Aligning Contrastive Multiple Clusterings with User Interests

Shan Zhang[1†], **Liangrui Ren**[1†], **Jun Wang**[1], **Yanyu Xu**[1], **Carlotta Domeniconi**[2], **Guoxian Yu**[1*]
[1]School of Software, Shandong University, Jinan, China
[2]Department of Computer Science, George Mason University, VA, USA
{szhangs, renliangrui}@mail.sdu.edu.cn, {kingjun, xu_yanyu}@sdu.edu.cn, carlotta@cs.gmu.edu, gxyu@sdu.edu.cn

Abstract

Multiple clustering approaches aim to partition complex data in different ways. These methods often exhibit a one-to-many relationship in their results, and relying solely on the data context may be insufficient to capture the patterns relevant to the user. User's expectation is key for the multiple clustering task. Two main challenges exist: identifying the significant features to represent user interests and aligning those interests with the clustering results. To address this issue, we propose Contrastive Multiple Clusterings (`CMClusts`), which extends contrastive learning to multiple clustering by elevating traditional instance-level contrast to clustering-level contrast. Furthermore, `CMClusts` integrates user expectations or interests by extracting desired features through tailored data augmentations, enabling the model to effectively capture user-relevant clustering features. Experimental results on benchmark datasets show that `CMClusts` can generate interpretable and high-quality clusterings, which reflect different user interests.

1 Introduction

Clustering is a mainstream unsupervised learning technique that groups samples into several disjoint clusters [Jain *et al.*, 1999]. As data grows more complex, relying on a single clustering may no longer be sufficient. In many situations, different orthogonal solutions may exist to meaningfully group a given dataset [Bailey, 2018]. For example, a face image dataset can be clustered based on identity and pose (as shown in Figure 4), respectively. To address this problem, **multiple clustering** approaches have been explored to partition a given dataset in various ways [Yu *et al.*, 2024].

The large majority of existing multiple clustering methods explore the data clustering structure by focusing on the **data**. One approach seeks alternative clusterings in a semi-supervised and sequential manner, where each clustering complements already explored ones [Bae and Bailey, 2006; Chang *et al.*, 2017], or simultaneously optimizes multiple sets of cluster centers or meta-clusterings in an unsupervised fashion [Tokuda *et al.*, 2021; Yao *et al.*, 2023]. In contrast, other techniques aim to explore non-redundant clusterings in different feature subspaces, including ISAAC [Ye *et al.*, 2016], MSC [Hu *et al.*, 2017], MISC [Wang *et al.*, 2019], MVMC [Yao *et al.*, 2019], and Nr-kmeans [Mautz *et al.*, 2020], thus discovering diverse clusterings in linear subspaces. Other methods aim to obtain alternative clusterings in nonlinear subspaces generated by deep neural networks [Miklautz *et al.*, 2020; Wei *et al.*, 2020b; Wei *et al.*, 2021; Ren *et al.*, 2023b; Yao *et al.*, 2023]. However, a critical point is that these approaches often assume the clusters are distinguishable as long as the data structure is known, which is not always the case. A one-to-many relationship exists in multiple clusterings, and relying solely on the data context may be insufficient to determine the appropriate clusterings. *User expectations* play a crucial role in the multiple clustering task, as they define the relevance and practical utility of the identified clusters.

Two main challenges exist in finding multiple clusterings that align with the user's interests. The first is identifying the significant features to represent user interests. This involves the extraction of effective and valuable information, typically embedded in high-dimensional spaces. The second challenge is how to align the user's interests with the clustering results. Although multiple clustering algorithms are capable of generating diverse clustering outcomes, users typically need to invest considerable effort to understand these results. Moreover, it is difficult to extract meaningful insights about the clustering semantics directly from the algorithmic framework.

To address these challenges, we propose the Contrastive Multiple Clusterings (`CMClusts`) approach, which leverages tailored data augmentation strategies to perform contrastive comparisons, thereby producing distinct clusterings aligned with user expectations. Targeted at the first challenge, we design an interest-guided data augmentation, to augment the original data to account for the user's different interests. As an example, the user's visual/semantic priors are typically readily available and easy to specify, e.g. we can intuitively select the most salient features in images, such as color, identity, and shape. To address the issue of insignificant features, the interest-guided data augmentation generates more diverse training samples by applying transformations on the original data. By capturing the invariant features under such transformations, we can effectively distill meaningful representations from the data. Feature compositions are complex and challenging to distinguish. By taking the original dataset as an anchor and contrasting it with the augmented data that preserves the desired features while altering others, we can guide the model to focus on the invariant features according with user interests. In addition, introducing augmented negative samples that break the invariance of the desired features and maximizing their distance from positive samples can yield more discriminative results. CMClusts formulates this process through the adoption of a *clustering-level contrast* that uses a triplet loss function to model the relationships between positive and negative samples (created through different data augmentations), aligning desired features with different interests. By minimizing the distance between anchors and positive samples, CMClusts ensures the effective extraction of desired features and the quality of clusterings, while maximizing the distance between anchors and negative samples guarantees the distinctiveness of different clusterings. As a result, CMClusts finds diverse clusterings of quality, and these clusterings align with user-interests.

The main contributions of our work are outlined below:
(i) Our CMClusts establishes a principled integration of user-specified clustering interests with data augmentation techniques, employing contrastive learning between augmented and original data to disentangle and extract the desired features from the original feature space, thereby effectively aligning with user interests.
(ii) CMClusts introduces the clustering-level contrast that treats augmented data for the other clustering as negative samples. This contrast not only enhances the discrimination between positive and negative samples but also boost the diversity between different clusterings.
(iii) We conduct experiments using real-world benchmark datasets, and compare CMClusts against representative and competitive multiple clustering methods [Cui *et al.*, 2007; Yang and Zhang, 2017; Miklautz *et al.*, 2020; Ren *et al.*, 2023b; Yao *et al.*, 2023; Yao *et al.*, 2024a]. Extensive results demonstrate the advantages of CMClusts in generating diverse clusterings aligning with user's interests.

