Consistency-Aware Padding for Incomplete Multi-Modal Alignment Clustering Based on Self-Repellent Greedy Anchor Search

Shubin Ma, Liang Zhao*, Mingdong Lu, Yifan Guo, Bo Xu
School of Software Technology, Dalian University of Technology
1369830844@mail.dlut.edu.cn, liangzhao@dlut.edu.cn
mingdonglu@mail.dlut.edu.cn, qiaoan412@mail.dlut.edu.cn, BoXu@dlut.edu.cn *

Abstract

Multimodal representation is faithful and highly effective in describing real-world data samples' characteristics by describing their complementary information. However, the collected data often exhibits incomplete and misaligned characteristics due to factors such as inconsistent sensor frequencies and device malfunctions. Existing research has not effectively addressed the issue of filling missing data in scenarios where multiview data are both imbalanced and misaligned. Instead, it relies on class-level alignment of the available data. Thus, it results in some data samples not being well-matched, thereby affecting the quality of data fusion. In this paper, we propose the Consistency-Aware Padding for Incomplete Multimodal Alignment Clustering Based on Self-Repellent Greedy Anchor Search(CAPIMAC) to tackle the problem of filling imbalanced and misaligned data in multimodal datasets. Specifically, we propose a self-repellent greedy anchor search module(SRGASM), which employs a self-repellent random walk combined with a greedy algorithm to identify anchor points for re-representing incomplete and misaligned multimodal data. Subsequently, based on noise-contrastive learning, we design a consistency-aware padding module (CAPM) to effectively interpolate and align imbalanced and misaligned data, thereby improving the quality of multimodal data fusion. Experimental results demonstrate the superiority of our method over benchmark datasets. The code will be publicly released at https://github.com/Autism-mm/CAPIMAC.git.

1 Introduction

Multimodal clustering [Zhao *et al.*, 2024a; Zhou *et al.*, 2024; Wang *et al.*, 2024] is an important research direction in machine learning and data mining, gaining significant attention due to its ability to leverage the features of multimodal data to enhance clustering results. To better describe the encoded features of each modality, multimodal clustering integrates each modality's complete information that is both aligned and complementary. However, a key challenge in multimodal clustering is addressing the incompleteness and misalignment [Tang *et al.*, 2023; Wang *et al.*, 2023] between modalities. Incompleteness refers to data being ignored or mislabeled as missing, often due to sensor failures, annotation errors, or preprocessing flaws. For example, In industrial monitoring, data may be missing due to sensor(infrared cameras and visual sensors) failures or environmental interference. Misalignment refers to the inability to establish a one-to-one correspondence between data items across different modalities, partially due to differing sensor collection frequencies or acquisition times. For example, in road traffic monitoring research [Dai *et al.*, 2021], the system uses multiple images to determine whether a vehicle is violating regulations or speeding. However, the images of the same vehicle are not captured consecutively. Between the capture of two images, the camera continuously captures images of other vehicles. During data acquisition, incompleteness and misalignment occur simultaneously, significantly affecting subsequent multimodal data fusion[Li *et al.*, 2023d; Liu *et al.*, 2024a]. Current research focuses on how to align and fuse multimodal incomplete data with misalignment characteristics.

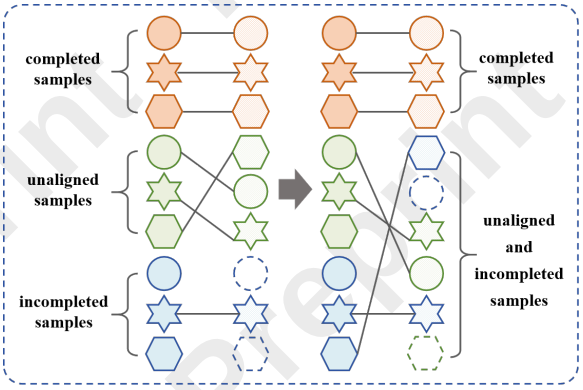

Figure 1: Incomplete partial alignment example graph.

A toy example regarding incomplete and misaligned multimodal data clustering is shown in Figure 1, and the same colour represents data of the same class, with solid lines in-

*Corresponding Author

dicating data matching. On the left, the matching of complete, misaligned, and incomplete multimodal data is shown; on the right, the matching of complete multimodal data and data with both incompleteness and misalignment is shown. From the right side of Figure 1, it can be seen that in the incomplete and misaligned, the data order between the two modalities is inconsistent, and the data sizes are imbalanced.

Incompleteness and misalignment present significant challenges to multimodal clustering. To address these challenges, researchers have proposed various methods. In terms of incompleteness [Zong *et al.*, 2021], Yang et al. [Yang *et al.*, 2023] introduced one seminal work, CGMIMC, which effectively integrates structures across different modalities through consistency graph learning, thereby facilitating the completion of missing data. Li et al. [Li *et al.*, 2023b] designed an IMvC method that simultaneously captures clustering information and modality-specific details through a dual-stream model and performs data recovery using prototypes from other modalities and data-prototype relationships in known modalities. Huang et al. [Huang *et al.*, 2024] restored missing data by dynamically learning feature similarity graphs and proposed a dynamic data quality assessment method, UNIFIER, based on semi-quadratic minimization to mitigate the impact of outliers and unreliable recovered data. In terms of misalignment, Huang et al. [Huang *et al.*, 2020] designed a PVC clustering model, introducing a differentiable alignment module that learns the optimal alignment strategy and can be integrated into deep learning models for alignment learning. Yang et al. [Yang *et al.*, 2021] designed a new noise-robust contrastive loss model, MVCLN, which mitigates the impact of noise on data while enhancing the similarity between features of data from the same class using contrastive learning. Zeng et al. [Zeng *et al.*, 2023] designed the clustering model SMILE using information theory, which addresses completely misaligned multimodal data. By combining autoencoders with two cross-entropy losses, it reduces cross-modal differences while enhancing semantic distinction between different classes. Zhao et al. [Zhao *et al.*, 2024b] proposed a progressive alignment learning model, DGPPVC, which, based on a Jaccard similarity variant and dynamic structural graph optimization, gradually identifies unknown correspondences between different modalities from simple to complex. He et al. [He *et al.*, 2024] proposed the VITAL model, which models each data sample as a Gaussian distribution in the latent space. By using variational inference and contrastive learning, the model preserves the commonality and specific semantics between modalities, enabling a comprehensive perception of the data. To address both incomplete and misaligned multimodal data clustering, Yang et al. [Yang *et al.*, 2022] designed a model called SURE. The model directly handles zeroed incomplete and misaligned multimodal data using Hungarian alignment and fills missing data by selecting and filling with the same data through KNN [Shi *et al.*, 2018; Li *et al.*, 2023a]. However, this method may lead to large similarity differences between aligned data pairs and lower-quality data pairs, affecting subsequent data fusion.

