Multi-view Granular-ball Contrastive Clustering

Peng Su[1,2], Shudong Huang[1,2*], Weihong Ma[3], Deng Xiong[4], Jiancheng Lv[1,2]

[1]College of Computer Science, Sichuan University, Chengdu 610065, China
[2]Engineering Research Center of Machine Learning and Industry Intelligence, Ministry of Education, China
[3]Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China
[4]Stevens Institute of Technology, 1 Castle Point Terrace, Hoboken, NJ 07030, USA
supeng@stu.scu.edu.cn, huangsd@scu.edu.cn, mawh@nercita.org.cn, dxiong@stevens.edu, lvjiancheng@scu.edu.cn

Abstract

Previous multi-view contrastive learning methods typically operate at two scales: instance-level and cluster-level. Instance-level approaches construct positive and negative pairs based on sample correspondences, aiming to bring positive pairs closer and push negative pairs further apart in the latent space. Cluster-level methods focus on calculating cluster assignments for samples under each view and maximize view consensus by reducing distribution discrepancies, e.g., minimizing KL divergence or maximizing mutual information. However, these two types of methods either introduce false negatives, leading to reduced model discriminability, or overlook local structures and cannot measure relationships between clusters across views explicitly. To this end, we propose a method named Multi-view Granular-ball Contrastive Clustering (MGBCC). MGBCC segments the sample set into coarse-grained granular balls, and establishes associations between intra-view and cross-view granular balls. These associations are reinforced in a shared latent space, thereby achieving multi-granularity contrastive learning. Granular balls lie between instances and clusters, naturally preserving the local topological structure of the sample set. We conduct extensive experiments to validate the effectiveness of the proposed method.

Code — https://github.com/Duo-laimi/mgbcc_main

Introduction

Multi-view data refers to data collected from different sensors or obtained by different feature extractors, often exhibiting heterogeneity (Fang et al. 2023). For example, a web page typically contains images, text, and videos, each of which can be considered a view, reflecting the same sample from different perspectives. Multi-view clustering has received continuous attention in recent years, aiming to partition multi-view data into different clusters in an unsupervised manner (Huang et al. 2019, 2022; Liu 2023; Deng et al. 2024). The key challenge in multi-view clustering is balancing the consistency and diversity between different views to learn the most comprehensive shared representation. Traditional multi-view learning methods mainly include subspace learning, graph learning, and multi-kernel

learning. These methods often involve matrix decomposition and fusion, resulting in high computational complexity, making them difficult to apply to large-scale datasets, which severely hinders their practical application.

In recent years, deep-based multi-view learning methods have gained significant attention due to their excellent representation capabilities. These methods extend deep single-view clustering and typically select appropriate feature extractors based on the properties of the views. DCCA (Andrew et al. 2013) projects data from two views into a common space using deep neural networks, where the representations of the two views are highly linearly correlated, making it a nonlinear extension of canonical correlation analysis. PARTY (Peng et al. 2016) is a deep subspace clustering method with sparse priors that projects input data into a latent space, maintaining local structure by minimizing reconstruction loss while introducing sparse prior information into the latent representation learning to preserve sparse reconstruction relationships across the entire dataset. DIMC (Wen et al. 2020) extracts high-level features of multiple views through view-specific autoencoders and introduces fusion graph-based constraints to preserve the local geometric structure of the data. DMCE (Zhao, Yang, and Nie 2023) applies ensemble clustering to fuse similarity graphs from different views, using graph autoencoders to learn a common spectral embedding. These methods combine deep learning with traditional multi-view learning ideas by introducing constraints from traditional methods such as neighbor graph constraints or self-expression constraints into the latent space projected by deep modules. This allows the models to learn concise but comprehensive representations that maximally preserve the structural information of the input data.

Multi-view contrastive learning is another important branch of deep multi-view clustering. It can generally be divided into two categories: instance-level contrastive learning and cluster-level contrastive learning. The basic idea of the former is that instances of the same sample from different views should be as close as possible in the latent space, typically forming positive pairs. In these methods (Liu et al. 2023; Su et al. 2024; Yang et al. 2023b; Xu et al. 2023a), the construction and handling of negative pairs is a key focus due to the unknown instance labels in unsupervised paradigms. Improper negative pairs can degrade the model's

---

*Corresponding author

discriminative ability. Cluster-level methods (Xu et al. 2022; Jin et al. 2023; Chen et al. 2023), on the other hand, align the clustering assignments of different views. They typically establish one-to-one correspondences between clusters across views and aim to make the distributions of corresponding clusters as consistent as possible. However, clusters are macro structures and do not effectively utilize the local structural information within views.

We propose a multi-granularity multi-view learning method. This method models the local structure of the sample set using granular balls and establishes intra-view and inter-view granular-ball connections based on overlap and intersection size, respectively. By bringing connected granular balls closer in the latent space, our model learns highly discriminative features. To the best of our knowledge, this is the pioneering work utilizing granular-ball methodology for multiview contrastive learning. Specifically, our contributions are summarized as follows:

- We propose a novel deep multi-view clustering method that performs contrastive learning at the granular-ball level. This method avoids directly using neighboring samples to construct negative pairs while preserving the local structural information of the sample set, addressing the shortcomings of instance-level and cluster-level methods.

- We introduce a simple yet effective granular-ball construction method. Unlike classical methods that continuously bisect the dataset until reaching the smallest granularity, our method directly partitions the sample set into multiple granular balls based on the granularity parameter, avoiding the drawback of non-adjacent samples being grouped into the same granular ball in boundary regions.

- Extensive experiments on seven typical multi-view datasets demonstrate that our method achieves comparable or superior performance compared to state-of-the-art methods.

Related Work

In this section, we briefly review the latest advancements in related topics, including multi-view contrastive learning and granular-ball computing.

