Adversarial Fair Incomplete Multi-View Clustering

Qianqian Wang[1], **Haiming Xu**[1*], **Wei Feng**[2], **Quanxue Gao**[1]

[1]School of Telecommunications Engineering, Xidian University, Xi'an, Shaanxi, China, 710071
[2]College of Information Engineering, Northwest A&F University, Yangling, China, 712100
qqwang@xidian.edu.cn, 24011211044@stu.xidian.edu.cn, wei.feng@nwafu.edu.cn, qxgao@xidian.edu.cn

Abstract

Incomplete multi-view clustering gains increasing attention by considering the view-missing issue in practical applications. However, existing methods impute the shared representation from a sample's unmissing views, which neglects the inter-view distribution difference and result in semantic bias. Additionally, to ensure model fairness, most methods sacrifices clustering performance and are difficult to trade off the relationships between clustering and fairness. In this paper, we propose a novel Adversarial Fair Incomplete Multi-View Clustering (AFIMVC). To accurately learn the shared representation, we leverage contextual information from view-complete representations to enhance view-incomplete ones with Cross-Sample Attention mechanism. Besides, we introduce feature-level and cluster-level alignment constraints to exploit consistent and discriminative information. Finally, we develop an adaptive fairness disentangled learning module, which is composed of adversarial disentangled mechanism and adaptive weight learning. The adversarial disentangled mechanism is implemented by an adversarial network with a gradient reversion. They make the shared representation independent of sensitive attributes and dynamically trade off clustering accuracy and fairness. Extensive experiments on three benchmark datasets demonstrate the superiority of the proposed method compared with the state-of-the-art methods.

Introduction

Multi-view data has become ubiquitous due to the development of data sensing technology, and multi-view learning receive considerable attention across various domains of pattern recognition and data mining. In order to process unlabeled multi-view data, multi-view clustering (MVC) (Yao et al. 2019; Liu and Fu 2018) and is a basic technique of multi-view learning. It can effectively exploit the consistent and complementary information of multiple views to separate them into different groups (Yu et al. 2024; Li et al. 2024c)

Although MVC attains outstanding clustering performance, it relies on view completeness to achieve optimal results. However, real-world data comes from multiple sources probably confronts the view-missing problem. For instance, remote sensing multi-view data composed of hyperspectral images and radar echo may suffer from missing partial spectral band due to sensor failure and cloud occlusion; In image-text retrieval, either images or texts might be missing due to transmission issues or storage medium damage. These scenarios undermine the efficacy of existing MVC methods. Consequently, numerous Incomplete Multi-View Clustering (IMVC) are developed to address the view-missing problem (Yan et al. 2023; Chen et al. 2023). One category of methods imputes missing samples via matrix completion or generative adversarial network (GAN) (Wang et al. 2021a). The main weaknesses of these methods lie in that the samples imputed by them generally deviate from the real values, and the generative models require massive paired data with complete views to train the network. The other category of methods imputes a shared representation from a sample's unmissing views by subspace learning (Li et al. 2024a), non-negative matrix factorization (NMF) (Wen et al. 2023), kernel-based learning (Xia et al. 2023), and spectral learning (Yin, Cai, and Sun 2022).

Nevertheless, current IMVC approaches often neglect the biases introduced by sensitive attributes like race and gender, resulting unfair clustering results (Backurs et al. 2019; Zhou et al. 2024; Kleindessner et al. 2019). The fairness issue becomes more critical in incomplete multi-view scenarios, which motivates several fair IMVC methods developed (Wang et al. 2025). For example, fair-aware IMVC method incorporates fairness constraints to ensure the sensitive attributes uniformly distributed across different clusters (Zheng, Zhu, and He 2023); FIMVC-DA maximizes mutual information and aligns the distributions of different views to ensure clustering results independent of sensitive features (Wang et al. 2025). Although these Fair IMVC methods effectively improve fairness, they still confront several fundamental limitations. First, enforcing the uniform distribution of sensitive attributes typically degrades clustering accuracy, and existing methods do not establish an optimal trade-off between clustering accuracy and fairness. Second, the methods of imputing shared representation leverage unmissing views' representation to impute the missing view's representation, which ignores the distribution difference in different views and thus leads to incorrect clustering assignment.

To address these limitations, we propose a novel Fair

*Corresponding Author

IMVC method named **A**dversarial **F**air **I**ncomplete **M**ulti-**V**iew **C**lustering (AFIMVC). We introduce a novel adaptive adversarial disentanglement module, which employs a adversarial network as sensitive attribute discriminator to enforce learning representations invariant to sensitive attributes. The module adopts gradient revision layer (GRL) to achieve end-to-end adversarial training, and it dynamically adjusts the influence of fairness constraint based on fairness performance. Therefore, the module addresses the trade-off problem between cluster accuracy and fairness. Additionally, to alleviate the semantic bias during imputing missing view's representation, our method utilizes contextual information from view-complete representations to impute missing view's representation with Cross-Sample Attention mechanism. Besides, our method introduces feature-level and cluster-level alignment constraints to exploit consistent and discriminative information. The main contributions of this paper are summarized as follows:

- We propose a novel Fair IMVC method, which aims to learn representations independent of sensitive attributes and establish an optimal trade-off between clustering accuracy and fairness.
- We design a context-aware representation imputation method that enhances the representations of missing-view by Cross-Sample Attention mechanism that leverages information from their complete-view representation.
- We introduce cluster-level contrastive learning strategy to enhance semantic consistency in the cluster assignments.

Related Work

Incomplete Multi-view Clustering

MVC exhibits impressive clustering results by leveraging information from multiple data representations, and IMVC further addresses the issue of view incompleteness, thus attracting extensive attention in academia. For instance, Wang et al. (2021a) proposed the Generative Partial Multi-View Clustering (GP-MVC) model, which extracts high-quality representations via adaptive fusion and cycle consistency constraints on the generation process, thus effectively addressing the IMVC problem. Lin et al. (2023) developed the Dual Contrastive Prediction (DCP) method, which enhances clustering consistency and the quality of missing view imputation by introducing a prototype-level contrastive loss. Xu et al. (2023) proposed the Adaptive Prototype-Aware Deep Clustering (APADC) method, dynamically handling the semantic importance of missing views through view-specific adaptive prototype fusion. Zhang et al. (2024) proposed an attention-driven direct contrastive learning framework which mitigates view-specific interference via self-attention losses and recovers missing data through cross-view prediction. However, existing methods still face challenges in exploring accurate representation for missing views.