2 Related Work

2.1 Multiple clusterings

Multiple clustering approaches aim to generate different alternative clusterings for more comprehensive analysis of the data [Bailey, 2018; Yu *et al.*, 2024]. It differs from the multi-view clustering [Zhang *et al.*, 2020] and subspace clustering [Elhamifar and Vidal, 2013], which generate only one clustering from diverse views or different cluster subspaces.

Early solutions aim to find alternative clusterings in the original feature space [Yu *et al.*, 2024]. Some of them take prior clusterings as the reference to sequentially generate the other alternative clustering in a semi-supervised manner [Yang and Zhang, 2017; Chang *et al.*, 2017]. Hence the quality of alternative clustering depends on referenced ones. In contrast, other methods simultaneously find multiple clusterings by jointly optimizing quality and diversity by cluster centroids [Jain *et al.*, 2008], mutual information [Dang and Bailey, 2010] and other criteria [Ren *et al.*, 2023a]. Recently, [Yao *et al.*, 2023; Yao and Hu, 2024] proposed two augmentation guided deep multiple clustering solutions (AugDMC and DDMC), which discover multiple clusterings through prototype-based representation learning and variational expectation-maximization framework, respectively. These methods learn feature consistency by disrupting the structural integrity of the original data, which is prone to introducing noise and semantic bias. Besides, [Yao *et al.*, 2024a; Yao *et al.*, 2024b] utilized the alignment capabilities of multi-modal large models (e.g. CLIP [Radford *et al.*, 2021]) to align users' textual descriptions of preferences with corresponding images, thereby generating clustering results that align with user expectations. However, these large model based solutions require additional textual descriptions, as they are unable to effectively extract user interested features solely from the data itself, and need a higher computational cost with demanding multi-modal data.

Another line of methods explore diverse feature subspaces and clusterings therein, where some methods [Cui *et al.*, 2007; Wang *et al.*, 2019; Mautz *et al.*, 2020] adopt orthogonal or independent subspace analysis to discover multiple clusterings, while others explore non-redundant clusterings by maximizing the feature statistical independence [Hu *et al.*, 2017; Niu *et al.*, 2013]. More recent solutions merge deep subspace learning to find multiple clusterings. ENRC [Miklautz *et al.*, 2020] trains a non-redundant clustering network leveraging deep autoencoders. iMClusts [Ren *et al.*, 2023b] diversifies multi-head attentions with redundancy control to generate diverse salient nonlinear subspaces and clusterings therein. [Wei *et al.*, 2020a; Wei *et al.*, 2020b; Wei *et al.*, 2021] found non-redundant clusterings from multi-view data and heterogeneous networks. These deep multiple clustering methods often obtain better results than shallow ones. However, they may produce meaningless clusterings, as they disregard user's interests or prior knowledge of the data, causing the relationships between subspaces and the embodied clusterings are ambiguous.

2.2 Contrastive clustering

As an emerging paradigm of self-supervised learning, contrastive learning aims to project data into a new space to reduce intra-class distances and increase inter-class distances [Chen *et al.*, 2020]. For this merit, it has been widely applied to clustering [Zhu *et al.*, 2022; Yin *et al.*, 2023]. For example, MiCE [Tsai *et al.*, 2020] introduces the instance-level contrastive learning to obtain discriminative representations to optimize the clustering, while [Zhong *et al.*, 2021] further proposed the cluster-level contrastive framework to get more compact cluster. Subsequently, [Xu *et al.*, 2022] proposed the contrastive multi-view clustering to solve the conflict between consistent public information and inconsistent private information of multi-view data.

These contrastive clustering methods can only generate a single clustering. In contrast, our CMClusts uses different data augmentations to align with user's interests, generating corresponding clustering results in the augmented data. By contrasting the clusterings across different augmented data, CMClusts facilitates clustering-level contrast to mine multiple clusterings with both quality and diversity.

3 Methodology

In this section, we elaborate on the proposed contrastive multiple clustering framework, including the user interest-guided data augmentation process, clustering-level contrastive feature extraction, and clustering loss function. The conceptual framework is presented in Fig. 1, and a notation table is included in the supplementary file.

3.1 User Interest-guided Data Augmentation

CMClusts aims to augment the original data to align with user's interests. Here, the user interests refer to the type of clusterings that users are expected or interested in. We consider image data as an example to illustrate the data augmentation process; other types of data can also be augmented using proper tools.

Given a dataset $\mathbf{X} \in \mathbb{R}^{N \times D}$, an image $\mathbf{x}_i$ can be augmented through a function $Aug(\cdot)$ based on the user's interest set $\mathcal{I} = \{I_1, \cdots, I_K\}$:

$$\mathbf{x}_{k,i} = Aug(\mathbf{x}_i, I_k), \tag{1}$$

where $\mathbf{x}_{k,i} \in \mathbb{R}^d$ corresponds to $\mathbf{x}_i$ augmented with interest type I_k. In essence, CMClusts frames the desired features as consistency information between the augmented and the original data. While the choice of data augmentation techniques is tailored to specific applications. For example, if $\mathcal{I} = \{\text{color}, \text{shape}\}$, CMClusts can take random rotation and color adjustment as augmentation techniques, which can preserve the invariant color features or shape context to align with user interests. Namely, rotation does not alter the color features, while color adjustment maintains the shape context. In this way, we can obtain K augmented representations $\{\mathbf{X}_k\}_{k=1}^K$.

Remarks. Compared to the hard-to-obtain reference labels (reference clusterings), these alike user interests provide valuable clues for generating different and meaningful clusterings. To pursue knowledge-driven non-redundant clusterings, most methods resort to the explicit division rules, such as the link constraints and known clustering labels [Bae and Bailey, 2006; Yang and Zhang, 2017; Ren *et al.*, 2023b]. These methods heavily depend on the quality of the reference information. Unfortunately, the generated clusterings maybe not of any interest to the users. In practice, the user's visual/semantic priors are generally available and more interesting. For example, we can intuitively select the most salient features in images, such as color, identity and shape. Similarly, by glancing conference papers, we can choose the interested clustering criteria (i.e., topics and authorship).

