

# EXPLAINABLE $K$ -MEANS NEURAL NETWORKS FOR MULTI-VIEW CLUSTERING

**Yalan Qin, Xinpeng Zhang, Guorui Feng** \*

School of Communication and Information Engineering, Shanghai University

{ylqin, xzhang, grfeng}@shu.edu.edu

## ABSTRACT

Despite multi-view clustering has achieved great progress in past decades, it is still a challenge to balance the effectiveness, efficiency, completeness and consistency of nonlinearly separable clustering for the data from different views. To address this challenge, we show that multi-view clustering can be regarded as a three-level optimization problem. To be specific, we divide the multi-view clustering into three sub-problems based on  $K$ -means or kernel  $K$ -means, i.e., linear clustering on the original multi-view dataset, nonlinear clustering on the set of obtained linear clusters and multi-view clustering by integrating partition matrices from different views obtained by linear and nonlinear clustering based on reconstruction. We propose Explainable  $K$ -means Neural Networks (EKNN) and present how to unify these three sub-problems into a framework based on EKNN. It is able to simultaneously consider the effectiveness, efficiency, completeness and consistency for the nonlinearly multi-view clustering and can be optimized by an iterative algorithm. EKNN is explainable since the effect of each layer is known. To the best of our knowledge, this is the first attempt to balance the effectiveness, efficiency, completeness and consistency by dividing the multi-view clustering into three different sub-problems. Extensive experimental results demonstrate the effectiveness and efficiency of EKNN compared with other methods for multi-view clustering on different datasets in terms of different metrics.

## 1 INTRODUCTION

As an important problem in machine learning and data mining, clustering aims to group a set of data points into clusters Qin et al. (2021; 2022b; 2023c;a; 2024a; 2023b; 2025f); Pu et al. (2023); Qin et al. (2025d;g;b;h); Qin & Qian (2024); Qin et al. (2024b;c; 2025e;a;c; 2023d); Li et al. (2023a;b); Liu et al. (2024; 2023a; 2025); Lu et al. (2024); Zhang et al. (2025); Li et al. (2025); Liu et al. (2023b; 2022b;a). The data points in different clusters are highly dissimilar but remarkably similar with data points in the same cluster. It has different applications in many fields including image processing Sandler & Lindenbaum (2011), bioinformatics Perner (2002) and signal processing Gupta & Xiao (2011). Various clustering algorithms have been presented Veldt et al. (2019); Gebru et al. (2016), which can be divided into four categories including hierachical, partitional, grid-based and density-based clustering Camastra & Verri (2005); Ng et al. (2001). However, some representative methods Fukunaga & Hostetler (1975); Mehmood et al. (2015) belonging to these categories need to calculate the distances between all data points in the dataset. Consequently, it needs an extremely high computational cost when dealing with large-scale datasets.

Many methods have been developed to improve the efficiency of clustering algorithms by reducing unnecessary calculations of distances between data points Bentley (1975); Viswanath & Pinkesh (2006); He et al. (2011), e.g., spatial index structure Bentley (1975), hybrid clustering Viswanath & Pinkesh (2006), grid-based clustering and parallel clustering He et al. (2011). However, most existing methods pay more attention to the effectiveness and the improvements in efficiency is not so significant as effectiveness. Recently, some methods have been proposed to directly improve the clustering speed of the existing work Mehmood et al. (2015) without considering extra conditions Wu & Wilamowski (2017); Zhang et al. (2016). These methods fail to fully explore how to reduce

\*Corresponding author

the calculations of distances between data points. As a well-known algorithm,  $K$ -means Queen (1966) belongs to the linearly separable clustering algorithm and has a low computational cost. It is not able to well recognize clusters which are nonlinearly separable, leading to unsatisfied clustering performance.

Nonlinear clustering is a widely exploited problem due to the complex structure of the data in the real world. Recently, various types of nonlinear clustering methods have been proposed Ng et al. (2001); Schölkopf et al. (1998); Cheng (1995) to better handle clusters with arbitrary shapes. Methods based on kernel extract nonlinear separable clusters by adopting a proper nonlinear mapping between the input space to a feature space with high dimensions. This nonlinear mapping is called the kernel function and the representative methods are spectral clustering Ng et al. (2001), kernel  $K$ -means Schölkopf et al. (1998) and mean-shift Cheng (1995). To determine the relation of each data point to the clusters, these methods need to compute the pairwise similarities of data points. The reason why they are able to recognize nonlinear clusters is that these approaches employ all data points to represent a cluster without choosing a center. However, their space or time costs are high, resulting in unsatisfied results for large-scale datasets. To effectively denote a nonlinear cluster, methods based on multi-clusters have been proposed Liu et al. (2009); Liang et al. (2012); Wang et al. (2013). These methods are able to reduce the computation complexity results by replacing data points with different centers. Therefore, their clustering performances are easily influenced by the chosen center and they also need high computational costs to obtain centers with high quality.

Information is usually presented in different forms simultaneously in the real world, including multiple types of features and multiple modalities Sui et al. (2018); Li & Tang (2017). We consider different types of information as multiple views describing objects. Given web pages, we can extract different types of features based on hyperlinks, text and possible visual information. Multiple types of features can be extracted based on text, edge and color for images. Various existing methods Zhang et al. (2020b;a) have demonstrated that obvious performance improvement can be achieved since different views complete each other. In multi-view learning, consistency and completeness are two well-known principles. To achieve the consistency of different views, the existing work assume that correlations among different views should be maximized Wang et al. (2015). The complete information is important to obtain comprehensive representations. However, existing work fails to balance the effectiveness, efficiency, consistency and completeness of multi-view clustering. Accordingly, a challenging problem appears - how to effectively and efficiently perform nonlinear clustering for multi-view datasets.

Motivated by the fact that different clusters in a complex dataset are nonlinearly separable in the global geometric space but linearly separable in the local geometric space, we assume that some small linearly separable clusters make up a nonlinearly separable cluster. Different from existing studies, we show that multi-view clustering can be treated as a three-level optimization problem based on this assumption, which is presented in Fig. 1. Our method is expected to have the following merits: effectiveness - the quality of multi-view clustering results is guaranteed, efficiency - the computational cost is able to be lowered, completeness and consistency - the clustering results incorporate information from different views and multiple views are consistent, where the symbol “-” is used for explanation. We divide the multi-view clustering into three different sub-problems: 1) the linear clustering on the set of original data points for every view, 2) the nonlinear clustering of each view on the set of linear clusters obtained by 1), and 3) multi-view clustering by incorporating partition matrices from different views obtained by linear and nonlinear clustering based on reconstruction. Here, the linear clustering makes data points similar with each other in the local geometric space into the same clusters, which reduces the size of the data for nonlinear clustering. Then, the efficiency is considered in this manner. Next, we perform nonlinear clustering to partition the linear clusters into different nonlinear clusters. In detail, we aim to project a given multi-view dataset into a feature space where the effectiveness, efficiency, completeness and consistency are guaranteed using three jointly learning objective functions. These three objective functions consist of linear clustering, nonlinear clustering and multi-view clustering, which are all defined based on the  $K$ -means or kernel  $K$ -means objective. As convolutional operation in Convolutional Neural Networks (CNN),  $K$ -means plays the same role in  $K$ -means Neural networks. The connections of different  $K$ -means components in Fig. 1 have the same usage as neural network in CNN. Therefore, we propose Explainable  $K$ -means Neural Networks (EKNN) based on  $K$ -means for multi-view clustering and use the iterative method in  $K$ -means to optimize the networks. Besides, it is explainable since the effect of each layer in EKNN is knowable, i.e., the layer of kernel  $K$ -means is adopted to obtain nonlinear

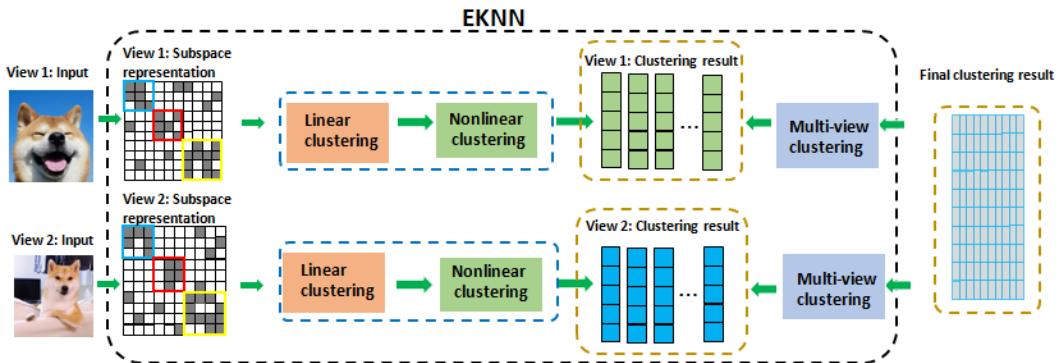


Figure 1: Framework of EKNN. Given the multi-view data as input, subspace representations from different views can be obtained based on self-representative coefficient matrix. The layer of linear clustering, nonlinear clustering and multi-view clustering are used to achieve the goal of efficiency, effectiveness, consistency and completeness, respectively. After linear and nonlinear clustering, we can obtain clustering result for single view. With EKNN, final clustering result can be achieved for the multi-view data.

clusters. We can also observe that  $K$ -means and kernel  $K$ -means are special cases of EKNN. We also extend EKNN for multi-view clustering to the case of multi-view subspace learning, which is able to learn a desired latent representation shared by different views. The shared subspace representation can be obtained by the latent representation based on self-expressiveness Elhamifar & Vidal (2013), which is used to achieve the final results by the existing clustering algorithms, i.e., spectral clustering algorithm.

The main contributions in this work are:

- We give a novel insight to the multi-view clustering community, i.e., the multi-view clustering can be regarded as a three-level optimization problem. We propose to divide the multi-view clustering into three sub-problems: the linear clustering on the set of original data points based on  $K$ -means, the nonlinear clustering of each view on the set of linear clusters based on kernel  $K$ -means, and multi-view clustering by integrating information from different views for the set of partitions in terms of  $K$ -means.
- We propose Explainable  $K$ -means Neural Networks (EKNN) based on  $K$ -means, which flexibly integrates the linear clustering, nonlinear clustering and multi-view clustering into a framework. It is explainable and able to balance the clustering quality, computation cost, completeness and consistency of multi-view clustering. The iterative method as in  $K$ -means is used to solve the optimization problem of EKNN for multi-view clustering.
- We further extend EKNN for multi-view clustering to the case of multi-view subspace learning, which is able to learn a desired latent representation shared by different views. Extensive experiments performed on different datasets demonstrate the efficiency and effectiveness of EKNN for multi-view clustering in terms of different metrics compared with the existing clustering algorithms.

The rest of this paper is organized as follows. Section 2 reviews some related work of  $K$ -means algorithm and multi-view representation learning. Section 3 describes the details of EKNN and the extension of EKNN. In Sections 4 and 5, we conduct experiments to validate the merits of EKNN and show the conclusion, respectively.

## 2 THE PROPOSED METHOD

In this section, we present that multi-view clustering can be seen as a three-level optimization problem and propose EKNN based on  $K$ -means for multi-view clustering. For multi-view dataset, we first explore the properties of each single view and then integrate them to realize multi-view clustering. For each view of a complex dataset, different clusters are nonlinearly separable in the global

geometric space but linearly separable in the local geometric space. Therefore, for each view, we assume that a linearly separable cluster is made up of the original data points in the dataset and several linearly separable clusters make up a nonlinearly separable cluster. Based on this assumption, the final shared clustering result is achieved by integrating partition matrices from different views obtained by linear and nonlinear clustering based on reconstruction, which is complete. The consistency among multiple views can be indirectly obtained by enforcing the final clustering result to be shared by different views. Built on these observations, we divide the multi-view clustering into three sub-problems based on  $K$ -means or kernel  $K$ -means. The proposed EKNN unifies these three sub-problems into a framework and we solve it by the iterative optimization. EKNN is explainable and consists of some layers including linear clustering based on  $K$ -means, nonlinear clustering based on kernel  $K$ -means and multi-view clustering based on  $K$ -means.

