Cross-view Topology Based Consistent and Complementary Information for Deep Multi-view Clustering

Zhibin Dong[1], Siwei Wang[2], Jiaqi Jin[1], Xinwang Liu[1,*], En Zhu[1,*]
[1]School of Computer, National University of Defense Technology, Changsha, China
[2] Intelligent Game and Decision Lab, Beijing, China
{dzb20, wangsiwei13, jinjiaqi}@nudt.edu.cn

Abstract

*Multi-view clustering aims to extract valuable information from different sources or perspectives. Over the years, the deep neural network has demonstrated its superior representation learning capability in multi-view clustering and achieved impressive performance. However, most existing deep clustering approaches are dedicated to merging and exploring the consistent latent representation across multiple views while overlooking the abundant complementary information in each view. Furthermore, finding correlations between multiple views in an unsupervised setting is a significant challenge. To tackle these issues, we present a novel **C**ross-view **T**opology based **C**onsistent and **C**omplementary information extraction framework, termed CTCC. In detail, deep embedding can be obtained from the bipartite graph learning module for each view individually. CTCC then constructs the cross-view topological graph based on the OT distance between the bipartite graph of each view. Utilizing the above graph, we maximize the mutual information across views to learn consistent information and enhance the complementarity of each view by selectively isolating distributions from each other. Extensive experiments on five challenging datasets verify that CTCC outperforms existing methods significantly.*

1. Introduction

With the proliferation and diversity of unlabeled data, multi-view clustering[26, 21, 13, 44, 7] has become an increasingly popular unsupervised paradigm. Its goal is to group data with similar features together by leveraging information from multiple views. Conventional multi-view clustering[41, 19, 42, 31, 55] methods commonly rely on shared information after multi-view fusion for clustering. Due to the limited ability of shallow methods to extract high-level information from data, their clustering performance is highly contingent upon the quality of the raw data. With the rapid development of deep learning[6, 25, 22], deep multi-view clustering(DMVC) approaches[49, 8, 15, 29, 10, 12] employ the powerful learning ability of neural networks to learn a high-level common representation from multiple views that is beneficial for clustering, thus overcoming the drawback of traditional methods. As a result, DMVC has made remarkable progress and attracted widespread attention in real-world applications.

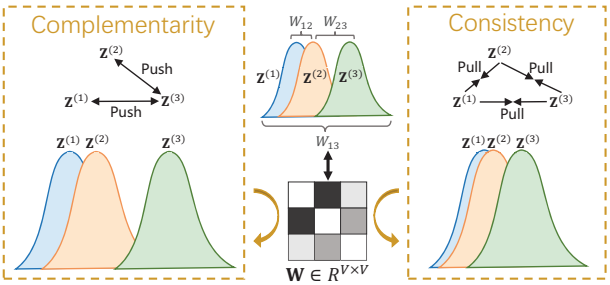


Figure 1: $\mathbf{W}$ represents the topology graph between views. We selectively combine consistent and complementary information between views using a cross-view topology graph $\mathbf{W}$. Specifically, we utilize the mutual information maximization module to obtain consistent information between different views. To selectively further the views incorporating view-specific information, we use the view topology graph to identify views with rich view-specific information. We pull them apart from other views to allow $\mathbf{Z}$ to obtain more complementary information.

Existing DMVC methods can be categorized into three types: graph-based methods[4, 45, 52, 20], subspace-based methods[37, 39, 53], and reconstruction-based methods[26, 51, 3, 40, 46]. These methods often utilize autoencoders or convolutional neural networks to learn the structural information by exploring the common representation or structure in the latent space[49, 8, 15]. Their fundamental con-

*Corresponding author

cepts revolve around fusing different views to uncover the common representation and achieve improved clustering effects. However, in real-world scenarios, multi-view data is collected in diverse ways, and each view contains a substantial amount of view-specific information. Solely focusing on the consistent information between multiple views can lead to significant information loss. Effectively utilizing the view-specific information in each view poses a pressing challenge. [50] has achieved remarkable results by decoupling the view-common information from the view-specific information. However, it overlooks the affinity between views, treating each view equally. In real multi-view data, different views contribute differently to the clustering task. Not all information within a view is equally essential. The correlation between pairwise views plays a crucial role in learning clustering-friendly representations, especially under unsupervised conditions. Thus, striking a balance between consistent and complementary information from different views is a challenging problem.

To tackle these challenges, we propose a novel multi-view deep clustering framework that leverages consistency and complementarity information, as well as cross-view topology. We establish a topological graph between views using bipartite graphs and balance the consistent and complementary information based on this graph. As illustrated in Fig. 1, we introduce the cross-view topology graph $\mathbf{W}$ into the framework to selectively learn consistent and complementary information. Specifically, to learn consistent representations across views, we maximize the mutual information between views and also between views and consistent representations. Since different views contribute differently to consistent representations, we generate weights based on cross-view topological graphs to constrain the mutual information between different views. Furthermore, to utilize view-specific information, we split the views into two sets via the topological graph and use the OT distance between views to pull the two sets farther in the latent space, thereby retaining more view-specific information. In general, we integrate consistent and complementary information in the same framework through the topological graph. We define the relationship between views according to the topological graph to obtain better clustering performance.

The contributions and novelties are summarized as follows:

- In the paradigm of unsupervised learning, we propose a multi-view deep clustering framework based on bipartite graphs. We employ OT distance to define the topological graph between views and selectively integrate the information from multiple views into a deep neural network framework.
- By introducing the topological graph between views, we can define semantic-level relationships between views and balance the consistent and complementary information from multiple views, thereby improving the clustering performance.
- Sufficient experiments demonstrate the effectiveness of selectively unifying consistent and complementary information of multi-view data into a deep clustering framework.

2. Related Work

In this section, We review and rethink the role of consistent and complementary information in multi-view data for clustering and the limitation of deep multi-view clustering in this regard. In addition, we also briefly introduce the latest research progress of mutual information in exploring the commonality and differences among views, which are closely related to our work.

2.1. Rethinking Consistency and Complementarity in MVC

Different from tasks such as cross-view retrieval and cross-view transformation[16], multi-view clustering(MVC)[19, 30, 18, 21] is a task that coordinates information from multiple views. MVC aims to coordinate and fuse two or more view information[23, 38] to achieve the purpose of information complementation and ultimately improve the clustering accuracy and generalization ability of the model. The information in multi-view data can be divided into consistent properties among views and view-specific properties[24, 48, 50, 37], which imply the commonality among views and individuality within each view, respectively.

Most traditional multi-view clustering methods analyze the differences between views while fusing consistent information to fully use all effective information in multiple views and make the clustering results more accurate. For example, CSMSC[24] explicitly splits the information of all views into low-rank common representations, intra-view specific representations, and noise. CDMGC[9] unifies measuring graph diversity and learning consistent cluster label assignments into a framework. Current deep multi-view clustering methods use neural networks to directly learn consistent representations for all views in the latent space[15, 35, 47], while ignoring the important role of view-specific discriminative information for clustering. For example, EAMC[57] uses adversarial learning to align the latent distribution between views to learn a consistent representation. DCP[17] achieves consistent learning by maximizing the mutual information of different views and does not mention the important function of complementary information of views.

