Deep Fair Multi-View Clustering with Attention KAN

HaiMing Xu[1], Qianqian Wang[1,*], Boyue Wang[2], Quanxue Gao[1]
[1]Xidian University, [2]Beijing University of Technology
24011211044@stu.xidian.edu.cn, qqwang@xidian.edu.cn, wby@bjut.edu.cn, qxgao@xidian.edu.cn

Abstract

Multi-view clustering is effective in unsupervised multi-view data analysis and has received considerable attention. However, most existing methods excessively emphasize certain attributes, resulting in unfair clustering outcomes, i.e.*, certain sensitive attributes dominate the clustering results. Moreover, existing methods struggle to effectively capture complex nonlinear relationships and interactions across views, limiting their ability to achieve optimal clustering performance. Therefore, in this work, we propose a novel method, **D**eep **F**air **M**ulti-**V**iew **C**lustering with **A**ttention **K**olmogorov-**A**rnold **N**etwork (DFMVC-AKAN), to generate fair clustering results while maintaining robust performance. DFMVC-AKAN integrates attention mechanisms into Kolmogorov-Arnold Networks (KAN) to exploit the complex nonlinear inter-view relationships. Specifically, KAN provides a nonlinear feature representation capable of efficiently approximating arbitrary multivariate continuous functions, augmented by a hybrid attention mechanism which enables the model to dynamically focus on the most relevant features. Finally, we refine the clustering assignments with a distribution alignment module to ensure fair outcomes across diverse groups while maintaining discriminative ability. Experimental results on four datasets containing sensitive attributes demonstrate that DFMVC-AKAN significantly improves fairness and clustering performance compared to state-of-the-art methods.*

1. Introduction

Multi-view data refers to the representation of the same object from multiple perspectives or different sources and have become increasingly ubiquitous in practice [6, 9, 24, 41]. For instance, in medical imaging, a patient's condition can be characterized through multiple scanning modalities, such as CT, MRI, and X-ray. Each imaging modality captures different aspects of the patient's health. Compared with single-view data, such multi-view data provides complementary information, enabling a more comprehensive understanding of underlying patterns [2, 31].

Multi-view clustering (MVC), as a widely-applied unsupervised multi-view data analysis tool, aims to partition unlabeled multi-view datasets into distinct categories by fusing the information from all accessible views [5, 7, 28, 44]. It effectively uncovers a wider range of latent patterns by integrating consistent and complementary information from multiple perspective, thereby garnering increasing attention in various fields, *e.g.,* bioinformatics, computer vision, *etc.* Existing MVC methods can be broadly categorized into traditional approaches and deep learning-based approaches[20, 42]. Traditional methods primarily focus on leveraging linear transformation to integrate multiple views, including matrix factorization-based, graph-based, and subspace-based methods [8, 45]. While these traditional methods have demonstrated effectiveness in various applications, they often rely on linear assumptions and imposes limitations on their overall effectiveness in real-world multi-view data. This limitation has prompted the advancement of deep learning-based methods, which utilize the capabilities of neural networks to capture complex patterns and interactions across different views.

Deep learning-based MVC methods are able to capture nonlinear relationships and high-level feature representations from multi-view data [18, 33]. These methods can be broadly categorized into three categories: (1) autoencoder-based methods [35, 37], which utilize deep autoencoders to extract and fuse multi-view representations; (2) generative adversarial network (GAN)-based methods [27, 29], which employ adversarial training to align the distributions of different views and generate unified representations for clustering; (3) attention-based methods [11, 34], which leverages attention mechanism to dynamically adjust the importance of different views or features. Deep MVC has demonstrated superior performance in capturing complex data patterns, thereby achieving impressive clustering results.

Despite the significant progress in deep MVC, they generally place excessive emphasis on certain attributes, such as gender, race, age, *etc.*, introducing extra bias and resulting in unfair results [15, 21, 49]. For instance, in a health-

*Corresponding authors.

care scenario where patient data includes gender attribute as sensitive attribute, a clustering model might disproportionately group individuals based on sensitive attributes rather than medical conditions, resulting in biased treatment recommendations. To mitigate this problem, several fair deep MVC methods have been proposed [46, 48]. For example, Zheng *et al.* developed a fairness-aware MVC method that achieves fairness by enforcing a uniform distribution of sensitive attributes with a fairness constraint [48]. However, these methods enforce sensitive attributes equal in each cluster to realize ideal fairness, which might not align with their real cluster distribution, resulting a trade-off between clustering performance and fairness insurance. In addition, these methods extract features via either CNN or MLP, which cannot efficiently approximate the complex continuous function of several variables.

To tackle these problems, we propose DFMVC-AKAN, a deep fair multi-view clustering method with attention Kolmogorov-Arnold network. DFMVC-AKAN utilizes KAN to efficiently approximate the complex nonlinear relationship. Besides, it integrates attention mechanism with KAN, enabling the model to select features most relevant to clustering. Finally, a distribution alignment module is introduce to align sensitive attributes with the target distribution, which ensures fairness without reducing the clustering accuracy. The main contributions of this paper can be summarized as follows:

- We propose a deep fair MVC method named DFMVC-AKAN that uses target clustering distributions to ensure fairness by mitigating the impact of sensitive attributes.
- We develop attention KAN encoder to learn complex data relationships and extract multi-view features most relevant to clustering, ensuring the clustering performance.
- Experiments on four datasets with sensitive attributes demonstrate that DFMVC-AKAN significantly improves both fairness and clustering accuracy compared to existing methods.

2. Related Work

2.1. Multi-view Clustering

Multi-view Clustering (MVC) methods improve clustering accuracy by integrating comprehensive information from multiple views. They generally extract the consistent latent representations shared by all the views. For example, Zhao *et al.* [47] leveraged deep matrix decomposition to extract shared latent representations; Kumar and Rai [13] optimized kernels to improve clustering accuracy; Li *et al.* [18] use GANs to generate unified latent representations across views. Deep Multi-view Subspace Clustering with Unified and Discriminative Learning (DMSC-UDL) [26] integrates global and local structures with a self-expression layer, improving intra-cluster coherence. Similarly, Self-supervised Information Bottleneck for Deep Multi-view Subspace Clustering (SIB-MSC) [30] utilizes cross-view features as mutually supervised signals, applying the information bottleneck principle to enhance consistency between views and learn a purer affinity matrix. However, the ubiquitous sensitive attributes might influence the clustering process and results in biased clustering outcomes, while the above methods ignore this issue and thus confronts unfairness problem [32].

