (0.08)	4.06 (0.07)	4.04 (0.06)	4.07 (0.04)	4.09 (0.08)	4.69 (3.26)	9.36 (0.46)
		Purity	10.71 (0.16)	10.73 (0.12)	10.70 (0.08)	10.66 (0.14)	10.64 (0.13)	10.66 (0.10)	10.74 (0.10)	11.04 (7.53)	11.34 (0.10)
		Fscore	1.09 (0.02)	1.09 (0.01)	1.06 (0.02)	1.09 (0.02)	1.08 (0.02)	1.07 (0.01)	1.09 (0.01)	1.77 (1.33)	62.83 (16.40)
	0.3	ACC	4.06 (0.09)	4.10 (0.09)	4.01 (0.07)	4.07 (0.09)	4.09 (0.09)	4.09 (0.08)	4.04 (0.06)	3.93 (2.60)	10.02 (0.12)
		Purity	10.77 (0.16)	10.72 (0.18)	10.69 (0.17)	10.71 (0.10)	10.71 (0.11)	10.75 (0.14)	10.63 (0.17)	9.72 (6.03)	11.14 (0.09)
		Fscore	1.08 (0.02)	1.09 (0.02)	1.06 (0.02)	1.11 (0.02)	1.12 (0.01)	1.11 (0.03)	1.06 (0.01)	1.41 (1.07)	65.66 (1.46)
0.5	ACC	4.10 (0.11)	4.32 (0.05)	4.01 (0.05)	4.08 (0.07)	4.23 (0.12)	4.07 (0.10)	4.02 (0.05)	3.97 (2.60)	10.03 (0.20)	
	Purity	10.65 (0.05)	10.60 (0.10)	10.71 (0.12)	10.63 (0.13)	10.65 (0.14)	10.66 (0.10)	10.70 (0.09)	9.79 (6.23)	11.08 (0.12)	
	Fscore	1.12 (0.03)	1.40 (0.05)	1.05 (0.02)	1.13 (0.03)	1.26 (0.08)	1.11 (0.02)	1.06 (0.01)	1.40 (1.01)	55.35 (1.34)	

Table 2: Partial results (mean(std)) of KOALA and other compared methods on seven datasets. ‘*’ represents ‘MKKM’.

1) Our proposed method achieves superior clustering results on the widely used datasets compared to eight other approaches, validating its ability to cluster uncoupled incomplete multi-view data. Specifically, the improvements of our algorithm over the second best method in terms of ACC are 4.07%, 3.64%, 6.70%, 0.49%, 23.09%, 5.92%, 5.92% for the missing ratio of 0.3 on all benchmarks, demonstrating the superior performance of KOALA.

2) Our approach outperforms existing kernel-based IMVC methods. For instance, KOALA outperforms the best alternatives by 0.37%, 0.25% and 0.28% in terms of purity, despite different missing ratios on UCL.Digits. The gaps for the other datasets have similar properties. The success of the

proposed approach is attributed to the utilization of cross-kernel coupling learning, which exploits existing consensus and complementary information to recouple kernel elements and improve clustering results.

3) UIMC also serves as an IMVC approach tailored for uncoupled incomplete multi-view data, which outperforms most IMVC methods in various circumstances due to its coupling learning. The comparison between our proposed UC-IMVC and UIMC is presented in Table 2. As observed from the table, our method shows better clustering performance than UIMC with different missing ratio in most cases. Specifically, in terms of ACC for the missing ratio of 0.5, our approach achieves better performance than UIMC on

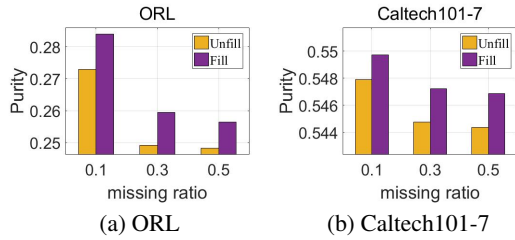


Figure 2: The ablation study of our missing kernel elements recovery strategy.

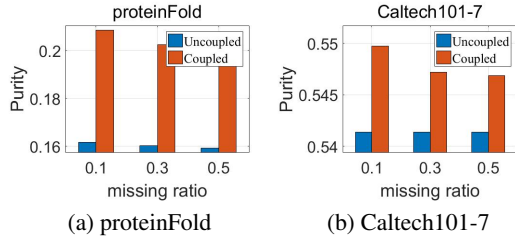


Figure 3: The ablation study of our cross-kernel coupling learning.

all datasets, i.e., 5.42%, 3.59%, 40.22%, 2.21%, 26.51%, 23.56% and 6.06% respectively, which demonstrates the effectiveness of KOALA in addressing the uncoupled incomplete multi-view data.