Multi-view Contrastive Learning

Contrastive learning (He et al. 2020; Chen et al. 2020; Zhang and Wang 2024) aims to learn a feature space with good discriminative properties, where positive pairs are pulled closer together, and negative pairs are pushed further apart. In a single-view setting, positive pairs are typically constructed through augmentations of the same sample, while negative pairs come from other samples within the same batch or dynamically constructed feature queues. This idea naturally extends to multi-view learning, as multi-view instances can be seen as natural augmentations of a sample: they are unique yet collectively describe the same sample, giving rise to multi-view contrastive learning.

Completer (Lin et al. 2021) uses conditional entropy and mutual information to measure the differences and consensus between different views. By maximizing the mutual in-

formation between views, they aims to learn rich and consistent representations. To resolve the conflict between maintaining multi-view semantic consistency and the reconstruction process that tends to preserve view-specific information, MFLVC (Xu et al. 2022) proposes a multi-level feature multi-view contrastive framework. This model learns low-level features, high-level features, and semantic labels from raw features in a fusion-free manner. Reconstruction is performed on low-level features, while consensus is explored through contrastive learning on high-level features. SURE (Yang et al. 2023a) addresses the issue of false negatives in multi-view contrastive learning, where two instances used to construct a negative pair might actually belong to the same cluster. It divides negative pairs into three intervals based on distance, treating those with distances below a certain threshold as potential positives for optimization.

Granular-ball Computing

Expanding on the theoretical foundations of traditional granularity computation and integrating the human cognitive mechanism of 'macro-first' (Chen 1982), Wang (Wang 2017) introduced the innovative concept of multi-granularity cognitive computation. Building on Wang's framework, Xia (Xia et al. 2019) developed an efficient, robust, and interpretable computational method known as granular-ball computing. Unlike traditional methods that process data at the most granular level of individual points, granular-ball computing encapsulates and represents data using granular-balls, thereby enhancing efficiency and robustness. Notable applications of granular-ball computing include granular-ball clustering (Cheng et al. 2024; Xie et al. 2024a,b,c), granular-ball classifiers (Xia et al. 2024b; Quadir and Tanveer 2024), granular-ball sampling methods (Xia et al. 2023b), granular-ball rough sets (Xia et al. 2022, 2023a; Zhang et al. 2023), granular-ball three-way decisions (Yang et al. 2024; Xia et al. 2024a), and advancements such as granular-ball reinforcement learning (Liu et al. 2024).
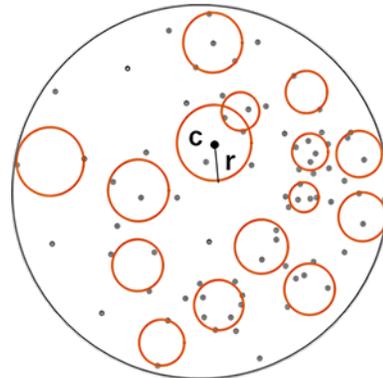

Figure 1: Examples of granular balls

Given a dataset $\{x_i\}_{i=1}^n$, let $\{GB_i\}_{i=1}^k$ denote the set of granular balls generated based on it, where k represents the total number of balls. As illustrated in Figure 1, one ball contains multiple neighboring samples or feature points, e.g., $GB_i = \{x_j\}_{j=1}^{n_i}$, which essentially reflects the local topo-

logical relationships among samples. The center c_i and the radius r_i of GB_i are defined as

$$c_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_j, \quad r_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \|c_i - x_i\|_2. \quad (1)$$

In granular-ball computation, the key lies in how to generate the granular-ball set, which involves two critical steps: partitioning and merging. Partitioning refers to the recursive process of dividing a large ball into two smaller ones. Initially, the entire dataset is initialized as a single granular ball. Granular balls that meet the split conditions will continue to split. The split conditions typically vary depending on the task. In clustering tasks, if the average radius of the original ball is greater than the weighted average radius of the two sub-ball combined, the ball will split. Otherwise, it will stop. This condition can lead to over-partitioning, such as having one ball per sample. To prevent this, a minimum capacity threshold η is introduced. If the number of samples in a ball is less than η, it will also stop splitting.

Merging refers to the process of combining two significantly overlapping ball into a single ball and recalculating the ball center and radius. Two balls are considered overlapping if they satisfy following conditions

$$\|c_i - c_j\|_2 - (r_i + r_j) < \omega, \omega = \frac{min(r_i, r_j)}{min(p_i, p_j)}. \quad (2)$$

where p_i and p_j denote the total number of overlaps with adjacent granular balls for GB_i and GB_j. The merging process continues until the ball set no longer changes.

Methodology

In this section, we introduce a deep multi-view clustering method called Multi-view Granular-ball Contrastive Clustering (MGBCC). MGBCC encompasses four crucial processes: within-view reconstruction, within-view granular-ball generation, and cross-view granular-ball asociation and granular-ball contrastive learning. The framework is shown in Figure 2.

Within-view Reconstruction

Given a multi-view dataset $\{\mathbf{X}^v\}_{v=1}^V$ with N samples, each sample has instances from V different views. Let d_v represent the feature dimension of the v-th view, which typically varies across different views. To standardize the dimensions across views for subsequent comparison and fusion, we project the features of different views into a common dimension d.