2.2. Fair Clustering

Fairness is a crucial issue in clustering, since traditional clustering methods might lead to sensitive attributes dominating clustering and biased results. Therefore, fair clustering was proposed. Traditional fair clustering is based on constraint optimization. For instance, Kleindessner *et al.* [12], directly incorporate fairness constraints into the objective function. Chierichetti *et al.* proposed Fair Clustering through Fairlets [4], which introduces fairness penalty terms during clustering and partition the dataset into smaller subsets, each satisfying predefined fairness constraints. However, this method has a high computational complexity as quadratic time complexity, which hinders its potential applicability. To address this issue, ScFC [1] introduced a tree-based metric approach to construct fairlet in near-linear time, which improves scalability and simultaneously guarantees fairness. Data reconstruction methods [3] achieves fairness by data pre-processing or resampling. There are also several deep fair clustering methods. For example, Li *et al.* [15] proposed a representation learning method that preserves group balance.

Some works concentrate on the fairness in MVC. For example, Wang *et al.* [32] introduced the iFiG method, which ensures fairness in multi-view graph clustering by balancing the distribution across views. Li *et al.* [16] developed a one-stage fair multi-view spectral clustering approach, optimizing a fair spectral objective. Yang *et al.* [40] introduced the Multi-view Fair-Augmentation Contrastive Graph Clustering method, which enhances fairness using contrastive graph clustering. Although these methods improve fairness in clustering, they always sacrifice clustering accuracy to achieve fairness. For another, they are on the basis of CNN or MLP, leading to them inefficiently approximating the inter-view relationship.

3. Methodology

Suppose a multi-view dataset comprising V distinct views, where each view is indexed by $v \in \{1, 2, \ldots, V\}$. The dataset is defined as $\mathcal{X} = \{\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^V, \mathbf{S}\}$, where $\mathbf{X}^v \in \mathbb{R}^{N \times d_v}$ represents the data matrix for the v-th view with N samples and d_v features, and $\mathbf{S} \in \mathbb{R}^{N \times d_s}$ represents sensitive attributes. Each individual sample from the v-th view is denoted as $\mathbf{x}_i^v$. DFMVC-AKAN aims to cate-

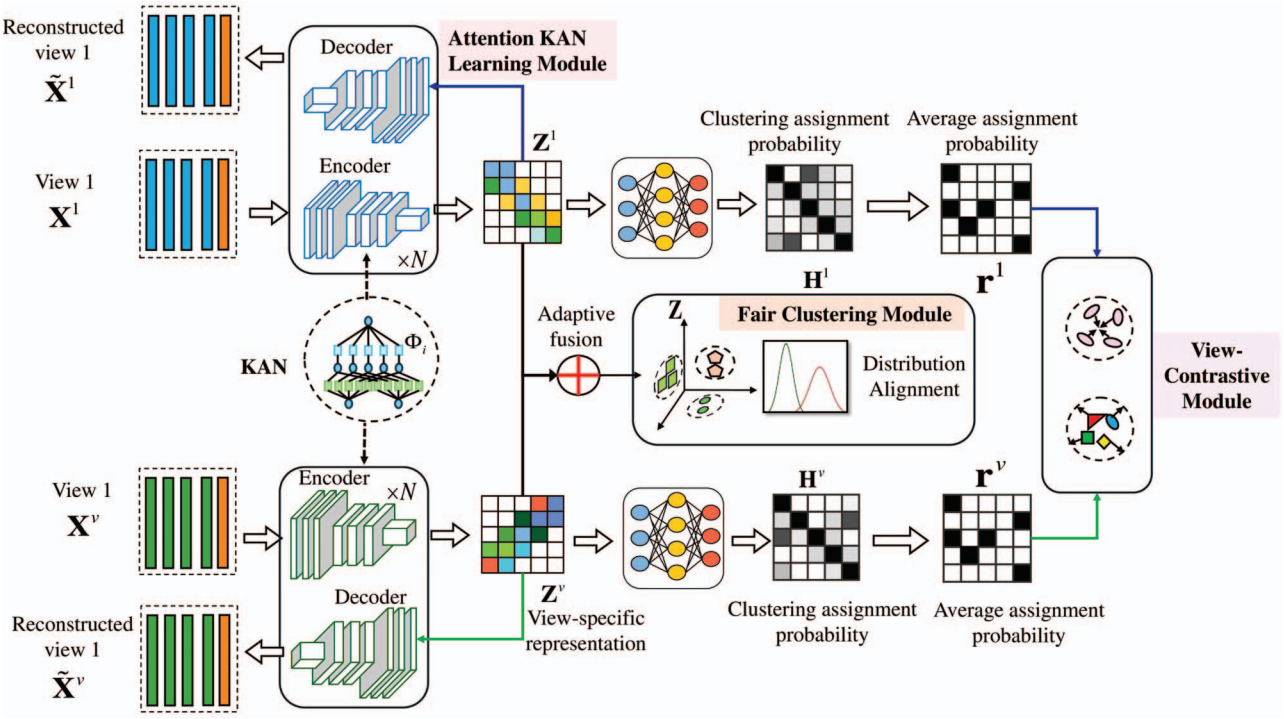


Figure 1. The architecture of our DFMVC-AKAN framework consists of three key modules: The Attention KAN Learning Module employs view-specific encoder-decoder pairs with hybrid attention mechanisms and KAN networks to extract robust features while minimizing reconstruction loss. The View-Contrastive Module generates label probabilities from view-specific representations and enforces semantic consistency across views. The Fair Clustering Module performs an adaptive fusion for multi-view representations and learns a unified embedding $\mathbf{Z}$, which is then processed through a distribution alignment Constraint that ensures fairness.

gorize these N samples into K distinct clusters while ensuring fairness by preventing sensitive attributes from disproportionately influencing the clustering outcomes, which is composed of three modules, *i.e.,* Attention KAN Learning Module responsible for extracting view-specific features, View-Contrastive Module enforcing semantic consistency across views, and Fair Clustering Module ensuring equitable representation across protected groups.

3.1. Attention KAN Learning Module

Extracting robust features from multi-view data with noise and redundancy is a persistent challenge in representation learning. We propose an Attention KAN Learning Module to address this issue, leveraging its ability to model complex distributions through learnable univariate functions. This module is composed of multiple view-specific encoders and decoders. For the v-th view encoder $\mathcal{E}_v$, it processes the input $\mathbf{x}_i^v$ through a series of KAN layers, augmented by our hybrid attention mechanism, and generates view-specific representations, *i.e.,*

$$\mathbf{z}_i^v = \mathcal{E}_v(\mathbf{x}_i^v), \tag{1}$$

Table 1. Basic notations used in the paper.