3.2 Clustering-level Contrastive Feature Extraction

After generating augmented data toward user interests, distinct deep neural networks $\{f_{\theta_1}(\cdot), \cdots, f_{\theta_K}(\cdot)\}$ are employed to extract feature representations $\mathbf{Z} \in \mathbb{R}^{N \times d}$ from the augmented samples as follows:

$$\mathbf{z}_{k,i} = f_{\theta_k}(\mathbf{x}_{k,i}). \tag{2}$$

Different deep networks facilitate the tailored design of feature extraction, ensuring adaptability to specific interests. Utilizing the same neural network for extracting diverse features may diminish the model's discriminative capability across different interests. Notably, CMClusts is not tied with any specific network architecture.

Unlike canonical contrastive clustering methods that minimizes the distance between different augmentations of the same instances [Li *et al.*, 2021] (instance-level contrast) or minimizes the overall distance between instances within the same cluster [Zhong *et al.*, 2021] (cluster-level contrast), CMClusts defines a **clustering-level contrast** that aims to maximize the diversity between clusterings of respective augmentations. This contrast can reduce the redundancy between clusterings. To compute the clustering-level contrast, we approximate it via a triplet loss as follows:

$$L_{tri} = \sum_{i}^{N} dist(\mathbf{z}_i, \mathbf{z}_{k,i}) - dist(\mathbf{z}_i, \mathbf{z}_{j,i}) + \alpha, \tag{3}$$

where $\mathbf{z}_i$ is the i-th anchor sample in $\mathbf{X}$, $\mathbf{z}_{j,i}$ ($j \neq k$) is the feature vector of the i-th sample in the j-th augmented data view. Typically, a single augmented view suffices. α is a margin that defines a minimum distance by which the negative sample should be farther from the anchor than the positive sample. The distance between $\mathbf{z}_i$ and $\mathbf{z}_{k,i}$, which can also be computed with other metrics, is computed as:

$$dist(\mathbf{z}_i, \mathbf{z}_{k,i}) = \|\mathbf{z}_i - \mathbf{z}_{k,i}\|_2^2. \tag{4}$$

For the first term of the triplet loss L_{tri}, as described in Section 3.1, the augmentation operations are determined based on the user's interests, ensuring the invariance of the desired features while altering other features. Therefore, by contrasting the original data $\mathbf{X}$ with the augmented data $\mathbf{X}_k$ and minimizing the distance between their feature embeddings, CMClusts learns the invariant features between the two, which correspond to the user's desired features.

For the second term of the triplet loss L_{tri}, increasing the distance between the anchor $\mathbf{z}_i$ and $\mathbf{z}_{j,i}$ in another augmented view increase the diversity between two clusterings. This is driven by the goal to explore different clusterings.

3.3 Clustering Loss Function

The clustering-level contrast mainly focuses on the diversity between different clusterings. CMClusts further designs a clustering loss to pursue the compact and separate relationship within and between clusters of a specific clustering, thereby improving the quality of respective clusterings. Specifically, the clustering loss contains the intra-cluster loss and inter-cluster loss. The intra-cluster loss pursues that samples within clusters are more compact, while the inter-cluster loss ensures that samples between clusters are more dispersed. The intra-cluster loss is computed as:

$$L_{intra} = \frac{1}{|\tilde{\mathcal{C}}_k^l|(|\tilde{\mathcal{C}}_k^l| - 1)} \sum_{i,j \in \tilde{\mathcal{C}}_k^l} \|\mathbf{x}_{k,i} - \mathbf{x}_{k,j}\|_2^2, \tag{5}$$

where $\mathbf{x}_{k,i}$ and $\mathbf{x}_{k,j}$ are samples of the l-th cluster in the k-th clustering. In addition, we ensure the separability between

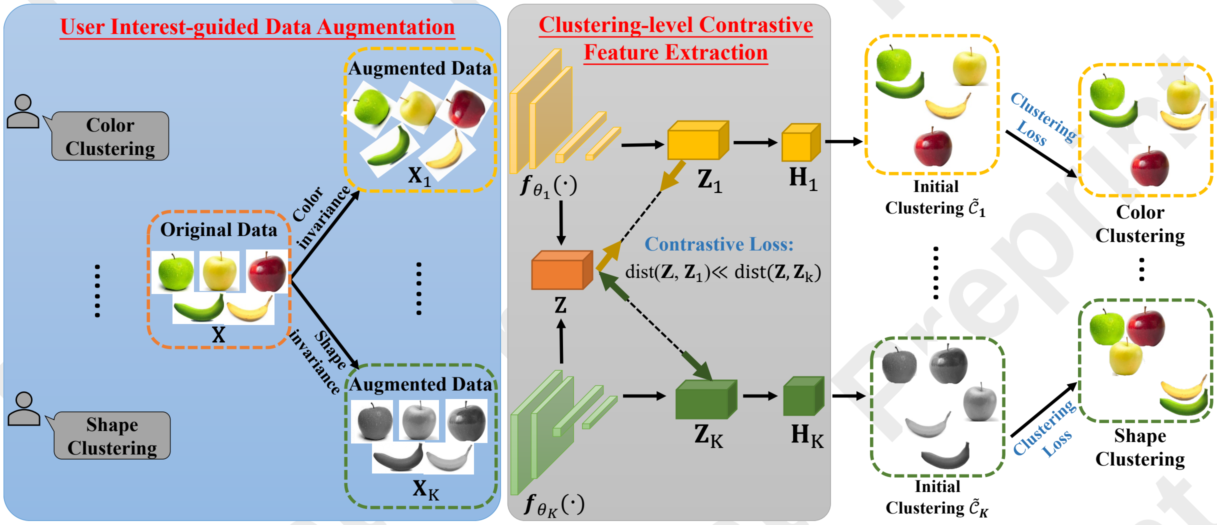


Figure 1: Framework overview of `CMClusts`. The data matrix $\mathbf{X}$ is augmented into K different views $\{\mathbf{X}_k\}_{k=1}^K$ based on user's interests, which specifically refer to the types of clustering or grouping criteria defined by users. Then, the augmented and original data are mapped into the latent feature spaces $\{\mathbf{Z}_k\}_{k=1}^K$ as positive pairs, and the augmented data in the other view are treated as negative ones. `CMClusts` leverages a clustering-level contrastive loss to minimize the distance between positive pairs while maximizing the distance between negative ones, guiding the model to focus on user-interest features. The trained model embeds data into feature spaces $\{\mathbf{H}_k\}_{k=1}^K$ for initial clusterings, which are further optimized by clustering loss to generate multiple clusterings $\{\mathcal{C}_k\}_{k=1}^K$ with diversity and quality.

