Reliable Conflictive Multi-View Learning

Cai Xu, Jiajun Si, Ziyu Guan, Wei Zhao*, Yue Wu, Xiyue Gao

School of Computer Science and Technology, Xidian University, China
{cxu@, jiajunsi@stu., zyguan@, ywzhao@mail., ywu@, xygao@}xidian.edu.cn

Abstract

Multi-view learning aims to combine multiple features to achieve more comprehensive descriptions of data. Most previous works assume that multiple views are strictly aligned. However, real-world multi-view data may contain low-quality conflictive instances, which show conflictive information in different views. Previous methods for this problem mainly focus on eliminating the conflictive data instances by removing them or replacing conflictive views. Nevertheless, real-world applications usually require making decisions for conflictive instances rather than only eliminating them. To solve this, we point out a new Reliable Conflictive Multi-view Learning (RCML) problem, which requires the model to provide decision results and attached reliabilities for conflictive multi-view data. We develop an Evidential Conflictive Multi-view Learning (ECML) method for this problem. ECML first learns view-specific evidence, which could be termed as the amount of support to each category collected from data. Then, we can construct view-specific opinions consisting of decision results and reliability. In the multi-view fusion stage, we propose a conflictive opinion aggregation strategy and theoretically prove this strategy can exactly model the relation of multi-view common and view-specific reliabilities. Experiments performed on 6 datasets verify the effectiveness of ECML. The code is released at https://github.com/jiajunsi/RCML.

Introduction

Artificial intelligence systems usually perceive and understand the world from multi-view data. For example, automated vehicle systems sense their surroundings through multiple sensors (e.g., camera, lidar, radar); recommender systems capture users' preferences from their multi-view generated content such as textual review and visual review. Integrating the consistent and complementary information of multiple views could obtain a more comprehensive description of data instances, which boosts various tasks such as clustering (Xu et al. 2019; Huang et al. 2022; Wen et al. 2022; Ektefaie et al. 2023), retrieval (Mostafazadeh et al. 2017; Qin et al. 2022) and recommendation (Fan et al. 2023; Tan et al. 2022).

*Corresponding author

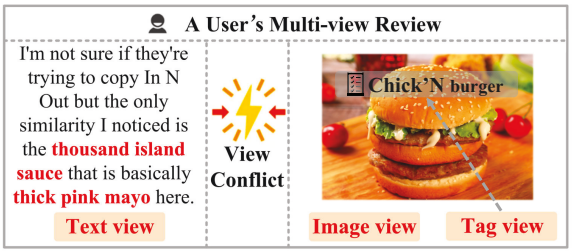

Figure 1: Visualization of the conflictive multi-view data: the text view is related to food "sauce"; however, the other views show conflictive information, i.e., food "burger".

Most of the previous studies on multi-view learning (Liang, Zadeh, and Morency 2022; Zhao et al. 2023; Zhang et al. 2023a) always assume that data of different views are strictly aligned. For example, different views consistently belong to the ground-truth category in a classification task. However, in real-world settings, this assumption cannot always be guaranteed. Fig. 1 visualizes a case of users' multi-view generated content: the text and image views show conflictive food categories. As a result, this conflictive information in different views makes most multi-view learning methods inevitably degenerate or even fail.

The prevalent solutions to this problem mainly aim to eliminate the conflictive data instance. Pioneer works regard the conflictive data as outliers (Marcos Alvarez et al. 2013; Hou et al. 2022). They usually consist of 3 steps: 1) measuring the consistency of views; 2) identifying outliers as instances with significant inconsistencies across views; 3) removing outliers to construct a clean dataset. Recently, some multi-view learning methods (Huang et al. 2020; Zhang et al. 2023d) dedicate to learning alignment relations for the original data and constructing new data instances accordingly. For example, the text view in Fig. 1 would be replaced with this view of another aligned instance. Therefore, the conflict in the original instances would be solved.