To the best of our knowledge, there is no deep method that explicitly unifies the view graph to balance consis-

tent information and complementary information into one framework so that the two can jointly promote the improvement of clustering performance.

2.2. Mutual Information in Multi-view Clustering

Over the years, information theory has been extensively applied in the field of multi-view representation learning, yielding remarkable results[36, 28, 5, 43]. Among these, the information bottleneck[33] is an approach grounded in information theory, offering a reliable theoretical explanation for related work. Ideally, the information bottleneck maximizes task information $I(\mathbf{Z}, \mathbf{Y})$ while minimizing raw feature information $I(\mathbf{Z}; \mathbf{X}^{(1)}, \mathbf{X}^{(2)})$ to obtain a high-quality representation. However, information bottleneck theory necessitates label information [1, 34], and even pseudo-labels constructed in unsupervised scenarios are not reliable, hindering the accurate removal of redundant information and the acquisition of robust representations. Consequently, in our work, rather than employing the information bottleneck to explore the nonlinear relationship between multiple views, we directly investigate the consistent information represented across views by maximizing the mutual information $I(\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)})$.

3. Method

In this section, we initially present the framework diagram of the proposed model, as illustrated in Fig. 2. Subsequently, we delve into the learning process of each module in detail.

3.1. Overview

In the realm of multi-view clustering, the majority of existing approaches primarily focus on achieving consistency across multiple views, often overlooking the efficacy of view-specific information. The interplay between these two types of information presents a conflict, prompting the need to strike a balance between consistency and specificity within multi-view settings. To address this challenge, we propose a novel model comprising three main components: bipartite graph learning, mutual information maximization between views for obtaining consistent information, and the utilization of optimal transport (OT) distance to measure the bipartite graph distance between views. This framework allows for the selective incorporation of view-specific information based on view graph.

3.2. View-Independent Bipartite Graph Learning

Given multi-view data $\mathbf{X} = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \cdots, \mathbf{X}^{(V)}\}$ with n samples and V views. Denote $\mathbf{G} = \{\mathbf{G}^{(1)}, \mathbf{G}^{(2)}, \cdots, \mathbf{G}^{(V)}\}$, where $\mathbf{G}^{(i)}$ is a bipartite graph on the i-th view. We introduce three spaces: the target space $\mathcal{Y}$, the raw data space $\mathcal{X}$, and the learned latent space $\mathcal{Z}$. In the proposed framework, we define three mappings: $f(\star) : \mathcal{X} \mapsto \mathcal{Z}$, which transforms the input space into the latent space; $g(\star) : \mathcal{Z} \mapsto \mathcal{X}$, which maps the latent space back to the input space; and $h(\star) : \mathcal{Z} \mapsto \mathcal{Y}$, which connects the latent space to the target space. To initialize the bipartite graph $\mathbf{G}^{(i)} \in \mathbb{R}^{n \times m^{(i)}}$ for the i-th view, we employ clustering algorithms to obtain anchor points within each view. Subsequently, we utilize an encoder-decoder architecture to map the initial $\mathbf{G}^{(i)}$ and obtain the refined graph $\hat{\mathbf{G}}^{(i)}$ as follows:

$$\hat{\mathbf{G}}^{(i)} = g(f(\mathbf{G}^{(i)})). \tag{1}$$

After the mapping process, we obtain a representation $\mathbf{Z}^i$ for the i-th view in the latent space, as well as the reconstructed graph $\hat{\mathbf{G}}^{(i)}$. In order to reconstruct the sample-to-sample relationships within each view, we proceed to reconstruct these relationships in the input space as follows:

$$L_{re} = \sum_{i=1}^{V} \left\| \hat{\mathbf{G}}^{(i)} \hat{\mathbf{G}^{(i)}}^{T} - \mathbf{G}^{(i)} \mathbf{G}^{(i)T} \right\|_F^2. \tag{2}$$

After the pre-training phase, we update the initial bipartite graph $\mathbf{G}^{(i)}$ by constructing a new bipartite graph using the representation of the latent space. This procedure corresponds to the mapping from the raw data space $\mathcal{X}$ to the learned latent space $\mathcal{Z}$.

3.3. Cross-view Topology Graph Construction

After pre-training, the latent representation $\mathbf{Z}^{(i)}$ for each view is acquired. Subsequently, these potential representations are mapped to generate the bipartite graph structure $\widetilde{\mathbf{G}}^{(i)}$ for each respective view. Notably, the bipartite graph structures on each view share a common dimensionality. In order to ascertain the comprehensive relationship between views, our objective is to individually assess the bipartite graph transition distances between them. Assume that for each node in $g_a^{(i)}$, it has $\mathbf{u}_a$ units transferred to $g_b^{(j)}$. For a node in $g_b^{(j)}$, it has $\mathbf{t}_b$ units to receive. Then, in this case, for a given pair of nodes $g_a^{(i)}$ and $g_b^{(j)}$, the cost of each unit transfer is $\mathbf{P}_{ab}$, and the total number of transfers is $\mathbf{R}_{ab}$. $\Pi(\mathbf{u}, \mathbf{r})$ denotes the set of all possible distributions whose marginal weights are $\mathbf{u}$ and $\mathbf{t}$. $\mathbf{u}$ and $\mathbf{t}$ are the marginal weights of $\mathbf{R}_{ab}$, respectively. Then, we can get the following optimization problem according to the above definition.

$$\begin{aligned} \mathbf{D}_w(\widetilde{\mathbf{G}}^{(i)}, \widetilde{\mathbf{G}}^{(j)}) &= \min_{\mathbf{R} \in \Pi(\mathbf{u}, \mathbf{r})} \sum_{a}^{n} \sum_{b}^{n} \mathbf{P}_{ab} \mathbf{R}_{ab}, \\ \text{s.t. } \mathbf{R}_{ab} &\geq 0, a, b = 1, 2, \ldots, n, \end{aligned} \tag{3}$$

where $\Pi(\mathbf{u}, \mathbf{t}) = \left\{ \mathbf{\Gamma} \in \mathbb{R}^{n \times n} \mid \mathbf{\Gamma} \mathbf{1}_n = \mathbf{u}, \mathbf{\Gamma}^T \mathbf{1}_n = \mathbf{t} \right\}$, $\mathbf{1}_n$ denotes an n-dimensional all-one vector. And $\mathbf{P}_{ab}$ is the cost function of transferring unit $g_a^{(i)}$ to $g_b^{(j)}$, as follows:

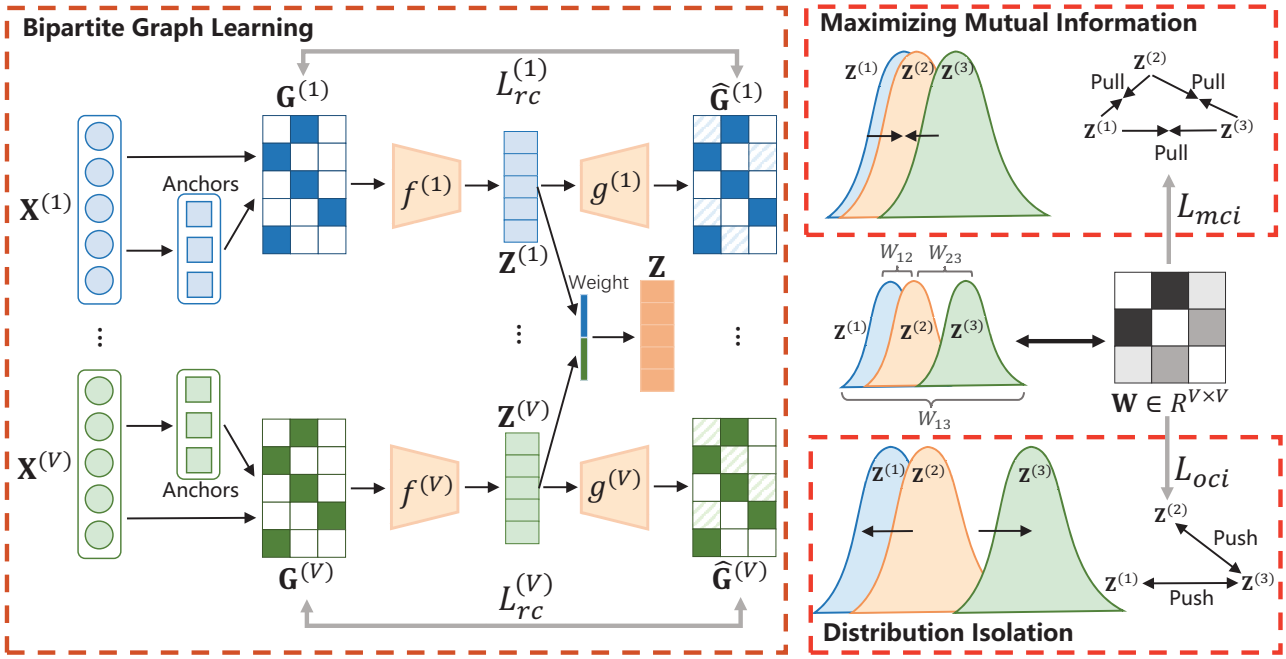


Figure 2: The framework of our proposed CTCC. The entire model consists of three main modules: bipartite graph learning, cross-view consistency maximization of information, and maximization of specific view Optimal Transport (OT) distance through the view topology graph $\mathbf{W}$ to obtain complementary information from the view. The goal of bipartite graph learning is to learn representative information in each view. Cross-view consistency information maximization aims to learn cross-view consistency, while distribution isolation seeks to preserve view-specific information.

$$\mathbf{P}_{ab} = 1 - \frac{{g_a^{(i)}}^T g_b^{(j)}}{\left\| g_a^{(i)} \right\| \left\| g_b^{(j)} \right\|}. \tag{4}$$

Then, we solve the Eq.(3) for every two views to get the view relation matrix $\mathbf{W} \in \mathbb{R}^{V \times V}$, the specific formula is listed as follows:

$$\mathbf{W}_{ij} = \mathbf{D}_w(\widetilde{\mathbf{G}}^{(i)}, \widetilde{\mathbf{G}}^{(j)}). \tag{5}$$

The matrix $\mathbf{W}_{ij}$ encapsulates the relationship between the i-th and j-th views. Consequently, the interrelationships among multiple views can be assessed via the matrix $\mathbf{W}$. It is evident that a smaller value of $\mathbf{W}_{ij}$ signifies a greater degree of shared information, thereby implying a higher similarity between the i-th and j-th views.

After optimizing the bipartite graph $\widetilde{\mathbf{G}}^{(i)}$, we proceed to fuse information from each view. Utilizing the view relation matrix $\mathbf{W}$, we can concentrate on and incorporate view-specific information. As illustrated in Fig. 2, based on the calculation method of $\mathbf{W}$, we can deduce that the $\mathbf{W}_{3j}$ between the third view and all other views is large, signifying that the third view contains more view-specific supplementary information. During the fusion process, we aim to incorporate this information.

However, the third view may also contain a significant amount of noise for some data. To address this issue, we introduce a hyperparameter δ, which represents the tolerance of the data to view-specific information. For each view, we perform the following operations according to the view relationship $\mathbf{W}$:

$$\overline{\mathbf{Z}}^{(v)} = \left(\frac{V \sum_{i=1}^{V} W_{vi}}{\sum_{i=1}^{V} \sum_{j=1}^{V} W_{ij}} \right)^{\delta} \mathbf{Z}^{(v)}. \tag{6}$$

Then, we concatenate the manipulated representations on each view to obtain a common representation $\mathbf{Z}$:

$$\mathbf{Z} = \overline{\mathbf{Z}}^{(1)} \oplus \overline{\mathbf{Z}}^{(2)} \oplus \cdots \overline{\mathbf{Z}}^{(v)}. \tag{7}$$

3.4. Consistency Maximization via Mutual Information

To obtain consistent information among multiple views, we utilize an approach that maximizes the mutual information between different views. The consistent information is optimized as the neural network refines the representation of each view. Initially, we define $I\left(\mathbf{Z}^{(1)}; \mathbf{Z}^{(2)}\right)$ as the mutual information between view 1 and view 2. Additionally, $I\left(\mathbf{Z}^{(i)}; \mathbf{Z}\right)$ denotes the mutual information between the i-th view and the shared view.

Definition 1. *Among the representations of multiple views* $\mathbf{Z}^{(v)}$, $\mathbf{Z} = \{\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \cdots, \mathbf{Z}^{(V)}\}$ *is considered consistent if and only if* $I\left(\mathbf{Z}^{(1)}; \mathbf{Z}^{(2)}; \cdots; \mathbf{Z}^{(i)}; \cdots; \mathbf{Z}^{(V)}\right) \geq I\left(\mathbf{Z}^{(1)}; \mathbf{Z}^{(2)}; \cdots; \hat{\mathbf{Z}}^{(i)}; \cdots; \mathbf{Z}^{(V)}\right)$, *for* $\forall i, \hat{\mathbf{Z}}^{(i)} \in \Omega(\mathbf{X}^{(i)})$, *where* $\Omega(\mathbf{X}^{(i)})$ *is the set of latent representations for the* i*-th view.*