clusters by increasing the distance between the cluster centers, and define the inter-cluster loss as follows:

$$L_{inter} = -\frac{1}{|\tilde{\mathcal{C}}_k|} \sum_{l_1,l_2 \in \tilde{\mathcal{C}}_k, l_1 \neq l_2} \|\mathbf{c}_{l_1} - \mathbf{c}_{l_2}\|_2^2, \tag{6}$$

where $\mathbf{c}_{l_1}$ and $\mathbf{c}_{l_2}$ are centers of cluster $\tilde{\mathcal{C}}_k^{l_1}$ and $\tilde{\mathcal{C}}_k^{l_2}$, respectively.

The total clustering loss incorporates both intra-cluster and inter-cluster losses, with a trade-off parameter β balancing these two losses as follows:

$$L_{clu} = L_{intra} + \beta \cdot L_{inter}. \tag{7}$$

Minimizing the intra-cluster variance and maximizing the inter-cluster separation can promote the formation of well-structured and semantically meaningful cluster representations, as our experiment will show.

3.4 Overall Objective

To learn the specific features of interest to the user across different augmented data views and generate high quality, diversified clusterings, we can jointly optimize `CMClusts` by combining the aforementioned losses as follows:

$$L_{\texttt{CMClusts}} = L_{tri} + L_{clu}. \tag{8}$$

However, the computation of clustering loss is dependent on the initial clustering labels. Therefore, we adopt a tailored alternative optimization strategy. Initially, `CMClusts` is trained using the triplet loss L_{tri} to extract the desired clustering features. Then, k-means is employed to generate the initial clustering labels. After this, the model parameters are frozen, and the clustering center positions are adjusted by the clustering loss L_{clu} to obtain the final clustering labels.

Algorithm 1 lists the procedure of `CMClusts`. Specifically, Line 1 initializes the augmented data views $\{\mathbf{X}\}_{k=1}^K$, Lines 2-6 iteratively optimize the parameters of θ_k and extract different sets of desired features $\{\mathbf{H}_k\}_{k=1}^K$ to align with user interests.It is worth noting that the contrastive loss is computed across clusterings, but only the feature extractor of the current clustering is updated; others remain fixed. Lines 9-11 optimize the clustering loss to generate K clusterings.

The time complexity of data augmentation is $O(wh)$, where w and h are the width and height of the image, respectively. The time complexity of an m-layer feature extraction network is $O(\sum_{i=1}^m d_{i-1}d_i)$, where d_i is the number of neurons in the i-th layer, $d_0 = D$ and $d_m = d$. The time complexity of the triplet loss L_{tri} and clustering loss L_{clu} are both $O(Nd)$. Therefore, the overall time complexity of `CMClusts` is $O(\sum_{i=1}^m d_{i-1}d_i + 2Nd + wh)$. Compared with Multi-Sub [Yao *et al.*, 2024a] using the CLIP pre-trained model, `CMClusts` does not require an additional text encoder and thus has a lower complexity and runtime.

4 Experiments

4.1 Experimental setup

Baselines. We compare `CMClusts` against representative multiple clustering methods, including OSC [Cui *et al.*, 2007], MNMF [Yang and Zhang, 2017], ENRC [Miklautz *et al.*, 2020], iMClusts [Ren *et al.*, 2023b], AugDMC [Yao *et al.*, 2023], and Multi-Sub [Yao *et al.*, 2024a]. These compared methods were discussed in Section 2, and their configurations are deferred into the supplementary file.

Datasets. Seven benchmark datasets (ALOI [Geusebroek *et al.*, 2005], Fruit [Yao *et al.*, 2023], CMUFace [Ren *et al.*, 2023b], COIL [Nayar, 1996], Cards [Yao *et al.*, 2023], We-

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

Algorithm 1 `CMClusts`: Contrastive Multiple Clustering

Input: Dataset $\mathbf{X}$; the number of clusterings K; the user's interest set $\mathcal{I} = \{I_k\}_{k=1}^{K}$; the training rounds $\{\tau_k\}_{k=1}^{K}$ for K clusterings;
Output: Clustering partition: $\{\mathcal{C}_k\}_{k=1}^{K}$.

Initialize augmented data $\{\mathbf{X}_k\}_{k=1}^{K}$ based on $\mathcal{I}$.
for $k = 1 \rightarrow K$ **do**
  for $\tau = 1 \rightarrow \tau_k$ **do**
    Update the latent representations $\mathbf{Z}_k$ by Eq. (2).
  end for
  Compute triplet loss L_{tri} by Eq. (3).
  Update the parameters of feature extraction network with Adam optimizer.
  Extract the desired feature $\mathbf{H}_k$: $\mathbf{H}_k = f_{\theta_k}(\mathbf{X}_k)$.
  Generate the k-th initial clustering $\tilde{\mathcal{C}}_k = kmeans(\mathbf{H}_k)$.
  Compute the intra-cluster loss L_{intra} via Eq. (5).
  Compute the inter-cluster loss L_{inter} via Eq. (6).
  Update the cluster centers via Eq. (7).
end for
Generate K clusterings $\{\mathcal{C}_k\}_{k=1}^{K}$.

bKB and Mice [Ren *et al.*, 2023b]) are used to evaluate the performance of `CMClusts` and other baselines. The first five are image datasets, the sixth is a textual dataset, and the last is a mouse single-oocyte transcriptome dataset with two different clusterings (2 or 4 clusters). These datasets have been widely used to validate multiple clustering methods [Bailey, 2018; Yu *et al.*, 2024]. Their statistical information are given in Table 1. More details are provided in the supplementary file. In experiments, we implement data augmentation by enriching features that are relevant to the desired clustering based on user's visual priors or interests. For example, we augment data with respect to color by varying shapes (i.e. rotation, flipping, random cropping) and ensure color invariance. For the expected shape clustering, we can augment data with respect to shape by varying color (i.e. hue, saturation, contrast changes) and keep the shape unchanged.