Nevertheless, real-world applications usually require making decisions for conflictive instances rather than only eliminating them. For instance, recommender systems need to predict users' preferences from their conflictive multi-view review. Considering the decision of a conflictive instance might be unreliable, we need the model can answer

"should the decision be reliable?". Therefore, we point out a new problem in this work, Reliable Conflictive Multi-view Learning (RCML) problem, which requires the model to provide the decision results and the attached reliabilities for conflictive multi-view data.

In this paper, we propose an Evidential Conflictive Multi-view Learning (ECML) method for the RCML problem. As shown in Fig. 2, we first construct the view-specific evidential Deep Neural Networks (DNNs) to learn view-specific evidence, which could be termed as the amount of support to each category collected from data. Then the view-specific distributions of the class probabilities are modeled by Dirichlet distribution, parameterized with view-specific evidence. From the distributions, we can construct opinions consisting of belief mass vector and decision reliability. Specifically, we calculate the conflictive degree according to the projected distance and conjunctive certainty among views. In the multi-view fusion stage, we propose a conflictive opinion aggregation strategy and establish a simple and effective average pooling fusion layer accordingly. We theoretically prove the final reliability would be less than view-specific reliabilities for conflictive instances.

The first contribution is recognizing the importance of explicitly providing decision results and associated reliabilities when dealing with conflicting multi-view data. Another contribution is that we develop a conflictive opinion aggregation strategy and theoretically prove it can exactly model the relation of multi-view common and view-specific reliabilities. Finally, we empirically compare ECML with state-of-the-art multi-view learning baselines on 6 publicly available datasets. Experiment results show that ECML outperforms baseline methods on accuracy, reliability and robustness.

Related Work

We briefly review related work about conflictive multi-view learning and uncertainty-aware deep learning.

Conflictive Multi-View Learning: The prevalent conflictive multi-view learning methods are mainly dedicated to eliminating the conflictive data instance. One line is based on multi-view outlier detection, which is developed to detect outliers with abnormal behavior in the multi-view context. There are two main categories for classifying these methods: cluster-based (Huang et al. 2023; Zhang et al. 2023b) and self-representation-based (Wang et al. 2019; Wen et al. 2023b). Cluster-based methods employ separate clustering in each view and generate affinity vectors for each instance accordingly (Marcos Alvarez et al. 2013; Zhao et al. 2018). Outliers are subsequently identified by comparing the affinity vectors across multiple views. Self-representation-based methods identify outliers by recognizing that they are difficult to represent using normal views (Hou et al. 2022).

Another line is based on partially view-aligned multi-view learning (Wen et al. 2023c; Zhang et al. 2021). Earlier work (Lampert and Krömer 2010) introduces weakly-paired maximum covariance analysis to overcome the limitations of unaligned data. Recently, Huang et al. (Huang et al. 2020) employ the differentiable agent of the Hungarian algorithm to establish alignment relationships for unaligned data. Along this line, researchers propose noise-robust contrastive learning (Yang et al. 2021), the self-focused mechanism (Zhang et al. 2023d), et al., to compute the alignment matrix. However, these methods aim to eliminate conflictive instances, while real-world applications usually require making decisions for them. Therefore, we propose to make reliable decisions for conflictive instances.

Uncertainty-aware Deep Learning: Deep neural networks have achieved remarkable success in various tasks, but often fail to capture the uncertainty of their predictions, especially for low-quality data (Wen et al. 2023a). Uncertainty can be categorized into aleatoric uncertainty (related to data uncertainty) and epistemic uncertainty (associated with model uncertainty). Deep learning for estimating uncertainty (Gawlikowski et al. 2023) can be classified into: single deterministic method (Sensoy, Kaplan, and Kandemir 2018), bayesian neural networks (Gal and Ghahramani 2016), ensemble methods (Lakshminarayanan, Pritzel, and Blundell 2017) and test-time augmentation methods (Lyzhov et al. 2020). Specifically, the representative single deterministic method, Evidential Deep Learning (EDL) (Sensoy, Kaplan, and Kandemir 2018) calculates the category-specific evidence according to a single DNN.