When addressing the misalignment of multimodal data, the number of data across modalities is consistent. However, multimodal incompleteness leads to an imbalance in the data across modalities, presenting a challenge for multimodal data alignment. To address the misalignment issue caused by data inconsistency and imbalances, enhance the coherence of data similarity—particularly when data missingness leads to misalignment between views, resulting in a sudden drop in similarity between data pairs (as illustrated in Figure 3)—and improve the quality of data fusion, we propose the Consistency-Aware Padding for Incomplete Multimodal Alignment Clustering Based on Self-Repellent Greedy Anchor Search, termed **CAPIMAC**. The proposed model primarily consists of three components: the self-repellent greedy anchor search module, the model training module, and the consistency-aware padding module. In the SR-GASM, we select anchors [Qin *et al.*, 2024; Liu *et al.*, 2024b; Li *et al.*, 2023c] using a self-repellent random walk [Doshi *et al.*, 2023], guided by a greedy algorithm. This method not only captures the complex local structures between data but also integrates the global information of the anchors. Then, we guide the learning of latent feature representations through a noise-contrastive loss [Zhao *et al.*, 2021a] function. Finally, in the CAPM, we introduce an innovative method by applying equal-weight Gaussian kernel interpolation to fill under-sampled modalities, preventing abrupt decreases in data-pair similarity and improving similarity coherence, thereby obtaining high-quality fused features. Experimental results demonstrate that the model effectively utilizes known data information for filling, thereby addressing the issues of incompleteness and misalignment in multimodal data.

The contributions of this paper can be summarized below.

- We design the self-repellent greedy anchor search module to select anchors. This module selects anchors that are more structurally representative, reducing the complexity of subsequent computations. At the same time, the anchor selection process avoids local cycles and enhances the diversity of exploration.
- We introduce noise-contrastive loss to reduce the impact of false negative pairs, thereby guiding the learning of latent feature representations.
- We innovatively design the consistency-aware padding module to address the padding and alignment issue under imbalanced and misordered multimodal data. In this module, the equal-weight Gaussian kernel interpolation enhances the coherence of data-pair similarity, ensuring one-to-one alignment and fusion of the two modality sample data.

2 Related Work

2.1 Random Walk

Random walk is a fundamental statistical model that tracks random activity. The recursive formula for random walk in a multidimensional space is given by,

$$\mathbf{p}(n) = P^n \mathbf{p}(0) \tag{1}$$

$\mathbf{p}(n) = [p_1(n), p_2(n) \ldots, p_N(n)]$ is the visit probability vector of the nodes at the *nth* step, and $\mathbf{p}(0)$ is the initial state vector, representing the visit probabilities of the nodes at the

0th step. P^n is the *nth* power of the transition matrix, representing the system's state after *n* steps of the random walk.

Random walk has various applications in machine learning. In adjacency graph construction, Wang et al. [Wang *et al.*, 2020] reduced the time overhead in multi-label classification tasks by constructing a KNN-based random walk graph of training data. In combination with anchors, Liu et al. [Liu *et al.*, 2010] adopted an anchor-based random walk strategy, using anchors as "central" nodes for information propagation, optimizing the flow of information and improving learning efficiency. In node embedding features , Aung et al. [Aung and Ohsaki, 2024] combined node embedding techniques with a random walk, enhancing the random walk process by using node embedding vectors. Additionally, random walk has wide applications in graph generation [Cai *et al.*, 2022], graph enhancement [Kim *et al.*, 2023], anomaly detection [Liu *et al.*, 2023], and graph sampling [Wang *et al.*, 2021].

2.2 Gaussian Kernel Interpolation

Gaussian kernel interpolation is commonly used for data smoothing, missing value imputation, or spatial data interpolation. The core idea is to estimate the value of a point using the weighted sum of Gaussian kernel functions. During Gaussian kernel interpolation, for a new interpolation point x_* , the interpolated value $\hat{y}_*$ can be expressed as,

$$\hat{y}_* = \sum_{i=1}^{n} w_i K(x_*, x_i) \tag{2}$$

where w_i is the weight, and $K(\cdot)$ denotes the Gaussian kernel function.

Gaussian kernel functions are widely used in multimodal clustering. For example, Liu et al. [Liu *et al.*, 2020] used Gaussian kernel functions to complete incomplete kernel matrices for learning consensus kernel matrices, addressing the issue of incomplete multimodal data. Ye et al. [Ye *et al.*, 2017]calculated the similarity between modalities based on the Gaussian kernel function and handled the nonlinear relationships in the data through the Gaussian kernel matrix, thereby increasing tolerance to missing data.

3 The Proposed Method

Given a dataset $\mathbf{X} = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \mathbf{X}^{(3)}, \ldots, \mathbf{X}^{(v)}\}$,$\mathbf{X}^{(v)} \in \mathbf{R}^{n \times dv}$, represents the data of the *vth* modality with *n* samples and dimensionality *dv*. To simulate the scenarios of misalignment and incompleteness, we introduce a shuffle matrix $\mathbf{U} = \{\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \mathbf{U}^{(3)}, \ldots, \mathbf{U}^{(v)}\}$ and a missing indicator matrix $\mathbf{C} = \{\mathbf{C}^{(1)}, \mathbf{C}^{(2)}, \mathbf{C}^{(3)}, \ldots, \mathbf{C}^{(v)}\}$ to process $\mathbf{X}$,where $\mathbf{U}^{(v)} \in \mathbf{R}^{n \times 1}$,$\mathbf{C}^{(v)} \in \mathbf{R}^{n \times 1}$.

To address the challenge of effectively filling missing values in incomplete and misaligned multimodal data, we propose the CAPIMAC framework, which consists of two key components:(1) the SRGASM, which selects critical data within each class through a self-repellent random walk strategy and utilizes a greedy algorithm to ensure that at least one key data is chosen from each class;(2) the CAPM, which employs an equal-weight Gaussian kernel padding strategy to ensure consistency in the data count across different modalities, thereby enhancing the coherence of data similarity.