## 2.1 FORMULATED PROBLEM

Given the dataset  $X = \{X^1, X^2, \dots, X^V\}$ , where  $X^v \in R^{n \times m}$ ,  $V$  is the number of views,  $n$  and  $m$  denote the number of total points and dimensions in the dataset, respectively. As abovementioned, the proposed EKNN consists of three parts: 1) linear clustering on the set of data points in the dataset based on  $K$ -means, 2) nonlinear clustering on the set of linear clusters obtained by 1) based on kernel  $K$ -means, and 3) multi-view clustering by incorporating partition matrices from different views obtained by linear and nonlinear clustering based on  $K$ -means. For each view, we first define the part of EKNN for linear clustering based on  $K$ -means as:

$$f'_{L^v} = \|X^v - W^v V^v\|_F^2, \quad (1)$$

where  $W^v = [w_{ij}^v] \in R^{n \times p}$  is the partition matrix of data points,  $w_{ij}^v$  denotes the membership of the  $i$ -th data point to the  $j$ -th linear cluster,  $p$  denotes the number of linear clusters in  $W$  and  $V^v \in R^{p \times m}$  is the cluster matrix of the  $v$ -th view. Note that the value of  $p$  is required to be larger than the number of real clusters  $k$  in the dataset. Then data points in the same linear clusters are close to each other in the original space of  $X$  when the value of  $f'_{L^v}$  is low.

Rather than integrating different views in the level of raw features as LMSC Zhang et al. (2017), we employ the subspace representation  $\Theta^v = [\theta_{ij}^v] \in R^{n \times n}$  for the  $v$ -th view based on the self-representative coefficient matrix, which is obtained by subspace clustering in terms of self-representation. Since the global or local space is built on the relative relationship between data points in the extracted feature representation of  $X^v$  for the  $v$ -th view, a linearly separable cluster is made up of the refined data points in the dataset and several linearly separable clusters make up a nonlinearly separable cluster. The subspace representation  $\Theta^v$  can reveal the underlying cluster structures of data. Then Eq. (1) is reformulated as:

$$f_{L^v} = \|\Theta^v - W^v V^v\|_F^2, \text{ s.t. } X^v = X^v \Theta^v, \quad (2)$$

where the dimensions of  $W$  and  $V$  are  $n \times p$  and  $p \times n$  here, respectively. The motivation that some clustering methods need to compute pairwise similarity between points is originated from the perspective of the final unified learning framework. The computation of the matrix  $\Theta$  with  $n \times n$  dimension is the pre-processing step, which is used as the input to the final unified learning framework. Considering the case that  $\text{diag}(\Theta^v) = 0$  can be removed based on the analysis in Lu et al. (2012), we do not impose the constraint of  $\text{diag}(\Theta^v) = 0$  to Eq. (2). It is observed that the part of EKNN for linear clustering consists of two layers, i.e., one layer is used for  $X^v = X^v \Theta^v$  and the other is employed to achieve  $f_{L^v} = \|\Theta^v - W^v V^v\|_F^2$ . Eq. (2) is the first part of EKNN for linear clustering and we then focus on another two parts. For each view, the second part of EKNN for nonlinear clustering is defined based on kernel  $K$ -means:

$$\begin{aligned} f_{N^v} &= \|\Phi(V^v) - U^v Z^v\|_F^2 \\ &= \text{Tr}(K^v) - \text{Tr}((\hat{U}^v)^T K^v \hat{U}^v), \end{aligned} \quad (3)$$

where  $U^v = [u_{ij}^v] \in R^{p \times k}$  is the partition matrix of  $V^v$ ,  $K^v \in R^{p \times p}$  denotes the kernel matrix of  $V^v$ ,  $\hat{U}^v = U^v (D^v)^{-\frac{1}{2}}$  is the normalized matrix of  $U^v$ ,  $D^v = [d_j^v] \in R^{k \times k}$  with  $d_j^v = \sum_{i=1}^p u_{ij}^v$ ,  $Z^v$  and  $\Phi(V^v)$  are cluster centers and the representations of  $V^v$  by a kernel function in the nonlinearly embedded space, respectively. The kernel used in Eq. (3) is Gaussian kernel. Based on Eq. (3),

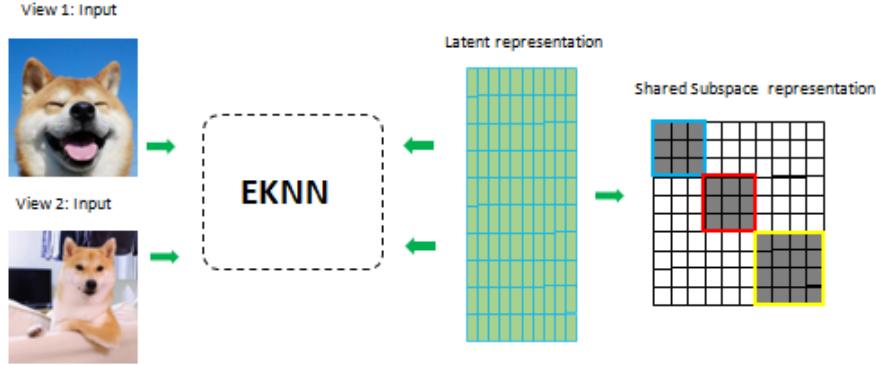


Figure 2: Framework of EKNN for multi-view subspace learning.

the difference of linear clusters belonging to the same nonlinear cluster can be minimized. We then define the last part of EKNN for multi-view clustering based on  $K$ -means as:

$$\begin{aligned} f_{M^v} &= \|W^v \hat{U}^v - \Psi^v(H)\|_F^2 \\ &= \|W^v \hat{U}^v - HG^v\|_F^2, \end{aligned} \quad (4)$$

where  $W^v$  and  $\hat{U}^v$  come from Eq. (2) and Eq. (3), respectively.  $H \in R^{n \times k}$  denotes the shared partition matrix of the final clustering,  $\Psi^v(\cdot)$  is the mapping for the  $v$ -th view based on reconstruction,  $G^v \in R^{k \times k}$  is the relation matrix between  $H$  and  $W^v \hat{U}^v$ . According to the physical meaning of different notations and their dimensions, we combine  $W^v$  and  $\hat{U}^v$  together in Eq (4). Here, we just consider the simple yet effective mapping, i.e., linear mapping with  $G^v$ . We can observe that  $f_{M^v}$  is the classical  $K$ -means algorithm when  $W^v \hat{U}^v$  is fixed. Based on reconstruction, properties from different views can be encoded into the shared partition matrix  $H$  in this manner.

Based on Eqs. (2)-(4), we define the final objective function  $F$  of EKNN as:

$$\begin{aligned} \min_H F &= \sum_{v=1}^V \alpha f_{L^v} + \beta f_{N^v} + \gamma f_{M^v}, \text{ s.t. } \in \{0, 1\}, 1 < \sum_{i=1}^n w_{ij}^v < n, X^v = X^v \Theta^v \\ &\sum_{j=1}^k u_{ij}^v = 1, 1 < \sum_{i=1}^p u_{ij}^v < p, w_{ij}^v, h_{ij}^v, u_{ij}^v \in \{0, 1\}, \sum_{j=1}^k h_{ij}^v = 1, 1 < \sum_{i=1}^n h_{ij}^v < n, \end{aligned} \quad (5)$$

where  $\alpha, \beta, \gamma > 0$  are parameters for balancing different terms. To alleviate the effect of noise, we apply the Frobenius norm for reconstruction loss based on  $X^v = X^v \Theta^v$  and use the nuclear norm as regularization term to ensure the high homogeneity within class. We then rewrite Eq. (5) as:

$$\begin{aligned} \min_H F &= \sum_{v=1}^V \alpha \| \Theta^v - W^v V^v \|_F^2 + \beta \text{Tr}(K^v) - \beta \text{Tr}((\hat{U}^v)^T K^v \hat{U}^v) + \gamma \| W^v \hat{U}^v - HG^v \|_F^2 \\ &+ \eta \| X^v - X^v \Theta^v \|_F^2 + \mu \| \Theta^v \|_*, \text{ s.t. } w_{ij}^v \in \{0, 1\}, \sum_{j=1}^p w_{ij}^v = 1, 1 < \sum_{i=1}^n w_{ij}^v < n, \\ &u_{ij}^v \in \{0, 1\}, \sum_{j=1}^k u_{ij}^v = 1, 1 < \sum_{i=1}^p u_{ij}^v < p, h_{ij}^v \in \{0, 1\}, \sum_{j=1}^k h_{ij}^v = 1, 1 < \sum_{i=1}^n h_{ij}^v < n, \end{aligned} \quad (6)$$

where  $\eta, \mu > 0$  are weight parameters. The solution for the above objective function, the corresponding algorithm and complexity analysis are shown in Appendix.

## 2.2 EXPLAINABILITY OF EKNN

Despite the remarkable progress has been achieved by eXplainable Artificial Intelligence (XAI), the explainability usually refers to the understandability of a model built on post-hoc explanations by

different methods, i.e., visual explanations, text explanations, and feature relevance explanations. In this work, we show the explainability from the perspective of the model design, which is more desired in real applications. We express the explainability of the model as transparency, which consists of algorithmic transparency and model decomposability. We then show why EKNN enjoys these two explainable properties in the following.

EKNN owns model decomposability since the input, learned variables and the loss function are all explainable. To be specific, the input to EKNN corresponds to the original dataset, the learned variables have their clear physical meanings and the loss function is built on the  $K$ -means and kernel  $K$ -means. EKNN also embraces algorithmic transparency and the dynamic behavior or error surface can be mathematically reasoned, which can make the user better know how the model works.

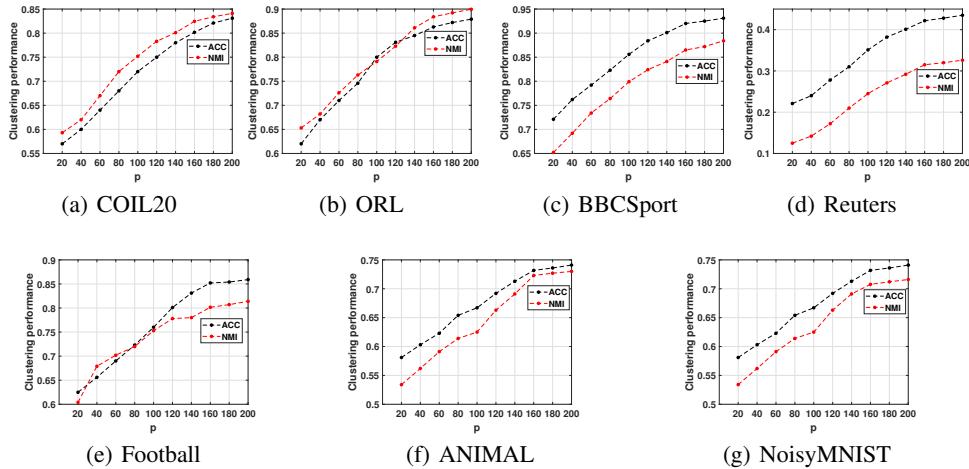


Figure 3: Clustering performance of EKNN with different  $p$  on different datasets.