Based on Definition 1, it is evident that to achieve consistency among multiple views, we aim to identify the maximum mutual information across these views in the latent space. For multiple views, we maximize the mutual information between any two views. As illustrated in Fig. 2, we obtain the fused $\mathbf{Z}$ through the relationship between views. To maximize the mutual information between views, we maximize the mutual information between each view and the shared $\mathbf{Z}$, thereby maximizing the mutual information between views. Consistent information between views can be maximized by maximizing the mutual information between each view representation and the shared view representation. However, to maximize mutual information, the dimensions of the two representations must be identical. Therefore, we employ an MLP to reduce the dimensionality of the shared $\mathbf{Z}$, making it the same dimension as the representation $\mathbf{Z}^{(i)}$ for each view: $\tilde{\mathbf{Z}} = \sigma(\mathbf{Z})$. Please refer to [17, 5] for the specific calculation method of mutual information. Nonetheless, assigning equal weight to more distant views will diminish the mutual information between views, as demonstrated in the third view in Fig. 1. Consequently, we utilize the view graph to impose varying penalties on the relationship between different views. For views similar to the third view, we apply a smaller weight when maximizing the mutual information to optimize the view information. The specific loss function is as follows:

$$\mathbf{L}_{mci} = -\sum_{i=1}^{V}\sum_{j\neq i}^{V}\left(C_{ij} I\left(\mathbf{Z}^{(i)}; \mathbf{Z}^{(j)}\right)\right) + \sum_{i=1}^{V} \varepsilon_i I\left(\mathbf{Z}^{(i)}; \tilde{\mathbf{Z}}\right), \tag{8}$$

where $C_{ij} \propto \frac{1}{\mathbf{W}_{ij}}$, ε_i is the row sum of the i-th row of $\mathbf{C}$.

3.5. Distribution Isolation via Cross-view Topology

After analyzing the view graph, it becomes apparent that solely maximizing the mutual information of latent representations between each view overlooks the specific information present within each view. As demonstrated in Fig. 2, much specific information of view three is neglected when maximizing the mutual information of view latent representations. Ideally, we would like to leverage the unique information from certain special views. In instances where this information is not noise, it can significantly contribute to clustering tasks. Consequently, we consider the optimal transport (OT) distance between the view farthest from all other views and the view closest to other views as the training objective, maintaining it within a specific range to effectively utilize the unique information of the view. We define Δ as the set $\Delta = \left\{\sigma | \sum_i^V w_{\sigma i} > \eta\right\}$. Additionally, we define Λ as the set $\Lambda = \left\{\varepsilon | \sum_i^V w_{\varepsilon i} < \rho\right\}$. We compute the distance $\mathbf{D}_w(\widetilde{\mathbf{G}}^{(\Delta)}, \widetilde{\mathbf{G}}^{(\Lambda)})$ between the bipartite graphs formed by the view latent representations in the two sets. However, for consistent information, we do not want to separate the views in these two collections too much. Therefore, we introduce a regularization term β for this loss. We aim to use this optimization formula to counterbalance Eq.(8), thereby learning both the consistency information of the view and the specific information within the view. The specific formula is as follows:

$$\mathbf{L}_{oci} = \sum_{i\in\Delta}\sum_{j\in\Lambda}\frac{1}{\mathbf{D}_w(\widetilde{\mathbf{G}}^{(i)}, \widetilde{\mathbf{G}}^{(j)})}. \tag{9}$$

In summary, the final loss is:

$$\mathbf{L}_{all} = \mathbf{L}_{re} + \varphi\mathbf{L}_{mci} + \beta\mathbf{L}_{oci}, \tag{10}$$

where φ and β is the regularization coefficient.The specific algorithm flow is shown in Algorithm 1.

Algorithm 1 The Proposed CTCC

Input: $\left\{\mathbf{X}^{(i)}\right\}_{i=1}^{V} \in \mathbb{R}^{n\times d_i}$, number of clusters k, learning rate α_t.

for $i = 1 : V$ do
  Obtain the bipartite graph $\mathbf{G}^{(i)}$ on the i-th view via k-means;
  Use the initial bipartite graph to pre-train the network through minimize Eq. (2);
  After pre-training, update the initial anchors;
end
while not reaching the maximal epochs **do**
  Obtain the bipartite graph $\widetilde{\mathbf{G}}^{(v)}$ corresponding to the latent space representation on each view;
  Obtain the view topology graph $\mathbf{W}$ between views by optimizing Eq.(5);
  Update $\overline{\mathbf{Z}}^{(v)}$ on each view through the view topology graph by Eq.(6);
  Obtain $\mathbf{Z}$ by minimizing Eq.(10);
end while
Output: Perform k-means on $\mathbf{Z}$ to achieve the final result.

Table 1: Multi-view datasets used in experiments.

Dataset	Samples	Clusters	Views	Dimensionality
MSRCV	210	7	6	256/48/100/512/210/1302
Leaves	1600	100	3	64/64/64
HandWritten	2000	10	6	216/76/64/6/240/47
UCI-digit	2000	10	3	64/76/216
ALOI	10800	100	4	77/13/64/125

4. Experiment

4.1. Datasets and Experimental Setting

In this section, we evaluate the performance of our proposed model in comparison with CTCC and nine state-of-the-art algorithms on five benchmark multi-view datasets. Table 1 presents the number of samples, number of clusters, and feature dimensionality for each dataset. MSRCV is an image dataset comprising 210 images with categories such as people, animals, buildings, and natural objects. UCI-digit is a handwritten digit dataset containing 2,000 samples. Leaves is a dataset of 100 plant species' leaves, consisting of 1,600 samples. For each feature, a 64-element vector is provided per leaf sample. ALOI is a color image collection of 1,000 small objects, resulting in a total of 10,800 images for the collection. In this model, there are several hyperparameters. The range of values for hyperparameters β and φ is [0.001, 1000], and the range for hyperparameter δ is [0.1, 1]. The specific settings for sets Δ and Λ are $\eta = 0.8$, $\rho = 0.3$. The learning rate (lr) employed in the model is 0.001. Our model is implemented using PyTorch 1.7.1 and trained on a desktop computer equipped with an NVIDIA GeForce RTX 3080 and 64GB RAM. We utilize the Adam optimizer with its default parameters.

4.2. Compared Methods

We conducted a comparison of CTCC with nine state-of-the-art multi-view clustering algorithms on five real-world multi-view datasets.

- **RMKM**[2] integrates heterogeneous representations of large-scale data to cluster large-scale multi-view data effectively.
- **LMVSC**[11] efficiently handles a vast number of views by exploiting the data structure, achieving linear time complexity.
- **FMR**[14] introduces a flexible multi-view representation learning approach for subspace clustering, which learns a shared latent space representation for multiple views while accommodating view-specific transformations.
- **BMVC**[56] is an innovative binary multi-view clustering method that employs graph-based techniques to infer the cluster structure of each view and subsequently integrates the results of different views into a final clustering outcome.
- **AE2-Nets**[54] utilizes a nested autoencoder architecture to encode data from diverse perspectives into a comprehensive embedding.
- **SDMVC**[51] proposes a self-supervised feature learning approach for deep multi-view clustering, leveraging a contrastive loss to learn discriminative features from multiple views.
- **MFLVC**[52] introduces a multi-level feature learning method for contrastive multi-view clustering, aiming to learn discriminative and complementary features from multiple views.
- **COMIC**[27] presents a parameter-free multi-view clustering method named COMIC, which leverages the consensus of multiple co-association matrices to achieve clustering results without the need for parameter tuning.
- **DSMVC**[32] incorporates a safe clustering module that uses the predictive variance of a deep neural network to identify potentially unsafe regions in the feature space.