Notation	Meaning
$\mathbf{X}^v$	Data matrix for view v
$\mathbf{x}_i^v$	Sample i from view v
$\mathbf{S}$	Sensitive attribute matrix
$\mathbf{Z}^v$	Latent representation for view v
$\mathbf{z}_i^v$	Latent features of sample i in view v
$\tilde{\mathbf{x}}_i^v$	Reconstructed sample i in view v
$\mathbf{H}^v$	Cluster assignment matrix for view v
$\mathbf{h}_i^v$	Cluster probabilities for sample i in view v
$\mathbf{Z}$	Unified latent representation
$\mathbf{c}_j$	Centroid of cluster j
$\mathbf{Q}$	Soft cluster assignment matrix
$\mathbf{Q}_{ij}$	Probability of sample i in cluster j
$\mathbf{P}$	Target distribution matrix
$\mathbf{P}_{ij}$	Target probability of sample i in cluster j
X_g	Subgroup with same sensitive attribute value

where $\mathbf{z}_i^v \in \mathbb{R}^{d_z}$ is the learned essential features of the input $\mathbf{x}_i^v$, with d_z typically smaller than d_v. Next, we will introduce the component of the encoder in detail.

Hybrid Attention Mechanism: The first component in our encoder is a novel hybrid attention mechanism that combines Squeeze-and-Excitation (SE)[10] with Multi-head Attention[25]. Given the input $\mathbf{x}_i^v$, the SE component recalibrates channel-wise feature responses:

$$\mathbf{SE}_i^v = \mathbf{x}_i^v \odot \sigma(\mathbf{W}_2\delta(\mathbf{W}_1\mathbf{x}_i^v)), \tag{2}$$

where σ is the sigmoid function, δ is ReLU, and $\mathbf{W}_1 \in \mathbb{R}^{\frac{d_v}{r}\times d_v}$, $\mathbf{W}_2 \in \mathbb{R}^{d_v\times\frac{d_v}{r}}$ are learnable weights with reduction ratio r, and $\odot$ denotes element-wise multiplication.

The Multi-head Attention component extends $\mathbf{SE}_i^v$ by capturing complex feature relationships. For each attention head $t \in \{1, ..., T\}$, we compute:

$$\mathbf{A}_t^v = \mathbf{W}_t^A\mathbf{SE}_i^v, \quad \mathbf{B}_t^v = \mathbf{W}_t^B\mathbf{SE}_i^v, \quad \mathbf{C}_t^v = \mathbf{W}_t^C\mathbf{SE}_i^v, \tag{3}$$

where $\mathbf{W}_t^A$, $\mathbf{W}_t^B$, and $\mathbf{W}_t^C$ are learnable projection matrices.

The attention output for each head is:

$$\mathbf{o}_t^v = \mathbf{A}_t^v \cdot \varphi(\mathbf{B}_t^v) \cdot \eta(\mathbf{C}_t^v), \tag{4}$$

where φ and η are normalization functions (implemented as softmax) that distribute attention weights across features. The outputs from all heads are concatenated and projected:

$$\mathbf{MHA}_i^v = \mathbf{W}_{recon}[\mathbf{o}_1^v; \mathbf{o}_2^v; ...; \mathbf{o}_T^v], \tag{5}$$

where W_{recon} is a learnable projection matrix.

The final attention output is:

$$\mathbf{a}_i^v = \alpha \cdot \mathbf{SE}_i^v + (1 - \alpha) \cdot \mathbf{MHA}_i^v, \tag{6}$$

where α is a learnable parameter balancing both attention mechanisms.

KAN Layer: After the attention mechanism, the enhanced features $\mathbf{a}_i^v$ are processed by KAN layers to generate the final representation $\mathbf{z}_i^v$. KAN is inspired by the Kolmogorov-Arnold theorem[22], which demonstrates that any multivariate function can be decomposed into univariate components. The computation in KAN layers follows:

$$\mathbf{z}_i^v = \phi\left(\sum_{m=1}^{d_v} \psi_m(\mathbf{a}_{i,m}^v)\right), \tag{7}$$

where $\mathbf{a}_{i,m}^v$ represents the m-th element of the attention-enhanced input vector $\mathbf{a}_i^v$ for view v. Here, ψ_m denotes a learnable function tailored to process each individual input dimension, while ϕ is another learnable function that acts as an activation, applied to the summed contributions of all dimensions. This formulation empowers KAN to capture intricate patterns in the data by dynamically adapting its functional components to the specific characteristics of each view.

The attention mechanism and KAN layers work in synergy: attention selectively emphasizes the most informative dimensions, while KAN models complex relationships between them, ensuring that $\mathbf{z}_i^v$ is both concise and expressive.

The decoder $\mathcal{D}_v$ mirrors the encoder structure to reconstruct the original input:

$$\tilde{\mathbf{x}}_i^v = \mathcal{D}_v(\mathbf{z}_i^v), \tag{8}$$

where $\tilde{\mathbf{x}}_i^v \in \mathbb{R}^{d_v}$ is the reconstructed sample. The decoder first applies attention to $\mathbf{z}_i^v$ to focus on the most relevant latent features, followed by KAN layers that map the latent representation back to the original input space.

To optimize the representation and reconstruction pipeline, we define the reconstruction loss across all V views and N samples:

$$\mathcal{L}_r = \sum_{v=1}^{V}\sum_{i=1}^{N} \|\mathbf{x}_i^v - \tilde{\mathbf{x}}_i^v\|_2^2, \tag{9}$$

where $\|\cdot\|_2$ is the Euclidean norm. This loss quantifies the discrepancy between the original input and its reconstruction, driving the model to preserve critical information while achieving an effective latent representation.

3.2. View-Contrastive Module

While reconstruction alone captures the structural fidelity of multi-view data, it falls short in ensuring semantic consistency across views and distinguishing between samples effectively. To address this, we introduce the View-Contrastive Module. This module leverages embedded features to enforce view-shared consistency and inter-cluster distinctiveness, thereby enhancing the robustness of representations for clustering

This module takes as input the data for each view v, denoted $\mathbf{X}^v$, and transforms it into an embedding $\mathbf{Z}^v = \mathcal{E}_v(\mathbf{X})$. These embeddings are subsequently processed through two linear layers followed by a softmax activation to yield the clustering assignment probability matrix $\mathbf{H}^v \in \mathbb{R}^{N\times K}$:

$$\mathbf{H}^v = f(\mathbf{Z}^v, \theta), \tag{10}$$

where N is the number of samples, K is the number of clusters, and θ denotes the parameters of the linear transformations. Each row $\mathbf{h}_i^v$ of $\mathbf{H}^v$ encapsulates the cluster assignment probability vector for the i-th sample in the v-th view, with the element $\mathbf{H}_{ij}^v$ representing the probability that sample i belongs to cluster j. The semantic label for each sample is inferred by selecting the cluster with the highest probability in $\mathbf{h}_i^v$.