Fair Multi-view Clustering

Fairness is a critical issue in MVC. Wang et al. (2022b) introduced the iFiG method, which ensures fairness in multi-view graph clustering by balancing the distribution across views. You et al. (2023) proposed a fairness-aware scalable multi-view bipartite graph clustering method, incorporating prior supervision for better fairness. Li et al. (2024b) developed a one-stage fair multi-view spectral clustering approach, optimizing a fair spectral objective. Yang et al. (2024) 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 do not address the challenges posed by missing data.

Methodology

Problem Formulation

Suppose $\mathcal{X}' = \{\mathbf{X}'^{(1)}, \mathbf{X}'^{(2)}, \ldots, \mathbf{X}'^{(V)}\}$ denotes a multi-view dataset comprising N samples from V views; $\mathbf{X}^{(v)} \in \mathbb{R}^{N\times d_v}$ is the data matrix of the v-th view, where d_v is the feature dimension of the v-th view; $\mathbf{x}_i^{(v)}$ denotes the i-th sample in the v-th view. To account for the missing-view issues, we introduce a binary mask matrix $\mathbf{M} \in \{0,1\}^{N\times V}$ with $\mathbf{M}_{iv} = 0$ indicating $\mathbf{x}_i^v$ is missing. Then, the incomplete multi-view data can be represented by $\mathcal{X} = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(V)}\}$, where $\mathbf{X}^{(v)} = \mathrm{diag}(\mathbf{m}_v) \cdot \mathbf{X}'^{(v)}$ and $\mathbf{m}_v \in \{0,1\}^{N\times 1}$ is the v-th column of $\mathbf{M}$. We further divide $\mathbf{X}^{(v)}$ into view-complete part $\mathbf{X}_{\mathrm{C}}^{(v)}$ and view-incomplete part $\mathbf{X}_{\mathrm{I}}^{(v)}$. Furthermore, we introduce a vector $\mathbf{s} \in \{0,1\}^{N\times 1}$, and $\mathbf{s}_i \in \{0,1\}$ denotes the binary sensitive attribute of the i-th sample. This attribute partitions the entire dataset into two disjoint protected groups: $\mathcal{G}_0 = \{i \mid s_i = 0\}$ and $\mathcal{G}_1 = \{i \mid s_i = 1\}$.

This work aims to learn a clustering function $f : (\mathcal{X}, \mathbf{M}) \to \mathbf{y}$ that maps the incomplete multi-view data to a cluster assignment vector $\mathbf{y} \in \{1, \ldots, K\}^N$, where K is the cluster number. The core challenge lies in learning a function that simultaneously optimizes two competing objectives: achieving high **clustering quality** by accurately reflecting the intrinsic data structure and ensuring **fairness** by guaranteeing that the resulting partition is not statistically contingent on the sensitive attribute $\mathbf{s}$. To overcome this challenge, we propose a unified optimization framework that learns robust representations from incomplete data while explicitly and adaptively mitigating fairness concerns. The subsequent sections detail the synergistic components designed to achieve this.

View-Dedicated Representation Learning

To learn a view-dedicated latent representation, we define a non-linear encoder $\mathcal{E}_v$ for each view. Specifically, we employ Kolmogorov-Arnold Networks (KAN) to build our encoder due to its capacity to approximate arbitrary multivariate continuous functions, which projects the input data matrix $\mathbf{X}^{(v)}$ into a latent representation matrix $\mathbf{H}^{(v)}$:

$$\mathbf{H}^{(v)} = \mathcal{E}_v(\mathbf{X}^{(v)}) \tag{1}$$

where $\mathbf{H}^{(v)} \in \mathbb{R}^{N\times d_h}$, and d_h is the dimension of the shared latent representation. Then, we employ a decoder $\mathcal{D}_v$ for

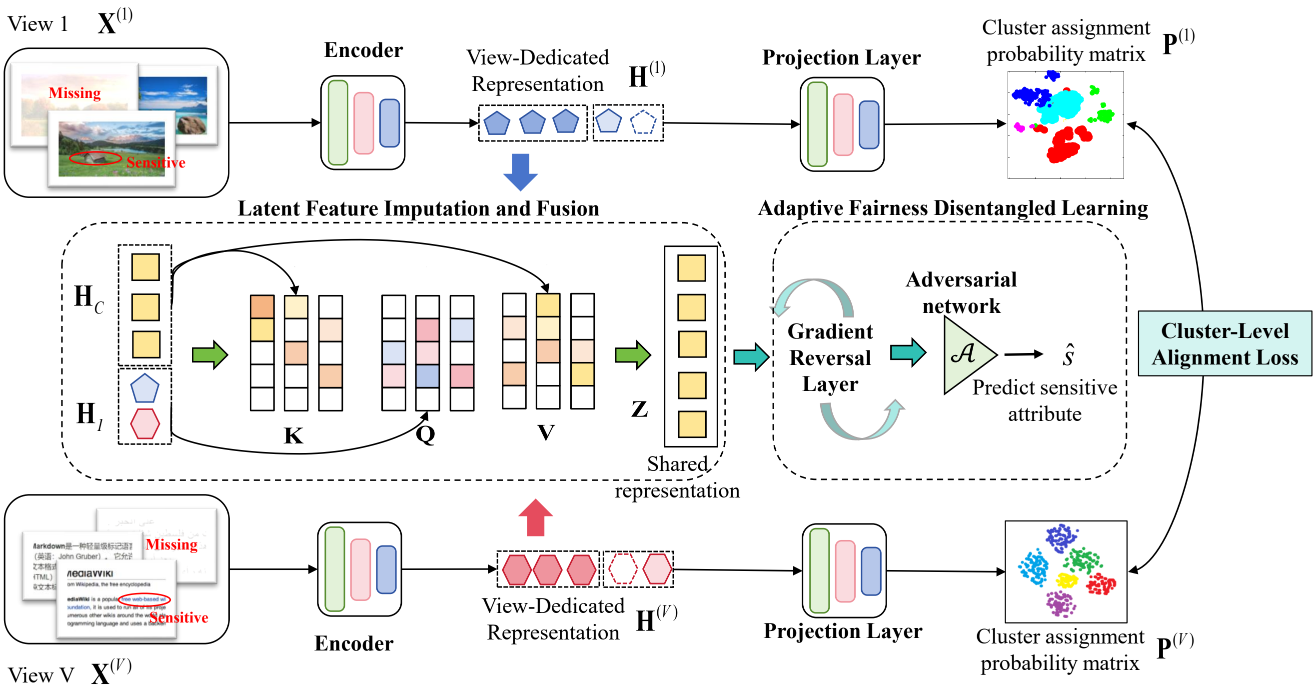


Figure 1: Our AFIMVC framework processes incomplete multi-view data (exemplified by two views) via view-dedicated encoders to learn latent representations ($\mathbf{H}^1, \mathbf{H}^2$). The latent representations undergo a Latent Feature Imputation and Fusion module that leverages complete-view information for imputation and fuse them into unified representation $\mathbf{Z}$ with feature-level and cluster-level alignment. The Adaptive Fairness Disentangled Learning eliminates the sensitive attributes in the unified representation $\mathbf{Z}$.