Recently, researchers extend EDL to the multi-view learning area. The pioneering work, Trusted Multi-View Classification (TMC) (Han et al. 2021) involves Dempster's combination rule, which assigns small weights to highly uncertain views. Following this line, multiple opinion aggregation methods (Jung et al. 2022; Liu et al. 2022, 2023; Zhang et al. 2023c) are proposed. However, an important characteristic of them is "After integrating another opinion into the original opinion, the obtained uncertainty mass will be reduced" (Han et al. 2023). We argue that this since when incorporating an unreliable or conflicting opinion, the uncertainty should increase. To solve this, we propose a conflictive opinion aggregation strategy and theoretically prove the uncertainty would increase for conflictive instances.

The Method

In this section, we first define the RCML problem, then present ECML in detail, together with the theoretical prove discussion, and analyses.

Problem Definition

In the RCML setting, suppose we are given a dataset with V views, $\bar{N}$ normal instances and $\tilde{N}$ conflictive instances as shown in Fig. 3. We use $\mathbf{x}_n^v \in \mathbb{R}^{D_v} (v = 1, ..., V)$ to denote the feature vector for the v-th view of the n-th instance $(n = 1, ..., N)$, where D_v is the dimensionality of the v-th view. The one-hot vector $\mathbf{y}_n \in \{0, 1\}^K$ denotes the ground-truth label of the n-th instance, where K is the total of all categories. The training tuples $\{\{\mathbf{x}_n^v\}_{v=1}^V, \mathbf{y}_n\}_{n=1}^{\bar{N}_{train}}$ contain $\bar{N}_{train}$ normal instances. The other $\bar{N} - \bar{N}_{train}$ normal instances and $\tilde{N}$ conflictive instances form the test set. The goal of RCML is to accurately predict $\mathbf{y}_n$ for the test instances and provide the attached prediction uncertainties

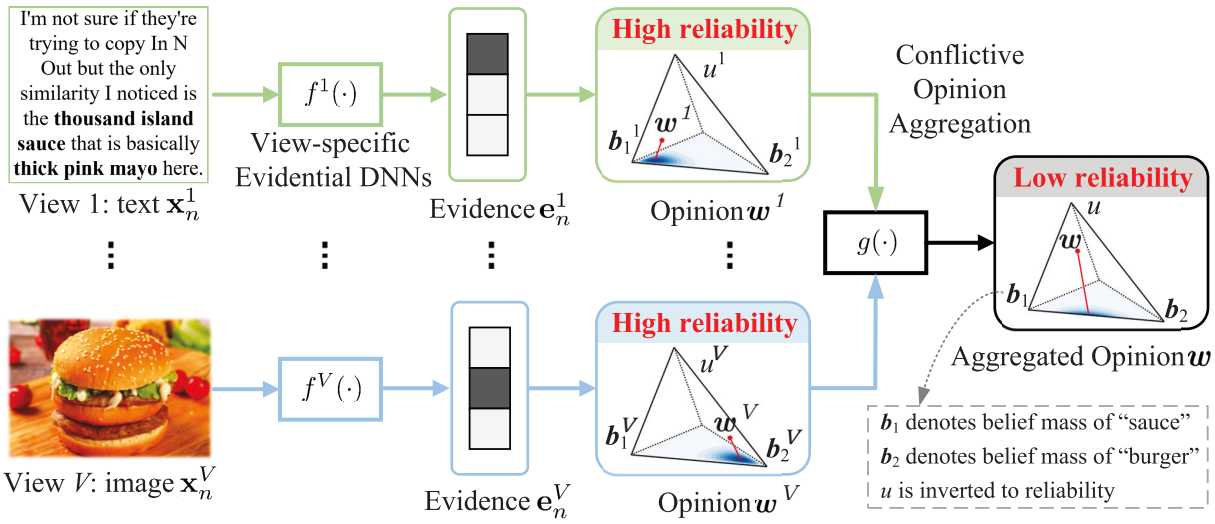

Figure 2: Illustration of ECML. View-specific DNNs collect evidence, which could be termed as the amount of support to each category. Then we form view-specific opinions consisting of belief masses of all categories and uncertainty (inverted to reliability). Finally, we integrate opinions by conflictive opinion aggregation. The uncertainty of the aggregated opinion might increase if view-specific opinions are conflictive.