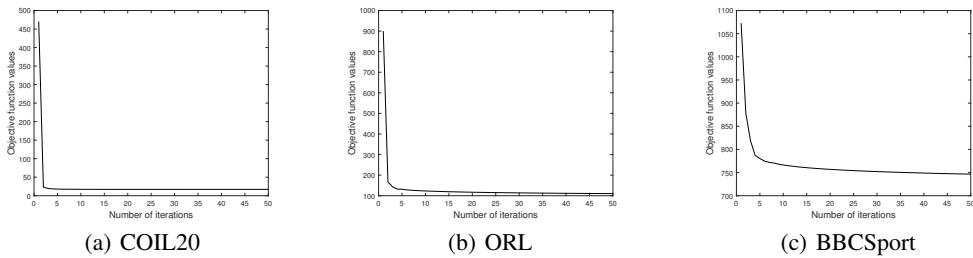


Figure 4: Convergence curve of EKNN on COIL20, ORL and BBCSport dataset.

### 2.3 EXTENSION OF THE PROPOSED METHOD

To learn a more effective latent representation shared by different views for multi-view clustering, we extend EKNN for multi-view clustering to the case of multi-view subspace learning, which is shown in Fig. 2. Note that we introduce shared subspace representation  $\Theta_C$  based on the latent representation  $H$  for existing clustering algorithms, i.e., spectral clustering algorithm.

Then the final objective function  $F$  of EKNN for multi-view subspace learning is formulated as:

$$\begin{aligned}
 \min_H F = & \sum_{v=1}^V \alpha \|\Theta^v - W^v V^v\|_F^2 + \beta \text{Tr}(K^v) + \mu' \|R'\|_* - \beta \text{Tr}((\hat{U}^v)^T K^v \hat{U}^v) + \eta' \|H - H\Theta_C\|_F^2 \\
 & + \eta \|X^v - X^v \Theta^v\|_F^2 + \mu \|R^v\|_* + \gamma \|W^v \hat{U}^v - HG^v\|_F^2, \\
 \text{s.t. } & w_{ij}^v \in \{0, 1\}, \sum_{j=1}^p w_{ij}^v = 1, 1 < \sum_{i=1}^n w_{ij}^v < n, u_{ij}^v \in \{0, 1\}, \sum_{j=1}^k u_{ij}^v = 1, 1 < \sum_{i=1}^p u_{ij}^v < p, \\
 & h_{ij}^v \in \{0, 1\}, \sum_{j=1}^k h_{ij}^v = 1, 1 < \sum_{i=1}^n h_{ij}^v < n, \Theta^v = R^v,
 \end{aligned} \tag{7}$$

where  $k_1 > k$ ,  $k_2 > k$ ,  $\mu' > 0$ ,  $\eta' > 0$  are parameters. Likewise, we can obtain the solution to update  $U^v$ ,  $W^v$ ,  $\Theta^v$ ,  $V^v$ ,  $H$  and  $G^v$  by employing Algorithm 1. As in  $\Theta^v$ -sub-problem, we also introduce the auxiliary variable  $R'$  and Lagrange multiplier  $Y'$  for solving  $\Theta_C$ . We define  $\Phi(Y', \Theta_C - R') = \frac{\rho}{2} \|\Theta_C - R'\|_F^2 + \langle Y', \Theta_C - R' \rangle$ . For  $\Theta_C$ -sub-problem,  $\Theta_C$ ,  $R'$  and  $Y'$  are updated by:

$$\Theta_C = \left[ 2\eta' H^T H + \rho I \right]^{-1} \cdot \left[ 2\eta' H^T H^v + \rho R' - Y' \right], \tag{8}$$

$$R' = \arg \min_{R'} \frac{\mu'}{\rho} \|R'\|_* + \frac{1}{2} \|R' - (\Theta_C + \frac{Y'}{\rho})\|_F^2, \tag{9}$$

$$Y' = Y' + \rho(\Theta_C - R'). \tag{10}$$

Based on the learned shared subspace representation  $\Theta_C$ , we use the spectral clustering algorithm to obtain the final clustering results.

### 3 EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we conduct experiments to demonstrate the effectiveness of the proposed EKNN and its extension by comparing the clustering performances of them with some representative subspace clustering approaches on different benchmark datasets. We show the details of the used datasets and compared methods in Appendix. The parameter selection regarding the parameters in the proposed EKNN is also presented in this Appendix.

Table 1: Clustering performance (ACC% $\pm$ STD%) on different datasets.

Data sets	COMIC	DMF	MVEC	BMVC	DiMSC	RMSL	SSSL-M	Ours	Ours(extension)
COIL20	74.23 $\pm$ 0.12	72.97 $\pm$ 0.20	70.45 $\pm$ 2.24	73.33 $\pm$ 0.0072.78 $\pm$ 1.4482.19 $\pm$ 1.3984.36 $\pm$ 0.0580.17 $\pm$ 0.18				85.22 $\pm$ 2.50	
ORL	81.29 $\pm$ 1.00	74.45 $\pm$ 2.29	71.06 $\pm$ 2.02	72.75 $\pm$ 0.00	83.84 $\pm$ 1.16	88.10 $\pm$ 1.2791.56 $\pm$ 0.0186.28 $\pm$ 1.50		91.73 $\pm$ 2.75	
BBCSprot	95.32 $\pm$ 0.18	76.84 $\pm$ 0.0096.88 $\pm$ 0.0080.70 $\pm$ 0.0095.10 $\pm$ 2.1797.61 $\pm$ 0.1893.25 $\pm$ 0.2092.00 $\pm$ 0.19						94.21 $\pm$ 0.32	
Reuters	46.27 $\pm$ 0.05	40.02 $\pm$ 2.40	50.10 $\pm$ 0.00	44.10 $\pm$ 0.01	43.70 $\pm$ 1.13	54.04 $\pm$ 0.5659.46 $\pm$ 1.0042.17 $\pm$ 0.35		60.55 $\pm$ 2.10	
Football	82.48 $\pm$ 1.01	78.98 $\pm$ 1.29	77.08 $\pm$ 2.55	77.42 $\pm$ 0.00	75.40 $\pm$ 2.26	91.57 $\pm$ 0.9389.65 $\pm$ 0.1085.21 $\pm$ 0.54		91.20 $\pm$ 1.90	
ANIMAL	58.17 $\pm$ 2.00	45.92 $\pm$ 1.80	57.96 $\pm$ 1.33	50.15 $\pm$ 0.00	32.61 $\pm$ 1.81	66.16 $\pm$ 0.5471.84 $\pm$ 1.2059.21 $\pm$ 0.68		73.41 $\pm$ 1.00	
NoisyMNIST	56.90 $\pm$ 0.05	53.42 $\pm$ 0.31	64.27 $\pm$ 0.50	60.43 $\pm$ 0.25	43.21 $\pm$ 1.20	80.50 $\pm$ 0.3181.47 $\pm$ 0.0573.18 $\pm$ 0.50		82.50 $\pm$ 2.50	

#### 3.1 EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we show the clustering accuracy of EKNN and comparison approaches on different datasets based on four metrics. Note that we list the clustering results of EKNN and its extension (EKNN for multi-view subspace learning) on different datasets. For comparison, we show the clustering performance of SSSL-M without supervisory information. According to Tables I-IV, we can draw some interesting insights as follows:

Table 2: Clustering performance (NMI% $\pm$ STD%) on different datasets.

Data sets	COMIC	DMF	MVEC	BMVC	DiMSC	RMSL	SSSL-M	Ours	Ours(extension)
COIL20	78.47 $\pm$ 1.00	85.80 $\pm$ 0.26	87.76 $\pm$ 0.89	80.07 $\pm$ 0.0084	61 $\pm$ 1.75	94.10 $\pm$ 1.32	87.43 $\pm$ 0.05	82.45 $\pm$ 0.50	89.20 $\pm$ 0.42
ORL	91.20 $\pm$ 0.05	86.27 $\pm$ 1.00	80.23 $\pm$ 1.16	85.20 $\pm$ 0.0094	40.02 $\pm$ 1.35	94.96 $\pm$ 0.47	91.47 $\pm$ 1.00	88.39 $\pm$ 0.25	95.10 $\pm$ 0.50
BBCSprot	90.23 $\pm$ 0.21	53.39 $\pm$ 0.08	90.32 $\pm$ 0.02	70.80 $\pm$ 0.0085	11.11 $\pm$ 0.13	91.73 $\pm$ 0.52	94.51 $\pm$ 0.50	86.50 $\pm$ 0.80	94.70 $\pm$ 1.00
Reuters	36.50 $\pm$ 2.31	22.88 $\pm$ 1.15	30.17 $\pm$ 0.02	25.41 $\pm$ 0.00	23.31 $\pm$ 0.33	37.49 $\pm$ 0.46	41.55 $\pm$ 0.05	31.50 $\pm$ 0.47	42.90 $\pm$ 1.28
Football	89.57 $\pm$ 0.01	83.38 $\pm$ 0.79	83.36 $\pm$ 1.11	80.22 $\pm$ 0.00	82.16 $\pm$ 1.45	92.29 $\pm$ 0.42	89.78 $\pm$ 0.05	80.15 $\pm$ 2.00	90.25 $\pm$ 0.79
ANIMAL	73.02 $\pm$ 0.22	56.64 $\pm$ 1.31	68.72 $\pm$ 0.50	66.76 $\pm$ 0.00	44.62 $\pm$ 0.89	73.19 $\pm$ 0.60	77.49 $\pm$ 1.00	70.70.76 $\pm$ 0.90	79.00 $\pm$ 1.92
NoisyMNIST	79.42 $\pm$ 0.01	62.48 $\pm$ 0.35	72.41 $\pm$ 0.31	71.23 $\pm$ 1.20	57.36 $\pm$ 0.05	84.28 $\pm$ 0.59	85.91 $\pm$ 1.21	72.30 $\pm$ 0.48	86.20 $\pm$ 0.95

Table 3: Clustering performance (F-score% $\pm$ STD%) on different datasets.