each view to reconstruct the original data from the latent representation, producing a reconstructed matrix $\tilde{\mathbf{X}}^{(v)}$:

$$\tilde{\mathbf{X}}^{(v)} = \mathcal{D}_v(\mathbf{H}^{(v)}) \tag{2}$$

To ensure the latent representation retains the most information in the original data, we introduce the **Reconstruction Loss** $\mathcal{L}_{\text{re}}$ as follows:

$$\mathcal{L}_{\text{re}} = \sum_{v=1}^{V} \left\| \mathbf{X}^{(v)} - \tilde{\mathbf{X}}^{(v)} \right\|_F^2 \tag{3}$$

By minimizing this objective, each encoder $\mathcal{E}_v$ is compelled to learn a representation $\mathbf{H}^{(v)}$ that preserves the essential information of $\mathbf{X}^{(v)}$.

Latent Feature Imputation and Fusion

While reconstruction loss retains important intra-view information, it treats each view independently and thus cannot align the representations from various views. To address this issue, we introduce **Feature-Level Alignment Loss** $\mathcal{L}_{\text{fla}}$ for the representation $\{\mathbf{H}_C^{(v)}\}$ of view-complete data, where $\mathbf{H}_C^{(v)} = \mathcal{E}_v(\mathbf{X}_C^{(v)})$. The guiding principle of this loss is that for any given sample, its representations across different views should be congruent. The loss is defined by the pair-wise squared Frobenius norm as follows:

$$\mathcal{L}_{\text{fla}} = \sum_{u=1}^{V-1} \sum_{v=u+1}^{V} \left\| \mathbf{H}_C^{(u)} - \mathbf{H}_C^{(v)} \right\|_F^2 \tag{4}$$

This loss enforces the distinct encoders to produce similar representations for cross-view alignment. Then, we employ a learnable weight γ_v to fuse the representations of complete data into a consistent representation by:

$$\mathbf{H}_C = \sum_{v=1}^{V} \gamma_v \mathbf{H}_C^{(v)} \tag{5}$$

For view-incomplete representations, we learn the fused representation by averaging the features of available views:

$$\mathbf{H}_I = \frac{1}{V'} \sum_{v'=1}^{V'} \mathbf{H}_I^{v'} \tag{6}$$

where V' is the number of available views for the i-th view-incomplete representation, which varies with i.

Most current works fuse view-complete representation and view-incomplete representation into a shared representation $\mathbf{H} \in \mathbb{R}^{N \times d_h}$, respectively, as described above. However, the view incompleteness leads to a semantic gap between $\mathbf{H}_C$ and $\mathbf{H}_I$. To address this problem, we propose a novel fusion method by leveraging contextual information from view-complete representations $\mathbf{H}_C$ to enhance view-incomplete ones $\mathbf{H}_I$.

First, we enhance the view-incomplete samples via a Cross-Sample Attention mechanism. This mechanism treats the shared representations of incomplete views $\mathbf{H}_I$ as queries $\mathbf{Q}$, and the shared representations of complete views $\mathbf{H}_C$ as keys $\mathbf{K}$ and values $\mathbf{V}$. For view-incomplete representations $\mathbf{H}_I$, the attention mechanism computes a context matrix $\mathbf{C}$, which aggregates weighted information from all view-complete shared representations:

$$\mathbf{C} = \text{MultiHeadAttn}(\mathbf{H}_I, \mathbf{H}_C, \mathbf{H}_C) \tag{7}$$

where $\text{MultiHeadAttn}(\cdot)$ denotes the standard multi-head attention operation. The context matrix $\mathbf{C}$ can be interpreted as a data-driven inference of the potential semantic gap between the shared representation of complete views and incomplete views.

Second, we fuse this inferred contextual information with the view-incomplete representation via a residual connection to obtain its augmented shared representation $\mathbf{Z}_I$:

$$\mathbf{H}_I^a = \mathbf{H}_I + \mathbf{C} \tag{8}$$

This augmentation process can be considered as the imputation of missing-view information based on the relationship with the complete view representations.

Finally, the augmented shared representations of all samples, including the view-complete shared representation $\mathbf{H}_C$ and augmented view-incomplete shared representation $\mathbf{H}_I^a$, are concatenated into unified representation $\mathbf{Z} = [\mathbf{H}_C; \mathbf{H}_I^a]$.

While $\mathcal{L}_{\text{fla}}$ provides a feature-level alignment for complete samples, a more robust mechanism is needed to ensure semantic consistency. We achieve it by introducing a **Cluster-Level Alignment Loss** $\mathcal{L}_{\text{cla}}$ that enforces cluster assignment probabilities of different views to be similar. Specifically, we input the representation matrix of each view $[\mathbf{H}_C^{(v)}; \mathbf{H}_I^a]$ into a projection head $\mathcal{F}(\cdot)$ to learn a cluster assignment probability matrix $\mathbf{P}^{(v)} \in \mathbb{R}^{N \times K}$. Suppose $\mathbf{p}_k^{(v)} \in \mathbb{R}^{N \times 1}$ is the k-th column of $\mathbf{P}^{(v)}$ and represents the probabilities of N representations assigned to the k-th cluster. Given two views u and v, we build the cross-view contrastive loss to realize cluster-level alignment:

$$\begin{aligned}\mathcal{L}_{\text{cla}} = -\frac{1}{K} \sum_{u=1}^{V} \sum_{v=1, v\neq u}^{V} \sum_{k=1}^{K} \log\{ \\ \frac{e^{(\text{sim}(\mathbf{p}_k^{(u)}, \mathbf{p}_k^{(v)})/\tau)}}{\sum_{v'=u,v} \sum_{j=1}^{K} e^{(\text{sim}(\mathbf{p}_k^{(u)}, \mathbf{p}_j^{(v')})/\tau)} - e^{1/\tau}}\}\end{aligned} \tag{9}$$

where $\text{sim}(\cdot, \cdot)$ is the cosine similarity, and τ is a temperature hyperparameter. The clustering alignment loss enforces the model to produce consistent and discriminative unified representation $\mathbf{Z}$.