3.1 Self-Repellent Greedy Anchor Search Module

In multimodal datasets, the high feature dimensionality of data leads to significant computational time during graph construction and training. To address this, we introduce anchors to simplify the representation of the original data. To better capture the complex local structures among the data and simultaneously identify the global optimal solution for the anchors, we propose the SRGASM. The module identifies anchors based on node density using a self-repellent random walk, with the number of steps and step size selected for each modality according to its characteristics,

$$(n_w, w_l) = \begin{cases} (20, 3) & \text{if } N < 100 \\ (10, 5) & \text{if } 100 \leq N < 1000 \\ (5, 10) & \text{if } 1000 \leq N < 10000 \\ (3, 20) & \text{if } N \geq 10000 \end{cases} \tag{3}$$

where n_w and w_l denote the number of steps and step size, respectively, and *N* represents the number of data. To describe the transition process between states, a state transition probability matrix is defined for each modality.

$$\begin{aligned} & P_{ij}^{(v)} = \cos(X_{ij}^{(v)}, X_{ij}^{(v)}) \\ & \text{s.t.} \quad \sum_{j=1}^{n} P_{ij}^{(v)} = 1, \quad P_{ii}^{(v)} = 0 \quad (i = 1, 2, \ldots, n) \end{aligned} \tag{4}$$

Specifically, **P** is a normalized probability matrix with zeros along the diagonal, and $cos(\cdot)$ represents the cosine similarity between data.

To avoid local cycles and enhance the diversity of exploration, we introduce a decay function r_{ui} to reduce the probability of visiting previously visited nodes, where

$$r_{\mu_i}(x_i) = \left(\frac{x_i}{\mu_i}\right)^{-\alpha}, \quad \alpha \geq 0 \tag{5}$$

Here, u_i represents the mean of the neighbouring data, and the parameter $\alpha \geq 0$ can be viewed as the strength of the self-repellent module in the self-repellent random walk transition kernel. When $\alpha > 0$, $r_{ui}(x_i)$ decreases as x_i increases, with larger α leading to faster decay. When $\alpha = 0$, no decay is applied. At the same time, $r_{ui}(x_i)$ ensures that the probability matrix **P** exhibits scale invariance.

In the self-random walk sampling process, a smaller value of α improves the mixing property of the sampling, whereas a larger value of α enhances its efficiency. To strike a balance between the two, we set $\alpha = 0.5$.Based on the transition probability matrix **P**, we obtain the visit matrix **V**,

$$V = \sum_{w=1}^{n_w} V_w, V_w \quad = \sum_{t=1}^{w_l} P^t \times \pi \tag{6}$$

Where *w* represents the number of random walks, and *t* denotes the number of steps in each random walk.

However, when all data in a particular class are significantly distant from data in other classes, it becomes challenging to select points from that isolated class during the self-repellent random walk, resulting in anchors that fail to effectively capture global information. To address this issue,

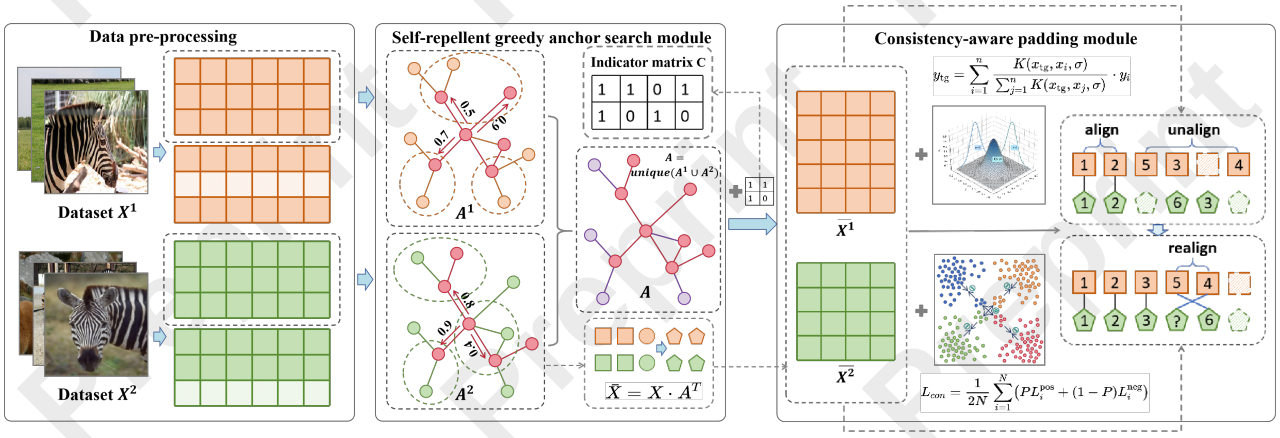

Figure 2: Incomplete and unaligned CAPIMAC model. In the Data Pre-processing, we handle incomplete and misaligned multimodal data. In the SRGASM, anchors are selected with a self-repelling random walk and greedy strategy, followed by data re-representation. In the CAPM, features learned via noise-contrastive loss are filled and aligned through Gaussian kernel interpolation.

we introduce a greedy algorithm in the self-repellent random walk process to expand the selection of anchors.

In the greedy algorithm, an anchor point $x_i \in X$ is first selected, and all nodes within a radius of d are marked.s

$$S_{k+1} = \{x_i\} \cup \{x_j \in X \mid d_{ij} \leq d\} \tag{7}$$

Next, when selecting the next anchor point $x_k \in X$, it is required that its distance to all nodes in the previously selected anchor set $\mathbf{S}$ exceeds d, meaning that a node outside the radius is chosen.

$$S_{k+2} = S_{k+1} \cup \{x_k \mid d_{kj} > d, \forall x_j \in S_{k+1}\} \tag{8}$$

Here, S_k denotes the set of all anchors selected at the *kth* step. Once all nodes have been marked, the marks are reset, and the selection process is repeated until the required number of anchor points is reached.

3.2 Model Training

Since the anchors for each modality are selected independently, this results in mismatched anchor indices in the aligned portions. To standardize the representation dimensions across modalities, we unify the indices of the multiple modalities.

$$I_u = unique(I^{(0)} \cup I^{(1)} \cup I^{(2)} \cdots I^{(v)}) \tag{9}$$

I_u denotes the deduplicated sum of the indices across multiple modalities.

To standardize the dimensional representation of multimodal data and facilitate improved training and graph construction, we re-represent the original data $\mathbf{X}^{(v)}$ using the selected consistency anchors $\mathbf{A}^{(v)} = \mathbf{X}^{(v)}[I_u]$,i.e.,

$$\bar{X} = X \cdot A^T \tag{10}$$

where $\mathbf{A} \in \mathbf{R}^{n_a \times dv}$, n_a represents the number of anchors, which is equal to the size of I_u.