To assess semantic consistency across views, the module computes the similarity between cluster assignment vectors from different views, $\mathbf{h}_j^u$ and $\mathbf{h}_j^v$ (u and v represent different views), using their dot product:

$$\text{sim}(\mathbf{h}_j^u, \mathbf{h}_j^v) = (\mathbf{h}_j^u)^\top \mathbf{h}_j^v. \tag{11}$$

This metric quantifies the alignment of cluster assignments for the same sample across views. Within this paradigm, pairs of assignment vectors from different views of the same sample are designated as positive pairs (reflecting their shared underlying identity), while pairs from distinct samples (whether within or across views) are treated as negative pairs. For a given view v, each positive pair $(\mathbf{h}_j^u, \mathbf{h}_j^v)$ is associated with $(V-1)$ positive pairs and $V(K-1)$ negative pairs (where V denotes the total number of views).

To reinforce the alignment of positive pairs, a specialized loss function is introduced. For a positive pair $(\mathbf{h}_j^u, \mathbf{h}_j^v)$, an intermediate term Ω is defined as:

$$\Omega = \sum_{k=1,k\neq j}^{K} e^{\text{sim}(\mathbf{h}_j^u, \mathbf{h}_k^v)/\tau} + \sum_{k=1}^{K} e^{\text{sim}(\mathbf{h}_j^u, \mathbf{h}_k^v)/\tau} - e^{1/\tau}, \quad (12)$$

where τ is a temperature parameter modulating the sharpness of the exponential distribution. The loss for an individual positive pair is then formulated as:

$$l(u,v) = -\frac{1}{K}\sum_{j=1}^{K} \log \frac{e^{\text{sim}(\mathbf{h}_j^u, \mathbf{h}_j^v)/\tau}}{\Omega}. \quad (13)$$

This loss function incentivizes the model to maximize the similarity between cluster assignments of the same sample across views, thereby enhancing consistency. The overarching Semantic Contrastive Loss L_c comprises two complementary components. The first, termed the Clustering Consistency Loss, enforces alignment by minimizing the divergence between cluster assignments of identical samples across views while maximizing separation from dissimilar clusters:

$$L_{c1} = \frac{1}{2}\sum_{u=1}^{V}\sum_{v=1,v\neq u}^{V} l(u,v). \quad (14)$$

The second component is a regularization term that prevents the degenerate case where all samples collapse into a single cluster. Defining $\mathbf{r}_j^v = \frac{1}{N}\sum_{i=1}^{N} \mathbf{H}_{ij}^v$ as the average assignment probability for cluster j in view v, the regularization term is expressed as:

$$L_{c2} = \sum_{v=1}^{V_n}\sum_{j=1}^{K} r_j^v \log r_j^v. \quad (15)$$

The total Semantic Contrastive Loss is then the summation of these terms:

$$L_c = L_{c1} + L_{c2}. \quad (16)$$

3.3. Fair Clustering Module

While the reconstruction and contrastive learning frameworks in the preceding sections effectively generate multi-view embeddings and enforce semantic consistency through clustering, they do not inherently guarantee fairness in the resulting cluster assignments with respect to sensitive attributes. To address this limitation, we propose a fair clustering module that refines these assignments to ensure equitable outcomes across diverse groups, preserving both discriminative power and impartiality.

After obtaining view-specific representations from each encoder, we fuse them into a unified embedding $\mathbf{Z} \in \mathbb{R}^{N\times d_z}$ using learned importance weights:

$$\mathbf{Z} = \frac{\sum_{v=1}^{V} a_v \mathbf{Z}^v}{\sum_{v=1}^{V} a_v} \quad (17)$$

where $a_v \in \mathbf{A} = [a_1, a_2, ..., a_V]$ are learnable parameters that weight each view according to its information content. Views containing more discriminative information receive higher weights in this fusion process.

The unified embedding $\mathbf{Z}$ is then passed to a clustering layer that computes soft assignments using a Student's t-distribution. For each sample i and cluster j, the assignment probability is:

$$\mathbf{Q}_{ij} = \frac{(1+\|\mathbf{z}_i - \mathbf{c}_j\|^2/\alpha)^{-\frac{\alpha+1}{2}}}{\sum_{j'=1}^{K}(1+\|\mathbf{z}_i - \mathbf{c}_{j'}\|^2/\alpha)^{-\frac{\alpha+1}{2}}} \quad (18)$$

where $\mathbf{c}_j$ represents the centroid of cluster j, and α is the degrees of freedom parameter.

Given a sensitive attribute that divides the dataset into protected subgroups X_g (e.g., males and females for gender), we define a balanced target distribution that prevents any cluster from being dominated by samples with the same sensitive attribute:

$$\mathbf{P}_{ij} = \frac{\mathbf{Q}_{ij}^2 / \sum_{i'\in X_g} \mathbf{Q}_{i'j}}{\sum_{j'=1}^{K}\left(\mathbf{Q}_{ij'}^2 / \sum_{i'\in X_g} \mathbf{Q}_{i'j'}\right)} \quad (19)$$

where X_g represents the subgroup that sample i belongs to, and $\sum_{i'\in X_g}$ denotes summation over all samples in that subgroup. For example, if the sensitive feature is gender, X_g could represent either the male or female subgroup. The squared term $\mathbf{Q}_{ij}^2$ amplifies high-confidence assignments, enhancing clustering decisiveness.

The squared term $\mathbf{Q}_{ij}^2$ amplifies high-confidence assignments, enhancing clustering decisiveness, while the group-specific normalization ensures that each protected subgroup's distribution is independently calibrated, preventing any single cluster from being dominated by samples with the same sensitive attribute.

The fairness-aware clustering objective is then defined as the KL divergence between the soft assignments $\mathbf{Q}$ and the target distribution $\mathbf{P}$:

$$\mathcal{L}_f = \sum_{g\in\{0,1\}}\sum_{i\in\mathcal{G}_g}\sum_{j=1}^{K} \mathbf{P}_{ij}\log\frac{\mathbf{P}_{ij}}{\mathbf{Q}_{ij}} \quad (20)$$

By minimizing this loss, we encourage the model to align cluster assignments with a distribution that inherently balances representation across sensitive attributes, yielding clusters that not only capture natural data patterns but also maintain equitable representation across protected groups.