Evaluation metrics. To quantitatively evaluate the performance of each method, we measure the quality and diversity using the Normalized Mutual Information (NMI) and Jaccard Index (JI) with reference to distinct ground-truth labels. A higher score indicates the generated clustering more consistent with the groundtruth, which is with higher quality and more diverse from the other clustering.

Implementation. For image data, `CMClusts` adopts the pre-trained ResNet50 [He *et al.*, 2016] as the backbone network for feature extraction, and defines a multilayer perceptron (MLP) with rectified linear units (ReLU) as the activate functions to project the learned features into a nonlinear representation. For text data, `CMClusts` integrates traditional text feature extraction techniques, such as the Bag-of-Words model and TF-IDF, to encode textual features, and then employs a multi-layer perception network (MLP) as the backbone for further feature extraction and representation learning. For single-cell Mice data, `CMClusts` obtains its feature matrix and maps it to a lower-dimensional space using an MLP. And the α and β are hyperparameters. For a detailed analysis, please refer to the supplementary file. Additionally, all the methods are implemented in PyTorch 2.4 and tested on a server with NVIDIA L40 GPUs. The source code of `CMClusts` is available at https://www.sdu-idea.cn/codes.php?name=CMClusts.

Datasets	Samples (N)	Dimension (d)	Clusters
ALOI	288	32×32	2; 2
COIL	648	32×32	3; 3
Fruit	4856	32×32	4; 4
CMUFace	640	32×32	20; 4; 2
Cards	8029	32×32	13; 4
WebKB	1041	500	4; 4
Mice	146	41092	4; 2

Table 1: Information of Used Benchmark Datasets

4.2 Results and analysis

Table 2 provides the average clustering results of 10 independent runs of each method on benchmark datasets, with the significantly best results are highlighted in bold (pairwise t-test with 95% confidence). We can find that `CMClusts` achieves the best results in almost all cases, proving its effectiveness. In addition, we have the following important observations:
(i) **Deep vs. Shallow methods:** In most cases, deep multiple clustering methods, ENRC, iMClusts, AugDMC, Multi-Sub and `CMClusts`, perform better than the shallow OSC and MNMF. This clear gap confirms the advantage of deep learning based methods, owing to the expressive representation capability. The performance of `CMClusts` surpasses that of other methods due to its capability to effectively incorporate user interests, encoding these interests through diverse data augmentations. By leveraging contrastive learning among augmented features, the original features, and other irrelevant augmented features, `CMClusts` enhances the salience of desired features while promoting diversity across different augmented features. This capability is critical for generating diverse and high-quality clusterings. ENRC and iMClusts obtain competitive results in simple task like ALOI, but struggle with more complex tasks, which may be due to their inability to generate multiple orthogonal or non-redundant subspaces from large data with complex structures. AugDMC outperforms ENRC and iMClusts in complex scenarios such as Fruit and CMUFace, because the discriminative data view obtained by data augmentation is beneficial for generating multiple clusterings. However, AugDMC's mediocre performance on other datasets indicates the limitation of its simple adoption of prototype representation learning. Multi-Sub demonstrates some improvements over the aforementioned methods, reflecting the strong representational power of large multimodal models in capturing clustering structures. Nevertheless, it cannot cluster text data and transcriptomic datasets, and clustering is also challenging for data that is hard to generate textual cues (e.g., identity-based clustering in CMUFace). `CMClusts` gains competitive performance on text dataset WebKB and transcriptomic dataset Mice, highlighting the versatility of its data augmentation and clustering-level

Dataset	Type	Metrics	OSC	MNMF	ENRC	iMClusts	AugDMC	Multi-Sub	CMClusts
ALOI	color	NMI	.344±.000●	.281±.006●	.982±.001●	.816±.001●	.344±.017●	**1.000±.000**	**1.000±.000**
		JI	.497±.000●	.361±.009●	.965±.000●	.873±.000●	.497±.105●	**1.000±.000**	**1.000±.000**
	shape	NMI	.344±.000●	.263±.008●	.992±.000○	.928±.000●	**1.000±.000**	**1.000±.000**	**1.000±.000**
		JI	.497±.000●	.362±.210●	.994±.000○	.937±.001●	**1.000±.000**	**1.000±.000**	**1.000±.000**
COIL	color	NMI	.022±.002●	.013±.001●	.084±.012●	.183±.001●	.094±.001●	.155±.001●	**.205±.000**
		JI	.267±.001●	.215±.002●	.232±.000●	.284±.001●	.276±.000●	.213±.002●	**.297±.000**
	species	NMI	.102±.001●	.055±.003●	.125±.013●	.126±.002●	.065±.003●	.251±.000●	**.269±.000**
		JI	.273±.001●	.225±.001●	.263±.004●	.273±.004●	.257±.002●	.352±.001●	**.382±.000**
Fruit	color	NMI	.184±.021●	.019±.112●	.402±.003●	.421±.002●	.466±.021●	.660±.001●	**.696±.000**
		JI	.206±.001●	.152±.022●	.301±.032●	.315±.041●	.304±.021●	.461±.001●	**.498±.000**
	species	NMI	.465±.031●	.018±.010●	.380±.110●	.410±.002●	.517±.101●	.610±.003●	**.619±.000**
		JI	.343±.001●	.163±.110●	.297±.131●	.311±.032●	.352±.071●	.401±.000●	**.408±.000**
CMUFace	identity	NMI	.495±.021●	.228±.010●	.504±.001●	.527±.004●	.512±.001●	.543±.007●	**.582±.000**
		JI	.192±.012●	.051±.023●	.164±.011●	.197±.002●	.170±.003●	.201±.005●	**.214±.000**
	pose	NMI	.013±.011●	.017±.022○	.023±.207○	.026±.004○	.015±.001●	**.032±.006**○	.017±.000
		JI	.160±.013●	.141±.112●	.153±.050●	.170±.001○	.156±.002●	**.201±.001**○	.172±.000
	glass	NMI	.006±.001●	.007±.002●	.007±.001●	.006±.010●	.005±.002●	.008±.001●	**.009±.000**
		JI	.360±.002●	.401±.001●	.362±.000●	.360±.001●	.359±.002●	.352±.001●	**.433±.000**
Cards	number	NMI	.133±.001●	.052±.031●	.100±.001●	.124±.003●	.106±.001●	.153±.000●	**.175±.000**
		JI	.109±.001●	.071±.010●	.092±.002●	.103±.001●	.082±.002●	.121±.000●	**.154±.000**
	suits	NMI	.019±.000●	.032±.010●	.090±.000●	.100±.001●	.081±.000●	.170±.001●	**.213±.000**
		JI	.170±.002●	.211±.010●	.140±.001●	.205±.001●	.182±.000●	.191±.002●	**.234±.000**
WebKB	university	NMI	.282±.003●	.012±.001●	.217±.000●	.451±.001●	.383±.010●	-	**.460±.000**
		JI	.290±.000●	.180±.011●	.221±.002●	.352±.001●	.331±.005●	-	**.359±.000**
	category	NMI	.201±.101●	.011±.010●	.133±.001●	.106±.002●	.155±.004●	-	**.242±.000**
		JI	.221±.026●	.142±.014●	.150±.002●	.137±.002●	.214±.000●	-	**.283±.000**
Mice	stage	NMI	.254±.000●	.018±.000●	.107±.002●	.123±.000●	.152±.001●	-	**.363±.000**
		JI	.232±.000●	.145±.000●	.190±.001●	.203±.000●	.218±.001●	-	**.267±.000**
	type	NMI	.768±.000●	.875±.000●	.695±.001●	.726±.000●	.703±.000●	-	**.909±.000**
		JI	.831±.000●	.921±.000●	.667±.001●	.676±.000●	.650±.001●	-	**.947±.000**