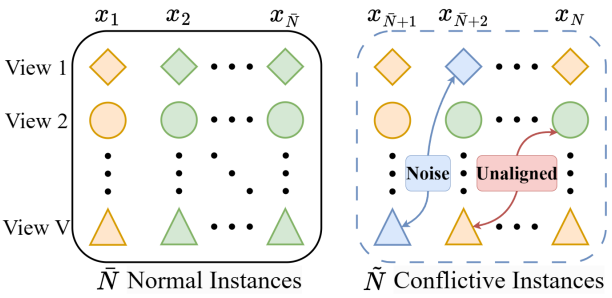

Figure 3: Notations for conflictive multi-view data. The two categories are marked as yellow and green respectively. Conflictive instances contain noise and unalignment views: noise views are marked as blue and do not belong to any ground-truth categories; unaligned views show different categories from other views.

$u_n \in [0,1]$[1] to measure the decision reliability $(1-u_n)$.

Evidential Conflictive Multi-view Learning

As shown in Fig. 2, the overall architecture consists of view-specific evidential learning and evidential multi-view fusion stages. In the first stage, we learn view-specific evidence by evidential DNNs, which could be termed as the amount of support to each category collected from data. Then the view-specific distributions of the class probabilities are modeled by Dirichlet distribution, parameterized with view-specific evidence. From the distributions, we can construct opinions consisting of belief mass vector and decision reliability. Specifically, we calculate the conflictive degree according to the projected distance and conjunctive certainty among views. By minimizing the conflictive degree, we force the view-specific DNNs to well capture multi-view common information in the training stage. This would reduce the decision conflict caused by view-specific DNNs, i.e., normal instances mistake for conflictive instance since view-specific models make wrong decisions. In the second stage, we propose a conflictive opinion aggregation strategy and establish a simple and effective average pooling fusion layer accordingly. Details will be elaborated as below.

View-specific Evidencial Deep Learning. Most existing deep multi-view learning methods commonly rely on employing a softmax layer atop deep DNNs for classification purposes. However, these softmax-based DNNs face limitations in accurately estimating predictive uncertainty. This is due to the fact that the softmax score essentially provides a single-point estimation of a predictive distribution, leading to over-confident outputs even in cases of false predictions.

To solve this, we employ EDL (Sensoy, Kaplan, and Kandemir 2018) in the view-specific learning stage. EDL was developed to address the above limitation by introducing the evidence framework of subjective logic (SL) (Jøsang 2016). In this context, evidence refers to the metrics collected from the input to support the classification process. We collect evidence, $\{\boldsymbol{e}_n^v\}$ by view-specific evidential DNNs $\{f^v(\cdot)\}_{v=1}^V$.

For K classification problems, a multinomial opinion over a specific view of an instance $(\mathbf{x}_n^v)$[2] can be represented as an ordered triplet $w = (\boldsymbol{b}, u, \boldsymbol{a})$, where the belief mass $\boldsymbol{b} = (b_1, ..., b_k)^\top$ assigns belief masses to possible values of

[1]The "ground truth" uncertainties are usually not available. We manually construct conflictive instances and expect large prediction uncertainties. We elaborate the construction approach in the experiment section.

[2]In this section, we omit the super- and sub-scripts for clarify.

the instance based on the evidence support for each value. The uncertainty mass u captures the degree of ambiguity or vacuity in the evidence, while the base rate distribution $\boldsymbol{a} = (a_1, ..., a_k)^\top$ represents the prior probability distribution over each class k. Subjective logic dictates that both $\boldsymbol{b}$ and u must be non-negative and their sum should equal one:

$$\sum_{k=1}^{K} b_k + u = 1, \forall k \in [1, ..., K], \tag{1}$$

where $b_k \geq 0$ and $u \geq 0$. The projected probability distribution of multinomial opinions is given by:

$$P_k = b_k + a_k u, \forall k \in [1, ..., K]. \tag{2}$$

Normally, the prior probabilities are manually set according to prior knowledge. For example, a common approach is to set the prior probabilities equal for each category, denoted as $a_k = 1/K$. This implies all categories have similar data instance numbers.