Data sets	COMIC	DMF	MVEC	BMVC	DiMSC	RMSL	SSSL-M	Ours	Ours(extension)
COIL20	70.05 $\pm$ 0.05	63.22 $\pm$ 1.03	64.05 $\pm$ 2.12	67.62 $\pm$ 0.0071	99.19 $\pm$ 0.50	81.20 $\pm$ 1.72	75.89 $\pm$ 0.05	76.39 $\pm$ 1.50	77.10 $\pm$ 0.50
ORL	75.98 $\pm$ 2.00	64.61 $\pm$ 2.56	61.94 $\pm$ 2.31	62.95 $\pm$ 0.0080	71.17 $\pm$ 1.38	84.22 $\pm$ 1.43	83.26 $\pm$ 0.59	82.60 $\pm$ 0.39	84.20 $\pm$ 1.40
BBCSprot	85.23 $\pm$ 0.01	62.46 $\pm$ 0.01	62.46 $\pm$ 0.01	62.46 $\pm$ 0.01	90.11 $\pm$ 0.00	91.01 $\pm$ 0.01	95.35 $\pm$ 0.41	91.45 $\pm$ 0.05	90.48 $\pm$ 0.27
Reuters	40.20 $\pm$ 1.01	34.63 $\pm$ 1.37	39.59 $\pm$ 0.02	37.93 $\pm$ 0.00	33.01 $\pm$ 0.39	44.20 $\pm$ 0.45	46.84 $\pm$ 0.01	40.29 $\pm$ 0.39	48.21 $\pm$ 2.34
Football	71.90 $\pm$ 2.00	70.35 $\pm$ 0.87	67.08 $\pm$ 3.70	63.72 $\pm$ 0.00	67.13 $\pm$ 1.19	83.87 $\pm$ 1.57	79.14 $\pm$ 1.45	75.28 $\pm$ 1.14	79.47 $\pm$ 0.62
ANIMAL	54.80 $\pm$ 1.00	32.86 $\pm$ 2.17	49.31 $\pm$ 2.49	41.47 $\pm$ 0.00	20.66 $\pm$ 1.10	57.29 $\pm$ 1.14	58.76 $\pm$ 0.05	42.16 $\pm$ 0.33	60.27 $\pm$ 2.84
NoisyMNIST	53.42 $\pm$ 0.01	33.42 $\pm$ 0.10	50.41 $\pm$ 0.04	47.31 $\pm$ 0.01	32.40 $\pm$ 0.10	70.05 $\pm$ 0.15	70.16 $\pm$ 0.01	52.50 $\pm$ 0.27	71.30 $\pm$ 1.18

- The proposed EKNN and its extension can achieve satisfied clustering performance on multi-view datasets. The clustering results of the proposed EKNN are not as well as its extension (EKNN for multi-view subspace learning). It can be explained by the fact that more complex mappings are able to result in more desired result, i.e., introducing the latent representation and shared subspace representation in the final objective function.
- On different multi-view datasets, the proposed EKNN still obtain relatively better results than some multi-view clustering methods, which can be explained by the effectiveness of dividing the multi-view clustering into three sub-problems based on  $K$ -means or kernel  $K$ -means. Besides, utilizing linear clustering as one part in the three-level optimization problem is able to reduce the overfitting of the final clustering to some degree and indirectly increase the clustering performance.
- The proposed EKNN and its extension consistently achieve satisfied clustering performances on all datasets in terms of different metrics, which demonstrate their robustness to multiple types of data and the effectiveness of dividing the multi-view clustering into three sub-problems based on clustering or learning, i.e., 1) linear clustering or learning on the set of data points in the dataset, 2) nonlinear clustering or learning on the set of linear clusters obtained by 1), and 3) multi-view clustering or learning by incorporating partition matrices from different views obtained by linear and nonlinear clustering or learning.

### 3.2 INFLUENCE FOR NUMBER OF LINER CLUSTERS

In order to evaluate the impact for the number of linear clusters  $p$  on different datasets, we vary it in the range of  $\{20, 40, 60, 80, 100, 120, 140, 160, 180, 200\}$  for the proposed EKNN. According to Fig. 3, we can observe that the clustering results in terms of different metrics increase with the increasing value of  $p$  and values of these four metrics increase slowly when they reach a certain value. However, the computing cost also increases with the increasing value of  $p$  based on the complexity analysis of our algorithm. We then choose  $p = 160$  for the proposed EKNN in the experiment.

### 3.3 RUNNING TIME

We also present the running time of the proposed EKNN and its extension on Table V, which is conducted on different datasets. Some conclusions can be obtained as follows:

- The proposed EKNN and its extension are able to own satisfied time costs on with different number of liner clusters, which validates their efficiencies in the experiment on different datasets. Due to the efficiency, EKNN uses much less time cost compared with RMSL and SSSL-M.

Table 4: Clustering performance (RI% $\pm$ STD%) on different datasets.

Data sets	COMIC	DMF	MVEC	BMVC	DiMSC	RMSL	SSSL-M	Ours	Ours(extension)
COIL20	96.20 $\pm$ 0.21	95.44 $\pm$ 0.19	95.35 $\pm$ 0.49	96.66 $\pm$ 0.0097	14 $\pm$ 0.11	97.90 $\pm$ 0.41	97.18 $\pm$ 0.15	97.28 $\pm$ 0.50	97.92 $\pm$ 1.20
ORL	97.84 $\pm$ 0.01	98.31 $\pm$ 0.14	97.16 $\pm$ 0.65	98.23 $\pm$ 0.0098	97 $\pm$ 0.05	99.15 $\pm$ 0.07	98.42 $\pm$ 0.20	98.30 $\pm$ 0.37	98.70 $\pm$ 0.90
BBCSport	95.14 $\pm$ 0.01	82.45 $\pm$ 0.0097	90.02 $\pm$ 0.01	90.23 $\pm$ 0.0095	72 $\pm$ 0.10	97.81 $\pm$ 0.19	94.56 $\pm$ 1.52	95.10 $\pm$ 0.59	97.90 $\pm$ 1.20
Reuters	68.99 $\pm$ 0.17	58.71 $\pm$ 1.29	74.17 $\pm$ 0.62	69.15 $\pm$ 0.00	67.49 $\pm$ 0.28	71.37 $\pm$ 0.35	73.81 $\pm$ 0.17	69.48 $\pm$ 1.35	74.12 $\pm$ 0.85
Football	97.00 $\pm$ 0.01	96.51 $\pm$ 0.44	96.22 $\pm$ 0.54	96.69 $\pm$ 0.00	96.74 $\pm$ 0.59	98.40 $\pm$ 0.16	97.68 $\pm$ 0.50	96.25 $\pm$ 0.47	97.95 $\pm$ 0.72
ANIMAL	97.25 $\pm$ 0.35	97.09 $\pm$ 0.22	97.12 $\pm$ 0.25	96.98 $\pm$ 0.00	96.30 $\pm$ 0.23	98.12 $\pm$ 0.05	98.87 $\pm$ 1.00	97.10 $\pm$ 0.91	98.95 $\pm$ 1.82
NoisyMNIST	84.00 $\pm$ 0.27	84.05 $\pm$ 0.01	84.20 $\pm$ 0.05	83.76 $\pm$ 0.10	83.52 $\pm$ 1.05	83.16 $\pm$ 0.35	84.20 $\pm$ 0.15	83.45 $\pm$ 1.20	85.10 $\pm$ 0.17

Table 5: Clustering speed (s) of EKNN and its extension on all dataset with different  $p$ .

Method	COIL20	ORL	BBCSport	Reuters	Football	ANIMAL	NoisyMNIST
RMSL	5210.0	3200.0	3670.5	3001.5	4200.0	1640.6	32000.0
SSSL-M	5560.6	3590.0	3700.5	3130.5	4250.0	1720.5	32406.7
Ours[p=120]	435.2	256.3	359.7	292.5	407.1	1432.2	3125.4
Ours(extension)[p=120]	672.4	432.8	589.2	789.0	639.0	2345.1	7639.0
Ours[p=140]	528.9	423.2	638.2	390.5	542.1	1934.2	4210.5
Ours(extension)[p=140]	830.4	672.9	723.8	920.5	798.2	3410.5	8219.8
Ours[p=160]	634.6	590.3	792.8	414.9	720.5	2319.5	4720.7
Ours(extension)[p=160]	926.6	825.9	904.2	690.3	1025.2	4021.6	8935.0
Ours[p=180]	890.4	746.2	880.5	530.2	843.1	2789.0	5290.5
Ours(extension)[p=180]	1204.3	1190.4	1068.0	725.6	1205.6	4529.0	9534.1

- The proposed EKNN needs less time costs than its extension on all datasets. It can be explained by the fact that its extension introduces more regularization terms for the effectiveness. Therefore, we can observe that the proposed EKNN and its extension can well balance the efficiency and effectiveness of multi-view clustering on different datasets.

### 3.4 CONVERGENCE ANALYSIS

We also show the convergence analysis of the proposed EKNN on COIL20, ORL and BBCSport dataset in Fig. 4. It can be observed that EKNN can achieve convergence in some iterations, which further demonstrates the feasibility of EKNN.

## 4 CONCLUSIONS

In this work, we have shown that multi-view clustering can be regarded as a three-level optimization problem. We flexibly divide the multi-view clustering into three sub-problems based on  $K$ -means or kernel  $K$ -means, i.e., the linear clustering on the set of original data points, the nonlinear clustering of each view on the set of linear clusters, and multi-view clustering by incorporating partition matrices from different views obtained by linear and nonlinear clustering. The proposed EKNN unifies three different sub-problems into a framework and we use the iterative algorithm to optimize the formulated problem. EKNN is able to effectively and rapidly cluster data points from different views. We also extend EKNN for multi-view clustering to the case of multi-view subspace learning, which is able to learn a satisfied latent representation shared by different views. Experiments on different datasets also demonstrate the effectiveness and efficiency of EKNN in terms of different metrics.

### ACKNOWLEDGMENTS

This work was supported by Eastern Talent Plan Leading Project under Grant BJKJ2024011 and National Natural Science Foundation of China (62402303).

### REFERENCES

Shotaro Akaho. A kernel method for canonical correlation analysis. *CoRR*, abs/cs/0609071, 2006.

Massih R. Amini, Nicolas Usunier, and Cyril Goutte. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 28–36, 2009.

Galen Andrew, Raman Arora, Jeff A. Bilmes, and Karen Livescu. Deep canonical correlation analysis. In *International Conference on Machine Learning (ICML)*, volume 28, pp. 1247–1255, 2013.

Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Commun. ACM*, 18(9):509–517, 1975.

Francesco Camastra and Alessandro Verri. A novel kernel method for clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(5):801–804, 2005. doi: 10.1109/TPAMI.2005.88.

X. Cao, C. Zhang, H. Fu, Si Liu, and Hua Zhang. Diversity-induced multi-view subspace clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586–594, 2015.

Xiaochun Cao, Changqing Zhang, Huazhu Fu, Si Liu, and Hua Zhang. Diversity-induced multi-view subspace clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586–594, 2015. doi: 10.1109/CVPR.2015.7298657.

Lluís Castrejón, Yusuf Aytar, Carl Vondrick, Hamed Pirsiavash, and Antonio Torralba. Learning aligned cross-modal representations from weakly aligned data. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2940–2949. IEEE Computer Society, 2016. doi: 10.1109/CVPR.2016.321.

Yizong Cheng. Mean shift, mode seeking, and clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(8):790–799, 1995.

Yu-An Chung, Wei-Hung Weng, Schrasing Tong, and James R. Glass. Unsupervised cross-modal alignment of speech and text embedding spaces. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 7365–7375, 2018.

Inderjit S. Dhillon, Yuqiang Guan, and Brian Kulis. Weighted graph cuts without eigenvectors A multilevel approach. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(11):1944–1957, 2007.

Ehsan Elhamifar and René Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11):2765–2781, 2013.

Keinosuke Fukunaga and Larry D. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. Inf. Theory*, 21(1):32–40, 1975.

Israel Dejene Gebru, Xavier Alameda-Pineda, Florence Forbes, and Radu Horaud. EM algorithms for weighted-data clustering with application to audio-visual scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 38(12):2402–2415, 2016.

Derek Greene and Padraig Cunningham. Practical solutions to the problem of diagonal dominance in kernel document clustering. In *2006 International Conference on Machine Learning (ICML)*, pp. 377–384, 2006.

Mithun Das Gupta and Jing Xiao. Non-negative matrix factorization as a feature selection tool for maximum margin classifiers. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2841–2848, 2011. doi: 10.1109/CVPR.2011.5995492.

Yaobin He, Haoyu Tan, Wuman Luo, Huajian Mao, Di Ma, Shengzhong Feng, and Jianping Fan. MR-DBSCAN: an efficient parallel density-based clustering algorithm using mapreduce. In *International Conference on Parallel and Distributed Systems (ICPADS)*, pp. 473–480, 2011.

Harold Hotelling. Relations between two sets of variates. *Biometrika*, 28:321–377, 1935. doi: 10.1093/biomet/28.3-4.321.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 1097–1105, 2012.