Table 2: Performance of baselines on generating multiple clusterings. ● or ○ indicates whether `CMClusts` is superior/inferior to the other method, with statistical significance checked by pairwise t-test at 95% level. The best results are highlighted in **bold** font.

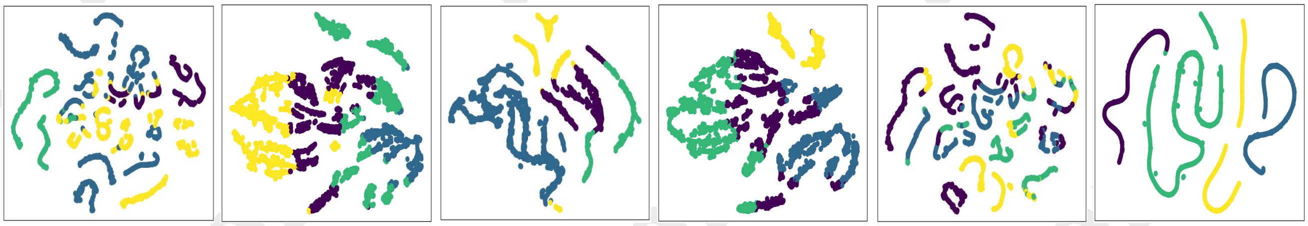
contrast.
(ii) **With interest vs. Without interest:** `CMClusts` outperforms other competitive baselines in most cases, which suggests the effectiveness of users' interest on boosting multiple clusterings. OSC and MNMF obtain the alternative clustering with reference to already explored clusterings, they have a poor overall quality and diversity. iMClusts rectifies each feature subspace by incorporating weakly-supervised prior knowledge, but these knowledge maybe not always accurate and available in complex data. Although ENRC projects data into different subspaces and provides alternative clusterings without requiring prior knowledge, the results are not good as `CMClusts`, due to the lack of redundancy control. Multi-Sub aligns image and textual features through a large multi-modal model, but it requires additional textual prompts, which may lead to suboptimal performance on datasets with a wide variety of categories and complex descriptions. In contrast, `CMClusts` leverages interest-guided data augmentations and clustering-level contrast to extract features aligned with user interests, without requiring additional textual prompts.

4.3 Visualization of Multiple Clusterings

To intuitively check the quality of clustering generated by `CMClusts`, we employ t-SNE [Van der Maaten and Hinton, 2008] to visualize the clustering space learnt by `CMClusts` and five baselines (excluding MNMF due to its poor performance) on the Fruit dataset as a two-dimensional scatter plot. The color clustering results are shown in Figure 2, and the species clustering results are provided in the supplementary, where different colors represent distinct cluster labels. We can observe that the cluster distributions produced by the baselines are more compact with small inter-cluster distances, making it difficult to obtain well-separated and distinguishable structures. In contrast, the clustering results of `CMClusts` exhibit clear inter-cluster boundaries and tightly distributed intra-cluster points, confirming that `CMClusts` can effectively generate higher-quality clusterings.

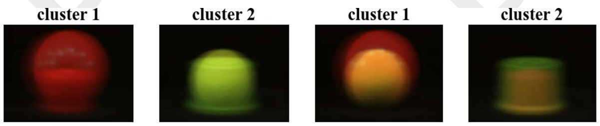
In addition, to vividly show the multiple clusterings generated by `CMClusts` and check the alignment between user-interests and the clustering outcomes, we present multiple clusterings guided by user-interests on the ALOI and CMU-Face datasets in Fig. 3∼Fig. 4. Each clustering outcome corresponds to a specific type of clustering expected by the user. In Figure 3, the first clustering reflects the outcome under the user's color preference, where the color differences between clusters are distinct, while the shapes are mixed. The second clustering coincides with the user's shape preference. Here, the colors are more chaotic, but the shapes are clearly distinct. Similarly, as shown in Figure 4, `CMClusts` effectively identifies three clusterings based on identity, pose, and glass or not, corresponding to different user-defined criteria.

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.



(a) OSC (b) ENRC (c) iMClusts (d) AugDMC (e) Multi-Sub (f) `CMClusts`

Figure 2: 2D scatter plot of the generated color clustering on Fruit by `CMClusts` and baselines. `CMClusts` better groups colorful images.



(a) Color (b) Shape

Figure 3: Two clusterings (color, shape) found by `CMClusts` on ALOI.

(a) Identity (b) Pose (c) Glass

Figure 4: Three clusterings (identity, pose, glass or not) found by `CMClusts` on CMUFace.