Adaptive Fairness Disentangled Learning

The unified representation $\mathbf{Z}$ may still implicitly encode the sensitive attribute $\mathbf{s}$, which can lead to biased clustering results. To address this issue, we propose an Adaptive Fairness Disentangled Learning module, which aims to learn a representation invariant to sensitive attributes.

First, we employ a Maximum Mean Discrepancy (MMD) loss(Gretton et al. 2006) $\mathcal{L}_{\text{MMD}}$ to align the distributions of sensitive attributes at a global level. It minimizes the distance between the mean embeddings of the feature sets from protected groups $\mathcal{G}_0$ and $\mathcal{G}_1$ in a Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}$:

$$\mathcal{L}_{\text{MMD}} = \left\| \frac{1}{|\mathcal{G}_0|} \sum_{i \in \mathcal{G}_0} \phi(\mathbf{z}_i) - \frac{1}{|\mathcal{G}_1|} \sum_{j \in \mathcal{G}_1} \phi(\mathbf{z}_j) \right\|_{\mathcal{H}}^2 \tag{10}$$

where $\mathbf{z}_i$ and $\mathbf{z}_j$ are the unified representation vectors (rows of matrix $\mathbf{Z}$) for samples belonging to the protected groups $\mathcal{G}_0$ and $\mathcal{G}_1$ respectively, and $\phi(\cdot)$ is the kernel-induced feature map. $\mathcal{L}_{\text{MMD}}$ ensures that the unified representations are statistically indistinguishable across groups.

Second, to remove sensitive information at the instance level, we employ an Adversarial Disentangled mechanism. This is realized through a minimax game between the encoder and an adversarial network $\mathcal{A}$ facilitated by a Gradient Reversal Layer (GRL) (Ganin et al. 2016). The adversarial network is trained to predict the sensitive attribute s_i from its corresponding representation vector $\mathbf{z}_i$. When the adversarial network minimizes this loss, the GRL reverses the gradient propagated back to the encoder, which compels the encoder to generate features that maximize the adversarial network's error. The adversarial disentangled loss $\mathcal{L}_{\text{AD}}$ is the standard cross-entropy of this prediction:

$$\mathcal{L}_{\text{AD}} = -\frac{1}{N} \sum_{i=1}^{N} (s_i \log \hat{s}_i + (1 - s_i) \log(1 - \hat{s}_i)) \tag{11}$$

where s_i is the ground-truth sensitive attribute for sample i, and $\hat{s}_i$ is the predicted probability that sample i belongs to the positive class ($s_i = 1$) produced by the adversarial network $\mathcal{A}$.

Finally, we integrate the global-level MMD loss $\mathcal{L}_{\text{MMD}}$ and the instance-level adversarial dsentangled loss $\mathcal{L}_{\text{AD}}$ into a single unified **Fairness Loss**:

$$\mathcal{L}_{\text{fair}} = \alpha_{\text{fair}} \cdot (\mathcal{L}_{\text{AD}} + \eta \mathcal{L}_{\text{MMD}}) \tag{12}$$

where η is a hyperparameter that balances the two losses, and α_{fair} is an adaptive weight that dynamically scales the fairness loss to intelligently navigate the accuracy-fairness trade-off.

Existing methods often improve fairness at the sacrifice of clustering accuracy, which is difficult to tune trade-off between fairness and accuracy. Our approach proposes an Adaptive Controller that dynamically balances clustering accuracy and fairness. We use the adversarial network's classification accuracy on the sensitive attribute prediction task, denoted as $\text{acc}_{\mathcal{A}}$, as a real-time proxy for the level of entanglement. An adaptive weight α_{fair} is then defined as a function of this accuracy:

$$\alpha_{\text{fair}} = \text{clamp}(2 \cdot (\text{acc}_{\mathcal{A}} - 0.5), 0, 1) \tag{13}$$

where $\text{clamp}(\cdot)$ is a function that constrains the output to the range $[0, 1]$. This formulation is able to adaptively balance

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

Table 1: Statistics summary of three datasets.

the fairness and clustering accuracy by dynamically adjusting the weight α_{fair}, which controls the influence of fairness loss based on the performance of the adversarial network $\mathcal{A}$. Specifically, when the adversarial network's accuracy is high ($\text{acc}_{\mathcal{A}} \to 1.0$) that indicates strong entanglement, α_{fair} will be close to 1; conversely, as accuracy approaches that of a random guess ($\text{acc}_{\mathcal{A}} \to 0.5$), α_{fair} decays towards 0 to relax the fairness constraint.

The Objective Function

We integrate the above reconstruction loss, feature-level alignment loss, cluster-level alignment loss, and fairness loss into the final objective $\mathcal{L}_{all}$ as follows:

$$\mathcal{L}_{\text{all}} = \mathcal{L}_{\text{re}} + \lambda \mathcal{L}_{\text{fla}} + \beta \mathcal{L}_{\text{cla}} + \mathcal{L}_{\text{fair}} \tag{14}$$

where λ, β are trade-off hyperparameters that balance the influence of the different terms.

The final cluster assignments are derived directly from this unified representation $\mathbf{Z}$. We employ a Clustering Layer to compute a soft assignment distribution, $\mathbf{q}_i \in \mathbb{R}^K$, by measuring the similarity between each representation $\mathbf{z}_i$ and a set of K learnable cluster centroids using the Student's t-distribution. The final prediction is then obtained by:

$$\hat{y}_i = \underset{k \in \{1,\ldots,K\}}{\arg\max} \; q_{ik} \tag{15}$$

Experimental Analysis

We validate our methods on three datasets and compare it with eight comparison methods.

Datasets

We evaluate our proposed AFIMVC model on three widely-used benchmark datasets for fair clustering: Banking Market, Zafar, and Credit Card (Zafar et al. 2017). A detailed statistical summary of these datasets is provided in Table 1. Since these datasets lack natural views, we synthetically generated two distinct views for each. For the Credit Card and Bank Marketing datasets, views were derived by applying different non-linear transformations to the standardized raw features: tanh/sigmoid for Credit Card and sigmoid/ReLU for Bank Marketing. For the Zafar dataset, View 1 was generated using a tanh-transformation of the original features to capture geometric properties, while View 2 was formed from the posterior probabilities of a Gaussian Mixture Model (GMM) to represent the probabilistic cluster structure. Finally, all generated views were standardized using a StandardScaler before being input to the model.