3.4. Unified Optimization Objective

The proposed framework harmonizes three distinct objectives: the reconstruction loss $\mathcal{L}_r$ from the KAN-based multi-view representation module, the semantic contrastive loss $\mathcal{L}_c$ from the feature alignment and consistency module, and the fairness loss $\mathcal{L}_f$ from the fair clustering module. The total optimization objective is formulated as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_r + \lambda \mathcal{L}_c + \gamma \mathcal{L}_f, \tag{21}$$

where λ and γ are hyperparameters that regulate the trade-off between semantic consistency across views and fairness with respect to sensitive attributes. This unified objective enables the model to strike a delicate balance among reconstructing faithful multi-view representations, enforcing view-shared semantic coherence, and ensuring equitable clustering outcomes. Upon convergence, the soft cluster assignments $\mathbf{h}_i^v$—the i-th row of the clustering probability matrix $\mathbf{H}^v$ for view v, with elements $\mathbf{H}_{ij}^v$ summing to 1 across K clusters—yield semantic labels via:

$$y_i = \arg\max_j \left(\frac{1}{V} \sum_{v=1}^{V} \mathbf{H}_{ij}^v \right), \quad 1 \le i \le N, \tag{22}$$

where V is the number of views and N the number of samples.

4. Experiment

4.1. Datasets & Metric

In this study, we evaluate the performance of the DFMVC-AKAN model using four diverse fairness datasets, which span various real-world application scenarios. These include credit card default prediction [43], banking product marketing [43], law school candidate qualification prediction [14], and a synthetic dataset. The datasets differ in sample size, feature dimensions, and the choice of sensitive attributes, such as gender, marital status, and binary values. To model the complexity of multi-view data, we apply multiple nonlinear functions to each dataset, generating two distinct views.

To evaluate the performance of the DFMVC-AKAN model, we use two key metrics: Normalized Mutual Information (NMI) and Balance Level (BAL). These metrics are commonly used to assess clustering performance and fairness. NMI measures the agreement between the predicted clusters and the true labels, with a higher value indicating better alignment. BAL evaluates the equitable distribution of sensitive attributes across clusters, with a higher value reflecting a more balanced representation. Our metrics refer to the methodology of this DFMVC[46] article, which ensures a comprehensive assessment of clustering effectiveness and fairness.

4.2. Experiment Setup

Experiments were run on an NVIDIA GeForce RTX 4090 GPU (driver 552.41, CUDA 12.4) in a Windows server environment. Implemented using Python 3.10.13 with PyTorch, and MATLAB was used as a supplementary tool for simulation and validation.

4.2.1. Comparison methods

Our proposed DFMVC-AKAN method was benchmarked against a suite of state-of-the-art clustering techniques to evaluate its effectiveness and efficiency:

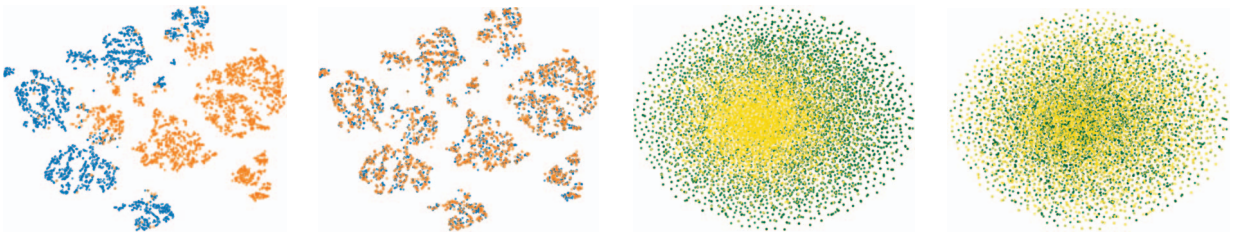
- **K-means** [23]: A classic clustering algorithm that partitions data into k clusters by assigning each point to the nearest cluster center based on Euclidean distance.
- **DEC** [36]: A deep learning method that combines feature representation learning with clustering by minimizing reconstruction loss in the embedded space.
- **CC** [17]: A clustering approach based on contrastive learning that optimizes instance-level and cluster-level contrastive losses to learn discriminative representations.
- **MvDSCN** [50]: A multi-view deep subspace clustering network that learns a self-representation matrix to capture the structure of multi-view data for enhanced clustering.
- **DCP** [19]: A multi-view clustering method leveraging contrastive learning and dual prediction modules to handle incomplete views and ensure consistent representations.
- **APADC** [39]: A subspace learning-based multi-view clustering method that aligns view distributions by minimizing disparity loss to improve clustering accuracy.
- **MFLVC** [38]: A multi-level feature learning method for contrastive multi-view clustering that captures both low-level and high-level features to improve the clustering process.
- **Fair-MVC** [48]: A fairness-aware multi-view clustering algorithm that integrates contrast regularization to ensure fair representation across protected groups.
- **DFMVC** [46]: A deep learning-based multi-view clustering approach that incorporates fairness constraints to ensure equal representation of different groups in the clustering results.

4.2.2. Parameter Settings

For the DFMVC-AKAN model, we used a training regimen combining standard training cycles and Progressive Training, a variant of Transfer Learning. We set a batch size of 100 and a learning rate of 0.0001. The model underwent two stages: a pre-training phase with 200 epochs us-

Table 2. Results on four datasets with sensitive features. The best results are highlighted in bold, while the second-best values are underlined. (A higher balance score indicates better fairness.)

Methods	Banking Market		Zafar		Credit Card		Law School	
Metrics (%)	NMI	BAL	NMI	BAL	NMI	BAL	NMI	BAL
K-means [23]	28.67 ± 1.44	37.64 ± 0.66	70.32 ± 0.78	17.06 ± 0.76	20.94 ± 1.14	35.53 ± 0.37	20.12 ± 1.25	43.22 ± 1.05
DEC [36]	30.93 ± 1.15	37.60 ± 0.96	72.55 ± 1.92	16.85 ± 0.73	21.03 ± 2.09	35.96 ± 0.60	21.23 ± 1.21	44.15 ± 1.24
CC [17]	36.23 ± 1.01	37.46 ± 0.97	78.95 ± 0.68	17.01 ± 0.71	23.87 ± 1.28	35.74 ± 0.47	23.02 ± 0.92	44.24 ± 1.16
MvDSCN [50]	36.24 ± 0.55	37.59 ± 0.67	76.91 ± 0.42	17.13 ± 0.65	21.92 ± 1.53	35.82 ± 0.41	22.06 ± 1.28	44.82 ± 0.98
DCP [19]	39.93 ± 1.84	26.75 ± 2.06	81.87 ± 1.57	21.65 ± 1.23	26.73 ± 0.26	24.19 ± 1.05	23.85 ± 1.26	35.76 ± 1.12
APADC [39]	40.62 ± 0.25	27.79 ± 2.59	72.38 ± 0.80	21.21 ± 0.57	23.07 ± 0.45	26.32 ± 0.22	22.08 ± 1.24	36.85 ± 0.56
MFLVC [38]	37.76 ± 1.15	38.64 ± 1.48	90.52 ± 1.42	27.16 ± 0.85	24.02 ± 0.42	36.09 ± 0.24	21.84 ± 1.52	43.87 ± 0.46
DFMVC [46]	54.62 ± 1.25	42.16 ± 0.82	93.93 ± 0.36	29.08 ± 0.28	**35.13 ± 0.62**	39.71 ± 0.38	**24.24 ± 0.32**	45.66 ± 0.38
Fair-MVC [48]	38.89 ± 0.91	40.75 ± 1.56	81.81 ± 0.57	28.32 ± 0.48	24.19 ± 0.51	37.23 ± 0.42	21.57 ± 0.92	44.79 ± 0.56
Our Method	**80.46 ± 0.51**	**42.52 ± 0.32**	**99.98 ± 0.02**	**29.39 ± 0.14**	34.84 ± 0.21	**39.98 ± 0.11**	23.64 ± 0.28	**46.01 ± 0.04**