Due to the presence of noisy data in real-world multimodal datasets, we introduce noise-contrastive learning to mitigate the impact of noise. In the traditional contrastive loss,

$$L_{con} = \frac{1}{n}\sum_{i=1}^{n} Y D_w^2 + (1-Y)\max(m - D_w, 0)^2 \tag{11}$$

where,

$$D_w(X_1, X_2) = \left(\sum_{i=1}^{n}(X_1^i - X_2^i)^2\right)^{\frac{1}{2}} \tag{12}$$

However, the traditional L_{con} loss function has two main issues: first, Euclidean distance struggles to effectively measure similarity between high-dimensional data; and second, in the presence of noisy data, false negative pairs may occur, where two data points in a noisy pair actually belong to the same class. To address these issues, we propose an improved noise-contrastive[Yang *et al.*, 2021] loss function,

$$L_{recon} = \frac{1}{2n}\sum_{i=1}^{n}\left(Y \cdot {D_c}^2 + (1-Y)\cdot\frac{1}{m} \max\left(a \cdot m \cdot D_c - {D_c}^3, 0\right)^2\right) \tag{13}$$

where $\mathbf{Y}$ represents the labels of positive and negative data pairs, $\mathbf{D_c}$ represents the cosine distance, and a is a hyperparameter that controls the distance range within which false negative pairs can be mitigated.

3.3 Consistency-Aware Padding Module

In scenarios with incomplete and misaligned data, incompleteness leads to an imbalance in the number of multimodal data, while misalignment disrupts the order of data, making data padding challenging. To address these issues, we propose the CAPM, as shown in Figure 3.

For $\mathbf{X}^{(v)}$, we sequentially apply misalignment and incompleteness processing to obtain the data $\hat{\mathbf{X}}^{(v)}$. Taking two modalities as an example,

$$\hat{X}^{(v)} = \check{X}^{(v)}[C^{(v)}], \check{X}^{(v)} = X^{(v)}[U^{(v)}] \tag{14}$$

Where, $\mathbf{U}$ is a shuffle index used to reorder the data in $\mathbf{X}$, and $\mathbf{C}$ is a **0**,**1** index, where **0** indicates a missing data and **1** indicates an existing data, used to remove missing data. The

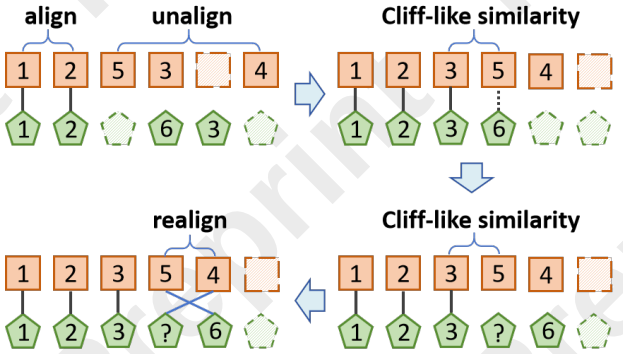

Figure 3: Flowchart of the Consistency-aware Padding Module: The process from 1 to 2 represents the sorting, from 2 to 3 represents the interpolation, and from 3 to 4 represents the realignment.

resulting $\mathbf{X}^{(1)} \in \mathbf{R}^{nv_1 \times dv}$, $\mathbf{X}^{(2)} \in \mathbf{R}^{nv_2 \times dv}$,and $nv_1 \neq nv_2$. Thus, the re-represented incomplete and misaligned data is,

$$\bar{X}^{(v)} = \hat{X}^{(v)} \cdot A^{T^{(v)}} \tag{15}$$

When the data is aligned, we need to compute a distance matrix $\mathbf{Z}$ for Hungarian alignment, where

$$Z = 1 - Z', Z' = \begin{cases} \cos(\bar{X}^{(1)}, \bar{X}^{(2)}) & \text{if } nv_1 > nv_2 \\ \cos(\bar{X}^{(2)}, \bar{X}^{(1)}) & \text{if } nv_1 \leq nv_2 \end{cases} \\ \text{s.t.} \quad \sum_{j=1}^{n} {Z'}_{ij} = 1 \tag{16}$$

For convenience, we use n_s and n_l to denote the smaller and larger data sizes, respectively.

$$Z \in R^{n_s \times n_l}, s.t. \begin{cases} n_s = nv_2, n_l = nv_1 & \text{if } nv_1 > nv_2 \\ n_s = nv_1, n_l = nv_2 & \text{if } nv_1 \leq nv_2 \end{cases} \tag{17}$$

We begin by using the Hungarian algorithm to find pairs of data with high similarity starting from the first row. However, due to class imbalance, some data do not form pairs with other data. Additionally, since missing data are randomly selected, there may be cases where data from different classes are aligned in the already aligned pairs, resulting in a significant disparity in similarity between data pairs. To maintain the coherence of data-pair similarity and reduce the occurrence of misaligned pairs from different classes, we rearrange the data-pair similarities, that is,

$$Z^r = U^r \cdot Z \tag{18}$$

Here, $\mathbf{U^r}$ is a row transformation indicator matrix that rearranges the rows of $\mathbf{Z}$ in ascending order based on the data-pair distances. Next, we select $n_k (n_k = n_l - n_s)$ index pairs with the largest similarity differences and perform equal-weight Gaussian kernel interpolation between the indexed data. The Gaussian kernel interpolation function is given by,

$$y_{\text{target}} = \sum_{i=1}^{n_s} \frac{K(x_{\text{target}}, x_i, \sigma)}{\sum_{j=1}^{n_s} K(x_{\text{target}}, x_j, \sigma)} \cdot y_i \tag{19}$$

where,

$$K(x, x_i, \sigma) = \exp\left(-\frac{1}{2}\left(\frac{x - x_i}{\sigma}\right)^2\right) \tag{20}$$

Through equal-weight interpolation, we obtain the new distance indicator matrix $\bar{\mathbf{Z}}^{\mathbf{r}} \in \mathbf{R}^{n_l \times n_l}$. Subsequently, Hungarian alignment is performed based on $\bar{\mathbf{Z}}^{\mathbf{r}^T}$ (Since no interpolation is performed on the modality with a length of n_l, virtual labels can be retrieved, which aids in performance evaluation.), followed by data fusion.

4 Experiments

In this section, we extensively evaluate the clustering property of the proposed method on five widely used multimodal benchmark datasets. The performance of CAPIMAC is compared with seven state-of-the-art multimodal methods (PVC[Huang *et al.*, 2020], MvCLN [Yang *et al.*, 2021], EGPVC [Zhao *et al.*, 2023b], SMILE [Zeng *et al.*, 2023], SURE[Yang *et al.*, 2022], AMCP [Zhao *et al.*, 2024c], DGPPVC [Zhao *et al.*, 2024b]) in terms of four clustering evaluation metrics.