Deep autoencoders are employed as the representation learning framework to effectively extract essential low-dimensional embeddings from raw features. We assign an autoencoder to each view. Specifically, for the v-th view, $E_v(\cdot; \theta^v)$ and $D_v(\cdot; \phi^v)$ denote its encoder and decoder, with θ^v and ϕ^v being their learnable parameters. As mentioned earlier, we set the output feature dimension of all encoders $\{E_v\}_{v=1}^V$ to d. After projection through encoders

$\{E_v\}^V$, we obtain the high-level features $\{\mathbf{H}^v\}_{v=1}^V$ of instances from the v-th view by minimizing

$$\mathcal{L}_{rec} = \sum_{v=1}^V \sum_{i=1}^N \|\mathbf{x}_i^v - D_v(E_v(\mathbf{x}_i^v; \theta^v); \phi^v)\|_2^2, \quad (3)$$

where $\mathbf{x}_i^v$ denotes the i-th sample of $\mathbf{X}^v$. The representation of i-th sample in v-th view is given by

$$\mathbf{h}_i^v = E_v(\mathbf{x}_i^v; \theta^v). \quad (4)$$

In the subsequent computations, we will use these instance representations for constructing granular balls instead of the original features.

Within-view Granular-ball Generation

In clustering tasks, the classic granular ball partitioning method controls the granularity of division using a minimum capacity threshold (e.g., η). Granular balls are recursively split until the number of samples in a granular ball is less than this threshold. The issue with this method arises when it reaches the edge regions of the sample space, where the number of samples within a granular ball may be below the threshold, but the samples are dispersed (e.g., outliers). In such cases, the ball center may experience significant deviation, and the radius may be overestimated, resulting in inappropriate overlapping relationships.

To address this, we designed an alternative granular ball generation method. First, we introduce a granularity control parameter p, which roughly reflects the granularity of granular ball generation. Let k denote the total number of granular balls to be generated. We set N and k to satisfy the following relationship

$$k = \max\left(\left\lfloor \frac{N}{p} \right\rfloor, 1\right). \quad (5)$$

We then directly apply the k-means (Lloyd 1982) to the entire dataset, dividing it into k clusters, with each cluster considered a granular ball. For each ball, we compute its center and radius. Intuitively, outliers in the sample space, being far from most samples, tend to form their own clusters. Non-outliers will cluster together approximately according to the granularity parameter p.

We construct granular balls for each view individually. For the v-th view, H^v denotes latent representations obtained through the projection by the encoding layer E^v. Using the aforementioned method, we obtain $S^v = \{GB_i^v\}_{i=1}^{k_v}$, where k_v represents the number of granular balls in the v-th view. If we set the granularity parameter p_v to the same value p for each view, the numbers $\{k_v\}_{v=1}^V$ across all views will be the same. For simplicity, we adopt this setting in the subsequent analysis and experiments.

Let $S = \{S^v\}_{v=1}^V$ denote granular ball sets constructed for all views, $\mathbf{C}^v$ be the center matrix of the v-th view, and $\mathbf{r}^v$ be the radius matrix. For a granular ball GB_i^v, its center is $\mathbf{c}_i^v$ and its radius is r_i^v. Note that the calculations of the centers and radii are gradient-preserving. In our granular-ball generation process, there is no merging process as in classical methods. However, we still need to consider the overlapping relationships of balls within each view.
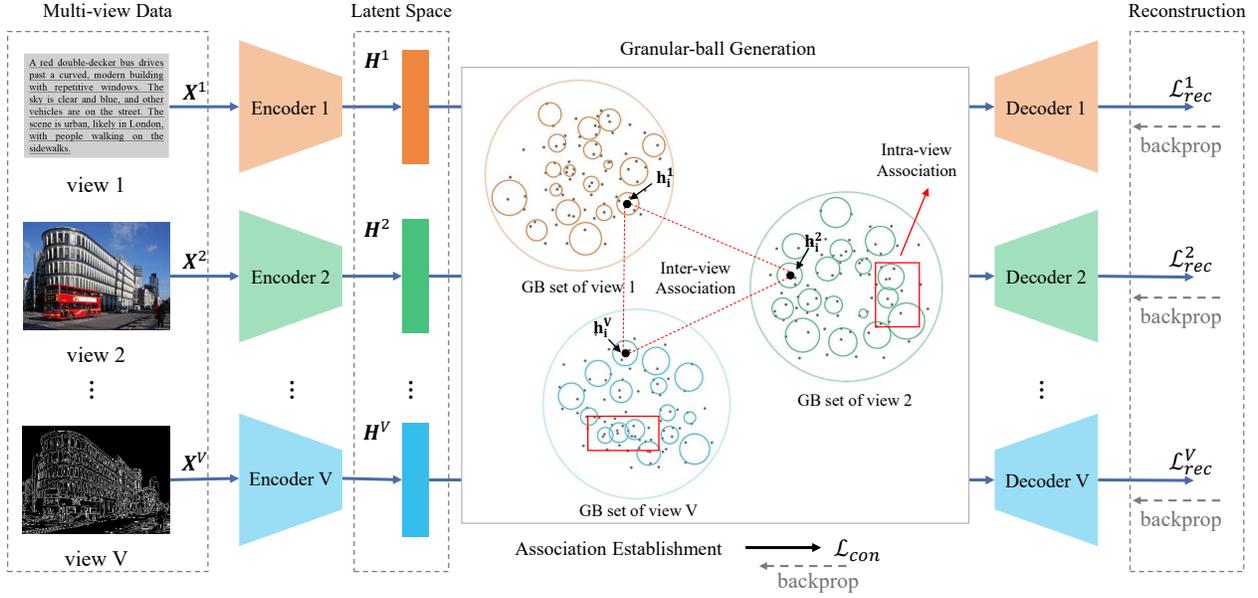
Figure 2: The framework of MGBCC. As shown, the overall loss function consists of two parts, e.g., reconstruction loss and granular-ball contrastive loss. We construct granular-ball sets $\{S^v\}_{v=1}^V$ for different views in the latent space and establish intra-view and cross-view associations based on overlap and intersection size respectively. Granular balls model the local structure of the dataset, and associated granular balls should be close to each other in the latent space.