The Dirichlet Probability Density Function (PDF) is used for forming category distribution. It can model second-order uncertainty, while the probability values in the softmax layer only capture first-order uncertainty. The probability density function of the Dirichlet distribution is given by:

$$D(\mathbf{p}|\boldsymbol{\alpha}) = \begin{cases} \frac{1}{B(\boldsymbol{\alpha})} \prod_{k=1}^{K} p_k^{\alpha_k - 1}, & for\ \mathbf{p} \in \mathcal{S}_K, \\ 0, & otherwise, \end{cases} \tag{3}$$

where $\boldsymbol{p} = (p_1, ..., p_k)^\top$ is the probability that the instance is assigned to k-th class, $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_k)^\top$ represents the Dirichlet parameters, $\mathcal{S}_K$ is the K-dimensional unit simplex, defined as:

$$\mathcal{S}_K = \left\{ \mathbf{p} | \sum_{k=1}^{K} p_k = 1\ and\ 0 \leq p_1, ..., p_k \leq 1 \right\}, \tag{4}$$

and $B(\boldsymbol{\alpha})$ is the K-dimensional multinomial beta function.

The Dirichlet PDF naturally reflects a random sampling of statistical events, which is the basis for the aleatory interpretation of opinions as statistical measures of likelihood. Then, uncertainty mass can be well expressed in the form of Dirichlet PDFs. Uncertainty mass in the Dirichlet model reflects the vacuity of evidence. Interpreting uncertainty mass as vacuity of evidence reflects the property that "the fewer observations the more uncertainty mass".

We calculate the Dirichlet distribution parameters $\boldsymbol{\alpha}$ by $\boldsymbol{\alpha} = \boldsymbol{e} + \mathbf{1}$ to guarantee the parameters are larger than one, and hence the Dirichlet distribution is non-sparse. A mapping between the multinomial opinion and Dirichlet distribution can be given by:

$$b_k = \frac{e_k}{S} = \frac{\alpha_k - 1}{S}, u = \frac{K}{S}, \tag{5}$$

where $S = \sum_{k=1}^{K} (e_k + 1) = \sum_{k=1}^{K} \alpha_k$ is the Dirichlet strength, $\boldsymbol{e} = (e_1, ..., e_K)^\top$. It is important to note that the level of uncertainty is inversely proportional to the amount of total evidence available. In the absence of any evidence, the belief for each view is 0, resulting in maximum uncertainty, i.e., 1. Specifically, the class probability p_k could be computed as $p_k = \alpha_k / S$.

Through the view-specific evidential learning stage, we obtain the view-specific opinion and the corresponding category distribution.

Evidential Multi-view Fusion via Conflictive Opinion Aggregation. In this subsection, we focus on multi-view fusion according to view-specific opinions. The noise views of the conflictive multi-view data would show high uncertainty. We would diminish their impact in the fusion stage. The unaligned views of the conflictive multi-view data would provide highly conflicting opinions with low uncertainty, which may indicate that one or more views are unreliable. In this case, we are hard to judge which view is high-quality. In fact, the uncertainty of multi-view learning results should not decrease with the increase of the number of views, but should be related to the quality of the perspectives to be fused, especially when the learning results of two views conflict. To solve this, we propose a new conflictive opinion aggregation method.