4.1 Experimental Settings

Five datasets are used in our experiments, including 3Sources, yale_mtv, BBCsports, Prokaryotic and 100leaves. In the experiments, we first perform data shuffling based on alignment rates, then apply missing data treatment to the shuffled portions. Since there are few models for handling incomplete and misaligned data, in the comparative experiments, we first apply missing data treatment to the models addressing misalignment, followed by Gaussian interpolation for filling and, finally, data shuffling.

The experiments were conducted in a Python 3.8 environment, using PyCharm 2022.2.3 as the integrated development environment. The implementation was based on the deep learning framework PyTorch 1.8.2 with CUDA version 11.1. In the experiments, the alignment rate for all methods was set to 0.3, 0.5, and 0.7, while the incompleteness rate was set to 0.5. Since the primary challenge of multimodal incomplete and misaligned problems is alignment, with data filling being secondary, we used these settings to assess the model's performance under low, normal, and high alignment rates in the presence of typical data incompleteness. To validate the effectiveness of CAPIMAC in clustering, we conducted comparative and ablation experiments using ACC, NMI, ARI and F_score(weighted) as performance metrics.

4.2 Experimental Results and Analysis

In this experiment, the best-optimized values for each metric are highlighted in bold, and the second-best values are underlined. BBC refers to the BBCsports dataset, Prok to the Prokaryotic dataset, and F1 to the F_score (weighted).

Tables 1,2,4 report the experimental results of eight different methods on five multimodal datasets, evaluated using metrics such as ACC, NMI, ARI and F1. These metrics assess the methods' performance in clustering tasks. From these results, we obtain the following conclusions.

Datasets		Methods							
		PVC 20'	MvCLN 21'	EGPVC 22'	SMILE 23'	SURE 23'	AMPC 24'	DGPPVC 24'	Ours
3Sources	ACC	0.4686	0.3645	0.4278	0.3462	0.2823	0.3604	0.4905	**0.5168**
	NMI	0.3354	0.1039	0.2182	0.0918	0.1256	0.1652	0.2270	**0.3707**
	ARI	0.2239	0.0134	0.1586	0.0114	0.0019	0.0763	0.2202	**0.2538**
	F1	0.4836	0.2809	0.4331	0.3091	0.3515	0.1692	0.4048	**0.5092**
yale_mtv	ACC	0.4333	0.3437	0.4588	0.4109	0.3721	0.2655	0.2751	**0.5229**
	NMI	0.5182	0.4044	**0.5335**	0.4796	0.4529	0.3611	0.3749	**0.5445**
	ARI	0.2490	0.1185	0.2656	0.1921	0.1564	0.0861	0.0976	**0.2735**
	F1	0.4189	0.3288	0.4465	0.4090	0.3611	0.0458	0.2198	**0.5412**
BBC	ACC	0.4475	0.3564	0.3575	0.3379	0.3060	0.4365	0.4248	**0.5119**
	NMI	0.1789	0.0571	0.0697	0.0492	0.891	0.2182	0.1139	**0.2871**
	ARI	0.1245	0.0151	0.0366	0.0098	0.0152	0.1210	0.0745	**0.1744**
	F1	0.4344	0.2593	0.3533	0.2852	0.3536	0.2275	0.3257	**0.531**
Prok	ACC	0.3425	0.4870	0.4011	0.4784	0.3428	0.3270	0.4182	**0.5823**
	NMI	0.0308	0.1204	0.0784	0.1618	0.0684	0.0110	0.0875	**0.2415**
	ARI1	0.0144	0.0469	0.0419	0.0731	0.0230	0.0042	0.0651	**0.2059**
	F1	0.3646	0.4541	0.4191	0.4788	0.5252	0.2793	0.4520	**0.5838**
100leaves	ACC	0.2329	0.2948	0.2958	0.3535	0.3195	0.2634	0.4306	**0.5345**
	NMI	0.5492	0.5750	0.6143	0.6605	0.1676	0.6320	0.6871	**0.7705**
	ARI	0.1034	0.1302	0.1613	0.2143	0.0587	0.1557	0.2824	**0.3827**
	F1	0.2203	0.2977	0.2832	0.3401	0.3876	0.0092	0.4171	**0.5295**

Table 1: The clustering performance(ACC, NMI, ARI and F1) of various algorithms on different datasets. The alignment rate is 0.5; the incomplete rate is 0.5.

- Tables 1,2,4 present the clustering results of eight models on five datasets. Among the four metrics evaluated on these datasets, our model achieved the best performance in the majority of the results, demonstrating its effectiveness in addressing the multimodal incomplete and misaligned data problem.
- As shown in Tables 1,2,4, our model achieves excellent results on the 100leaves dataset under incomplete data conditions with low, medium, and high alignment rates. This demonstrates that our model can maintain strong clustering performance even when dealing with datasets with a large number of classes.
- When addressing the multimodal misalignment and incompleteness problem, the clustering results of our model show a slight decrease as the alignment rate decreases, yet it still achieves the best performance in the majority of clustering results. This demonstrates that our model has strong robustness and can handle incomplete alignment fusion problems under varying alignment rates.
- Since our model requires selecting anchors from complete data for data re-representation, it is unable to handle incomplete and misaligned datasets with excessively low alignment rates. This is why the model's performance shows a slight decline when the alignment rate is 0.3.
- The experimental results demonstrate that the selected anchors effectively capture both the local and global information of the modality data. Additionally, equal-weight Gaussian kernel interpolation addressed the issue of incoherent similarity, further improving clustering performance.