Abhishek Kumar and Hal Daumé III. A co-training approach for multi-view spectral clustering. In *International Conference on Machine Learning (ICML)*, pp. 393–400, 2011.

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

C. H. Lampert, H. Nickisch, and S. Harmeling. Attribute-based classification for zero-shot visual object categorization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(3):453–465, 2014.

Haobin Li, Yijie Lin, Peng Hu, Mouxing Yang, and Xi Peng. Community-aware multi-view representation learning with incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

Ruihuang Li, Changqing Zhang, Huazhu Fu, Xi Peng, Joey Tianyi Zhou, and Qinghua Hu. Reciprocal multi-layer subspace learning for multi-view clustering. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8171–8179. IEEE, 2019. doi: 10.1109/ICCV.2019.00826.

Xingfeng Li, Zhenwen Ren, Quansen Sun, and Zhi Xu. Auto-weighted tensor schatten p-norm for robust multi-view graph clustering. *Pattern Recognit.*, 134:109083, 2023a.

Xingfeng Li, Yinghui Sun, Quansen Sun, and Zhenwen Ren. Consensus cluster center guided latent multi-kernel clustering. *IEEE Trans. Circuits Syst. Video Technol.*, 33(6):2864–2876, 2023b.

Zechao Li and Jinhui Tang. Weakly supervised deep matrix factorization for social image understanding. *IEEE Trans. Image Process.*, 26(1):276–288, 2017.

Jiye Liang, Liang Bai, Chuangyin Dang, and Fuyuan Cao. The K -means-type algorithms versus imbalanced data distributions. *IEEE Trans. Fuzzy Syst.*, 20(4):728–745, 2012.

Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. Completer: Incomplete multi-view clustering via contrastive prediction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11169–11178, 2021.

Zhouchen Lin, Risheng Liu, and Zhixun Su. Linearized alternating direction method with adaptive penalty for low-rank representation. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 612–620, 2011.

Chengliang Liu, Zhihao Wu, Jie Wen, Yong Xu, and Chao Huang. Localized sparse incomplete multi-view clustering. *IEEE Transactions on Multimedia*, 25:5539–5551, 2023a.

Chengliang Liu, Jie Wen, Zhihao Wu, Xiaoling Luo, Chao Huang, and Yong Xu. Information recovery-driven deep incomplete multiview clustering network. *IEEE Transactions on Neural Networks and Learning Systems*, 35(11):15442–15452, 2024.

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, 2025.

Jiyuan Liu, Xinwang Liu, Jian Xiong, Qing Liao, Sihang Zhou, Siwei Wang, and Yuexiang Yang. Optimal neighborhood multiple kernel clustering with adaptive local kernels. *IEEE Trans. Knowl. Data Eng.*, 34(6):2872–2885, 2022a.

Jiyuan Liu, Xinwang Liu, Yuexiang Yang, Xifeng Guo, Marius Kloft, and Liangzhong He. Multi-view subspace clustering via co-training robust data representation. *IEEE Trans. Neural Networks Learn. Syst.*, 33(10):5177–5189, 2022b.

Jiyuan Liu, Xinwang Liu, Yuexiang Yang, Qing Liao, and Yuanqing Xia. Contrastive multi-view kernel learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(8):9552–9566, 2023b.

Manhua Liu, Xudong Jiang, and Alex C. Kot. A multi-prototype clustering algorithm. *Pattern Recognit.*, 42(5):689–698, 2009.

Can-Yi Lu, Hai Min, Zhong-Qiu Zhao, Lin Zhu, De-Shuang Huang, and Shuicheng Yan. Robust and efficient subspace segmentation via least squares regression. In *European Conference on Computer Vision*, pp. 347–360, 2012.

Yiding Lu, Haobin Li, Yunfan Li, Yijie Lin, and Xi Peng. A survey on deep clustering: from the prior perspective. *Vicinagearth*, 1(1):4, 2024.

Rashid Mehmood, Rongfang Bie, Hussain Dawood, and Haseeb Ahmad. Fuzzy clustering by fast search and find of density peaks. In *International Conference on Identification, Information, and Knowledge (IIKI)*, pp. 258–261, 2015.

Sameer A. Nene, Shree K. Nayar, and Hiroshi Murase. Columbia object image library (coil-20). In *Technical Report CUCS-006-96*, 1996.

Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In Thomas G. Dietterich, Suzanna Becker, and Zoubin Ghahramani (eds.), *Advances in Neural Information Processing Systems (NIPS)*, pp. 849–856, 2001.

Xi Peng, Zhenyu Huang, Jiancheng Lv, Hongyuan Zhu, and Joey Tianyi Zhou. COMIC: multi-view clustering without parameter selection. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, volume 97, pp. 5092–5101, 2019.

Petra Perner. Data mining - concepts and techniques. *Künstliche Intell.*, 16(1):77, 2002.

Nan Pu, Zhun Zhong, Xinyuan Ji, and Nicu Sebe. Federated generalized category discovery. *CoRR*, abs/2305.14107, 2023.

Yalan Qin and Li Qian. Fast elastic-net multi-view clustering: A geometric interpretation perspective. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 10164–10172, 2024.

Yalan Qin, Hanzhou Wu, and Guorui Feng. Structured subspace learning-induced symmetric non-negative matrix factorization. *Signal Process.*, 186:108115, 2021.

Yalan Qin, Hanzhou Wu, Xinpeng Zhang, and Guorui Feng. Semi-supervised structured subspace learning for multi-view clustering. *IEEE Trans. Image Process.*, 31:1–14, 2022a.

Yalan Qin, Hanzhou Wu, Jian Zhao, and Guorui Feng. Enforced block diagonal subspace clustering with closed form solution. *Pattern Recognit.*, 130:108791, 2022b.

Yalan Qin, Guorui Feng, Yanli Ren, and Xinpeng Zhang. Block-diagonal guided symmetric non-negative matrix factorization. *IEEE Trans. Knowl. Data Eng.*, 35(3):2313–2325, 2023a.

Yalan Qin, Chuan Qin, Xinpeng Zhang, Donglian Qi, and Guorui Feng. Nim-nets: Noise-aware incomplete multi-view learning networks. *IEEE Trans. Image Process.*, 32:175–189, 2023b.

Yalan Qin, Xinpeng Zhang, Liquan Shen, and Guorui Feng. Maximum block energy guided robust subspace clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(2):2652–2659, 2023c.

Yalan Qin, Nan Pu, and Hanzhou Wu. Elastic multi-view subspace clustering with pairwise and high-order correlations. *IEEE Trans. Knowl. Data Eng.*, 36(2):556–568, 2024a.

Yalan Qin, Chuan Qin, Xinpeng Zhang, and Guorui Feng. Dual consensus anchor learning for fast multi-view clustering. *IEEE Trans. Image Process.*, 33:5298–5311, 2024b.

Yalan Qin, Zhenjun Tang, Hanzhou Wu, and Guorui Feng. Flexible tensor learning for multi-view clustering with markov chain. *IEEE Trans. Knowl. Data Eng.*, 36(4):1552–1565, 2024c.

Yalan Qin, Guorui Feng, and Xinpeng Zhang. Fast incomplete multi-view clustering by flexible anchor learning. In *Forty-second International Conference on Machine Learning*, pp. 1–1, 2025a.

Yalan Qin, Guorui Feng, and Xinpeng Zhang. Scalable one-pass incomplete multi-view clustering by aligning anchors. In *AAAI Conference on Artificial Intelligence*, pp. 20042–20050, 2025b.

Yalan Qin, Nan Pu, Guorui Feng, and Nicu Sebe. Robust consensus anchor learning for efficient multi-view subspace clustering. In *Forty-second International Conference on Machine Learning*, pp. 1–1, 2025c.

Yalan Qin, Nan Pu, Nicu Sebe, and Guorui Feng. Latent space learning-based ensemble clustering. *IEEE Trans. Image Process.*, 34:1259–1270, 2025d.

Yalan Qin, Nan Pu, Hanzhou Wu, and Zhaoxin Fan. Flexible multi-view clustering with dynamic views generation. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 1072–1081, 2025e.

Yalan Qin, Nan Pu, Hanzhou Wu, and Nicu Sebe. Discriminative anchor learning for efficient multi-view clustering. *IEEE Trans. Multim.*, 27:1386–1396, 2025f.

Yalan Qin, Nan Pu, Hanzhou Wu, and Nicu Sebe. Margin-aware noise-robust contrastive learning for partially view-aligned problem. *ACM Trans. Knowl. Discov. Data*, 19(1):26:1–26:20, 2025g.

Yalan Qin, Xinpeng Zhang, Shui Yu, and Guorui Feng. A survey on representation learning for multi-view data. *Neural Networks*, 181:106842, 2025h.

Yang Qin, Yuan Sun, Dezhong Peng, Joey Tianyi Zhou, Xi Peng, and Peng Hu. Cross-modal active complementary learning with self-refining correspondence. In *Advances in Neural Information Processing Systems*, 2023d.

J. Queen. Some methods for the classification and analysis of multivariate observations. *Proceedings of the Fifth Berkely Symposium on Mathematical Statistics and Probability*, 1, 01 1966.

Nikhil Rasiwasia, José Costa Pereira, Emanuele Coviello, Gabriel Doyle, Gert R. G. Lanckriet, Roger Levy, and Nuno Vasconcelos. A new approach to cross-modal multimedia retrieval. In *International Conference on Multimedia (ACM MM)*, pp. 251–260. ACM, 2010. doi: 10.1145/1873951.1873987.

F. S. Samaria and A. C. Harter. Parameterisation of a stochastic model for human face identification. In *1994 IEEE Workshop on Applications of Computer Vision*, pp. 138–142, 1994.

Roman Sandler and Michael Lindenbaum. Nonnegative matrix factorization with earth mover’s distance metric for image analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(8):1590–1602, 2011.

Bernhard Schölkopf, Alexander J. Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.*, 10(5):1299–1319, 1998.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv*, September 2014.

Jing Sui, Shile Qi, Theo van Erp, Juan Bustillo, Rongtao Jiang, Dongdong Lin, Jessica Turner, Eswar Damaraju, Andrew Mayer, Yue Cui, Zening Fu, Yuhui Du, Jiayu Chen, Steven Potkin, A. Preda, Daniel Mathalon, Judith Ford, James Voyvodic, Bryon Mueller, and Vince Calhoun. Multimodal neuromarkers in schizophrenia via cognition-guided mri fusion. *Nature Communications*, 9, 2018. doi: 10.1038/s41467-018-05432-w.

Zhiqiang Tao, Hongfu Liu, Sheng Li, Zhengming Ding, and Yun Fu. From ensemble clustering to multi-view clustering. In *2017 International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2843–2849, 2017.

Nate Veldt, David F. Gleich, Anthony Wirth, and James Saunderson. Metric-constrained optimization for graph clustering algorithms. *SIAM J. Math. Data Sci.*, 1(2):333–355, 2019. doi: 10.1137/18M1217152.

P. Viswanath and Rajwala Pinkesh. 1-dbscan : A fast hybrid density based clustering method. In *International Conference on Pattern Recognition (ICPR)*, pp. 912–915, 2006.

Chang-Dong Wang, Jian-Huang Lai, Ching Y. Suen, and Jun-Yong Zhu. Multi-exemplar affinity propagation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(9):2223–2237, 2013.

Weiran Wang, Raman Arora, Karen Livescu, and Jeff A. Bilmes. On deep multi-view representation learning. In *International Conference on Machine Learning (ICML)*, volume 37, pp. 1083–1092, 2015.