4.3. Clustering Performance

We assessed the performance of our proposed method (CTCC) by comparing it to nine other baseline algorithms, utilizing four widely accepted clustering evaluation metrics: accuracy (ACC), normalized mutual information (NMI), Purity, and F-score. The clustering performance of our method, as well as the other baseline algorithms, is presented in Table 2. Based on the findings in Table 2, we can draw the following conclusions: (1) Our method exhibits a substantial improvement over the other nine multi-view clustering algorithms across all datasets and evaluation metrics. With respect to ACC, our method outshines the baseline algorithms, particularly on the MSRCV, 100Leaves, and ALOI-100 datasets. Our algorithm surpasses the second-best algorithm by 9.6%, 15.07%, and 11.75%, respectively. These results suggest that emphasizing the consistency and complementarity among multiple views, as well as optimizing based on the relationships between views, has achieved remarkable results. (2) As observed in Table 2, the clustering performance of our model notably enhances as the number of data views increases. This is particularly evident in the MSRCV and HandWritten datasets, which demonstrate the effectiveness of view relationships.

Table 2: ACC, NMI, Purity and F-score comparison of different clustering algorithms on benchmark datasets. The best result is highlighted in bold, while the second best is marked with an underlined number. 'O/M' refers to out-of-memory failure. '-' indicates the error of the method itself.

Datasets	Conventional Methods				Deep Methods					
	RMKM	LMVSC	FMR	BMVC	AE^2-Nets	SDMVC	MFLVC	COMIC	DSMVC	**CTCC**
ACC(%)										
MSRCV	71.42	<u>83.73</u>	76.16	26.66	32.10	48.10	38.57	46.19	64.29	**93.33**
HandWritten	67.10	85.10	59.62	75.50	87.82	49.45	86.20	45.15	<u>95.25</u>	**97.45**
UCI	81.85	89.35	54.44	85.45	80.57	63.00	<u>92.00</u>	73.55	85.45	**95.65**
Leaves	48.81	66.31	48.21	<u>74.38</u>	47.94	67.31	13.00	40.69	51.81	**81.38**
ALOI-100	33.74	42.40	O/M	<u>63.54</u>	26.90	O/M	23.12	9.940	14.89	**75.29**
NMI(%)										
MSRCV	63.03	<u>78.93</u>	68.38	8.290	27.44	49.24	32.78	66.52	54.29	**87.27**
HandWritten	65.33	80.56	49.19	77.52	80.18	5.190	85.53	67.85	<u>91.19</u>	**94.52**
UCI	76.04	83.27	45.57	82.52	69.51	64.29	<u>85.40</u>	81.80	80.67	**91.95**
Leaves	78.13	85.22	71.41	<u>89.37</u>	75.15	88.28	60.20	74.16	79.24	**91.99**
ALOI-100	63.55	55.82	O/M	<u>76.99</u>	49.47	O/M	67.88	30.92	40.57	**82.87**
Purity(%)										
MSRCV	74.76	<u>85.25</u>	77.77	27.14	33.10	51.90	38.57	85.24	64.29	**93.33**
HandWritten	75.95	85.10	60.98	79.45	87.82	49.85	86.20	<u>96.75</u>	95.25	**97.45**
UCI	81.85	89.35	56.96	85.45	80.57	67.05	<u>92.00</u>	81.35	85.40	**95.65**
Leaves	74.19	70.09	50.66	<u>78.06</u>	49.94	70.44	13.00	45.44	53.12	**84.88**
ALOI-100	64.02	44.09	O/M	<u>65.49</u>	27.58	O/M	23.12	10.44	15.79	**76.49**
F-score(%)										
MSRCV	59.98	<u>77.43</u>	66.69	16.01	32.50	44.84	33.04	—	60.12	**93.58**
HandWritten	59.21	77.23	41.49	71.13	79.35	42.10	85.82	—	<u>95.27</u>	**97.45**
UCI	81.85	89.35	35.77	85.45	67.65	62.77	<u>92.05</u>	—	86.09	**95.66**
Leaves	38.70	57.07	35.54	<u>66.61</u>	45.82	64.46	5.290	—	44.13	**81.43**
ALOI-100	28.82	31.68	O/M	<u>50.98</u>	24.59	O/M	12.02	—	0.360	**80.81**

Table 3: Ablation study on MSRCV. ✓ denotes CTCC with the component.

Components			Metrics(%)			
$\mathbf{L}_{re}$	$\mathbf{L}_{mci}$	$\mathbf{L}_{oci}$	ACC	NMI	Purity	F-score
✓			74.29	67.20	74.29	76.71
✓	✓		90.00	82.19	90.00	89.93
✓		✓	87.62	82.38	87.62	87.44
✓	✓	✓	**93.33**	**87.27**	**93.33**	**93.58**

Table 4: Ablation study on HandWritten. ✓ denotes CTCC with the component.

Components			Metrics(%)			
$\mathbf{L}_{re}$	$\mathbf{L}_{mci}$	$\mathbf{L}_{oci}$	ACC	NMI	Purity	F-score
✓			84.80	84.23	84.80	84.09
✓	✓		91.75	88.17	91.75	91.62
✓		✓	90.40	87.41	90.40	90.48
✓	✓	✓	**97.45**	**94.52**	**97.45**	**97.45**

4.4. Ablation of Mutual Information Maximization Module

To substantiate the efficacy of our mutual information maximization module, we carried out ablation experiments on both the MSRCV and Handwritten datasets. The specific results are delineated in Tables 3 and 4. Our observations reveal that maximizing mutual information between views can considerably enhance clustering performance. This underscores the pivotal role that shared information between views plays in clustering tasks. Notably, our mutual information maximization module surpassed models that solely focused on specific information within each view.

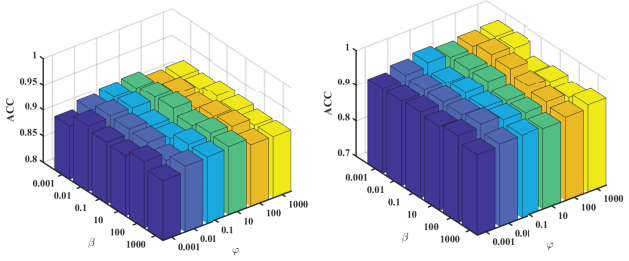
As depicted in Fig. 6, on the MSRCV and Handwritten datasets, maximizing mutual information across multiple views yielded a more consistent representation that outperformed clustering based on representations from individual views. This demonstrates that maximizing mutual information can effectively integrate shared information across multiple views, resulting in representations that are advantageous for clustering tasks. This confirmation supports the effectiveness of the mutual information maximization module. By maximizing mutual information between views, we

were able to obtain more consistent representations in the latent space, leading to improved multi-view clustering performance.