Ablation Study. In order to analyze two primary contributions in our algorithm, we conduct two ablation experiments to illustrate the validity of our strategies separately.

Missing kernel elements recovery strategy. The purpose of this ablation experiment is to verify the effectiveness of the imputation of uncoupled incomplete multi-view data. We present the experimental results of the ablation study in Fig. 2, where "Unfill" indicates not using our incomplete multi-view imputation strategy. As shown in Fig. 2, our method has significant advantages in clustering performance, which fully proves the validity of the strategy of kernel completion.

Cross-kernel coupling learning Strategy. To further verify the superiority of the cross-kernel mapping strategy, we recommend conducting the ablation study experiment presented in Fig. 3, where "Uncoupled" manifests the correlation relationship among different kernels is unknown. As we can see from Fig. 3, our proposed cross-kernel coupling learning strategy enhances the performance under various missing ratios on proteinFold and Caltech101-7 datasets.

Convergence. Convergence to a local optimum of our proposed method is guaranteed. The objective value in Eq. (6) consistently decreases in each iteration when optimizing one variable while keeping the others fixed. This provides proof that the entire optimization algorithm converges to a local optimum. To demonstrate this point in practice, we generate plots of the objective value curves of our approach, specifi-

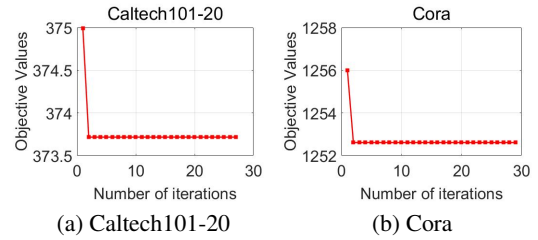


Figure 4: Caltech101-20 and Cora objective values from KOALA with missing ratio 0.1.

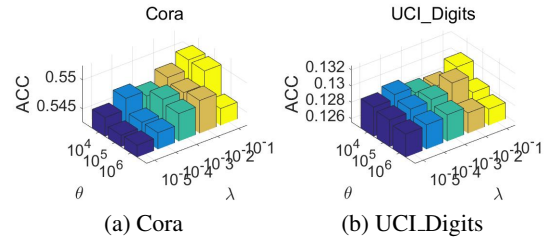


Figure 5: Sensitivity analysis of λ and θ of our method on two benchmark datasets.

cally showing the number of iterations with varying missing ratios on partial datasets. As observed in Fig. 4, the objective value of KOALA decreases monotonically with each iteration and typically converges quickly.

Parameter Sensitivity. KOALA introduces two parameters λ and θ to trade off better clustering performance. λ denotes the low-rank tensor constraint balance parameter and θ is the kernel alignment balance parameter. We perform experiments to investigate the impact and sensitivity of these parameters on clustering performance across all datasets. We set λ in a range of $[10^{-5}, 10^{-1}]$ and θ in $[10^4, 10^6]$. Fig. 5 shows the ACC of KOALA on two datasets with different missing ratios. From the observation, our approach achieves stable clustering performance over a wide range of λ and θ .

Conclusion

In this paper, we present a novel coupling-imputation-clustering framework termed Kernel Coupling and Element Imputation Induced Multi-View Clustering (KOALA). The proposed method constructs a kernel for each uncoupled incomplete multiple view and uses cross-kernel coupling learning to refine the correspondence among multiple kernels. Meanwhile, KOALA explores high-order correlations to guide the imputation of absent kernel elements. We jointly optimize the reconstruction of incomplete kernels, cross-kernel alignment, and clustering within an integrated framework to achieve superior clustering performance. Finally, a six-step alternative algorithm is designed to ensure KOALA convergence. The effectiveness of the model is demonstrated through extensive experiments on benchmark datasets.

Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities (No. 2022JBZY019) and the Beijing Natural Science Foundation (No. 4242046). Ji-azheng Yuan is the co-corresponding author.