Bo Wu and Bogdan M. Wilamowski. A fast density and grid based clustering method for data with arbitrary shapes and noise. *IEEE Trans. Ind. Informatics*, 13(4):1620–1628, 2017.

Chang Xu, Dacheng Tao, and Chao Xu. Multi-view learning with incomplete views. *IEEE Trans. Image Process.*, 24(12):5812–5825, 2015. doi: 10.1109/TIP.2015.2490539.

Mouxing Yang, Yunfan Li, Zhenyu Huang, Zitao Liu, Peng Hu, and Xi Peng. Partially view-aligned representation learning with noise-robust contrastive loss. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1134–1143, 2021.

Zhiyong Yang, Qianqian Xu, Weigang Zhang, Xiaochun Cao, and Qingming Huang. Split multiplicative multi-view subspace clustering. *IEEE Trans. Image Process.*, 28(10):5147–5160, 2019. doi: 10.1109/TIP.2019.2913096.

C. Zhang, H. Fu, S. Liu, G. Liu, and X. Cao. Low-rank tensor constrained multiview subspace clustering. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1582–1590, 2015.

Changqing Zhang, Qinghua Hu, Huazhu Fu, Pengfei Zhu, and Xiaochun Cao. Latent multi-view subspace clustering. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4333–4341, 2017. doi: 10.1109/CVPR.2017.461.

Changqing Zhang, Yajie Cui, Zongbo Han, Joey Tianyi Zhou, Huazhu Fu, and Qinghua Hu. Deep partial multi-view learning. *CoRR*, abs/2011.06170, 2020a.

Changqing Zhang, Huazhu Fu, Jing Wang, Wen Li, Xiaochun Cao, and Qinghua Hu. Tensorized multi-view subspace representation learning. *Int. J. Comput. Vis.*, 128(8):2344–2361, 2020b.

Yanfeng Zhang, Shimin Chen, and Ge Yu. Efficient distributed density peaks for clustering large data sets in mapreduce. *IEEE Trans. Knowl. Data Eng.*, 28(12):3218–3230, 2016.

Yuanyang Zhang, Yijie Lin, Weiqing Yan, Li Yao, Xinhang Wan, Guangyuan Li, Chao Zhang, Guanzhou Ke, and Jie Xu. Incomplete multi-view clustering via diffusion contrastive generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 22650–22658, 2025.

Handong Zhao, Zhengming Ding, and Yun Fu. Multi-view clustering via deep matrix factorization. In *AAAI Conference on Artificial Intelligence (AAAI)*, pp. 2921–2927, 2017.

Joey Tianyi Zhou, Ivor W. Tsang, Sinno Jialin Pan, and Mingkui Tan. Multi-class heterogeneous domain adaptation. *J. Mach. Learn. Res.*, 20:57:1–57:31, 2019.

## A APPENDIX

### A.1 RELATED WORK OF $K$ -MEANS ALGORITHM AND ANALYSIS

As one of the most efficient algorithms for clustering,  $K$ -means algorithm uses an initial set of cluster centers as the start and decreases the sum of squared errors by iteratively refining this set, which has gained considerable attentions in the related literature. The  $K$ -means algorithm aims to minimize the objective function  $P$  as:

$$P(S, V) = \sum_{l=1}^k \sum_{x_i \in S_l} d(x_i, v_l), \quad (11)$$

where  $S = \{S_1, \dots, S_k\}$  denotes a partition of the dataset  $X$ ,  $k$  is the number of real clusters in  $X$ ,  $S_l \subset X$ ,  $\cup_{l=1}^k S_l = X$ ,  $S_l \cap S_j = \emptyset$  for  $1 \leq l \neq j \leq k$ ,  $V = \{v_l\}_{l=1}^k$  and  $v_l$  indicates the  $l$ -th center of  $S_l$ . According to Eq. (11), we can observe that  $K$ -means algorithm is related to two updated equations based on  $V$  and  $S$ . The alternative optimization method is used to seek for the optimal minimized value of  $P$  and the updated equations are shown in the following. Given  $V$ , we update  $S$  by

$$x_i \in S_l, \quad \text{if } l^* = \arg \min_l d(x_i, v_l), \quad (12)$$

where  $1 \leq i \leq n$ ,  $n$  is the number of data points and  $1 \leq l \leq k$ . Given  $S$ , we update  $V$  by

$$v_l = \frac{\sum_{x_i \in S_l} x_i}{|S_l|}, \quad (13)$$

where  $1 \leq l \leq k$ . The time complexity is  $O(nkt)$  for  $K$ -means algorithm with  $t$  being the number of total iterations.

Based on Eq. (11), we can find that the cost of  $K$ -means algorithm is much lower than that of calculating the distances between all data points since it just needs to calculate the distances between centers and data points. We also observe that the clustering performance of  $K$ -means is influenced by some factors, i.e., the initial cluster centers. Different initial cluster centers usually result in different results. However,  $K$ -means algorithm often fails to achieve satisfied clustering performance when clusters are nonlinearly separable.

### A.2 MULTI-VIEW REPRESENTATION LEARNING

Multi-view representation learning aims to effectively explore the complete and consistent information from different views. Various multi-view learning methods Kumar et al. (2011); Kumar & III (2011); Xu et al. (2015); Zhang et al. (2017); Yang et al. (2019) have been proposed by finding the consistent hypotheses for clustering across different views. For example, canonical correlation analysis (CCA) Hotelling (1935), kernelized CCA Akaho (2006), deep neural networks based CCA Andrew et al. (2013), and deep canonically correlated autoencoder (DCCAE) Wang et al. (2015) are representative methods in multi-view representation learning, which have achieved great success. To be specific, CCA minimizes the correlation between two different views to find a shared representation. CCA is formulated as:

$$\begin{aligned} (P^1, P^2) &= \arg \min_{P^1, P^2} \text{tr}((P^1)^T X^1 (X^2)^T P^2), \\ \text{s.t. } (P^v)^T X^v (X^v)^T P^v &= I, \quad v = 1, 2, \end{aligned} \quad (14)$$

where  $X^v = [x_1^v, x_2^v, \dots, x_n^v] \in R^{m \times n}$  denotes the representation of the  $v$ -th view.  $n$  and  $m$  represent the size of dataset and the dimension of the  $v$ -th view, respectively.  $k$  indicates the dimension for the shared representation.  $P^v$  and  $I$  denote the mapping matrix and the identity matrix of the  $v$ -th view, respectively. As one method based on CCA, Zhang et al. Cao et al. (2015) presented to exploit the correlations among multiple views as:

$$\begin{aligned} (P^1, \dots, P^V) &= \arg \min_{P^1, \dots, P^V} \sum_{v=1}^V \text{tr}(P^v X^v L^v (X^v)^T (P^v)^T) \\ &+ \lambda \sum_{\neq u} \text{HSIC}(P^v X^v, P^u X^u), \\ \text{s.t. } (P^v)^T X^v (X^v)^T P^v &= I, \quad v = 1, 2, \dots, V, \end{aligned} \quad (15)$$

where  $V$  denotes the number of total views and  $L^v$  is the graph Laplacian matrix for the  $v$ -th view, respectively. Li *et al.* Li et al. (2019) studied complex relations among multiple views and learned the underlying latent representation. It is formulated by:

$$\begin{aligned} & \min_{\Theta, H} \alpha L_E(\{\Theta_S^v\}_{v=1}^V, H; \{\Theta_E^v\}_{v=1}^V) + \gamma R(\{\Theta_E^v\}_{v=1}^V) \\ & + L_S(\{X^v\}_{v=1}^V, H; \{\Theta_S^v\}_{v=1}^V, \Theta_C) + \beta R(\{\Theta_S^v\}_{v=1}^V, \Theta_C), \end{aligned} \quad (16)$$

where  $L_E(\cdot, \cdot)$  is the loss of reconstruction to update  $H$ ,  $R(\cdot)$  is adopted for representing the term for regularization,  $L_S(\cdot, \cdot)$  denotes the loss based on self-representation,  $\Theta$  including  $\Theta_{E^v}$ ,  $\Theta_{S^v}$  and  $\Theta_C$  are parameters of the model.

Different from multi-view representation learning, cross-view representation learning targets at searching for mappings between two different views. Numerous methods based on cross-view representation learning have been proposed Rasiwasia et al. (2010); Chung et al. (2018); Castrejón et al. (2016); Zhou et al. (2019). The embedding spaces of two different views are simultaneously learned and aligned by adversarial training Chung et al. (2018). To solve partially view-aligned problem, the literature in Yang et al. (2021) simultaneously learns the representation and aligns the data by a contrastive loss which is robust to noise. The major difference between the above methods and our work is that they do not take the efficiency for multi-view clustering into consideration while our method is able to achieve final clustering results with desired efficiency.

### A.3 SOLUTION FOR THE FORMULATED PROBLEM

To make the objective function of EKNN in Eq. (6) separable, we introduce the auxiliary variable  $R^v$  to replace  $\Theta^v$  and obtain the equivalent objective function as follows:

$$\begin{aligned} & \min_H F = \sum_{v=1}^V \alpha \|\Theta^v - W^v V^v\|_F^2 + \beta \text{Tr}(K^v) \\ & - \beta \text{Tr}((\hat{U}^v)^T K^v \hat{U}^v) + \gamma \|W^v \hat{U}^v - H G^v\|_F^2 \\ & + \eta \|X^v - X^v \Theta^v\|_F^2 + \mu \|R^v\|_*, \\ & \text{s.t. } w_{ij}^v \in \{0, 1\}, \sum_{j=1}^p w_{ij}^v = 1, 1 < \sum_{i=1}^n w_{ij}^v < n, \\ & u_{ij}^v \in \{0, 1\}, \sum_{j=1}^k u_{ij}^v = 1, 1 < \sum_{i=1}^p u_{ij}^v < p, \\ & h_{ij}^v \in \{0, 1\}, \sum_{j=1}^k h_{ij}^v = 1, 1 < \sum_{i=1}^n h_{ij}^v < n, \\ & \Theta^v = R^v. \end{aligned} \quad (17)$$

We then divide the problem in Eq. (17) into different sub-problems to minimize:

- $\min_{W^v, V^v} F$  with fixed  $U^v, R^v, \Theta^v, H$  and  $G^v$ .
- $\min_{U^v} F$  with fixed  $W^v, R^v, V^v, H, \Theta^v$  and  $G^v$ .
- $\min_{H, G^v} F$  with fixed  $W^v, R^v, \Theta^v, V^v$  and  $U^v$ .
- $\min_{\Theta^v} F$  with fixed  $U^v, R^v, W^v, V^v, H$  and  $G^v$ .
- $\min_{R^v} F$  with fixed  $U^v, \Theta^v, W^v, V^v, H$  and  $G^v$ .

These optimization sub-problems are solved in the following.

#### (1) $W^v, V^v$ -sub-problem:

Given  $U^v, R^v, \Theta^v, H$  and  $G^v$ ,  $\min_{W^v, V^v} F$  is formulated as:

$$\min_{W^v, V^v} \alpha \|\Theta^v - W^v V^v\|_F^2 + \gamma \|W^v \hat{U}^v - H G^v\|_F^2. \quad (18)$$

We then simplify Eq. (18) as:

$$\min_{W^v, \tilde{V}^v} \|\tilde{\Theta}^v - W^v \tilde{V}^v\|_F^2, \quad (19)$$

where  $\tilde{\Theta}^v = [\alpha^{\frac{1}{2}} \Theta^v, \gamma^{\frac{1}{2}} HG^v]$  is obtained by concatenating  $\Theta^v$  and  $HG^v$ , and  $\tilde{V}^v = [\alpha^{\frac{1}{2}} V^v, \gamma^{\frac{1}{2}} \hat{U}^v]$ . Therefore, we can regard  $\tilde{\Theta}^v$  and  $\tilde{V}^v$  as the new representation of  $\Theta^v$  and the corresponding cluster center matrix, respectively. Then the optimization problem in Eq. (18) is transformed into a  $K$ -means clustering problem regarding  $\tilde{\Theta}^v$  and we solve it by the update rules of  $W^v$  and  $\tilde{V}^v$ . The details are shown in the following.