Definition 1 Conflictive Opinion Aggregation. Let $\boldsymbol{w}^A = (\boldsymbol{b}^A, u^A, \boldsymbol{a}^A)$ and $\boldsymbol{w}^B = (\boldsymbol{b}^B, u^B, \boldsymbol{a}^B)$ be the opinions of view A and B over the same instance, respectively. The conflictive aggregated opinion $\boldsymbol{w}^{A\underline{\Diamond}B}$ is calculated in the following manner:

$$\boldsymbol{w}^{A\underline{\Diamond}B} = \boldsymbol{w}^A \underline{\Diamond} \boldsymbol{w}^B = (\boldsymbol{b}^{A\underline{\Diamond}B}, u^{A\underline{\Diamond}B}, \boldsymbol{a}^{A\underline{\Diamond}B}), \tag{6}$$

$$b_k^{A\underline{\Diamond}B} = \frac{b_k^A u^B + b_k^B u^A}{u^A + u^B}, \tag{7}$$

$$u^{A\underline{\Diamond}B} = \frac{2u^A u^B}{u^A + u^B}, \boldsymbol{a}_k^{A\underline{\Diamond}B} = \frac{\boldsymbol{a}_k^A + \boldsymbol{a}_k^B}{2}. \tag{8}$$

The opinion $\boldsymbol{w}^{A\underline{\Diamond}B}$ represents the combination of the dependent opinions of A and B. This combination is achieved by mapping the belief opinions to evidence opinions using a bijective mapping between multinomial opinions and the Dirichlet distribution. Essentially, the combination rule ensures that the quality of the new opinion is proportional to the combined one. In other words, when a highly uncertain opinion is combined, the uncertainty of the new opinion is larger than the original opinion. The averaging belief fusion can be computed simply by averaging the evidence. A more detailed explanation is shown in Proposition 1.

Following Definition 1, we can fusion the final joint opinions $\boldsymbol{w}$ from different views with the following rule:

$$\boldsymbol{w} = \boldsymbol{w}^1 \underline{\Diamond} \boldsymbol{w}^2 \underline{\Diamond} ... \underline{\Diamond} \boldsymbol{w}^V. \tag{9}$$

According to the above fusion rules, we can get the final multi-view joint opinion, and thus get the final probability of each class and the overall uncertainty.

We also aim to: 1) ensure the consistency of the model in different views during the training stage (using normal instances); 2) get an intuitive sense of the level of conflict. Therefore, we introduce a measure named the conflictive degree in Definition 2, which is established according to opinion entropy.

Definition 2 Conflictive Degree. Given two opinions $\boldsymbol{w}^A$ and $\boldsymbol{w}^B$ over an instance, the conflictive degree between $\boldsymbol{w}^A$ and $\boldsymbol{w}^B$ is defined as:

$$c(\boldsymbol{w}^A, \boldsymbol{w}^B) = c_p(\boldsymbol{w}^A, \boldsymbol{w}^B) \cdot c_c(\boldsymbol{w}^A, \boldsymbol{w}^B), \tag{10}$$

where $c_p(\boldsymbol{w}^A, \boldsymbol{w}^B)$ is the projected distance between $\boldsymbol{w}^A$ and $\boldsymbol{w}^B$, $c_c(\boldsymbol{w}^A, \boldsymbol{w}^B)$ is the conjunctive certainty between

$\boldsymbol{w}^A$ and $\boldsymbol{w}^B$, which can be formulated as follows:

$$c_p(\boldsymbol{w}^A, \boldsymbol{w}^B) = \frac{\sum_{k=1}^{K} \left| p_k^A - p_k^B \right|}{2}, \tag{11}$$

$$c_c(\boldsymbol{w}^A, \boldsymbol{w}^B) = (1 - u^A)(1 - u^B). \tag{12}$$

Intuitively, this metric ensures two things: (1) The scenario where $c = 0$ arises when the same projected probability distributions are observed, indicating non-conflicting opinions; (2) $c = 1$ arises when absolute opinions are present but with different projected probabilities. Specifically, when $c_c = 0$, it indicates that the vacuous condition is present in one or both opinions. On the other hand, when $c_c = 1$, it signifies that the opinions are considered credible, meaning they have zero uncertainty mass.