Method		3Sources	yale_mtv	BBC	Prok	100leaves
PVC	ACC	0.3710	0.4364	0.4096	0.3998	0.1848
	NMI	0.2110	0.5036	0.1376	0.0613	0.4875
	ARI	0.1096	0.2212	0.0893	0.0522	0.0521
	F1	0.3757	0.4339	0.3869	0.4328	0.1812
MvCLN	ACC	0.3538	0.3297	0.3638	0.5015	0.2964
	NMI	0.0996	0.4106	0.0610	0.1446	0.5832
	ARI	0.0081	0.1278	0.0234	0.0577	0.1351
	F1	0.2066	0.3080	0.2838	0.4671	0.2988
EGPVC	ACC	0.4414	0.4424	0.3645	0.4726	0.3025
	NMI	0.2247	0.5228	0.0696	0.1740	0.6054
	ARI	0.1373	**0.2539**	0.0320	0.1166	0.1574
	F1	0.4286	0.4270	0.3238	0.4766	0.2891
SMILE	ACC	0.3643	0.4388	0.4004	0.4885	0.3487
	NMI	0.1122	0.4995	0.1211	0.1555	0.6632
	ARI	0.0016	0.2183	0.0980	0.1213	0.2104
	F1	0.3222	0.4336	0.3383	0.4970	0.3421
SURE	ACC	0.3426	0.3636	0.3596	0.3733	0.3092
	NMI	0.1834	0.4361	0.1558	0.0866	0.1539
	ARI	0.0417	0.1768	0.0831	0.0415	0.0351
	F1	0.3971	0.4103	0.3982	0.4986	0.3417
AMPC	ACC	0.3266	0.2497	0.4674	0.3323	0.2019
	NMI	0.1418	0.3302	**0.2767**	0.0112	0.5370
	ARI	0.0426	0.0580	**0.1753**	0.0016	0.0763
	F1	0.1662	0.0640	0.2323	0.2871	0.0058
DGPPVC	ACC	0.4651	0.1830	0.3840	0.4145	0.3838
	NMI	0.2386	0.2480	0.0714	0.0748	0.6607
	ARI	0.1706	0.0205	0.0389	0.0551	0.2404
	F1	0.3747	0.1640	0.2555	0.4413	0.3677
Ours	ACC	**0.5299**	**0.4830**	**0.4953**	**0.5979**	**0.4887**
	NMI	**0.3926**	**0.5277**	0.2399	**0.2755**	**0.7487**
	ARI	**0.2720**	0.2358	0.1712	**0.2637**	**0.3426**
	F1	**0.5416**	**0.5068**	**0.5066**	**0.6068**	**0.4847**

Table 2: The clustering performance(ACC, NMI, ARI and F1) of various algorithms on different datasets. The alignment rate is 0.3; the incomplete rate is 0.5.

Metric		ACC		NMI		ARI		F1	
Rate	Datasets	No IPT	IPT	No IPT	IPT	No IPT	IPT	No IPT	IPT
0.3align	3sources	0.4745	**0.5299**	0.3201	**0.3926**	0.2344	**0.2720**	0.4967	**0.5416**
	BBC	0.4554	**0.5073**	0.1768	**0.2485**	0.1246	**0.1827**	0.4604	**0.5179**
	yale_mtv	0.4741	**0.4830**	0.5144	**0.5277**	0.2064	**0.2358**	0.4966	**0.5068**
	Prokaryotic	0.4703	**0.6101**	0.0517	**0.2842**	0.0748	**0.2860**	0.4722	**0.6201**
	100leaves	0.3942	**0.4887**	0.6771	**0.7487**	0.2272	**0.3426**	0.3869	**0.4847**
0.5align	3sources	0.4897	**0.5168**	0.3078	**0.3707**	0.2396	**0.2538**	0.5092	**0.5273**
	BBC	0.4527	**0.4722**	0.1393	**0.2406**	0.1246	**0.1597**	0.4701	**0.4834**
	yale_mtv	0.4786	**0.5229**	0.5074	**0.5445**	0.2085	**0.2735**	0.4917	**0.5427**
	Prokaryotic	0.4937	**0.5823**	0.0648	**0.2415**	0.0186	**0.2059**	0.4191	**0.5838**
	100leaves	0.3977	**0.5345**	0.6771	**0.7705**	0.2347	**0.3827**	0.3951	**0.5295**
0.7align	3sources	0.4808	**0.5531**	0.4279	**0.4330**	0.2598	**0.3209**	0.4982	**0.5719**
	BBC	0.4864	**0.5027**	0.2173	**0.2768**	0.1825	**0.1901**	0.4861	**0.5222**
	yale_mtv	0.5033	**0.5411**	0.5392	**0.5614**	0.2623	**0.3033**	0.5254	**0.5629**
	Prokaryotic	0.5520	**0.6259**	0.1035	**0.2976**	0.1290	**0.2930**	0.5374	**0.6386**
	100leaves	0.3744	**0.5749**	0.6703	**0.7959**	0.2210	**0.4424**	0.3716	**0.5648**

Table 3: Ablation Experiment of CAPIMAC on Whether to Use the Consistency-Aware Padding Module at Alignment Rates of 0.3, 0.5, and 0.7.

Method		3Sources	yale_mtv	BBC	Prok	100leaves
PVC	ACC	0.4414	0.4497	0.4113	0.3386	0.2504
	NMI	0.3303	0.5287	0.1775	0.0442	0.5736
	ARI	0.2053	0.2607	0.1021	0.0162	0.1227
	F1	0.4603	0.4394	0.3772	0.3630	0.2282
MvCLN	ACC	0.3539	0.3224	0.3521	0.5158	0.2971
	NMI	0.0874	0.3860	0.0495	0.0986	0.5715
	ARI	0.0069	0.1034	0.0082	0.0343	0.1319
	F1	0.2530	0.3101	0.2445	0.4629	0.3015
EGPVC	ACC	0.4201	0.4412	0.3862	0.4688	0.3296
	NMI	0.2183	0.5170	0.0994	0.1057	0.6261
	ARI	0.1380	0.2484	0.0539	0.0535	0.1811
	F1	0.4229	0.4244	0.3784	0.4526	0.3205
SMILE	ACC	0.3118	0.4115	0.3521	0.4554	0.3800
	NMI	0.0863	0.4768	0.0595	0.1598	0.6812
	ARI	0.0126	0.1974	0.0165	0.0567	0.2395
	F1	0.2896	0.4032	0.2960	0.4515	0.3725
SURE	ACC	0.2858	0.3951	0.3177	0.3434	0.3240
	NMI	0.1134	0.4712	0.0856	0.0796	0.1627
	ARI	0.0152	0.1848	0.0160	0.0375	0.0333
	F1	0.3805	0.4085	0.3674	0.5363	0.3935
AMPC	ACC	0.3053	0.2976	0.3574	0.3438	0.2250
	NMI	0.1099	0.4159	0.1123	0.0079	0.5725
	ARI	0.0376	0.1210	0.0486	0.0028	0.1049
	F1	0.1842	0.0715	0.1955	0.2968	0.0108
DGPPVC	ACC	0.4716	0.2497	0.3874	0.3924	0.4341
	NMI	0.2445	0. 3825	0.0884	0.0735	0.6982
	ARI	0.1915	0.0903	0.0394	0.0599	0.3030
	F1	0.3794	0.2030	0.2607	0.4293	0.4157
Ours	ACC	**0.5531**	**0.5411**	**0.5027**	**0.6259**	**0.5749**
	NMI	**0.4279**	**0.5614**	**0.2768**	**0.2976**	**0.7959**
	ARI	**0.3209**	**0.3033**	**0.1901**	**0.2930**	**0.4424**
	F1	**0.5719**	**0.5629**	**0.5222**	**0.6386**	**0.5648**

Table 4: The clustering performance(ACC, NMI, ARI and F1) of various algorithms on different datasets. The alignment rate is 0.7; the incomplete rate is 0.5.