4.4 Ablation Study

To investigate the contributing factors of `CMClusts`, we introduce four variants: w/oAug, w/oContrast, w/oClusteringLoss, and wJointOpt, which respectively remove data augmentation in Eq. (1), contrastive learning in Eq. (3), clustering loss in Eq. (7), and the separate optimization strategy for the feature extraction networks in Eq. (8). Figure 5 presents the average NMI values of `CMClusts` and its variants. Similar trends are observed in JI, whose results are provided in the supplementary file. We find that `CMClusts` outperforms its variants, demonstrating that data augmentation, contrastive learning, clustering loss, and the separate optimization strategy are indispensable for generating high-quality clusterings with diversity.

The NMI of w/oContrast are generally the lowest, confirming the critical role of clustering-level contrast in capturing discriminative latent features from augmented data views. Solely the augmented data and feature extraction networks fail to produce satisfactory clustering results. Additionally, the performance of wJointOpt is noticeably inferior to `CMClusts`, highlighting that, compared to jointly optimizing the triplet loss and clustering loss, optimizing the network with the triplet loss to obtain effective initial clustering labels and then fine-tuning with the clustering loss gives better clustering performance.

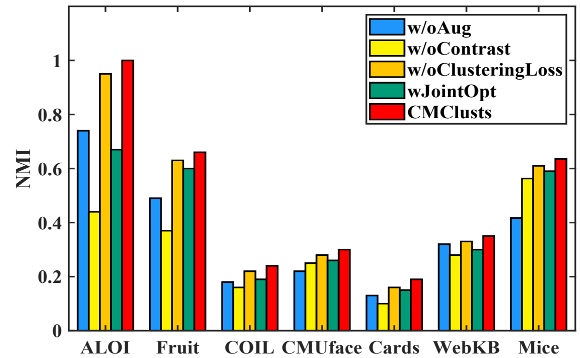


Figure 5: NMI of `CMClusts` and its variants.

We also observe that removing data augmentation leads to a significant drop in NMI. This is because data augmentation generates diverse enhanced views from complex data, reflecting user's respective interests. Without data augmentation, only using original data considerably reduces the diversity of clustering results. Furthermore, the clustering loss also plays an important role in fine-tuning the clustering outcome with moderate improvement. In summary, these factors enable `CMClusts` to achieve high-quality multiple clusterings.

In addition, we investigate the hyperparameter sensitivity of `CMClusts` and provide the results in the supplementary file. The results indicate that the performance of `CMClusts` is not greatly impacted by hyperparameters, and the fluctuation range of NMI is less than 0.1.

5 Conclusion and Future Work

This paper introduces `CMClusts`, a novel contrastive multiple clustering solution aligning with user interests. `CMClusts` fulfils user interests by extracting desired features from augmented data, and explores multiple clusterings with diversity by clustering-level contrast. `CMClusts` outperforms competitive baselines in both quality and diversity on benchmark datasets. The explored distinctive clusterings indeed accord with different user interests. Next, we will integrate foundational models with multiple clusterings to enhance the applicability of multiple clustering across broader domains and validate its generalizability in real-world tasks.

Acknowledgments

This work is supported by National Key Research and Development Program of China (No. 2023YFF0725500), NSFC (62272276 and 62432006), Shandong Provincial Natural Science Foundation (No. ZR2024JQ001), Taishan Scholars Program (No. tsqn202306007 and tsqn202408317).

Contribution Statement

Shan Zhang and Liangrui Ren contributed equally to this work and are designated as co-first authors (indicated by †). Guoxian Yu, identified as the corresponding author (denoted by ∗), provided overall conceptual guidance, project supervision, and coordination. Jun Wang, Yanyu Xu, and Carlotta Domeniconi made substantial contributions to the study design, experimental implementation, and manuscript refinement.

References

[Bae and Bailey, 2006] Eric Bae and James Bailey. Coala: A novel approach for the extraction of an alternate clustering of high quality and high dissimilarity. In *ICDM*, pages 53–62, 2006.

[Bailey, 2018] James Bailey. Alternative clustering analysis: A review. *Data Clustering*, pages 535–550, 2018.

[Chang *et al.*, 2017] Yale Chang, Junxiang Chen, Michael H Cho, Peter J Castaldi, Edwin K Silverman, and Jennifer G Dy. Multiple clustering views from multiple uncertain experts. In *ICML*, pages 674–683, 2017.

[Chen *et al.*, 2020] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, pages 1597–1607, 2020.

[Cui *et al.*, 2007] Ying Cui, Xiaoli Z Fern, and Jennifer G Dy. Non-redundant multi-view clustering via orthogonalization. In *ICDM*, pages 133–142, 2007.

[Dang and Bailey, 2010] Xuanhong Dang and James Bailey. Generation of alternative clusterings using the cami approach. In *ICDM*, pages 118–129, 2010.

[Elhamifar and Vidal, 2013] Ehsan Elhamifar and René Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *TPAMI*, 35(11):2765–2781, 2013.

[Geusebroek *et al.*, 2005] Jan-Mark Geusebroek, Gertjan J Burghouts, and Arnold WM Smeulders. The amsterdam library of object images. *IJCV*, 61:103–112, 2005.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

[Hu *et al.*, 2017] Juhua Hu, Qi Qian, Jian Pei, Rong Jin, and Shenghuo Zhu. Finding multiple stable clusterings. *KAIS*, 51:991–1021, 2017.

[Jain *et al.*, 1999] Anil K Jain, M Narasimha Murty, and Patrick J Flynn. Data clustering: a review. *ACM CSUR*, 31(3):264–323, 1999.

[Jain *et al.*, 2008] Prateek Jain, Raghu Meka, and Inderjit S Dhillon. Simultaneous unsupervised learning of disparate clusterings. *SADM*, 1(3):195–210, 2008.

[Li *et al.*, 2021] Yunfan Li, Peng Hu, Zitao Liu, Dezhong Peng, Joey Tianyi Zhou, and Xi Peng. Contrastive clustering. In *AAAI*, pages 8547–8555, 2021.