Given  $\tilde{V}^v$ , we update  $W^v$  by

$$w_{ij}^v = \begin{cases} 1, & \text{for } j = \arg \min \|\tilde{\theta}_i^v - \tilde{v}_l^v\|^2, \\ 0, & \text{otherwise,} \end{cases} \quad (20)$$

where  $l \in [1, p]$ ,  $\tilde{v}_l^v$  is the  $l$ -th row of  $\tilde{V}^v$  and  $\tilde{\theta}_i^v$  denotes the  $i$ -th row of  $\tilde{\Theta}^v$ . Given  $W^v$ , we update  $\tilde{V}^v$  as

$$\tilde{v}_j^v = \frac{\sum_{i=1}^n w_{ij}^v \tilde{\theta}_i^v}{\sum_{i=1}^n w_{ij}^v}. \quad (22)$$

Then the update rules of  $W^v$  and  $\tilde{V}^v$  can be obtained in this manner.

**(2)  $U^v$ -sub-problem:**

Given  $\Theta^v, R^v, W^v, V^v, H$  and  $G^v$ ,  $\min_{U^v} F$  is formulated by

$$\min_{U^v} \beta(Tr(K^v) - Tr(\hat{U}^v K^v \hat{U}^v)) + \gamma \|W^v \hat{U}^v - HG^v\|_F^2. \quad (23)$$

We further rewrite Eq. (23) by

$$\min_{U^v} \beta Tr(K^v) - Tr((\hat{U}^v)^T L^v \hat{U}^v), \quad (24)$$

where  $L^v = \beta K^v + \gamma (W^v)^T (I_n - \hat{H}^v (\hat{H}^v)^T) W^v$  and  $I_n \in R^{n \times n}$  is the identity matrix. It has been proved that the equivalence between kernel  $K$ -means and spectral clustering algorithm Dhillon et al. (2007). Thus, Eq. (24) is equivalent to the problem as follows:

$$\max_{\hat{U}^v} Tr((\hat{U}^v)^T L^v \hat{U}^v), \quad \text{s.t.} \quad (\hat{U}^v)^T \hat{U}^v = I_k, \quad (25)$$

Therefore, the spectral clustering method can be used to solve this problem.

**(3)  $H, G^v$ -sub-problem:**

Given  $\Theta^v, R^v, W^v, V^v$  and  $U^v$ ,  $\min_{H, G^v} F$  is formulated as:

$$\min_{H, G^v} \sum_{v=1}^V \|W^v \hat{U}^v - HG^v\|_F^2. \quad (26)$$

Likewise, we can solve this problem as the paradigm of  $K$ -means. Given  $\Theta^v, W^v, R^v, V^v, H$  and  $U^v$ , we update  $G^v$  as

$$g_i^v = \frac{\sum_{i=1}^n h_{il} q_i^v}{\sum_{i=1}^n h_{il}}, \quad (27)$$

where  $q_i^v$  denotes the  $i$ -th row of  $W^v \hat{U}^v$ . Given  $\Theta^v, R^v, W^v, V^v, G^v$  and  $U^v$ , we update  $H$  as

$$h_{il} = \begin{cases} 1, & \text{for } l = \arg \min \sum_{v=1}^V \|q_i^v - g_e^v\|^2 \\ 0, & \text{otherwise,} \end{cases} \quad (28)$$

$$(29)$$

where  $e \in [1, k]$ ,  $l \in [1, k]$  and  $i \in [1, n]$ .

**(4)  $\Theta^v$ -sub-problem:**

Given  $U^v, W^v, R^v, V^v, H$  and  $U^v, \min_{H, G^v} F$  is formulated by:

$$\begin{aligned} \min_{\Theta^v} F &= \sum_{v=1}^V \alpha \|\Theta^v - W^v V^v\|_F^2 + \eta \|X^v - X^v \Theta^v\|_F^2, \\ \text{s.t. } \Theta^v &= R^v. \end{aligned} \quad (30)$$

We then adopt the Augmented Lagrange Multiplier (ALM) Lin et al. (2011) to tackle this problem as follows:

$$\begin{aligned} \min_{\Theta^v} F &= \sum_{v=1}^V \alpha \|\Theta^v - W^v V^v\|_F^2 + \eta \|X^v - X^v \Theta^v\|_F^2 \\ &\quad + \Phi(Y^v, \Theta^v - R^v), \end{aligned} \quad (31)$$

where  $\Phi(Y^v, \Theta^v - R^v) = \frac{\rho}{2} \|\Theta^v - R^v\|_F^2 + \langle Y^v, \Theta^v - R^v \rangle$ ,  $\langle \cdot, \cdot \rangle$  denotes the Frobenius inner product,  $\rho > 0$  and  $Y^v$  is the Lagrange multiplier. We take the derivative with respect to  $\Theta^v$  and then set it to be zero. Then the closed-form solution can be obtained as:

$$\begin{aligned} \Theta^v &= [2\eta(X^v)^T X^v + (2\alpha + \rho)I]^{-1} \\ &\quad \cdot [2\eta(X^v)^T X^v + \rho R^v - Y^v + 2\alpha W^v V^v], \end{aligned} \quad (32)$$

where  $I$  denotes the identity matrix.

#### (5) $R^v$ -sub-problem:

Given  $\Theta^v, W^v, H, V^v, H$  and  $U^v, \min_{R^v} F$  is formulated by:

$$\min_{\Theta^v} F = \sum_{v=1}^V \mu \|R^v\|_*, \quad \text{s.t. } \Theta^v = R^v. \quad (33)$$

Then we can obtain

$$R^v = \arg \min_{R^v} \frac{\mu}{\rho} \|R^v\|_* + \frac{1}{2} \|R^v - (\Theta^v + \frac{Y^v}{\rho})\|_F^2. \quad (34)$$

The Lagrange multiplier  $Y^v$  is updated by:

$$Y^v = Y^v + \rho(\Theta^v - R^v). \quad (35)$$

Since subproblems for  $\Theta^v, R^v$  and  $Y^v$  are adopted to obtain effective subspace representation of  $X^v$ , we can regard them as the same subproblem for learning  $W^v$  and  $V^v$  based on linear clustering. Then the objective function  $F$  of EKNN can be approximately minimized by iteratively solving these sub-problems. The whole algorithm is presented in **Algorithm 1**.

---

#### Algorithm 1: Algorithm of the proposed method

---

**Input:** Multi-view dataset  $X$ , number of real clusters  $k$  and max iteration  $s$

**Output:** Clustering result  $H$

**Initialize:** Randomly initialize  $\Theta, W, U, V, R, Y, H, \rho = 10^{-6}$  and  $G, i = 0$

**repeat**

Update  $W, V$  with the fixed  $\Theta, R, U, H$  and  $G$ ;  
 Update  $U$  with the fixed  $\Theta, R, W, V, H$  and  $G$ ;  
 Update  $H, G$  with the fixed  $\Theta, W, R, V$  and  $U$ ;  
 Update  $\Theta$  with the fixed  $U, W, R, V, H$  and  $G$ ;  
 Update  $R, Y$  with the fixed  $U, W, V, \Theta, H$  and  $G$ ;

$i = i + 1$ ;

**until**  $i > s$ ;

---

It can be observed that the efficiency, effectiveness, completeness and consistency of EKNN are determined by  $\alpha, \beta$  and  $\gamma$ . The following two cases can be obtained according to the choices of these three parameters. (1) We can regard the clustering result as linear multi-view clustering based on  $K$ -means if  $V > 1, \alpha \neq 0, \beta = 0$  and  $\gamma \neq 0$ . (2) The clustering result can be seen as multi-view clustering based on kernel  $K$ -means if  $V > 1, \beta \neq 0$  and  $\gamma \neq 0$ .

#### A.4 COMPLEXITY ANALYSIS

The time complexity of EKNN consists of different parts according to the objective function in Eq. (6). To be specific, the complexity of solving  $W^v$ ,  $V^v$ -sub-problem is  $O(npsV(m+k))$ , where  $s$  is the number of total iterations for this clustering. It needs  $O((k+m)p^2V)$  to find the solution of  $U^v$ -sub-problem. The complexity to solve  $H$ ,  $G^v$ -sub-problem is  $O(nk^2sV)$ . Therefore, the total time complexity of solving these three subproblems is  $O(npsV(m+k) + (k+m)p^2V + nk^2sV)$ . The complexities to update  $\Theta^v$ ,  $Y^v$  and  $R^v$  are  $O(n^3)$ . In general, the total computational complexity including employing subspace representation  $\Theta^v$  is  $O(n^3)$ .

#### A.5 DATASETS

We adopt seven different datasets in the experiment for comprehensive evaluation, i.e., ORL face images Samaria & Harter (1994), COIL20 object images Nene et al. (1996), Reuters multilingual dataset Amini et al. (2009), Football Li et al. (2019), BBCSport documents Greene & Cunningham (2006), ANIMAL Lampert et al. (2014) and NoisyMNIST Lin et al. (2021).

- **ORL** consists of 40 diverse subjects and each subject has ten face images. It has total three views, which are obtained by different types of features.
- **COIL20** contains over 20 objects and it has 1440 images obtained by the camera with multiple angles. This dataset is characterized by three views.
- **Reuters** has six classes and it is a textual dataset with 1000 different newswire, which are expressed in five languages, i.e., French, Spanish, German, Italian and English.
- **Football** has 20 clubs from English Premier League in BBC website and there are total 248 football players in the league. This dataset has total nine different views.
- **BBCSport** is taken from sports articles and it has 544 documents, which consists of two different views in five topical areas.
- **ANIMAL** consists of 50 different animal classes and it has total 30475 images. We use VGG19 Simonyan & Zisserman (2014) and DECAF Krizhevsky et al. (2012) to extract two types of deep features, which are used as two different views.
- **NoisyMNIST** has two views with selected MNIST images containing Gaussian noise in one view and 70k images from MNIST dataset in the second view. We use 10k validation images and 10k testing image in NoisyMNIST for the experiment as in Lin et al. (2021).

#### A.6 COMPARED METHODS

We conduct experiments to demonstrate the effectiveness of EKNN for multi-view clustering by comparing ours with some existing representative methods in the following.

- **DiMSC** Cao et al. (2015) incorporates complementary information from multiple views by making use of the diversity among different subspace representations.
- **LT-MSC** Zhang et al. (2015) employs the low-rank tensor for exploring high-order relationships among different views.
- **DMF** Zhao et al. (2017) enforces the nonnegative representation of each view in the last layer to be consistent by maximizing the mutual information from different views.
- **MVEC** Tao et al. (2017) produces the basic clustering result for each view and then learns a shared one based on the obtained basic clustering results.
- **RSML** Li et al. (2019) comprehensively describes the data and flexibly encodes complementary information from multiple views. It utilizes backward encoding networks to encode specific subspace representations from different views into the latent representation.
- **COMIC** Peng et al. (2019) maps the data into a space, where cluster assignment consistency and geometric consistency are simultaneously considered.
- **SSSL-M** Qin et al. (2022a) builds an indicator matrix which is anti-block-diagonal for pursuing the block-diagonal structure of the shared affinity matrix with small amount of supervisory information.