(a) **Raw features** (b) **Fairness feature** (c) **Raw features** (d) **Fairness feature**

Figure 2. In this visualization, we apply the t-SNE algorithm to explore sensitive features in the Bank Marketing and Zafar datasets. For the Bank Marketing dataset, blue points represent unmarried individuals, while orange points represent married individuals. In the Zafar dataset, green points correspond to binary 0 values, and yellow points represent binary 1 values. The visualizations on the left show the original data, and the right visualizations display the clustering results after training.

Table 3. Statistics summary of four datasets.

Dataset	Samples	Sensitive Feature	Clusters
Bank Marketing	5000	Marital status	2
Zafar	10000	Binary value	2
Credit Card	5000	Gender	5
Law School	3600	Gender	2

ing autoencoder and reconstruction loss, followed by 200 epochs of fine-tuning. For the Credit Card and Law School datasets, we applied Progressive Training, saving model weights after pre-training and fine-tuning, and retraining from the checkpoint until performance was optimized.

4.3. Ablation Experimental Analysis

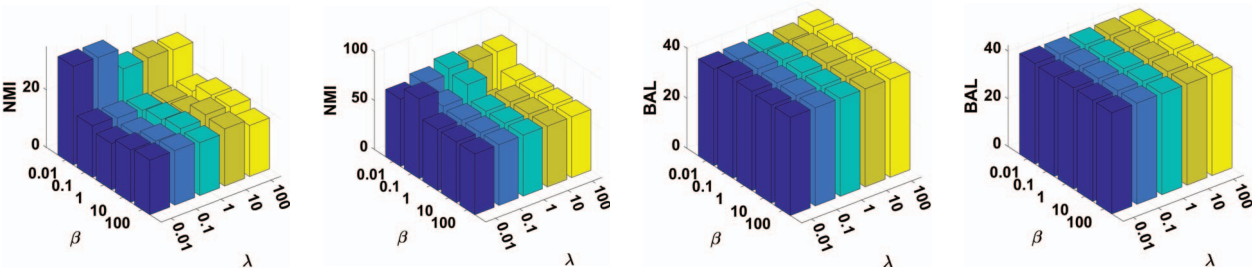
This section evaluates the impact of the semantic contrastive learning module (L_c) and the fairness learning module (L_f) on DFMVC-AKAN's performance, as shown in Table 4. The full DFMVC-AKAN model achieves the highest NMI and BAL scores. Removing either module results in significant performance drops. Excluding the fairness learning module decreases BAL, particularly on the "Banking Market" dataset (BAL = 41.59), highlighting its importance for group fairness. Removing the semantic contrastive learning module negatively affects clustering accuracy, emphasizing its role in capturing semantic relationships in the data.

4.4. Experimental Analysis

Visualization Analysis: Figure 2 presents t-SNE visualizations comparing raw features and fairness-processed features. In the Bank dataset, Fig.2a shows raw features where blue points (unmarried individuals) and orange points (married individuals) form distinct clusters with clear separation between groups, indicating that marital status strongly influences the feature distribution. After processing through DFMVC-AKAN (Fig.2b), these points become more intermixed with reduced separation, demonstrating the model's ability to diminish the influence of this sensitive attribute while maintaining overall data structure. Similarly, in Fig.2c and 2d, the raw features show some separation based on sensitive attributes, while the fairness-processed features exhibit more uniform distribution. This visual evidence confirms that DFMVC-AKAN effectively mitigates the impact of sensitive attributes in the feature representation, producing more equitable embeddings where clustering outcomes are less influenced by protected characteristics.

Table 4. Ablation study of the primary components in the proposed DFMVC-AKAN model across all datasets. The terms "Excl. Semantic" and "Excl. Fairness" denote model variants without the semantic contrastive learning and cluster distribution-guided fairness learning modules, respectively.

Model Variants	L_r	L_c	L_f	Banking Market		Zafar		Credit Card		Law School	
				NMI	BAL	NMI	BAL	NMI	BAL	NMI	BAL
Excl. Fairness	✓	✓	-	<u>76.36</u>	41.59	<u>95.23</u>	28.60	<u>24.69</u>	38.06	<u>21.44</u>	44.29
Excl. Semantic	✓	-	✓	59.73	<u>42.25</u>	86.66	<u>28.91</u>	19.98	<u>39.75</u>	17.85	<u>45.87</u>
DFMVC-AKAN Full	✓	✓	✓	**80.46**	**42.52**	**99.98**	**29.39**	**34.84**	**39.98**	**23.64**	**46.01**

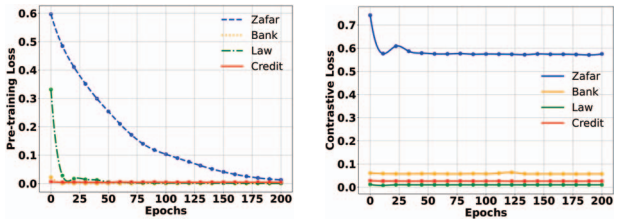


(a) **Credit Card NMI** (b) **Bank Marketing NMI** (c) **Credit Card Balance** (d) **Bank Marketing Balance**

Figure 3. The NMI and balance values yielded by the DFMVC-AKAN method with different α and β combinations on the datasets.