Let $\mathbf{D}^v$ represent the center distance matrix of the v-th view. The distance between the i-th ball and the j-th ball is calculated as $\mathbf{d}_{ij}^v = \|\mathbf{c}_i^v - \mathbf{c}_j^v\|_2$. Based on $\mathbf{D}^v$, we compute the granular balls overlapping matrix $\mathbf{A}^v$, which satisfies

$$a_{ij}^v = \begin{cases} 1, & \text{if Eq(2) is satisfied} \\ 0, & \text{otherwise} \end{cases}. \quad (6)$$

Note that $\mathbf{A}^v$ will serve as part of the mask matrix for contrastive learning.

Cross-view Granular-ball Association

Through matrices $\{\mathbf{A}^v\}_{v=1}^V$, we have established the intra-view relationships between granular balls. Then we need to consider how to establish connections between cross-view granular balls. An intuitive approach is to consider two balls from different views as neighbors in the latent space if they each contain instances of the same sample from their respective views. However, this method lacks robustness. When p is relatively large, granular balls also become larger, and two cross-view balls might contain a very small number of common samples due to randomness.

To address this, we modified the method. Let GB_i^m and GB_j^n be two granular balls from views m and n, respectively, containing t_i and t_j samples. First, we identify the common sample set between the two balls based on the stored sample indices:

$$\text{Id}_{\text{both}} = \text{Id}(GB_i^m) \cap \text{Id}(GB_j^n) \quad (7)$$

where $\text{Id}(\cdot)$ represents the sample indices contained in a granular ball. Next, we count the number of samples in Id_{both}:

$$t_{\text{both}} = \text{length}(\text{Id}_{\text{both}}) \quad (8)$$

Let $\mathbf{P}^{(m,n)}$ be the cross-view granular ball association matrix, which satisfies the following condition

$$\mathrm{p}_{ij}^{(m,n)} = \begin{cases} 1, & \text{if } t_{\text{both}}/\min(t_i, t_j) \geq \tau \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

Here, τ is a threshold parameter that determines the minimum proportion of common samples required to consider two cross-view granular balls as associated.

Granular-ball Contrastive Learning

Matrices $\{\mathbf{A}^v\}_{v=1}^V$ reflects whether any two granular balls within a view overlap, while matrices $\{\mathbf{P}^{(m,n)}\}_{\forall m \neq n}$ indicates whether any two balls across views have sufficient intersection. We aim for these associated granular ball pairs to be as close as possible in the latent space, while unrelated pairs should be far apart.

We use the granular-ball centers to represent the entire granular balls during the calculations. To facilitate computation, for any two views m and n, we define the combined center matrix as

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}^m \\ \mathbf{C}^n \end{bmatrix} \quad (10)$$

then concatenate $\{\mathbf{A}^m, \mathbf{A}^n\}$ and $\{\mathbf{P}^{(m,n)}, \mathbf{P}^{(n,m)}\}$ into a unified mask matrix

$$\mathbf{M} = \begin{bmatrix} \mathbf{A}^m & \mathbf{P}^{(m,n)} \\ \mathbf{P}^{(n,m)} & \mathbf{A}^n \end{bmatrix} \quad (11)$$

where matrix $\mathbf{P}^{(n,m)}$ is the transpose of $\mathbf{P}^{(m,n)}$. This unified mask matrix $\mathbf{M}$ ensures that granular balls within the

Dataset	Samples	Clusters	Views	Dimensionality
BBCSport	544	5	2	3183/3203
Caltech101-20	2386	20	2	1984/512
Cora	2708	7	2	2708/1433
Scene-15	4485	15	3	20/59/40
MNIST-USPS	5000	10	2	784/784
ALOI-100	10800	100	4	77/13/64/125
NoisyMNIST	50000	10	2	784/784

Table 1: Description of the seven multi-view datasets.

same view and across different views are appropriately considered in the contrastive learning process. we calculate the contrastive loss at the granular-ball level

$$\mathcal{L}^{(m,n)} = \frac{1}{k}\sum_i^k \sum_{j\in\Omega_i} \frac{exp(cos(\mathbf{c}_i,\mathbf{c}_j))}{\sum_{z\in\Phi_i} exp(cos(\mathbf{c}_i,\mathbf{c}_z))} \quad (12)$$

where $\Omega_i = \{j|\mathbf{M}_{ij}=1,\forall j\}$ and $\Phi_i = \{z|\mathbf{M}_{iz}=0,\forall z\}$. k represents the total number of granular balls and $cos(\cdot,\cdot)$ denotes the cosine similarity between two vectors.

$$\mathcal{L}_{con} = \frac{2}{V(V-1)}\sum_{\forall m\neq n} \mathcal{L}^{(m,n)} \quad (13)$$

We perform the same calculation process between any two views and take the average of all the losses as the final contrastive loss.

Overall Loss And Optimization

Combining the above two loss functions with a regularization parameter λ, the overall loss is formulated as

$$\mathcal{L} = \mathcal{L}_{con} + \lambda\mathcal{L}_{rec}. \quad (14)$$

Any gradient-based optimization method can be used to minimize this objective function. We will further discuss the implementation details later.

Experiments

In this section, we analyze the experimental results of the proposed method on seven widely used multi-view datasets and compare it with several state-of-the-art methods to demonstrate its effectiveness.