References

- Dong, Z.; Jin, J.; Xiao, Y.; Wang, S.; Zhu, X.; Liu, X.; and Zhu, E. 2023a. Iterative Deep Structural Graph Contrast Clustering for Multiview Raw Data. *IEEE Transactions on Neural Networks and Learning Systems*, 1–13.
- Dong, Z.; Wang, S.; Jin, J.; Liu, X.; and Zhu, E. 2023b. Cross-view Topology Based Consistent and Complementary Information for Deep Multi-view Clustering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 19440–19451.
- Gönen, M.; and Margolin, A. A. 2014. Localized data fusion for kernel k-means clustering with application to cancer biology. *Advances in Neural Information Processing Systems*, 27.
- Gu, Z.; and Feng, S. 2023. Individuality Meets Commonality: A Unified Graph Learning Framework for Multi-View Clustering. *ACM Transactions on Knowledge Discovery from Data*, 17(1): 1–21.
- Gu, Z.; Feng, S.; Li, Z.; Yuan, J.; and Liu, J. 2024. NOODLE: Joint Cross-View Discrepancy Discovery and High-Order Correlation Detection for Multi-View Subspace Clustering. *ACM Transactions on Knowledge Discovery from Data*, 18(6): 1–23.
- Gu, Z.; Li, Z.; and Feng, S. 2024. Topology-Driven Multi-View Clustering via Tensorial Refined Sigmoid Rank Minimization. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 920–931.
- Hu, M.; and Chen, S. 2019a. Doubly aligned incomplete multi-view clustering. *ArXiv Preprint ArXiv:1903.02785*.
- Hu, M.; and Chen, S. 2019b. One-pass incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 3838–3845.
- Jin, J.; Wang, S.; Dong, Z.; Liu, X.; and Zhu, E. 2023. Deep Incomplete Multi-view Clustering with Cross-view Partial Sample and Prototype Alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11600–11609.
- Li, L.; Wan, Z.; and He, H. 2021. Incomplete multi-view clustering with joint partition and graph learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(1): 589–602.
- Li, S.-Y.; Jiang, Y.; and Zhou, Z.-H. 2014. Partial multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28.
- Li, X.; Sun, Y.; Sun, Q.; Dai, J.; and Ren, Z. 2023. Distribution Consistency based Fast Anchor Imputation for Incomplete Multi-view Clustering. In *Proceedings of the 31st ACM International Conference on Multimedia*, 368–376.
- Lin, J.-Q.; Li, X.-L.; Chen, M.-S.; Wang, C.-D.; and Zhang, H. 2022a. Incomplete Data Meets Uncoupled Case: A Challenging Task of Multiview Clustering. *IEEE Transactions on Neural Networks and Learning Systems*.
- Lin, Y.; Gou, Y.; Liu, X.; Bai, J.; Lv, J.; and Peng, X. 2022b. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 4447–4461.
- Lin, Y.; Gou, Y.; Liu, Z.; Li, B.; Lv, J.; and Peng, X. 2021. Completer: Incomplete multi-view clustering via contrastive prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11174–11183.
- Lin, Z.; Liu, R.; and Su, Z. 2011. Linearized alternating direction method with adaptive penalty for low-rank representation. *Advances in Neural Information Processing Systems*, 24.
- Liu, J.; Liu, X.; Zhang, Y.; Zhang, P.; Tu, W.; Wang, S.; Zhou, S.; Liang, W.; Wang, S.; and Yang, Y. 2021. Self-representation subspace clustering for incomplete multi-view data. In *Proceedings of the 29th ACM International Conference on Multimedia*, 2726–2734.
- Liu, X.; Li, M.; Tang, C.; Xia, J.; Xiong, J.; Liu, L.; Kloft, M.; and Zhu, E. 2020. Efficient and effective regularized incomplete multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(8): 2634–2646.
- Liu, X.; Wang, L.; Zhu, X.; Li, M.; Zhu, E.; Liu, T.; Liu, L.; Dou, Y.; and Yin, J. 2019a. Absent multiple kernel learning algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(6): 1303–1316.
- Liu, X.; Zhu, X.; Li, M.; Tang, C.; Zhu, E.; Yin, J.; and Gao, W. 2019b. Efficient and effective incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 4392–4399.
- Liu, X.; Zhu, X.; Li, M.; Wang, L.; Tang, C.; Yin, J.; Shen, D.; Wang, H.; and Gao, W. 2018. Late fusion incomplete multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(10): 2410–2423.
- Liu, X.; Zhu, X.; Li, M.; Wang, L.; Zhu, E.; Liu, T.; Kloft, M.; Shen, D.; Yin, J.; and Gao, W. 2019c. Multiple kernel k -means with incomplete kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(5): 1191–1204.
- Tang, H.; and Liu, Y. 2022. Deep safe incomplete multi-view clustering: Theorem and algorithm. In *International Conference on Machine Learning*, 21090–21110. PMLR.
- Wan, X.; Liu, J.; Liang, W.; Liu, X.; Wen, Y.; and Zhu, E. 2022. Continual multi-view clustering. In *Proceedings of the 30th ACM International Conference on Multimedia*, 3676–3684.
- Wang, H.; Zong, L.; Liu, B.; Yang, Y.; and Zhou, W. 2019. Spectral perturbation meets incomplete multi-view data. *ArXiv Preprint ArXiv:1906.00098*.
- Wang, J.; Li, Z.; Tang, C.; Liu, S.; Wan, X.; and Liu, X. 2024. Multiple Kernel Clustering with Adaptive Multi-scale Partition Selection. *IEEE Transactions on Knowledge and Data Engineering*.