4.3 Time Complexity Analysis

After obtaining the similarity graph, KNN needs to find the k nearest neighbors for each data point x_i. The k smallest distances are identified through sorting, which has a complexity of $O(nlogn)$, resulting in an overall complexity of $O(n^2logn)$. However, random walk typically explores only within the local neighbourhood, with a complexity of $O(nlogn)$ or $O(n \cdot k)$.

4.4 Ablation Experiment

Table 3 presents the ablation experiment on whether the model incorporates the CAPM. No IPT indicates that missing data is discarded directly without applying equal-weight Gaussian kernel interpolation; IPT indicates that equal-weight Gaussian kernel interpolation was applied.

As shown in Table 3, the CAPM has improved clustering performance to some extent under low, medium, and high alignment rates. Essentially, the CAPM not only utilizes all available data information, increasing data utilization, but also addresses the issue of sharp similarity gaps in imbalanced data, ultimately enhancing the quality of multimodal data fusion.

5 Conclusion

In this paper, we propose CAPIMAC to address the feature fusion problem under the conditions of incomplete and misaligned multimodal data. By designing the SRGASM, we select anchors that capture both local and global information from the modalities. We resolve the alignment and padding issue under misaligned and imbalanced multimodal data through the CAPM, thus improving the quality of the fused features. Our method provides a new perspective for solving the filling problem of misordered and imbalanced data. In the future, we plan to extend CAPIMAC to explore the interpolation and alignment issues in cases where multimodal data are completely misaligned and partially incomplete.

Acknowledgments

This work is supported by the Science and Technology Project of Liaoning Province (2024JH2/102600027, 2023JH2/101700363) and the Science and Technology

Project of Dalian City (2024JJ12GX025, 2023JJ12SN029 and 2023JJ11CG005).

References

[Aung and Ohsaki, 2024] Han Nay Aung and Hiroyuki Ohsaki. Node embedding accelerates randoms walk on a graph. In *2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 537–545. IEEE, 2024.

[Cai *et al.*, 2022] Wenyu Cai, Gilbert Chen Ye, and Hao Zhou. Graph generation and diffusion using random walks. *Highlights in Science, Engineering and Technology*, 16:490–494, 2022.

[Dai *et al.*, 2021] Shaojie Dai, Jinshuai Wang, Chao Huang, Yanwei Yu, and Junyu Dong. Temporal multi-view graph convolutional networks for citywide traffic volume inference. In *2021 IEEE International Conference on Data Mining (ICDM)*, pages 1042–1047. IEEE, 2021.

[Dong *et al.*, 2017] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 135–144, 2017.

[Doshi *et al.*, 2023] Vishwaraj Doshi, Jie Hu, et al. Self-repellent random walks on general graphs-achieving minimal sampling variance via nonlinear markov chains. In *International Conference on Machine Learning*, pages 8403–8423. PMLR, 2023.

[Grover and Leskovec, 2016] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 855–864, 2016.

[He *et al.*, 2024] Changhao He, Hongyuan Zhu, Peng Hu, and Xi Peng. Robust variational contrastive learning for partially view-unaligned clustering. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 4167–4176, 2024.

[Huang *et al.*, 2020] Zhenyu Huang, Peng Hu, Joey Tianyi Zhou, Jiancheng Lv, and Xi Peng. Partially view-aligned clustering. *Advances in Neural Information Processing Systems*, 33:2892–2902, 2020.

[Huang *et al.*, 2024] Yanyong Huang, Zongxin Shen, Tianrui Li, and Fengmao Lv. Unified view imputation and feature selection learning for incomplete multi-view data. *arXiv preprint arXiv:2401.10549*, 2024.

[Kang *et al.*, 2020] 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.

[Kim *et al.*, 2023] Jun Hee Kim, Jaeman Son, Hyunsoo Kim, and Eunjo Lee. Node embedding for homophilous graphs with argew: Augmentation of random walks by graph edge weights. *arXiv preprint arXiv:2308.05957*, 2023.

[Li *et al.*, 2023a] Chao Li, Shifei Ding, Xiao Xu, Haiwei Hou, and Ling Ding. Fast density peaks clustering algorithm based on improved mutual k-nearest-neighbor and sub-cluster merging. *Information Sciences*, 647:119470, 2023.

[Li *et al.*, 2023b] Haobin Li, Yunfan Li, Mouxing Yang, Peng Hu, Dezhong Peng, and Xi Peng. Incomplete multi-view clustering via prototype-based imputation. *arXiv preprint arXiv:2301.11045*, 2023.

[Li *et al.*, 2023c] Xingfeng Li, Yinghui Sun, Quansen Sun, Jia Dai, and Zhenwen Ren. Distribution consistency based fast anchor imputation for incomplete multi-view clustering. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 368–376, 2023.

[Li *et al.*, 2023d] Xingyuan Li, Yang Zou, Jinyuan Liu, Zhiying Jiang, Long Ma, Xin Fan, and Risheng Liu. From text to pixels: a context-aware semantic synergy solution for infrared and visible image fusion. *arXiv preprint arXiv:2401.00421*, 2023.

[Liu *et al.*, 2010] Wei Liu, Junfeng He, and Shih-Fu Chang. Large graph construction for scalable semi-supervised learning. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 679–686. Citeseer, 2010.

[Liu *et al.*, 2020] Xinwang Liu, Miaomiao Li, Chang Tang, Jingyuan Xia, Jian Xiong, Li Liu, Marius Kloft, and En Zhu. Efficient and effective regularized incomplete multi-view clustering. *IEEE transactions on pattern analysis and machine intelligence*, 43(8):2634–2646, 2020.

[Liu *et al.*, 2023] Chang Liu, Zhong Yuan, Baiyang Chen, Hongmei Chen, and Dezhong Peng. Fuzzy granular anomaly detection using markov random walk. *Information Sciences*, 646:119400, 2023.

[Liu *et al.*, 2024a] Jinyuan Liu, Xingyuan Li, Zirui Wang, Zhiying Jiang, Wei Zhong, Wei Fan, and Bin Xu. Promptfusion: Harmonized semantic prompt learning for infrared and visible image fusion. *IEEE/CAA Journal of Automatica Sinica*, 2024.