Experimental Settings

Datasets. Seven multi-view benchmark datasets are employed in this work. BBCSport (Greene and Cunningham 2006) includes 544 sports news articles in 5 subject areas, with 3183-dimensional MTX features and 3203-dimensional TERMS features, forming 2 views. Caltech101-20 (Li et al. 2015) contains 101 classes in total. We select 20 widely used classes with 2 views and 2386 samples for our experiments. Cora (Bisson and Grimal 2012) contains 4 views, including content, inbound, outbound, and citations, extracted from the documents. Scene-15 (Fei-Fei and Perona 2005) consists of 4568 natural scenes

categorized into 15 groups. Each scene is described by three types of features: GIST, SIFT, and LBP. MNIST-USPS (Peng et al. 2019) is a popular handwritten digit dataset containing 5000 samples with two different styles of digital images. ALOI-100 (Schubert and Zimek 2010) consists of 10800 object images, with each image described by 4 different features. NoisyMNIST (Wang et al. 2015) uses the original images as view 1 and randomly selects within-class images with white Gaussian noise as view 2. Table 1 lists the important information of all datasets.

Compared Methods. We compared the proposed method with seven classical or state-of-the-art methods including Completer (Lin et al. 2021), MFLVC (Xu et al. 2022), DealMVC (Yang et al. 2023b), DMCE (Zhao, Yang, and Nie 2023), CSPAN (Jin et al. 2023), ADPAC (Xu et al. 2023b), SURE (Yang et al. 2023a). All compared methods are implemented according to the source codes released by the authors, and the hyper parameters are set according to the suggestion in the corresponding paper.

Evaluation Metrics. To perform a fair comparison, we adopt the commonly used metrics, e.g., clustering accuracy (ACC), normalized mutual information (NMI), and purity (PUR).

Implementation Details

The network structure follows a standard autoencoder architecture. For each view, the encoder consists of several linear layers with ReLU activation functions between each pair of layers. Except for the BBCSport and Cora datasets, all other datasets use the same encoder structure with dimensions set as $\{d_v, 2000, 500, 500, d\}$, where d_v is the input feature dimension of each view. d is the projection feature dimension, which is the same for all views. After encoding, inputs undergo standardization. The decoder mirrors the encoder structure. For the Cora dataset, we use the same dimensions but without activation functions between layers, resulting in a linear projection. For BBCSport, given its small sample size of 544, we use a single-layer linear projection with the encoder dimensions set to $\{d_v, d\}$.

Our implementation of MGBCC is carried out using Py-Torch 2.3 (Paszke et al. 2019) on a Windows 10 operating system, powered by an NVIDIA GeForce GTX 1660 Ti GPU. We employ the Adam optimizer with learning rate of 0.0001 and weight decay of 0. The batch size is typically set to either 256 or 1024, depending on the dataset size. The regularization parameter λ is generally set to 1 across most datasets, except for BBCSport and Cora, where it is adjusted to 0 due to differences in the projection approach (e.g., linear or nonlinear). The threshold parameter τ is uniformly set to 0.1 across all datasets. The granularity parameter p significantly impacts the experimental results, which will be analyzed later.

During the clustering phase, we equally weight and fuse the projected features $\{\mathbf{H}^v\}_{v=1}^V$ from each view and then apply the k-means algorithm to obtain clustering labels.

Experimental Results

In Table 2, we present the experimental results of the proposed method in comparison with other methods, leading to

Dataset \Method	Completer	MFLVC	DealMVC	DMCE	CSPAN	ADPAC	SURE	**MGBCC**
			ACC(%)					
BBCSport	35.11	60.11	80.70·	37.13	58.27	35.29	55.33	**95.77**
Caltech101-20	71.42	36.92	40.36	61.48	44.72	**77.16**	50.21	72.63
Cora	22.12	41.84	49.07·	26.48	42.54	30.87	42.76	**65.44**
Scene-15	39.29	32.87	32.71	34.85	33.73	41.49	42.01	**43.72**
MNIST-USPS	67.02	99.50·	80.92	62.18	80.14	98.16	99.56	**99.60**
ALOI-100	62.70	47.26	15.80	75.39	14.86	26.24	**90.37**	88.39
NoisyMNIST	87.45	98.91	**99.42**	-	55.68	97.27	99.14	98.48
			NMI(%)					
BBCSport	2.62	43.04	65.59·	4.39	48.27	3.94	38.78	**87.21**
Caltech101-20	70.96	52.68	59.41	57.93	63.04	**73.85**	60.92	71.57
Cora	1.42	30.41	37.75·	1.19	22.21	7.65	22.49	**46.86**
Scene-15	43.46	33.93	32.46	32.09	31.23	**44.21**	42.87	41.62
MNIST-USPS	81.83	98.50·	90.24	72.90	78.27	95.61	98.68	**98.96**
ALOI-100	84.00	75.72	56.12	83.67	39.88	51.59	94.04	**94.78**
NoisyMNIST	86.70	96.79	**98.15**	-	60.07	93.34	97.30	96.17
			PUR(%)					
BBCSport	35.85	69.12	80.70·	37.13	69.30	38.62	63.60	**95.77**
Caltech101-20	78.58	69.74	70.12	71.12	76.19	80.68	73.60	**81.89**
Cora	30.28	51.51	60.67·	30.32	49.89	36.74	48.15	**68.02**
Scene-15	42.88	34.02	34.34	36.77	36.39	45.08	44.93	**47.07**
MNIST-USPS	73.00	99.50·	80.92	66.80	81.70	98.16	99.56	**99.60**
ALOI-100	66.74	48.58	15.80	76.97	18.53	30.74	**90.67**	89.31
NoisyMNIST	87.45	98.91	**99.42**	-	60.69	97.27	99.14	98.48

Table 2: The clustering results on seven datasets (%). The best results are bolded, and the second-best results are underlined. Results marked with a dot are directly quoted from the original papers. - indicates unavailable results due to out of memory.