- Wen, J.; Xu, G.; Liu, C.; Fei, L.; Huang, C.; Wang, W.; and Xu, Y. 2023. Localized and Balanced Efficient Incomplete Multi-view Clustering. In *Proceedings of the 31st ACM International Conference on Multimedia*, 2927–2935.
- Wen, J.; Yan, K.; Zhang, Z.; Xu, Y.; Wang, J.; Fei, L.; and Zhang, B. 2020. Adaptive graph completion based incomplete multi-view clustering. *IEEE Transactions on Multimedia*, 23: 2493–2504.
- Wen, J.; Zhang, Z.; Xu, Y.; Zhang, B.; Fei, L.; and Liu, H. 2019. Unified embedding alignment with missing views inferring for incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 5393–5400.
- Wen, J.; Zhang, Z.; Xu, Y.; Zhang, B.; Fei, L.; and Xie, G.-S. 2021. Cdimc-net: Cognitive deep incomplete multi-view clustering network. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 3230–3236.
- Wen, Z.; and Yin, W. 2013. A feasible method for optimization with orthogonality constraints. *Mathematical Programming*, 142(1-2): 397–434.
- Wu, T.; Feng, S.; and Yuan, J. 2024. Low-Rank Kernel Tensor Learning for Incomplete Multi-View Clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 15952–15960.
- Xie, Y.; Tao, D.; Zhang, W.; Liu, Y.; Zhang, L.; and Qu, Y. 2018. On unifying multi-view self-representations for clustering by tensor multi-rank minimization. *International Journal of Computer Vision*, 126: 1157–1179.
- Xu, J.; Chen, S.; Ren, Y.; Shi, X.; Shen, H. T.; Niu, G.; and Zhu, X. 2024a. Self-weighted contrastive learning among multiple views for mitigating representation degeneration. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, 1119 – 1131.
- Xu, J.; Li, C.; Ren, Y.; Peng, L.; Mo, Y.; Shi, X.; and Zhu, X. 2022. Deep incomplete multi-view clustering via mining cluster complementarity. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 8761–8769.
- Xu, J.; Ren, Y.; Wang, X.; Feng, L.; Zhang, Z.; Niu, G.; and Zhu, X. 2024b. Investigating and Mitigating the Side Effects of Noisy Views for Self-Supervised Clustering Algorithms in Practical Multi-View Scenarios. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 22957–22966.
- Xu, N.; Guo, Y.; Zheng, X.; Wang, Q.; and Luo, X. 2018. Partial multi-view subspace clustering. In *Proceedings of the 26th ACM International Conference on Multimedia*, 1794–1801.
- Yan, W.; Xu, J.; Liu, J.; Yue, G.; and Tang, C. 2022. Bipartite graph-based discriminative feature learning for multi-view clustering. In *Proceedings of the 30th ACM International Conference on Multimedia*, 3403–3411.
- Yu, S.; Tranchevent, L.; Liu, X.; Glanzel, W.; Suykens, J. A.; De Moor, B.; and Moreau, Y. 2011. Optimized data fusion for kernel k-means clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5): 1031–1039.
- Zhang, C.; Cui, Y.; Han, Z.; Zhou, J. T.; Fu, H.; and Hu, Q. 2020. Deep partial multi-view learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5): 2402–2415.
- Zhang, C.; Li, H.; Lv, W.; Huang, Z.; Gao, Y.; and Chen, C. 2023a. Enhanced tensor low-rank and sparse representation recovery for incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 11174–11182.
- Zhang, C.; Wei, J.; Wang, B.; Li, Z.; Chen, C.; and Li, H. 2023b. Robust Spectral Embedding Completion Based Incomplete Multi-view Clustering. In *Proceedings of the 31st ACM International Conference on Multimedia*, 300–308.
- Zhang, Y.; Liu, X.; Liu, J.; Dai, S.; Zhang, C.; Xu, K.; and Zhu, E. 2022. Fusion multiple kernel k-means. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 9109–9117.
- Zhang, Y.; Liu, X.; Wang, S.; Liu, J.; Dai, S.; and Zhu, E. 2021. One-stage incomplete multi-view clustering via late fusion. In *Proceedings of the 29th ACM International Conference on Multimedia*, 2717–2725.
- Zhang, Z.; Wang, Q.; Tao, Z.; Gao, Q.; and Feng, W. 2023c. Dropping pathways towards deep multi-view graph subspace clustering networks. In *Proceedings of the 31st ACM International Conference on Multimedia*, 3259–3267.
- Zhao, S.; Fei, L.; Wen, J.; Zhang, B.; and Zhao, P. 2023. Incomplete Multi-View Clustering with Regularized Hierarchical Graph. In *Proceedings of the 31st ACM International Conference on Multimedia*, 3060–3068.