Hyper-parameter Analysis λ and β: In DFMVC-AKAN, we analyze how λ (reconstruction loss weight) and β (fairness loss weight) balance clustering accuracy (NMI) and fairness (BAL) across two datasets. For Bank Marketing, NMI peaks at 81.34% with $\lambda = 0.01, \beta = 0.1$, but stabilizes around 62% when $\lambda > 1$, indicating potential overfitting. BAL increases with β from 41.59% to 42.40% without significantly affecting NMI. The Credit Card dataset shows higher sensitivity: NMI reaches 34.16% at $\lambda = 0.01, \beta = 0.01$ but drops to 15.88% at $\lambda = 0.1, \beta = 0.1$, while BAL improves from 38.64% to 40.24% as β increases to 100. These findings suggest that dataset-specific tuning with fine-tuned λ and elevated β effectively optimizes the fairness-accuracy trade-off.

Convergence Analysis: Figure 4 shows the pre-training loss and comparison loss curves of the DFMVC-AKAN method on different datasets, reflecting the convergence performance of the model. The pre-training loss curves show that the model loss decreases and stabilizes with the training cycle on all datasets, indicating that the model is able to learn the low-dimensional representation of the data effectively. Similarly, the comparison loss curve gradually decreases during the training process, indicating that the model shows good convergence properties with increased consistency of feature representations across different views. These results validate the stability and effectiveness of the DFMVC-AKAN model in multi-view clustering tasks, which is able to achieve fast and stable convergence in the pre-training and contrast learning stages.



(a) **The pre-training loss curve** (b) **The contrastive loss curve**

Figure 4. Convergence results achieved through the DFMVC-AKAN method across all the datasets

5. Conclusions

In conclusion, our proposed fair multi-view clustering method utilizing the Kolmogorov-Arnold network (KAN) attention mechanism effectively addresses the limitations of existing approaches that enforce uniform distribution of sensitive attributes within clusters. By leveraging contrastive learning, we achieve consistent and differentiated representations across multiple views, allowing for a more nuanced understanding of the data. The incorporation of a fair loss function ensures that the distribution of sensitive attributes aligns with the target clustering distribution, thus enhancing both fairness and overall clustering performance. Our experimental results across four datasets demonstrate the significant improvements achieved by our method, establishing it as a robust solution for fair multi-view clustering in sensitive contexts.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant 62176203 and Grant 62102306, the Fundamental Research Funds for the Central Universities, the Natural Science Basic Research Program of Shaanxi Province (Grant 2023-JC-YB-534), and the Science and Technology Project of Xi'an (Grant 2022JH-JSYF-0009), Initiative Postdocs Supporting Program (Grant BX20190262), Open Project of Anhui Provincial Key Laboratory of Multimodal Cognitive Computation, Anhui University (No. MMC202416), Selected Support Project for Scientific and Technological Activities of Returned Overseas Chinese Scholars in Shaanxi Province 2023-02 and the Innovation Fund of Xidian University (Project NoYJSJ25007).

References

[1] Arturs Backurs, Piotr Indyk, Krzysztof Onak, Baruch Schieber, Ali Vakilian, and Tal Wagner. Scalable fair clustering. In *International Conference on Machine Learning*, pages 405–413. PMLR, 2019. 2

[2] Jie Chen, Shengxiang Yang, Hua Mao, and Conor Fahy. Multiview subspace clustering using low-rank representation. *IEEE Transactions on Cybernetics*, 52(11):12364–12378, 2021. 1

[3] Anshuman Chhabra, Peizhao Li, Prasant Mohapatra, and Hongfu Liu. Robust fair clustering: A novel fairness attack and defense framework. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. 2

[4] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. Fair clustering through fairlets. *Advances in neural information processing systems*, 30, 2017. 2

[5] Anal Roy Chowdhury, Avisek Gupta, and Swagatam Das. Deep multi-view clustering: A comprehensive survey of the contemporary techniques. *Information Fusion*, page 103012, 2025. 1

[6] F Dornaika, S El Hajjar, J Charafeddine, and N Barrena. Unified multi-view data clustering: Simultaneous learning of consensus coefficient matrix and similarity graph. *Cognitive Computation*, 17(1):38, 2025. 1

[7] Wei Feng, Guoshuai Sheng, Qianqian Wang, Quanxue Gao, Zhiqiang Tao, and Bo Dong. Partial multi-view clustering via self-supervised network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 11988–11995, 2024. 1

[8] Yifan Gao, Zhan Cao, and Shanshan Ma. A multi-view fuzzy compactness and separation co-clustering algorithm. In *Proceedings of the ACM International Conference on Knowledge Discovery and Data Mining*, pages 1842–1850, 2022. 1

[9] Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. Trusted multi-view classification with dynamic evidential fusion. *IEEE transactions on pattern analysis and machine intelligence*, 45(2):2551–2566, 2022. 1

[10] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018. 4

[11] Zongmo Huang, Yazhou Ren, Xiaorong Pu, Shudong Huang, Zenglin Xu, and Lifang He. Self-supervised graph attention networks for deep weighted multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, pages 7936–7943, 2023. 1

[12] Matthäus Kleindessner, Samira Samadi, Pranjal Awasthi, and Jamie Morgenstern. Guarantees for spectral clustering with fairness constraints. In *International conference on machine learning*, pages 3458–3467. PMLR, 2019. 2

[13] Abhishek Kumar, Piyush Rai, and Hal Daumé. Co-regularized multi-view spectral clustering. In *Advances in Neural Information Processing Systems*, pages 1413–1421, 2011. 2

[14] Tai Le Quy, Arjun Roy, Vasileios Iosifidis, Wenbin Zhang, and Eirini Ntoutsi. A survey on datasets for fairness-aware machine learning. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(3):e1452, 2022. 6

[15] Peizhao Li, Han Zhao, and Hongfu Liu. Deep fair clustering for visual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9070–9079, 2020. 1, 2

[16] Rongwen Li, Haiyang Hu, Liang Du, Jiarong Chen, Bingbing Jiang, and Peng Zhou. One-stage fair multi-view spectral clustering. In *ACM Multimedia 2024*, 2024. 2

[17] Yunfan Li, Peng Hu, Zitao Liu, Dezhong Peng, Joey Tianyi Zhou, and Xi Peng. Contrastive clustering. In *Proceedings of the AAAI conference on artificial intelligence*, pages 8547–8555, 2021. 6, 7

[18] Zhilong Li, Qi Wang, Zhiqiang Tao, Qiyu Gao, and Zongming Yang. Deep adversarial multi-view clustering network. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 3995–4001, 2019. 1, 2

[19] Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 4447–4461, 2022. 6, 7

[20] Hong Liu and Yun Fu. Consensus guided multi-view clustering. *Proceedings of the ACM on Knowledge Discovery and Data Mining*, pages 933–941, 2018. 1