the following conclusions: (1) Across the three given metrics, the proposed method achieves the best or second-best results for most case. Even on the NoisyMNIST datasets, the proposed method ranks approximately third, with only slight differences from the top two methods. Using Cora dataset as an example, our proposed method achieves an accuracy of 65.44%, significantly surpassing the best comparative result of 49.07%. This demonstrates the effectiveness and competitiveness of our method. (2) Compared with classical multi-view contrastive learning methods (e.g., Completer, DealMVC, SURE), the proposed method consistently achieves more favorable clustering results across the majority of datasets. As a representative, SURE is an instance-level contrastive learning method that focuses on the issue of false negatives, achieving the best or second-best results on the Scene-15, MNIST-USPS, ALOI-100, and NoisyMNIST datasets. It's noteworthy that the proposed method does not show significant performance degradation on these four datasets and clearly outperforms other methods. Moreover, on the remaining three datasets, the proposed method significantly outperforms SURE, highlighting the effectiveness of contrastive learning at the granular-ball level. (3) We visualize the clustering results of the proposed method on the MNIST-USPS dataset using t-SNE. As shown in Fig-

ure 3, the clustering structure becomes progressively clearer with the increase of the optimization epoch. This suggests that the proposed method is effective at revealing the underlying cluster structure.

Ablation Studies

To accurately validate the effectiveness of contrastive learning at the granular-ball level, we conduct an ablation study on the Caltech101-20 dataset. The feature dimension d is set to 128. Based on this, we adopted three experimental settings. The first setting trains the model solely based on reconstruction loss. The second setting includes both the reconstruction loss and instance-level contrastive loss (i.e., $p = 1$). The third setting incorporates reconstruction loss and granular-ball contrastive loss, with the granularity parameter set to 2. It is important to emphasize that the parameter p reflects the average granularity rather than the absolute granularity.

Table 3 presents the corresponding experimental results. As can be seen, the instance-level contrastive method performed poorly on this dataset, whereas the granular-ball contrastive method achieved significant improvements. This further demonstrates the feasibility and effectiveness of contrastive learning at the granular-ball level.
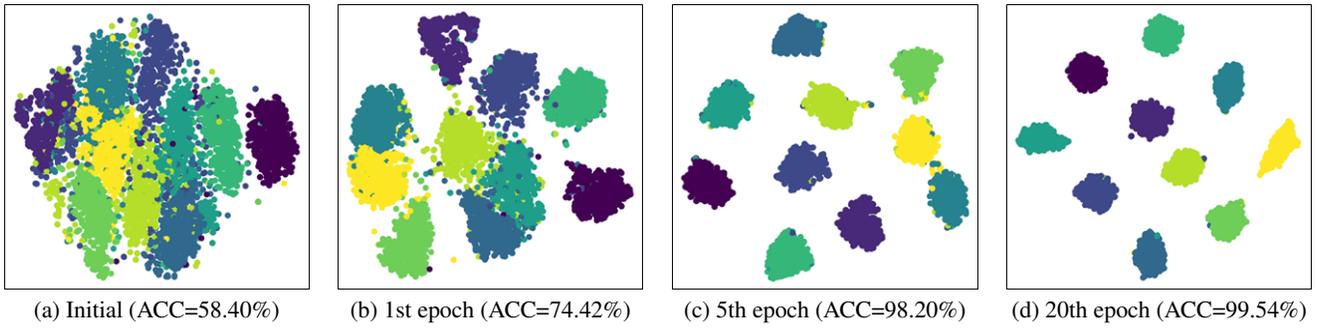
| (a) Initial (ACC=58.40%) | (b) 1st epoch (ACC=74.42%) | (c) 5th epoch (ACC=98.20%) | (d) 20th epoch (ACC=99.54%) |

Figure 3: The t-SNE visualization of the clustering results on MNIST-USPS dataset.

Setting	ACC (%)	NMI (%)	PUR (%)
$\mathcal{L}_{rec}$	40.28	60.03	75.40
$\mathcal{L}_{rec} + \mathcal{L}_{con} \sim p = 1$	44.47	62.38	78.54
$\mathcal{L}_{rec} + \mathcal{L}_{con} \sim p = 2$	**72.63**	**71.57**	**81.89**

Table 3: Ablation studies on Caltech101-20.

Parameter Analysis

The model has two important hyperparameters: the granularity parameter p and the dimension d of the projection features. The former essentially reflects the average size of the granular balls. When p is set to 1, it is equivalent to instance-level contrastive learning. The latter affects the amount of original feature information contained in the latent representation. If d is too small, important information might be lost, whereas if d is too large, it increases the complexity of optimization and memory requirements. To explore the optimal parameter settings, we conduct experiments on Caltech101-20 and Cora. We varied the parameter p within the range [1, 2, 4, 8, 16], and the parameter d within the range [8, 16, 32, 64, 128, 256].

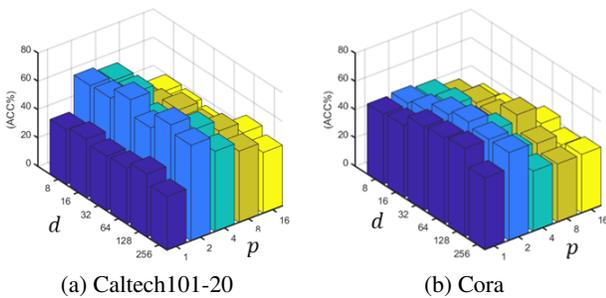| (a) Caltech101-20 | (b) Cora |

Figure 4: The clustering accuracy (%) with different parameters p and d on Caltech101-20 and Cora.