[Liu *et al.*, 2024b] Suyuan Liu, Qing Liao, Siwei Wang, Xinwang Liu, and En Zhu. Robust and consistent anchor graph learning for multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

[Qin *et al.*, 2024] Yalan Qin, Chuan Qin, Xinpeng Zhang, and Guorui Feng. Dual consensus anchor learning for fast multi-view clustering. *IEEE Transactions on Image Processing*, 2024.

[Shi *et al.*, 2018] Bing Shi, Lixin Han, and Hong Yan. Adaptive clustering algorithm based on knn and density. *Pattern Recognition Letters*, 104:37–44, 2018.

[Tang *et al.*, 2023] Chuan Tang, Kun Sun, Chang Tang, Xiao Zheng, Xinwang Liu, Jun-Jie Huang, and Wei Zhang. Multi-view subspace clustering via adaptive graph learning and late fusion alignment. *Neural Networks*, 165:333–343, 2023.

[Wang *et al.*, 2020] Zhen-Wu Wang, Si-Kai Wang, Ben-Ting Wan, and William Wei Song. A novel multi-label classification algorithm based on k-nearest neighbor and random walk. *International Journal of Distributed Sensor Networks*, 16(3):1550147720911892, 2020.

[Wang *et al.*, 2021] Pengyu Wang, Chao Li, Jing Wang, Taolei Wang, Lu Zhang, Jingwen Leng, Quan Chen, and Minyi Guo. Skywalker: Efficient alias-method-based graph sampling and random walk on gpus. In *2021 30th International Conference on Parallel Architectures and Compilation Techniques (PACT)*, pages 304–317. IEEE, 2021.

[Wang *et al.*, 2022] 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.

[Wang *et al.*, 2023] Chenxu Wang, Zhenhao Huang, Yue Wan, Junyu Wei, Junzhou Zhao, and Pinghui Wang. Fualign: Cross-lingual entity alignment via multi-view representation learning of fused knowledge graphs. *Information Fusion*, 89:41–52, 2023.

[Wang *et al.*, 2024] Siwei Wang, Xinwang Liu, Suyuan Liu, Wenxuan Tu, and En Zhu. Scalable and structural multi-view graph clustering with adaptive anchor fusion. *IEEE Transactions on Image Processing*, 2024.

[Xu *et al.*, 2022] 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.

[Yang *et al.*, 2021] Mouxing Yang, Yunfan Li, Zhenyu Huang, Zitao Liu, Peng Hu, and Xi Peng. Partially view-aligned representation learning with noise-robust contrastive loss. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1134–1143, 2021.

[Yang *et al.*, 2022] Mouxing Yang, Yunfan Li, Peng Hu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Robust multi-view clustering with incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):1055–1069, 2022.

[Yang *et al.*, 2023] Jing-Hua Yang, Le-Le Fu, Chuan Chen, Hong-Ning Dai, and Zibin Zheng. Cross-view graph matching for incomplete multi-view clustering. *Neurocomputing*, 515:79–88, 2023.

[Yang *et al.*, 2024] Zequn Yang, Han Zhang, Yake Wei, Zheng Wang, Feiping Nie, and Di Hu. Geometric-inspired graph-based incomplete multi-view clustering. *Pattern Recognition*, 147:110082, 2024.

[Ye *et al.*, 2017] Yongkai Ye, Xinwang Liu, Qiang Liu, and Jianping Yin. Consensus kernel k-means clustering for incomplete multiview data. *Computational intelligence and neuroscience*, 2017(1):3961718, 2017.

[Yu *et al.*, 2024] Shengju Yu, Siwei Wang, Pei Zhang, Miao Wang, Ziming Wang, Zhe Liu, Liming Fang, En Zhu, and Xinwang Liu. Dvsai: Diverse view-shared anchors based incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 16568–16577, 2024.

[Zeng *et al.*, 2023] Pengxin Zeng, Mouxing Yang, Yiding Lu, Changqing Zhang, Peng Hu, and Xi Peng. Semantic invariant multi-view clustering with fully incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[Zhao *et al.*, 2021a] Han Zhao, Xu Yang, Zhenru Wang, Erkun Yang, and Cheng Deng. Graph debiased contrastive learning with joint representation clustering. In *IJCAI*, pages 3434–3440, 2021.

[Zhao *et al.*, 2021b] Shuping Zhao, Lunke Fei, Jie Wen, Jigang Wu, and Bob Zhang. Intrinsic and complete structure learning based incomplete multiview clustering. *IEEE Transactions on Multimedia*, 25:1098–1110, 2021.

[Zhao *et al.*, 2023a] Liang Zhao, Xiao Wang, Zhenjiao Liu, Hong Yuan, Jingyuan Zhao, and Shuang Zhou. Deep probability multi-view feature learning for data clustering. *Expert Systems with Applications*, 217:119458, 2023.

[Zhao *et al.*, 2023b] Liang Zhao, Qiongjie Xie, Sontao Wu, and Shubin Ma. An end-to-end framework for partial view-aligned clustering with graph structure. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023.

[Zhao *et al.*, 2023c] Mingyu Zhao, Weidong Yang, and Feiping Nie. Deep graph reconstruction for multi-view clustering. *Neural Networks*, 168:560–568, 2023.

[Zhao *et al.*, 2024a] Liang Zhao, Xiao Wang, Zhenjiao Liu, Ziyue Wang, and Zhikui Chen. Learnable graph guided deep multi-view representation learning via information bottleneck. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[Zhao *et al.*, 2024b] Liang Zhao, Qiongjie Xie, Zhengtao Li, Songtao Wu, and Yi Yang. Dynamic graph guided progressive partial view-aligned clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[Zhao *et al.*, 2024c] Liang Zhao, Yukun Yuan, Qiongjie Xie, and Ziyue Wang. Anchor based multi-view clustering for partially view-aligned data. In *2024 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–5. IEEE, 2024.

[Zhou *et al.*, 2024] Lihua Zhou, Guowang Du, Kevin Lü, Lizheng Wang, and Jingwei Du. A survey and an empirical evaluation of multi-view clustering approaches. *ACM Computing Surveys*, 56(7):1–38, 2024.

[Zong *et al.*, 2021] Linlin Zong, Faqiang Miao, Xianchao Zhang, Xinyue Liu, and Hong Yu. Incomplete multi-view clustering with partially mapped instances and clusters. *Knowledge-Based Systems*, 212:106615, 2021.