We adopt four commonly used metrics for comprehensive investigations, i.e., Accuracy (ACC), F-score, Normalized Mutual Information (NMI), and Rand Index (RI), which are adopted to reflect the clustering performance of the proposed method. The higher values of these metrics indicate more satisfied clustering performance. We conduct experiments on AMD Ryzen5 2600 with 16G RAM. To reduce the randomness, we run each experiment for 30 times and list the mean as well as the standard deviation (STD) as the final result. We adopt Gaussian kernel in the experiment for comparison and analysis.

### A.7 PARAMETER SELECTION

The proposed EKNN consists of some parameters, i.e., balance weights  $\alpha, \beta, \gamma, \eta$  and  $\mu$ . The grid search strategy is employed to find the optimal parameters in this work. To study the influences of balance weights  $\alpha, \beta, \gamma, \eta$  and  $\mu$ , we change their values in  $\{0.2, 0.4, 0.6, 0.8, 1\}$ . Based on Figs. 5-9, it is observed that our method is relatively robust to the choices of  $\alpha, \beta, \gamma, \eta$  and  $\mu$  when their values are in  $\{0.6, 0.8, 1\}$ . Moreover, we can remarkably improve the final clustering performance of the proposed method with good choices of  $\alpha, \beta, \gamma, \eta$  and  $\mu$ . The optimal clustering performance can be obtained when  $\alpha = 0.8, \beta = 0.8, \gamma = 0.8, \eta = 0.8$  and  $\mu = 0.6$ . Likewise, the values of parameters in the extension of EKNN can also be obtained in the same manner and we omit here for simplicity. For parameters in the extension version, we set  $\alpha = 0.8, \beta = 0.8, \gamma = 0.8, \eta = 0.8, \eta' = 0.8, \mu = 0.6$  and  $\mu = 0.6$ .

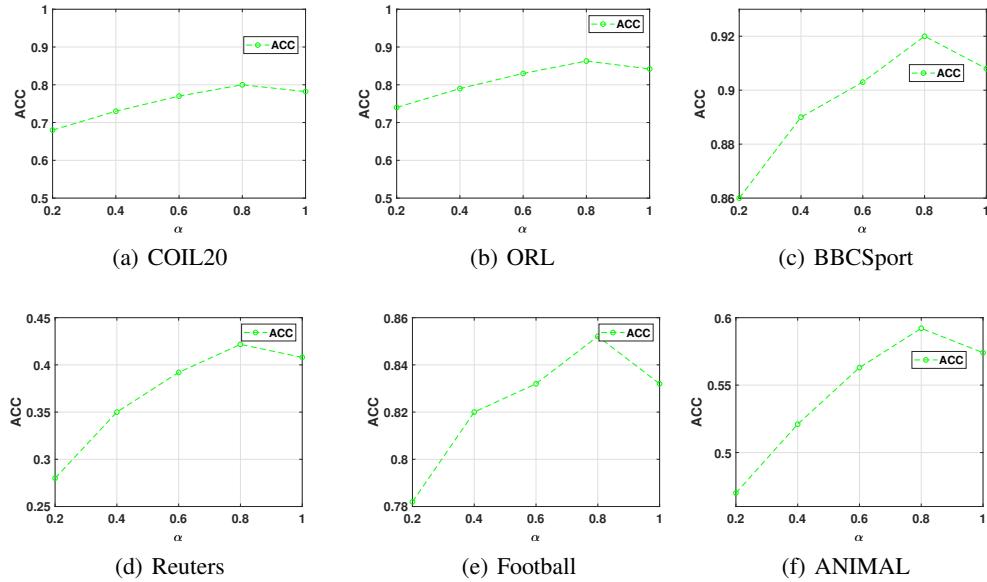
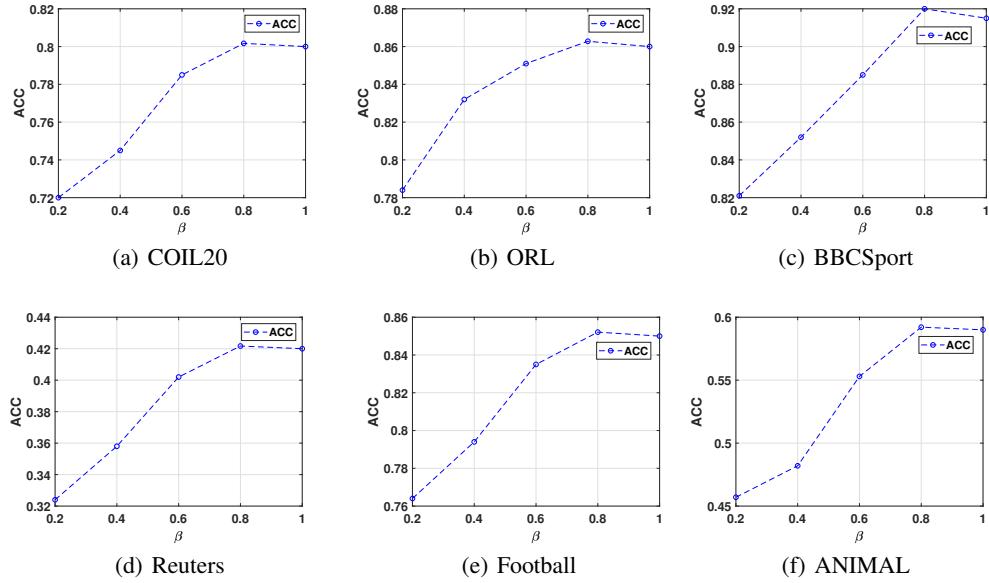
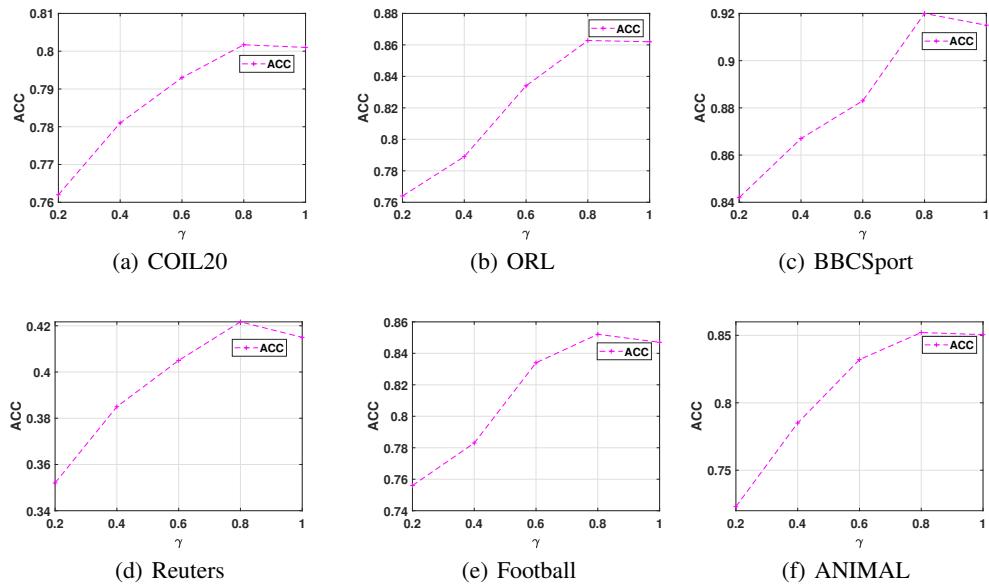
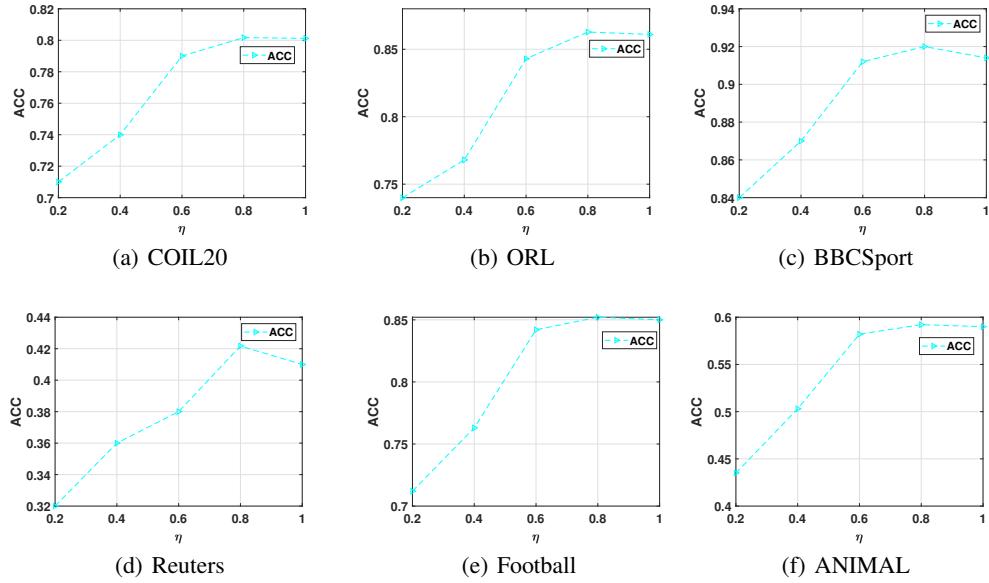
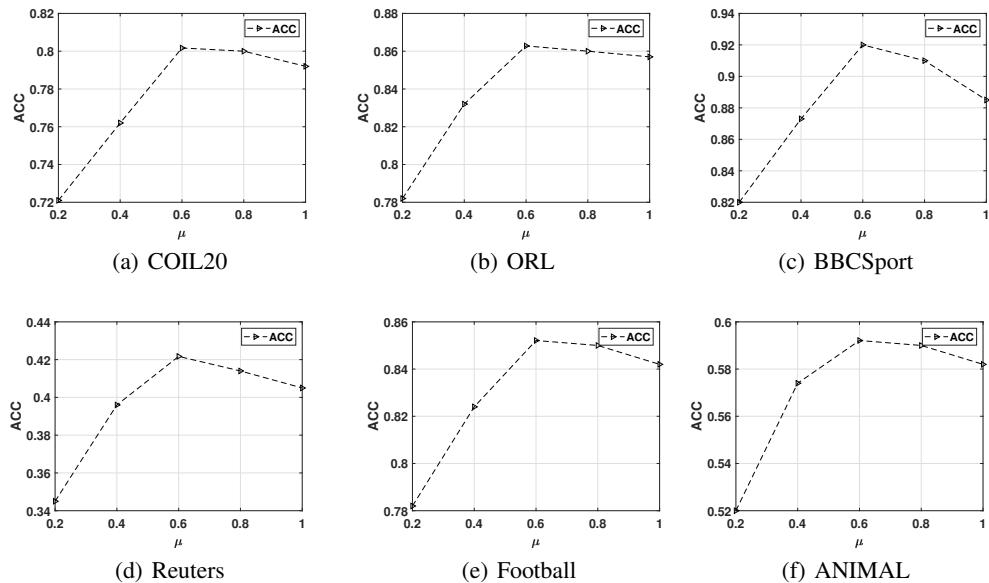


Figure 5: Clustering performance of EKNN with different  $\alpha$  on different datasets.

Figure 6: Clustering performance of EKNN with different  $\beta$  on different datasets.Figure 7: Clustering performance of EKNN with different  $\gamma$  on different datasets.

Figure 8: Clustering performance of EKNN with different  $\eta$  on different datasets..Figure 9: Clustering performance of EKNN with different  $\mu$  on different datasets.