Evaluation Metrics

We assess the performance of our framework using two standard metrics. Clustering quality is evaluated using **Normalized Mutual Information (NMI)**(Strehl and Ghosh 2002), which measures the agreement between the predicted cluster assignments and the ground-truth labels.

$$\text{NMI}(X, Y) = \frac{I(X;Y)}{H(X) + H(Y)} \tag{16}$$

where $I(X;Y)$ is the mutual information and $H(\cdot)$ represents entropy. The Balance Level evaluates demographic parity across clusters. For fairness, we employ the **Balance Level (BAL)**(Chierichetti et al. 2017; Wang et al. 2021b) to quantify demographic parity. BAL is defined as the minimum proportion of samples with a sensitive attribute within any given cluster:

$$\text{BAL} = \min_i \frac{|C_i \cap S|}{|C_i|} \tag{17}$$

where C_i represents the i-th cluster and S denotes samples with sensitive attributes.

Comparing Methods

We validate our proposed AFIMVC against eight state-of-the-art baselines. These include single-view traditional clustering (**k-means** (Macqueen 1967)) and deep clustering (**DEC** (Xie, Girshick, and Farhadi 2016)) methods. For MVC, we compare it against **MvDSCN** (Wang et al. 2022a) and **CC** (Li et al. 2020); for IMVC, we compare it with **DCP** (Lin et al. 2023) and **APADC** (Xu et al. 2023)). Besides, we compare with a fair MVC (**DFMVC** (Zhao et al. 2024)) and a fair IMVC (**Fair-MVC** (Zheng, Zhu, and He 2023)).

Implementation Details

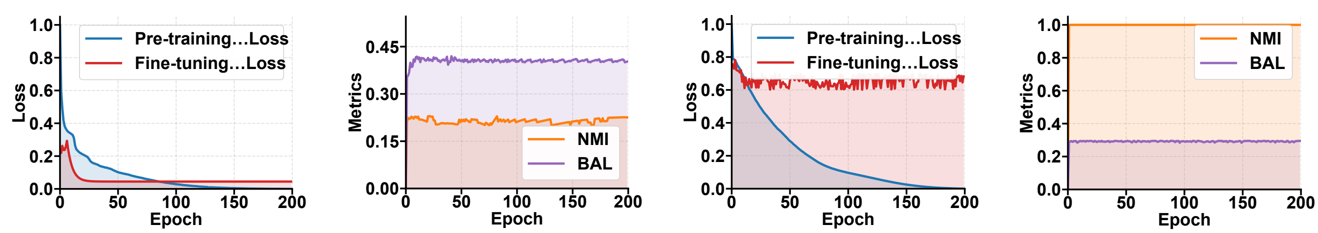
All experiments were conducted on a workstation equipped with an NVIDIA GeForce RTX 4090 GPU (CUDA 12.4) and 64 GB of RAM. The proposed model as well as all baseline methods were implemented in PyTorch using Python 3.10.13. We employed the Adam optimizer (Kingma and Ba 2014) to minimize the overall objective function in an end-to-end training manner.

Experimental Analysis

Comparsion Results with State-of-The-Arts. Table 2 presents a comparative performance analysis of AFIMVC against eight baseline methods across varying missing rates. The results robustly demonstrate that AFIMVC achieves a state-of-the-art balance between clustering quality (NMI) and fairness (Balance). From Table 2, we can make the following key observations. First, both IMVC and MVC methods outperform single-view methods in NMI, owing to their ability to fully exploit complementary information for enhanced clustering performance. Second, on incomplete multi-view data, IMVC methods generally outperform MVC and single-view methods in NMI, as they employ imputation techniques to tackle view-missing issues effectively. Third, regarding the Balance metric, Fair-MVC,