[21] Zixin Liu, Xin Zhang, and Bin Jiang. Active learning with fairness-aware clustering for fair classification considering multiple sensitive attributes. *Information Sciences*, 2023. 1

[22] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljai, Thomas Y. Hou, and Max Tegmark. Kan: Kolmogorov-arnold networks. *ICLR 2025 Oral*, 2024. 4

[23] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, pages 281–297. Oakland, CA, USA, 1967. 6, 7

[24] Chang Tang, Zhenglai Li, Jun Wang, Xinwang Liu, Wei Zhang, and En Zhu. Unified one-step multi-view spectral clustering. *IEEE Transactions on Knowledge and Data Engineering*, 35(6):6449–6460, 2022. 1

[25] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 4

[26] Qianqian Wang, Jiafeng Cheng, Quanxue Gao, Guoshuai Zhao, and Licheng Jiao. Deep multi-view subspace clustering with unified and discriminative learning. *IEEE Transactions on Multimedia*, 23:3483–3493, 2020. 2

[27] Qianqian Wang, Zhengming Ding, Zhiqiang Tao, Quanxue Gao, and Yun Fu. Generative partial multi-view clustering with adaptive fusion and cycle consistency. *IEEE Transactions on Image Processing*, 30:1771–1783, 2021. 1

[28] Qianqian Wang, Zhiqiang Tao, Quanxue Gao, and Licheng Jiao. Multi-view subspace clustering via structured multi-pathway network. *IEEE Transactions on Neural Networks and Learning Systems*, 35(5):7244–7250, 2022. 1

[29] Qianqian Wang, Zhiqiang Tao, Wei Xia, Quanxue Gao, Xiaochun Cao, and Licheng Jiao. Adversarial multiview clustering networks with adaptive fusion. *IEEE transactions on neural networks and learning systems*, 34:7635–7647, 2023. 1

[30] Shiye Wang, Changsheng Li, Yanming Li, Ye Yuan, and Guoren Wang. Self-supervised information bottleneck for deep multi-view subspace clustering. *IEEE TIP*, 32:1555–1567, 2023. 2

[31] Siwei Wang, Xinwang Liu, Qing Liao, Yi Wen, En Zhu, and Kunlun He. Scalable multi-view graph clustering with cross-view corresponding anchor alignment. *IEEE Transactions on Knowledge and Data Engineering*, 2025. 1

[32] Yian Wang, Jian Kang, Yinglong Xia, Jiebo Luo, and Hanghang Tong. ifig: Individually fair multi-view graph clustering. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 329–338. IEEE, 2022. 2

[33] Jie Wen, Chengliang Liu, Shijie Deng, Yicheng Liu, Lunke Fei, Ke Yan, and Yong Xu. Deep double incomplete multi-view multi-label learning with incomplete labels and missing views. *IEEE transactions on neural networks and learning systems*, 2023. 1

[34] Danyang Wu, Xia Dong, Feiping Nie, Rong Wang, and Xuelong Li. An attention-based framework for multi-view clustering on grassmann manifold. *Pattern Recognition*, 128: 108610, 2022. 1

[35] Si-Jia Xiang, Heng-Chao Li, Jing-Hua Yang, and Xin-Ru Feng. Dual auto-weighted multi-view clustering via autoencoder-like nonnegative matrix factorization. *Information Sciences*, 667:120458, 2024. 1

[36] Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*, pages 478–487. PMLR, 2016. 6, 7

[37] Jie Xu, Yazhou Ren, Guofeng Li, Lili Pan, Ce Zhu, and Zenglin Xu. Deep embedded multi-view clustering with collaborative training. *Information Sciences*, 573:279–290, 2021. 1

[38] Jie Xu, Huayi Tang, Yazhou Ren, Liang Peng, Xiaofeng Zhu, and Lifang He. Multi-level feature learning for contrastive multi-view clustering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16051–16060, 2022. 6, 7

[39] Jie Xu, Chao Li, Liang Peng, Yazhou Ren, Xiaoshuang Shi, Heng Tao Shen, and Xiaofeng Zhu. Adaptive feature projection with distribution alignment for deep incomplete multi-view clustering. *IEEE Transactions on Image Processing*, 32:1354–1366, 2023. 6, 7

[40] Shaochen Yang, Zhaorun Liao, Runyu Chen, Yuren Lai, and Wei Xu. Multi-view fair-augmentation contrastive graph clustering with reliable pseudo-labels. *Information Sciences*, 674:120739, 2024. 2

[41] Yan Yang and Hao Wang. Multi-view clustering: A survey. *Big data mining and analytics*, 1(2):83–107, 2018. 1

[42] Shixin Yao, Guoxian Yu, Jun Wang, Carlotta Domeniconi, and Xiangliang Zhang. Multi-view multiple clustering. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 4121–4127, 2019. 1

[43] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rogriguez, and Krishna P Gummadi. Fairness constraints: Mechanisms for fair classification. In *Artificial intelligence and statistics*, pages 962–970. PMLR, 2017. 6

[44] Chaoyang Zhang, Zhengzheng Lou, Qinglei Zhou, and Shizhe Hu. Multi-view clustering via triplex information maximization. *IEEE Transactions on Image Processing*, 32: 4299–4313, 2023. 1

[45] Xianliang Zhang, Xiaotong Zhang, and Shuhui Liu. Multi-task multi-view clustering for non-negative data. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 4043–4049, 2015. 1

[46] Bowen Zhao, Qianqian Wang, Zhiqiang Tao, Wei Feng, and Quanxue Gao. Dfmvc: Deep fair multi-view clustering. In *ACM Multimedia 2024*, 2024. 2, 6, 7

[47] Handong Zhao, Zhengming Ding, and Yun Fu. Multi-view clustering via deep matrix factorization. In *Proceedings of the AAAI conference on artificial intelligence*, 2017. 2

[48] Lecheng Zheng, Yada Zhu, and Jingrui He. Fairness-aware multi-view clustering. In *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*, pages 856–864. SIAM, 2023. 2, 6, 7

[49] Zhuoping Zhou, Davoud Ataee Tarzanagh, Bojian Hou, Boning Tong, Jia Xu, Yanbo Feng, Qi Long, and Li Shen. Fair canonical correlation analysis. *Advances in Neural Information Processing Systems*, 36, 2024. 1

[50] Pengfei Zhu, Binyuan Hui, Changqing Zhang, Dawei Du, Longyin Wen, and Qinghua Hu. Multi-view deep subspace clustering networks. *CoRR*, abs/1908.01978, 2019. 6, 7