4.5. Ablation of cross-view topology module

To demonstrate the effectiveness of the cross-view topology module, we conducted an ablation experiment in which we eliminated the mutual information maximization module. Additionally, we conducted clustering on the representations obtained from each view, as illustrated in Tables 5 and 6, and Fig. 6. Our findings reveal that the clustering performance of the representations derived from the cross-view topology graph is markedly superior to that of each individual view. This implies that view-specific information can be advantageous to multi-view clustering. These observations provide compelling evidence for the effectiveness of the cross-view topology graph module within our proposed framework.

Nonetheless, our results demonstrated that complementarity information alone did not enhance clustering performance as substantially as consistency information. However, when the view consistency module and the complementarity module were amalgamated, we discovered that the combination exerted a significant impact on clustering performance, signifying that complementary information is valuable for improving clustering performance. From Tables 3 and 4, We can also see impressive clustering performance by combining the two modules. Consequently, to a certain extent, integrating both consistency and complementary information can substantially augment clustering performance.

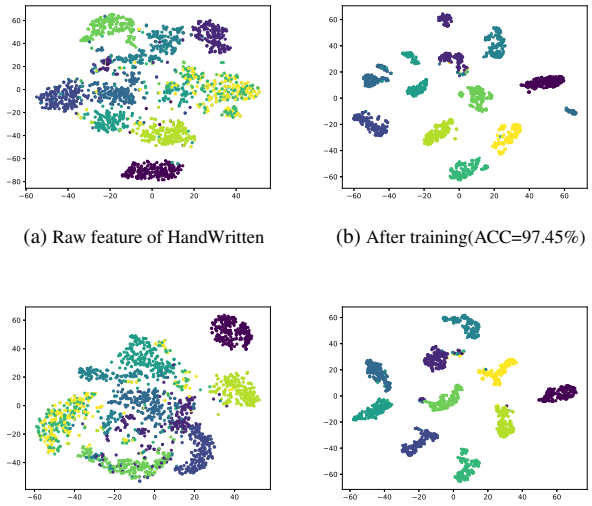


(a) MSRCV (b) HandWritten

Figure 3: Sensitivity analysis of φ and β for our method over MSRCV and HandWritten.

4.6. Visualization of bipartite graphs

To visually demonstrate the impact of cross-view topology on bipartite graph fusion for each view, we depicted the bipartite graphs learned on each view of the HandWritten dataset, as shown in Fig. 5. Upon scrutinizing the figure, it becomes apparent that the best view of the bipartite graph results in poor performance. However, a more effective bipartite graph outcome was achieved under the guidance of view topology, as demonstrated in (b) and (c). The influence of the bipartite graph, molded by consistent and complementary information, was markedly amplified. In (b) and (c), the likelihood of a sample being associated with a specific cluster increased after incorporating supplementary information. This observation can be ascribed to the acquisition of particular supplementary information in the view, which corroborates the validity of the supplementary information.



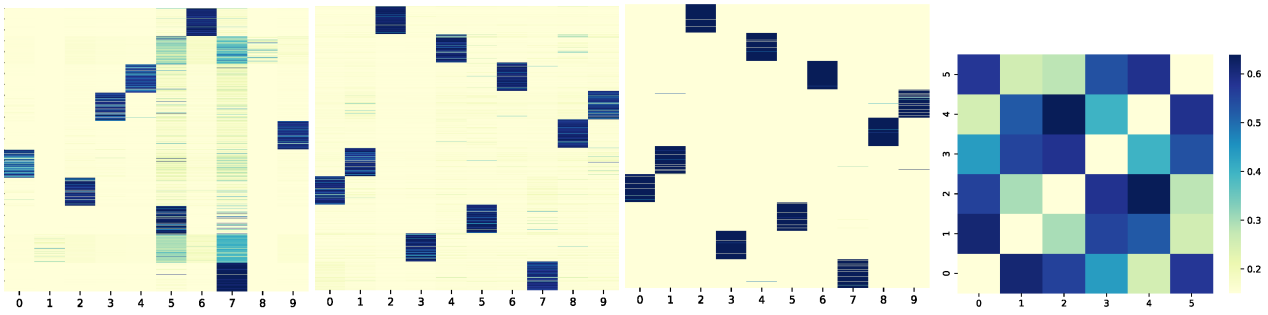
(a) Raw feature of HandWritten (b) After training(ACC=97.45%)

(c) Raw feature of UCI-digit (d) After training(ACC=95.65%)

Figure 4: The visualization results on the HandWritten and UCI-digit datasets.

4.7. Parameter Sensitivity Analysis

In our parameter sensitivity analysis, we maintained a learning rate of 0.001 for all experiments. Our loss function includes two hyperparameters: a regularization parameter φ to explore consistency information between views and a regularization parameter β for complementarity information. To evaluate the impact of these two parameters, we varied their values within the range of [0.001, 0.01, 0.1, 10, 100, 1000], as depicted in Fig. 3. Our findings revealed an initial increase followed by a decrease in performance. Notably, alterations in parameter φ exerted a substantial impact on the outcomes, demonstrating the effectiveness of complementary information in improving clustering performance. Specifically, our model attained maximum accuracy on the MSRCV dataset when φ was assigned a value of 10 and β was set to 0.1, whereas on the HandWritten dataset,



(a) best view (b) consistent representation (c) consistent and complementary (d) view graph

Figure 5: The visualization of the bipartite graph on the best view, consistent representation of all views, adding complementary information to the bipartite graph, and view topology graph on the HandWritten dataset.

the maximum accuracy was achieved when φ was assigned a value of 100 and β was set to 0.01.

4.8. Visualization of clustering performance

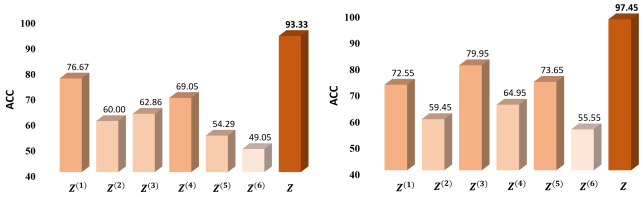
Additionally, we performed visualization of the clustering effect of CTCC on the HandWritten and UCI-digit datasets. We generated visualizations for both the original features and the final results obtained through our deep network. As illustrated in Fig. 4, it is evident that the results achieved through our proposed network framework surpass those of the original features. The final representations successfully cluster similar samples together, resulting in accuracy values of 97.45% and 95.65% on the HandWritten and UCI-digit datasets, respectively. This serves as compelling evidence for the effectiveness of our proposed framework, which integrates consistency and complementarity information through the cross-view topological graph.