Datasets	Missing Rate	0		0.25		0.5		0.75	
	Metrics(%)	NMI	Balance	NMI	Balance	NMI	Balance	NMI	Balance
Credit Card	k-means(Macqueen 1967)	21.82 ± 0.94	35.21 ± 0.47	17.10 ± 1.25	37.12 ± 0.04	13.30 ± 0.57	36.12 ± 0.21	9.62 ± 0.37	36.10 ± 0.35
	DEC(Xie, Girshick, and Farhadi 2016)	22.61 ± 0.45	36.12 ± 0.55	19.88 ± 0.37	35.61 ± 0.97	14.21 ± 2.07	36.15 ± 0.37	11.04 ± 1.65	36.42 ± 0.55
	MvDSCN(Wang et al. 2022a)	22.72 ± 1.53	35.60 ± 0.38	20.88 ± 0.75	35.75 ± 0.19	16.99 ± 2.03	36.40 ± 0.56	12.35 ± 0.91	36.55 ± 0.67
	CC(Li et al. 2020)	24.57 ± 1.28	35.28 ± 0.38	22.84 ± 2.04	36.09 ± 0.57	17.95 ± 0.88	36.57 ± 0.21	12.74 ± 0.96	36.95 ± 0.87
	DCP(Lin et al. 2023)	30.13 ± 2.27	25.29 ± 0.88	28.58 ± 0.45	30.45 ± 1.20	20.25 ± 0.79	33.87 ± 1.02	17.35 ± 0.97	38.38 ± 0.41
	APADC(Xu et al. 2023)	27.91 ± 1.30	26.80 ± 0.31	23.17 ± 1.01	34.27 ± 0.44	16.75 ± 1.17	36.96 ± 0.56	12.15 ± 1.55	33.15 ± 0.27
	Fair-MVC(Zheng, Zhu, and He 2023)	25.12 ± 0.39	**42.08 ± 0.28**	22.24 ± 0.73	38.35 ± 0.51	18.88 ± 0.82	39.21 ± 0.69	14.43 ± 0.64	39.76 ± 0.87
	DFMVC(Zhao et al. 2024)	30.12 ± 0.42	40.11 ± 0.68	25.46 ± 0.47	39.59 ± 0.12	19.64 ± 0.54	39.39 ± 0.59	13.83 ± 0.61	39.27 ± 0.61
	AFIMVC	**49.48 ± 0.02**	41.69 ± 0.71	**47.75 ± 0.58**	**39.92 ± 0.37**	**23.34 ± 0.41**	**40.41 ± 0.27**	**19.79 ± 0.56**	**41.18 ± 0.57**
Zafar	k-means(Macqueen 1967)	71.12 ± 0.31	19.26 ± 0.27	65.17 ± 0.76	17.06 ± 0.39	61.55 ± 0.87	17.51 ± 0.91	50.19 ± 1.27	15.21 ± 1.25
	DEC(Xie, Girshick, and Farhadi 2016)	72.48 ± 1.47	16.92 ± 0.66	70.75 ± 1.38	17.05 ± 0.82	64.85 ± 1.32	16.89 ± 0.95	58.82 ± 1.04	17.10 ± 0.64
	MvDSCN(Wang et al. 2022a)	76.85 ± 0.45	17.20 ± 0.62	74.85 ± 1.05	18.25 ± 0.92	69.45 ± 1.62	16.95 ± 0.95	64.15 ± 1.38	17.15 ± 1.12
	CC(Li et al. 2020)	79.10 ± 0.95	16.90 ± 0.82	75.35 ± 1.15	18.30 ± 0.95	72.25 ± 1.05	17.85 ± 1.12	68.15 ± 0.98	17.75 ± 1.35
	DCP(Lin et al. 2023)	81.70 ± 1.85	21.50 ± 1.55	73.90 ± 2.75	22.40 ± 2.15	65.65 ± 2.25	19.75 ± 2.80	57.25 ± 2.50	21.15 ± 2.65
	APADC(Xu et al. 2023)	72.50 ± 1.10	21.10 ± 0.85	66.50 ± 0.45	21.60 ± 0.65	61.40 ± 1.25	21.90 ± 1.55	55.40 ± 1.15	21.35 ± 1.25
	Fair-MVC(Zheng, Zhu, and He 2023)	88.85 ± 0.65	28.40 ± 0.72	79.75 ± 0.88	30.50 ± 1.15	76.85 ± 1.05	29.20 ± 1.10	73.05 ± 1.25	29.45 ± 1.18
	DFMVC(Zhao et al. 2024)	94.33 ± 0.47	29.01 ± 0.45	81.27 ± 0.32	29.28 ± 0.58	77.64 ± 0.33	29.27 ± 0.59	74.02 ± 0.18	29.13 ± 0.65
	AFIMVC	**99.99 ± 0.01**	**29.98 ± 0.25**	**99.84 ± 0.11**	**30.85 ± 0.69**	**99.75 ± 0.13**	**29.51 ± 0.45**	**99.70 ± 0.23**	**29.92 ± 0.63**
Banking Market	k-means(Macqueen 1967)	28.67 ± 1.44	37.65 ± 0.66	24.77 ± 1.71	36.66 ± 0.45	21.08 ± 1.17	36.32 ± 0.89	16.83 ± 1.46	36.25 ± 0.92
	DEC(Xie, Girshick, and Farhadi 2016)	30.93 ± 1.15	37.60 ± 0.96	28.97 ± 1.80	36.42 ± 0.69	24.37 ± 1.02	37.09 ± 0.78	19.43 ± 1.22	36.68 ± 0.85
	MvDSCN(Wang et al. 2022a)	36.24 ± 0.55	37.59 ± 0.67	35.02 ± 1.21	36.11 ± 0.57	31.33 ± 1.76	36.96 ± 0.77	27.15 ± 1.54	37.24 ± 0.86
	CC(Li et al. 2020)	36.23 ± 1.01	37.46 ± 0.97	34.88 ± 1.33	37.69 ± 0.95	31.13 ± 1.63	37.80 ± 0.85	27.12 ± 1.54	37.62 ± 0.94
	DCP(Lin et al. 2023)	39.93 ± 1.84	26.75 ± 2.06	34.20 ± 2.87	20.49 ± 1.54	28.12 ± 2.31	27.29 ± 1.24	21.54 ± 2.46	26.58 ± 1.36
	APADC(Xu et al. 2023)	40.62 ± 0.25	27.79 ± 2.59	32.21 ± 2.79	28.41 ± 2.11	33.14 ± 2.77	27.82 ± 2.28	26.32 ± 2.56	27.92 ± 2.12
	Fair-MVC(Zheng, Zhu, and He 2023)	38.99 ± 0.91	42.40 ± 0.75	36.66 ± 1.01	41.76 ± 0.83	32.90 ± 1.03	41.61 ± 0.67	28.34 ± 0.98	41.82 ± 0.78
	DFMVC(Zhao et al. 2024)	54.62 ± 1.25	42.16 ± 0.82	46.42 ± 0.42	39.67 ± 0.51	39.32 ± 0.38	39.02 ± 0.50	30.26 ± 0.17	38.79 ± 0.53
	AFIMVC	**82.67 ± 0.15**	**44.08 ± 1.14**	**80.06 ± 0.24**	**44.15 ± 0.57**	**50.83 ± 0.31**	**43.07 ± 0.28**	**45.55 ± 0.18**	**42.77 ± 0.69**

Table 2: Clustering NMI and balance scores on the three datasets containing sensitive features. The best results are marked in bold. Suboptimal results are represented in blue.



(a) Loss curve. (b) Evaluation curve. (c) Loss curve. (d) Evaluation curve.

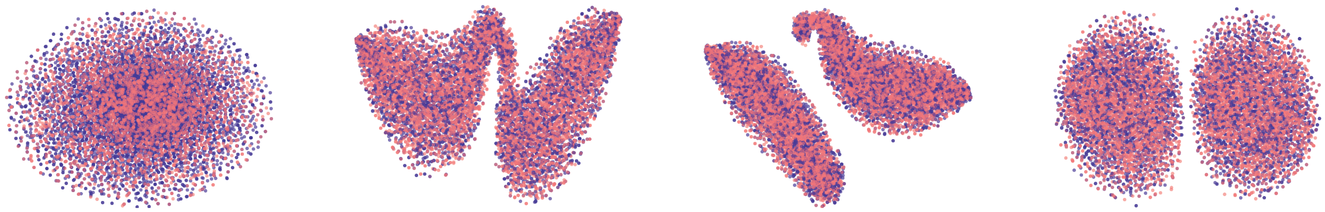
Figure 2: The graphs show the variation of training loss and evaluation metrics for the Credit Card (left) and Zafar (right) datasets with loss curves versus evaluation metrics at 50% missing rate.

DFMVC and AFIMVC outperform counterparts without fairness consideration, demonstrating their superior fairness performance. Finally, our proposed AFIMVC achieves the optimal NMI and Balance across most scenarios. This is primarily attributed to its context-aware representation imputation that alleviates potential semantic bias in imputation, while the adaptive weight learning mechanism achieves an ideal balance between clustering performance and fairness.