Figure 4 illustrates the experimental results on the aforementioned datasets. When p is set to 2, the proposed method performs well. However, as p increases, performance gradually declines. This may be because larger granular balls can no longer be effectively associated based solely on overlap and intersection size. When p becomes too large, the

method essentially degrades into a cluster-level contrastive approach, where reducing cluster assignment differences might be a more reasonable strategy. In experiments, we typically set the parameter p to 1, 2, or 4. The parameter d has a relatively minor impact on the experimental results and is generally set to 64.
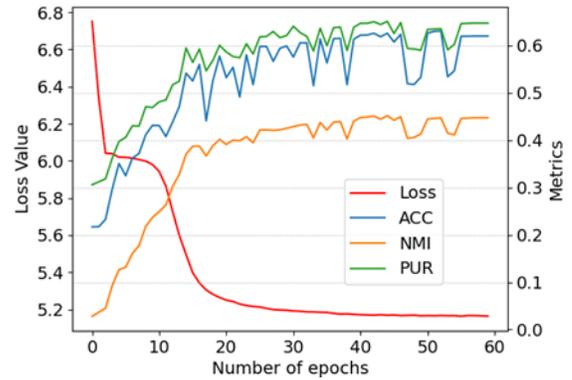

Figure 5: Loss vs. Metrics on Cora.

Convergence Analysis

We evaluated the convergence of the proposed method on Cora dataset by tracking the loss values and corresponding clustering performance over increasing epochs. As shown in Figure 5, the total loss values gradually decrease and converge within 100 epochs. These results demonstrate the strong convergence performance of the proposed method.

Conclusion

In this paper, we propose a multi-view granular-ball contrastive clustering method. Specifically, we model the local structure of the sample set using granular balls in the latent space, resulting in respective granular-ball sets for each view. We establish intra-view and cross-view associations between granular balls based on their overlap and intersection size, encouraging associated granular balls to be close to each other in the latent space. Extensive experiments have been conducted to validate the effectiveness of the proposed method. In the future, we will extend the proposed method to handle incomplete multi-view data.

Acknowledgments

References

Andrew, G.; Arora, R.; Bilmes, J.; and Livescu, K. 2013. Deep canonical correlation analysis. In *Proceedings of the 30th International Conference on International Conference on Machine Learning*, volume 28, III–1247–III–1255.

Bisson, G.; and Grimal, C. 2012. Co-clustering of Multi-view Datasets: A Parallelizable Approach. In *2012 IEEE 12th International Conference on Data Mining*, 828–833.

Chen, J.; Mao, H.; Woo, W. L.; and Peng, X. 2023. Deep Multiview Clustering by Contrasting Cluster Assignments. In *2023 IEEE/CVF International Conference on Computer Vision*, 16706–16715.

Chen, L. 1982. Topological Structure in Visual Perception. *Science*, 218(4573): 699–700.

Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020. A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning*.

Cheng, D.; Li, Y.; Xia, S.; Wang, G.; Huang, J.; and Zhang, S. 2024. A Fast Granular-Ball-Based Density Peaks Clustering Algorithm for Large-Scale Data. *IEEE Transactions on Neural Networks and Learning Systems*, 35(12): 17202–17215.

Deng, S.; Wen, J.; Liu, C.; Yan, K.; Xu, G.; and Xu, Y. 2024. Projective Incomplete Multi-View Clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 35(8): 10539–10551.

Fang, U.; Li, M.; Li, J.; Gao, L.; Jia, T.; and Zhang, Y. 2023. A Comprehensive Survey on Multi-View Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 35(12): 12350–12368.

Fei-Fei, L.; and Perona, P. 2005. A Bayesian hierarchical model for learning natural scene categories. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, 524–531.

Greene, D.; and Cunningham, P. 2006. Practical Solutions to the Problem of Diagonal Dominance in Kernel Document Clustering. In *Proceedings of International Conference on Machine learning*, 377–384.

He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum Contrast for Unsupervised Visual Representation Learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9726–9735.

Huang, S.; Kang, Z.; Tsang, I. W.; and Xu, Z. 2019. Auto-weighted multi-view clustering via kernelized graph learning. *Pattern Recognition*, 88: 174–184.

Huang, S.; Tsang, I. W.; Xu, Z.; and Lv, J. 2022. Measuring Diversity in Graph Learning: A Unified Framework for Structured Multi-View Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 34(12): 5869–5883.

Jin, J.; Wang, S.; Dong, Z.; Liu, X.; and Zhu, E. 2023. Deep Incomplete Multi-View Clustering with Cross-View Partial Sample and Prototype Alignment. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11600–11609.

Li, Y.; Nie, F.; Huang, H.; and Huang, J. 2015. Large-scale multi-view spectral clustering via bipartite graph. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2750–2756.

Lin, Y.; Gou, Y.; Liu, Z.; Li, B.; Lv, J.; and Peng, X. 2021. COMPLETER: Incomplete Multi-view Clustering via Contrastive Prediction. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11169–11178.

Liu, J.; Hao, J.; Ma, Y.; and Xia, S. 2024. Unlock the Cognitive Generalization of Deep Reinforcement Learning via Granular Ball Representation. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, 31062–31079.

Liu, J.; Liu, X.; Yang, Y.; Liao, Q.; and Xia, Y. 2023. Contrastive Multi-View Kernel Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8): 9552–9566.

Liu, X. 2023. SimpleMKKM: Simple Multiple Kernel K-Means. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 5174–5186.

Lloyd, S. 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2): 129–137.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Kopf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; and Chintala, S. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*, 32.

Peng, X.; Huang, Z.; Lv, J.; Zhu, H.; and Zhou, J. T. 2019. COMIC: Multi-view Clustering Without Parameter Selection. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, 5092–5101.

Peng, X.; Xiao, S.; Feng, J.; Yau, W.-Y.; and Yi, Z. 2016. Deep subspace clustering with sparsity prior. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 1925–1931.

Quadir, A.; and Tanveer, M. 2024. Granular Ball Twin Support Vector Machine With Pinball Loss Function. *IEEE Transactions on Computational Social Systems*, 1–10.

Schubert, E.; and Zimek, A. 2010. ELKI Multi-View Clustering Data Sets Based on the Amsterdam Library of Object Images (ALOI) (1.0). Data set.