Table 5: Clustering performance comparison of single-view and multi-view on MSRCV dataset.

Metrics	V1	V2	V3	V4	V5	V6	**CTCC**
ACC	76.67	60.00	62.86	69.05	54.29	49.05	**93.33**
NMI	67.47	55.30	60.12	59.25	47.41	43.15	**87.27**
F-score	77.43	58.68	63.53	68.54	56.48	49.20	**93.58**

Table 6: Clustering performance comparison of single-view and multi-view on HandWritten dataset.

Metrics	V1	V2	V3	V4	V5	V6	**CTCC**
ACC	72.55	59.45	79.95	64.95	73.65	55.55	**97.45**
NMI	70.50	65.74	83.18	65.61	73.18	65.97	**94.52**
F-score	70.86	57.85	78.48	63.78	73.19	52.43	**97.45**



(a) MSRCV (b) HandWritten

Figure 6: The clustering performance of latent representations on each view and representations containing consistency and complementarity information is compared on datasets MSRCV and HandWritten.

5. Conclusion

To make full use of the complementary information of each view while learning consistent representation and to explore the correlation between multiple views, this paper present CTCC, which unifies the two seemingly opposite pieces of information into a framework. In brief, the proposed CTCC utilize the constructed cross-view topology to guide the maximization of the mutual information between views and the isolation of distributions to tackle the above issues. Comprehensive experiments demonstrate our proposed CTCC's superiority compared with conventional and deep SOTA methods. We believe that our motivation is worthy to be discussed and will bring some new insights to the multi-view clustering community.

Acknowledgments

This work is supported by the National Key R&D Program of China under Grant No.2022ZD0209103, the National Natural Science Foundation of China (project no. 62325604, 62276271).

References

[1] Rana Ali Amjad and Bernhard C Geiger. Learning representations for neural network-based classification using the information bottleneck principle. *IEEE transactions on pattern analysis and machine intelligence*, 42(9):2225–2239, 2019.

[2] Xiao Cai, Feiping Nie, and Heng Huang. Multi-view k-means clustering on big data. In *Twenty-Third International Joint conference on artificial intelligence*. Citeseer, 2013.

[3] Rui Chen, Yongqiang Tang, Wensheng Zhang, and Wenlong Feng. Deep multi-view semi-supervised clustering with sample pairwise constraints. *Neurocomputing*, 500:832–845, 2022.

[4] Guowang Du, Lihua Zhou, Zhongxue Li, Lizhen Wang, and Kevin Lü. Neighbor-aware deep multi-view clustering via graph convolutional network. *Information Fusion*, 2023.

[5] Marco Federici, Anjan Dutta, Patrick Forré, Nate Kushman, and Zeynep Akata. Learning robust representations via multi-view information bottleneck. *arXiv preprint arXiv:2002.07017*, 2020.

[6] Xianxu Hou, Linlin Shen, Ke Sun, and Guoping Qiu. Deep feature consistent variational autoencoder. In *2017 IEEE winter conference on applications of computer vision (WACV)*, pages 1133–1141. IEEE, 2017.

[7] Xingchen Hu, Xinwang Liu, Witold Pedrycz, Qing Liao, Yinghua Shen, Yan Li, and Siwei Wang. Multi-view fuzzy classification with subspace clustering and information granules. *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[8] Shudong Huang, Zhao Kang, and Zenglin Xu. Auto-weighted multi-view clustering via deep matrix decomposition. *Pattern Recognition*, 97:107015, 2020.

[9] Shudong Huang, Ivor W Tsang, Zenglin Xu, and Jiancheng Lv. Measuring diversity in graph learning: a unified framework for structured multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 34(12):5869–5883, 2021.

[10] Zhenyu Huang, Joey Tianyi Zhou, Xi Peng, Changqing Zhang, Hongyuan Zhu, and Jiancheng Lv. Multi-view spectral clustering network. In *IJCAI*, pages 2563–2569, 2019.

[11] Zhao Kang, Wangtao Zhou, Zhitong Zhao, Junming Shao, Meng Han, and Zenglin Xu. Large-scale multi-view subspace clustering in linear time. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 4412–4419, 2020.

[12] Guanzhou Ke, Zhiyong Hong, Zhiqiang Zeng, Zeyi Liu, Yangjie Sun, and Yannan Xie. Conan: contrastive fusion networks for multi-view clustering. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 653–660. IEEE, 2021.

[13] Liang Li, Junpu Zhang, Siwei Wang, Xinwang Liu, Kenli Li, and Keqin Li. Multi-view bipartite graph clustering with coupled noisy feature filter. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–13, 2023.

[14] Ruihuang Li, Changqing Zhang, Qinghua Hu, Pengfei Zhu, and Zheng Wang. Flexible multi-view representation learning for subspace clustering. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 2916–2922, 2019.

[15] Zhaoyang Li, Qianqian Wang, Zhiqiang Tao, Quanxue Gao, Zhaohua Yang, et al. Deep adversarial multi-view clustering network. In *IJCAI*, pages 2952–2958, 2019.

[16] Ke Liang, Sihang Zhou, Yue Liu, Lingyuan Meng, Meng Liu, and Xinwang Liu. Structure guided multi-modal pre-trained transformer for knowledge graph reasoning. *arXiv preprint arXiv:2307.03591*, 2023.

[17] Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[18] Jiyuan Liu, Xinwang Liu, Siwei Wang, Sihang Zhou, and Yuexiang Yang. Hierarchical multiple kernel clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 8671–8679, 2021.

[19] Jiyuan Liu, Xinwang Liu, Yuexiang Yang, Xifeng Guo, Marius Kloft, and Liangzhong He. Multiview subspace clustering via co-training robust data representation. *IEEE Transactions on Neural Networks and Learning Systems*, 33(10):5177–5189, 2021.

[20] Liang Liu, Peng Chen, Guangchun Luo, Zhao Kang, Yonggang Luo, and Sanchu Han. Scalable multi-view clustering with graph filtering. *Neural Computing and Applications*, 34(19):16213–16221, 2022.

[21] Suyuan Liu, Siwei Wang, Pei Zhang, Kai Xu, Xinwang Liu, Changwang Zhang, and Feng Gao. Efficient one-pass multi-view subspace clustering with consensus anchors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 7576–7584, 2022.

[22] Xuanwu Liu, Guoxian Yu, Carlotta Domeniconi, Jun Wang, Yazhou Ren, and Maozu Guo. Ranking-based deep cross-modal hashing. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 4400–4407, 2019.

[23] Xinwang Liu, Xinzhong Zhu, Miaomiao Li, Lei Wang, Chang Tang, Jianping Yin, Dinggang Shen, Huaimin Wang, and Wen Gao. Late fusion incomplete multi-view clustering. *IEEE transactions on pattern analysis and machine intelligence*, 41(10):2410–2423, 2018.