Ablation Study. As shown in Table 3, each component of our framework demonstrates a clear and significant contribution. The baseline model with only the reconstruction loss ($\mathcal{L}_{\text{re}}$) performs poorly, establishing the necessity of further objectives. A substantial performance leap is observed with the inclusion of the feature-level alignment and cluster-level alignment losses ($\mathcal{L}_{\text{fla}}$ and $\mathcal{L}_{\text{cla}}$), with the cluster-level alignment $\mathcal{L}_{\text{cla}}$ showing the most significant impact on clustering quality (e.g., NMI on Bank Market jumps from 23.23% to 79.92%). Most critically, the final integration of our adaptive fairness loss ($\mathcal{L}_{\text{fair}}$) not only yields the best fairness scores (BAL) across all datasets but also further improves clustering performance. This synergistic effect strongly validates our core hypothesis: learning a representation that is explicitly disentangled from sensitive attributes can lead to a more robust and generalized data structure, benefiting both fairness and clustering quality.

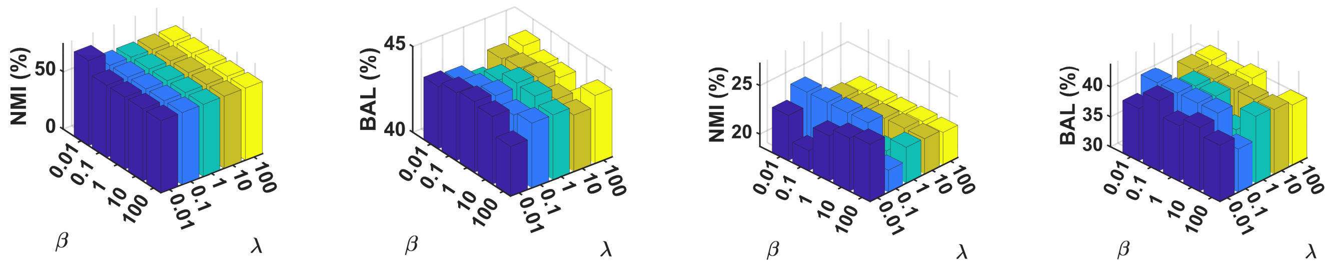
Convergence Analysis. Figure 2 illustrates the stable and efficient convergence of our model. The pre-training loss rapidly decreases, providing a robust initialization. Subsequently, the fine-tuning loss converges smoothly to a low value, while key metrics concurrently rise to a high-performance plateau. This behavior confirms the effectiveness of our two-stage training strategy and the stability of the joint optimization process.

Visualization Analysis. The t-SNE(Maaten and Hinton 2008) visualizations in Figure 3 provide an intuitive confirmation of our model's dual success. As training progresses, the unified representation $\mathbf{Z}$ evolves from an undifferentiated mass into two highly compact and well-separated clus-



(a) Epoch 10 (b) Epoch 50 (c) Epoch 100 (d) Epoch 200

Figure 3: Visualization of t-SNE for the Zafar dataset (25% missing rate). As training epochs increase, both clustering quality and fairness are gradually optimized.



(a) NMI (b) BAL (c) NMI (d) BAL

Figure 4: Parameter sensitivity analysis on two datasets at a 50% missing rate. (a) and (b) show the NMI and BAL performance on the **Banking Market** dataset, while (c) and (d) illustrate the corresponding metrics on the **Credit Card** dataset.

Dataset	$\mathcal{L}_{\text{re}}$	$\mathcal{L}_{\text{fla}}$	$\mathcal{L}_{\text{cla}}$	$\mathcal{L}_{\text{fair}}$	NMI	BAL
Credit Card	✓	×	×	×	16.51±0.07	29.87±0.71
	✓	✓	×	×	19.21±0.46	30.17±0.35
	✓	✓	✓	×	33.39±0.26	32.15±0.47
	✓	✓	✓	✓	**47.75±0.58**	**39.92±0.37**
Bank Market	✓	×	×	×	0.42±1.15	22.05±0.33
	✓	✓	×	×	23.23±0.14	24.86±0.57
	✓	✓	✓	×	79.92±0.09	35.03±0.42
	✓	✓	✓	✓	**80.06±0.24**	**44.15±0.57**
Zafar	✓	×	×	×	0.2±0.01	9.01±0.83
	✓	✓	×	×	5.15±0.47	10.34±1.22
	✓	✓	✓	×	98.71±0.05	25.52±0.14
	✓	✓	✓	✓	**99.84±0.11**	**30.85±0.69**

Table 3: Ablation study of our method's main components on three datasets with a 25% missing rate.

ters, demonstrating potent clustering capability. Crucially, the sensitive attributes (indicated by color) remain uniformly mixed within each cluster, visually verifying that our framework learns a high-quality partition while effectively disentangling the representation from sensitive information.

Hyper-parameter Analysis λ and β. To investigate the sensitivity of our model to the key hyperparameters, we analyze the impact of the alignment weights λ and β on the Banking Market dataset. As illustrated in Figure 4, the model's performance exhibits distinct sensitivities to these parameters. NMI remain stale on most values of β, which indicates that the model is not s This indicates that a sufficient probabilistic semantic alignment is crucial for achieving high clustering quality. Notably, the BAL metric's stability across a wide range of α and β highlights the robustness of our adaptive disentanglement module, which autonomously regulates fairness and reduces dependency on hyperparameter tuning.

Conclusions

In this paper, we proposed a novel framework termed Adversarial Fair Incomplete Multi-View Clustering (AFIMVC), to simultaneously address the challenges of view incompleteness, view inconsistency, and fairness in multi-view clustering. Specifically, we impute missing-view information based on the relationship learned from the complete view representations. To learn representations invariant to sensitive attributes, AFIMVC introduces a novel adaptive adversarial disentangled mechanism that dynamically controls adversarial intensity based on model's real-time bias. Experimental results demonstrate that AFIMVC exhibits effectiveness and superiority in balancing clustering performance and fairness.

Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grant 62176203, the Fundamental Research Funds for the Central Universities (ZYTS25267, QTZX25004), and the Science and Technology Project of Xi'an (Grant 2022JH-JSYF-0009), Selected Support Project for Scientific and Technological Activities of Returned Overseas Chinese Scholars in Shaanxi Province 2023-02.