[Mautz *et al.*, 2020] Dominik Mautz, Wei Ye, Claudia Plant, and Christian Böhm. Non-redundant subspace clusterings with nr-kmeans and nr-dipmeans. *ACM TKDD*, 14(5):1–24, 2020.

[Miklautz *et al.*, 2020] Lukas Miklautz, Dominik Mautz, Muzaffer Can Altinigneli, Christian Böhm, and Claudia Plant. Deep embedded non-redundant clustering. In *AAAI*, pages 5174–5181, 2020.

[Nayar, 1996] S Nayar. Columbia object image library (coil100). *http://www1. cs. columbia. edu/CAVE/software/softlib/coil-100. php*, 1996.

[Niu *et al.*, 2013] Donglin Niu, Jennifer G Dy, and Michael I Jordan. Iterative discovery of multiple alternativeclustering views. *TPAMI*, 36(7):1340–1353, 2013.

[Radford *et al.*, 2021] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *ICML*, pages 8748–8763, 2021.

[Ren *et al.*, 2023a] Liangrui Ren, Jun Wang, Zhao Li, Qingzhong Li, and Guoxian Yu. scmcs: a framework for single-cell multi-omics data integration and multiple clusterings. *Bioinformatics*, 39(4):btad133, 2023.

[Ren *et al.*, 2023b] Liangrui Ren, Guoxian Yu, Jun Wang, Lei Liu, Carlotta Domeniconi, and Xiangliang Zhang. A diversified attention model for interpretable multiple clusterings. *TKDE*, 35(9):8852–8864, 2023.

[Tokuda *et al.*, 2021] Tomoki Tokuda, Okito Yamashita, and Junichiro Yoshimoto. Multiple clustering for identifying subject clusters and brain sub-networks using functional connectivity matrices without vectorization. *Neural Net.*, 142:269–287, 2021.

[Tsai *et al.*, 2020] Tsung Wei Tsai, Chongxuan Li, and Jun Zhu. Mice: Mixture of contrastive experts for unsupervised image clustering. In *ICLR*, 2020.

[Van der Maaten and Hinton, 2008] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *JMLR*, 9(11):1–27, 2008.

[Wang *et al.*, 2019] Xing Wang, Jun Wang, Carlotta Domeniconi, Guoxian Yu, Guoqiang Xiao, and Maozu Guo. Multiple independent subspace clusterings. In *AAAI*, pages 5353–5360, 2019.

[Wei *et al.*, 2020a] Shaowei Wei, Jun Wang, Guoxian Yu, Carlotta Domeniconi, and Xiangliang Zhang. Deep incomplete multi-view multiple clusterings. In *ICDM*, pages 651–660, 2020.

[Wei *et al.*, 2020b] Shaowei Wei, Jun Wang, Guoxian Yu, Carlotta Domeniconi, and Xiangliang Zhang. Multi-view multiple clusterings using deep matrix factorization. In *AAAI*, pages 6348–6355, 2020.

[Wei *et al.*, 2021] Shaowei Wei, Guoxian Yu, Jun Wang, Carlotta Domeniconi, and Xiangliang Zhang. Multiple clusterings of heterogeneous information networks. *Mach. Learn.*, 110(6):1505–1526, 2021.

[Xu *et al.*, 2022] Jie Xu, Huayi Tang, Yazhou Ren, Liang Peng, Xiaofeng Zhu, and Lifang He. Multi-level feature learning for contrastive multi-view clustering. In *CVPR*, pages 16051–16060, 2022.

[Yang and Zhang, 2017] Sen Yang and Lijun Zhang. Non-redundant multiple clustering by nonnegative matrix factorization. *Mach. Learn.*, 106(5):695–712, 2017.

[Yao and Hu, 2024] Jiawei Yao and Juhua Hu. Dual-disentangled deep multiple clustering. In *SDM*, pages 679–687, 2024.

[Yao *et al.*, 2019] Shixin Yao, Guoxian Yu, Jun Wang, Carlotta Domeniconi, and Xiangliang Zhang. Multi-view multiple clustering. In *IJCAI*, pages 4121–4127, 2019.

[Yao *et al.*, 2023] Jiawei Yao, Enbei Liu, Maham Rashid, and Juhua Hu. Augdmc: Data augmentation guided deep multiple clustering. *Proc. Comp. Sci.*, 222:571–580, 2023.

[Yao *et al.*, 2024a] Jiawei Yao, Qi Qian, and Juhua Hu. Customized multiple clustering via multi-modal subspace proxy learning. In *NeurIPS*, 2024.

[Yao *et al.*, 2024b] Jiawei Yao, Qi Qian, and Juhua Hu. Multi-modal proxy learning towards personalized visual multiple clustering. In *CVPR*, pages 14066–14075, 2024.

[Ye *et al.*, 2016] Wei Ye, Samuel Maurus, Nina Hubig, and Claudia Plant. Generalized independent subspace clustering. In *ICDM*, pages 569–578, 2016.

[Yin *et al.*, 2023] Jun Yin, Haowei Wu, and Shiliang Sun. Effective sample pairs based contrastive learning for clustering. *Inf. Fusion*, 99:101899, 2023.

[Yu *et al.*, 2024] Guoxian Yu, Liangrui Ren, Jun Wang, Carlotta Domeniconi, and Xiangliang Zhang. Multiple clusterings: Recent advances and perspectives. *Comp. Sci. Rev.*, 52:100621, 2024.

[Zhang *et al.*, 2020] Changqing Zhang, Huazhu Fu, Qinghua Hu, Xiaochun Cao, Yuan Xie, Dacheng Tao, and Dong Xu. Generalized latent multi-view subspace clustering. *TPAMI*, 42(1):86–99, 2020.

[Zhong *et al.*, 2021] Huasong Zhong, Jianlong Wu, Chong Chen, Jianqiang Huang, Minghua Deng, Liqiang Nie, Zhouchen Lin, and Xian-Sheng Hua. Graph contrastive clustering. In *CVPR*, pages 9224–9233, 2021.

[Zhu *et al.*, 2022] Pengfei Zhu, Jialu Li, Yu Wang, Bin Xiao, Shuai Zhao, and Qinghua Hu. Collaborative decision-reinforced self-supervision for attributed graph clustering. *TNNLS*, 34(12):10851–10863, 2022.