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

[25] Muhammad Mateen, Junhao Wen, Sun Song, and Zhouping Huang. Fundus image classification using vgg-19 architecture with pca and svd. *Symmetry*, 11(1):1, 2018.

[26] Xi Peng, Zhenyu Huang, Jiancheng Lv, Hongyuan Zhu, and Joey Tianyi Zhou. Comic: Multi-view clustering without parameter selection. In *International conference on machine learning*, pages 5092–5101. PMLR, 2019.

[27] 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*, volume 97 of *Proceedings of Machine Learning Research*, pages 5092–5101, Long Beach, California, USA, 09–15 Jun 2019. PMLR.

[28] Zhen Peng, Minnan Luo, Wenbing Huang, Jundong Li, Qinghua Zheng, Fuchun Sun, and Junzhou Huang. Learning representations by graphical mutual information estimation and maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):722–737, 2022.

[29] Yazhou Ren, Kangrong Hu, Xinyi Dai, Lili Pan, Steven CH Hoi, and Zenglin Xu. Semi-supervised deep embedded clustering. *Neurocomputing*, 325:121–130, 2019.

[30] Zhenwen Ren, Quansen Sun, and Dong Wei. Multiple kernel clustering with kernel k-means coupled graph tensor learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 9411–9418, 2021.

[31] Mengjing Sun, Pei Zhang, Siwei Wang, Sihang Zhou, Wenxuan Tu, Xinwang Liu, En Zhu, and Changjian Wang. Scalable multi-view subspace clustering with unified anchors. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 3528–3536, 2021.

[32] Huayi Tang and Yong Liu. Deep safe multi-view clustering: Reducing the risk of clustering performance degradation caused by view increase. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 202–211, 2022.

[33] Naftali Tishby, Fernando C Pereira, and William Bialek. The information bottleneck method. *arXiv preprint physics/0004057*, 2000.

[34] Naftali Tishby and Noga Zaslavsky. Deep learning and the information bottleneck principle. In *2015 ieee information theory workshop (itw)*, pages 1–5. IEEE, 2015.

[35] Daniel J Trosten, Sigurd Lokse, Robert Jenssen, and Michael Kampffmeyer. Reconsidering representation alignment for multi-view clustering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1255–1265, 2021.

[36] Zhibin Wan, Changqing Zhang, Pengfei Zhu, and Qinghua Hu. Multi-view information-bottleneck representation learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 10085–10092, 2021.

[37] Qianqian Wang, Jiafeng Cheng, Quanxue Gao, Guoshuai Zhao, and Licheng Jiao. Deep multi-view subspace clustering with unified and discriminative learning. *IEEE Transactions on Multimedia*, 23:3483–3493, 2020.

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

[39] Qianqian Wang, Zhiqiang Tao, Quanxue Gao, and Licheng Jiao. Multi-view subspace clustering via structured multi-pathway network. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[40] Shiping Wang, Xincan Lin, Zihan Fang, Shide Du, and Guobao Xiao. Contrastive consensus graph learning for multi-view clustering. *IEEE/CAA Journal of Automatica Sinica*, 9(11):2027–2030, 2022.

[41] Siwei Wang, Xinwang Liu, Li Liu, Sihang Zhou, and En Zhu. Late fusion multiple kernel clustering with proxy graph refinement. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[42] Siwei Wang, Xinwang Liu, Xinzhong Zhu, Pei Zhang, Yi Zhang, Feng Gao, and En Zhu. Fast parameter-free multi-view subspace clustering with consensus anchor guidance. *IEEE Transactions on Image Processing*, 31:556–568, 2021.

[43] Yifei Wang, Zhengyang Geng, Feng Jiang, Chuming Li, Yisen Wang, Jiansheng Yang, and Zhouchen Lin. Residual relaxation for multi-view representation learning. *Advances in Neural Information Processing Systems*, 34:12104–12115, 2021.

[44] Yang Wang, Lin Wu, Xuemin Lin, and Junbin Gao. Multi-view spectral clustering via structured low-rank matrix factorization. *IEEE transactions on neural networks and learning systems*, 29(10):4833–4843, 2018.

[45] Wei Xia, Qianqian Wang, Quanxue Gao, Xiangdong Zhang, and Xinbo Gao. Self-supervised graph convolutional network for multi-view clustering. *IEEE Transactions on Multimedia*, 24:3182–3192, 2021.

[46] Shunxin Xiao, Shide Du, Zhaoliang Chen, Yunhe Zhang, and Shiping Wang. Dual fusion-propagation graph neural network for multi-view clustering. *IEEE Transactions on Multimedia*, 2023.

[47] Yuan Xie, Bingqian Lin, Yanyun Qu, Cuihua Li, Wensheng Zhang, Lizhuang Ma, Yonggang Wen, and Dacheng Tao. Joint deep multi-view learning for image clustering. *IEEE Transactions on Knowledge and Data Engineering*, 33(11):3594–3606, 2020.

[48] Jie Xu, Chao Li, Yazhou Ren, Liang Peng, Yujie Mo, Xiaoshuang Shi, and Xiaofeng Zhu. Deep incomplete multi-view clustering via mining cluster complementarity. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8761–8769, 2022.

[49] Jie Xu, Yazhou Ren, Guofeng Li, Lili Pan, Ce Zhu, and Zenglin Xu. Deep embedded multi-view clustering with collaborative training. *Information Sciences*, 573:279–290, 2021.

[50] Jie Xu, Yazhou Ren, Huayi Tang, Xiaorong Pu, Xiaofeng Zhu, Ming Zeng, and Lifang He. Multi-vae: Learning disentangled view-common and view-peculiar visual representations for multi-view clustering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9234–9243, 2021.

[51] Jie Xu, Yazhou Ren, Huayi Tang, Zhimeng Yang, Lili Pan, Yang Yang, Xiaorong Pu, S Yu Philip, and Lifang He. Self-supervised discriminative feature learning for deep multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[52] Jie Xu, Huayi Tang, Yazhou Ren, Liang Peng, Xiaofeng Zhu, and Lifang He. Multi-level feature learning for contrastive multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16051–16060, 2022.

[53] Zhe Xue, Junping Du, Dawei Du, and Siwei Lyu. Deep low-rank subspace ensemble for multi-view clustering. *Information Sciences*, 482:210–227, 2019.

[54] Changqing Zhang, Yeqing Liu, and Huazhu Fu. Ae2-nets: Autoencoder in autoencoder networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[55] Pei Zhang, Xinwang Liu, Jian Xiong, Sihang Zhou, Wentao Zhao, En Zhu, and Zhiping Cai. Consensus one-step multi-view subspace clustering. *IEEE Transactions on Knowledge and Data Engineering*, 34(10):4676–4689, 2020.

[56] Zheng Zhang, Li Liu, Fumin Shen, Heng Tao Shen, and Ling Shao. Binary multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7):1774–1782, 2019.

[57] Runwu Zhou and Yi-Dong Shen. End-to-end adversarial-attention network for multi-modal clustering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14619–14628, 2020.