References

Backurs, A.; Indyk, P.; Onak, K.; Schieber, B.; Vakilian, A.; and Wagner, T. 2019. Scalable fair clustering. In *International Conference on Machine Learning*, 405–413. PMLR.

Chen, Z.; Li, Y.; Lou, K.; and Zhao, L. 2023. Incomplete multi-view clustering with complete view guidance. *IEEE Signal Processing Letters*.

Chierichetti, F.; Kumar, R.; Lattanzi, S.; and Vassilvitskii, S. 2017. Fair clustering through fairlets. *Advances in neural information processing systems*, 30.

Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; March, M.; and Lempitsky, V. 2016. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59): 1–35.

Gretton, A.; Borgwardt, K.; Rasch, M.; Schölkopf, B.; and Smola, A. 2006. A kernel method for the two-sample-problem. *Advances in neural information processing systems*, 19.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kleindessner, M.; Samadi, S.; Awasthi, P.; and Morgenstern, J. 2019. Guarantees for spectral clustering with fairness constraints. In *International conference on machine learning*, 3458–3467. PMLR.

Li, A.; Feng, C.; Cheng, Y.; Zhang, Y.; and Yang, H. 2024a. Incomplete multiview subspace clustering based on multiple kernel low-redundant representation learning. *Information Fusion*, 103: 102086.

Li, R.; Hu, H.; Du, L.; Chen, J.; Jiang, B.; and Zhou, P. 2024b. One-Stage Fair Multi-View Spectral Clustering. In *ACM Multimedia 2024*.

Li, X.; Pan, Y. P.; Sun, Y.; Sun, Q. S.; Tsang, I. W.; and Ren, Z. 2024c. Fast Unpaired Multi-view Clustering.

Li, Y.; Hu, P.; Liu, Z.; Peng, D.; Zhou, J. T.; and Peng, X. 2020. Contrastive Clustering. arXiv:2009.09687.

Lin, Y.; Gou, Y.; Liu, X.; Bai, J.; Lv, J.; and Peng, X. 2023. Dual Contrastive Prediction for Incomplete Multi-View Representation Learning. *IEEE TPAMI*, 45(4): 4447–4461.

Liu, H.; and Fu, Y. 2018. Consensus Guided Multi-View Clustering. *Proceedings of the ACM on Knowledge Discovery and Data Mining*, 933–941.

Maaten, L. v. d.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov): 2579–2605.

Macqueen, J. 1967. Some methods for classification and analysis of multivariate observations. *Proc. Symp. Math. Statist. and Probability, 5th*, 1.

Strehl, A.; and Ghosh, J. 2002. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of machine learning research*, 3(Dec): 583–617.

Wang, Q.; Ding, Z.; Tao, Z.; Gao, Q.; and Fu, Y. 2021a. Generative partial multi-view clustering with adaptive fusion and cycle consistency. *IEEE Transactions on Image Processing*, 30: 1771–1783.

Wang, Q.; Tao, Z.; Gao, Q.; and Jiao, L. 2022a. Multi-View Subspace Clustering via Structured Multi-Pathway Network. *IEEE TNNLS*, 1–7.

Wang, Q.; Xu, H.; Liu, M.; Feng, W.; and Zhang, X. 2025. Fair incomplete multi-view clustering via distribution alignment. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, 6379–6387.

Wang, S.; Liu, X.; Liu, L.; Zhou, S.; and Zhu, E. 2021b. Late fusion multiple kernel clustering with proxy graph refinement. *IEEE Transactions on Neural Networks and Learning Systems*, 34(8): 4359–4370.

Wang, Y.; Kang, J.; Xia, Y.; Luo, J.; and Tong, H. 2022b. ifig: Individually fair multi-view graph clustering. In *2022 IEEE International Conference on Big Data (Big Data)*, 329–338. IEEE.

Wen, J.; Xu, G.; Tang, Z.; Wang, W.; Fei, L.; and Xu, Y. 2023. Graph regularized and feature aware matrix factorization for robust incomplete multi-view clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(5): 3728–3741.

Xia, D.; Yang, Y.; Yang, S.; and Li, T. 2023. Incomplete multi-view clustering via kernelized graph learning. *Information Sciences*, 625: 1–19.

Xie, J.; Girshick, R.; and Farhadi, A. 2016. Unsupervised Deep Embedding for Clustering Analysis. arXiv:1511.06335.

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

Yan, W.; Zhang, Y.; Lv, C.; Tang, C.; Yue, G.; Liao, L.; and Lin, W. 2023. Gcfagg: Global and cross-view feature aggregation for multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19863–19872.

Yang, S.; Liao, Z.; Chen, R.; Lai, Y.; and Xu, W. 2024. Multi-view fair-augmentation contrastive graph clustering with reliable pseudo-labels. *Information Sciences*, 674: 120739.

Yao, S.; Yu, G.; Wang, J.; Domeniconi, C.; and Zhang, X. 2019. Multi-View Multiple Clustering. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 4093–4099.

Yin, J.; Cai, R.; and Sun, S. 2022. Anchor-based incomplete multi-view spectral clustering. *Neurocomputing*, 514: 526–538.

You, J.; Ren, Z.; You, X.; Li, H.; and Yao, Y. 2023. Priori anchor labels supervised scalable multi-view bipartite graph clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 10972–10979.

Yu, X.; Jiang, Y.; Chao, G.; and Chu, D. 2024. Deep contrastive multi-view subspace clustering with representation and cluster interactive learning. *IEEE Transactions on Knowledge and Data Engineering*.

Zafar, M. B.; Valera, I.; Rodriguez, M. G.; and Gummadi, K. P. 2017. Fairness Constraints: Mechanisms for Fair Classification. arXiv:1507.05259.

Zhang, K.; Du, S.; Wang, Y.; and Deng, T. 2024. Deep incomplete multi-view clustering via attention-based direct contrastive learning. *Expert Systems with Applications*, 255: 124745.

Zhao, B.; Wang, Q.; Tao, Z.; Feng, W.; and Gao, Q. 2024. DFMVC: Deep Fair Multi-view Clustering. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 8090–8099.

Zheng, L.; Zhu, Y.; and He, J. 2023. Fairness-aware multi-view clustering. In *Proceedings of the 2023 SIAM International Conference on Data Mining (SDM)*, 856–864. SIAM.

Zhou, Z.; Ataee Tarzanagh, D.; Hou, B.; Tong, B.; Xu, J.; Feng, Y.; Long, Q.; and Shen, L. 2024. Fair canonical correlation analysis. *Advances in Neural Information Processing Systems*, 36.