Su, P.; Liu, Y.; Li, S.; Huang, S.; and Lv, J. 2024. Robust Contrastive Multi-view Kernel Clustering. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 4938–4945.

Wang, G. 2017. DGCC: data-driven granular cognitive computing. *Granular Computing*, 2(4): 343–355.

Wang, W.; Arora, R.; Livescu, K.; and Bilmes, J. 2015. On deep multi-view representation learning. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, volume 37, 1083–1092.

Wen, J.; Zhang, Z.; Zhang, Z.; Wu, Z.; Fei, L.; Xu, Y.; and Zhang, B. 2020. DIMC-net: Deep Incomplete Multi-view Clustering Network. In *Proceedings of the 28th ACM International Conference on Multimedia*, 3753–3761. New York, NY, USA.

Xia, D.; Wang, G.; Zhang, Q.; Yang, J.; and Xia, S. 2024a. Three-Way Approximations Fusion With Granular-Ball Computing to Guide Multigranularity Fuzzy Entropy for Feature Selection. *IEEE Transactions on Fuzzy Systems*, 32(10): 5963–5977.

Xia, S.; Lian, X.; Wang, G.; Gao, X.; Chen, J.; and Peng, X. 2024b. GBSVM: An Efficient and Robust Support Vector Machine Framework via Granular-Ball Computing. *IEEE Transactions on Neural Networks and Learning Systems*, 1–15.

Xia, S.; Liu, Y.; Ding, X.; Wang, G.; Yu, H.; and Luo, Y. 2019. Granular ball computing classifiers for efficient, scalable and robust learning. *Information Sciences*, 483: 136–152.

Xia, S.; Wang, C.; Wang, G.; Gao, X.; Ding, W.; Yu, J.; Zhai, Y.; and Chen, Z. 2023a. GBRS: A Unified Granular-Ball Learning Model of Pawlak Rough Set and Neighborhood Rough Set. *IEEE Transactions on Neural Networks and Learning Systems*, 1–15.

Xia, S.; Zhang, H.; Li, W.; Wang, G.; Giem, E.; and Chen, Z. 2022. GBNRS: A Novel Rough Set Algorithm for Fast Adaptive Attribute Reduction in Classification. *IEEE Transactions on Knowledge and Data Engineering*, 34(3): 1231–1242.

Xia, S.; Zheng, S.; Wang, G.; Gao, X.; and Wang, B. 2023b. Granular Ball Sampling for Noisy Label Classification or Imbalanced Classification. *IEEE Transactions on Neural Networks and Learning Systems*, 34(4): 2144–2155.

Xie, J.; Dai, M.; Xia, S.; Zhang, J.; Wang, G.; and Gao, X. 2024a. An Efficient Fuzzy Stream Clustering Method Based on Granular-Ball Structure. In *Proceedings of the 40th International Conference on Data Engineering*, 901–913.

Xie, J.; Hua, C.; Xia, S.; Cheng, Y.; Wang, G.; and Gao, X. 2024b. W-GBC: An Adaptive Weighted Clustering Method Based on Granular-Ball Structure. In *Proceedings of the 40th International Conference on Data Engineering*, 914–925.

Xie, J.; Xiang, X.; Xia, S.; Jiang, L.; Wang, G.; and Gao, X. 2024c. MGNR: A Multi-Granularity Neighbor Relationship and Its Application in KNN Classification and Clustering Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12): 7956–7972.

Xu, J.; Chen, S.; Ren, Y.; Shi, X.; Shen, H.; Niu, G.; and Zhu, X. 2023a. Self-Weighted Contrastive Learning among Multiple Views for Mitigating Representation Degeneration. *Advances in Neural Information Processing Systems*, 36: 1119–1131.

Xu, J.; Li, C.; Peng, L.; Ren, Y.; Shi, X.; Shen, H. T.; and Zhu, X. 2023b. Adaptive Feature Projection With Distribution Alignment for Deep Incomplete Multi-View Clustering. *IEEE Transactions on Image Processing*, 32: 1354–1366.

Xu, J.; Tang, H.; Ren, Y.; Peng, L.; Zhu, X.; and He, L. 2022. Multi-Level Feature Learning for Contrastive Multi-View Clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16051–16060.

Yang, J.; Liu, Z.; Xia, S.; Wang, G.; Zhang, Q.; Li, S.; and Xu, T. 2024. 3WC-GBNRS++: A Novel Three-Way Classifier With Granular-Ball Neighborhood Rough Sets Based on Uncertainty. *IEEE Transactions on Fuzzy Systems*, 32(8): 4376–4387.

Yang, M.; Li, Y.; Hu, P.; Bai, J.; Lv, J.; and Peng, X. 2023a. Robust Multi-View Clustering With Incomplete Information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1): 1055–1069.

Yang, X.; Jiaqi, J.; Wang, S.; Liang, K.; Liu, Y.; Wen, Y.; Liu, S.; Zhou, S.; Liu, X.; and Zhu, E. 2023b. DealMVC: Dual Contrastive Calibration for Multi-view Clustering. In *Proceedings of the 31st ACM International Conference on Multimedia*, 337–346.

Zhang, B.; and Wang, L. 2024. False Negative Sample Detection for Graph Contrastive Learning. *Tsinghua Science and Technology*, 29(2): 529–542.

Zhang, Q.; Wu, C.; Xia, S.; Zhao, F.; Gao, M.; Cheng, Y.; and Wang, G. 2023. Incremental Learning Based on Granular Ball Rough Sets for Classification in Dynamic Mixed-Type Decision System. *IEEE Transactions on Knowledge and Data Engineering*, 35(9): 9319–9332.

Zhao, M.; Yang, W.; and Nie, F. 2023. Deep multi-view spectral clustering via ensemble. *Pattern Recognition*, 144: 109836.