Loss Function. In this subsection, we will introduce the training DNN to obtain the multi-view joint opinion. Traditional DNN can be easily converted into evidential DNN with minimal modifications, as demonstrated in (Sensoy, Kaplan, and Kandemir 2018). This transformation primarily involves replacing the softmax layer with an activation layer (e.g., ReLU) and considering the non-negative output of this layer as evidence. By doing so, we can obtain the parameters of the Dirichlet distribution.

For instance $\{\mathbf{x}_n^v\}_{v=1}^V$, $\mathbf{e}_n^v = f^v(\mathbf{x}_n^v)$ represent the evidence vector predicted by the network for the classification. $\boldsymbol{\alpha}_n^v = \mathbf{e}_n^v + \mathbf{1}$ is the parameters of the corresponding Dirichlet distribution. In the case of conventional neural network-based classifiers, the cross-entropy loss is typically employed. However, we need to adapt the cross-entropy loss to account for the evidence-based approach:

$$\begin{aligned} L_{ace}(\boldsymbol{\alpha}_n) &= \int \left[\sum_{j=1}^{K} -y_{nj} \log p_{nj} \right] \frac{\sum_{j=1}^{K} p_{nj}^{\alpha_{nj}-1}}{B\left(\boldsymbol{\alpha}_n\right)} d\mathbf{p}_n \\ &= \sum_{j=1}^{K} y_{nj} (\psi(S_n) - \psi(\alpha_{nj})), \end{aligned} \tag{13}$$

where $\psi(\cdot)$ is the digamma function.

The above loss function does not guarantee that the evidence generated by the incorrect labels is lower. To address this issue, we can introduce an additional term in the loss function, namely the Kullback-Leibler (KL) divergence:

$$\begin{aligned} L_{KL}(\boldsymbol{\alpha}_n) &= KL\left[D(\boldsymbol{p}_n|\tilde{\boldsymbol{\alpha}}_n) \parallel D(\boldsymbol{p}_n|\mathbf{1})\right] \\ &= \log\left(\frac{\Gamma(\sum_{k=1}^{K} \tilde{\alpha}_{nk})}{\Gamma(K) \prod_{k=1}^{K} \Gamma(\tilde{\alpha}_{nk})}\right) \\ &\quad + \sum_{k=1}^{K} (\tilde{\alpha}_{nk} - 1) \left[\psi(\tilde{\alpha}_{nk}) - \psi(\sum_{j=1}^{K} \tilde{\alpha}_{nj}) \right], \end{aligned} \tag{14}$$

where $D(\mathbf{p}_n|\mathbf{1})$ is the uniform Dirichlet distribution, $\tilde{\boldsymbol{\alpha}}_n = \mathbf{y}_n + (\mathbf{1} - \mathbf{y}_n) \odot \boldsymbol{\alpha}_n$ is the Dirichlet parameters after removal of the non-misleading evidence from predicted parameters $\boldsymbol{\alpha}_n$ for the n-th instance, and $\Gamma(\cdot)$ is the gamma function.

Therefore, given the Dirichlet distribution with parameter α_n for the n-th instance, the loss is:

$$L_{acc}(\boldsymbol{\alpha}_n) = L_{ace}(\boldsymbol{\alpha}_n) + \lambda_t L_{KL}(\boldsymbol{\alpha}_n), \tag{15}$$

where $\lambda_t = min(1.0, t/T) \in [0, 1]$ is the annealing coefficient, t is the index of the current training epoch, and T is the annealing step. By gradually increasing the influence of KL divergence in loss, premature convergence of misclassified instances to uniform distribution can be avoided.

In order to ensure the consistency of results between different opinions during training, minimizing the degree of conflict between opinions was adopted. The consistency loss for the instance $\{\mathbf{x}_n^v\}_{v=1}^V$ is calculated as:

$$L_{con} = \frac{1}{V-1} \sum_{p=1}^{V} \left(\sum_{q \neq p}^{V} c(\boldsymbol{w}_n^p, \boldsymbol{w}_n^q) \right). \tag{16}$$

To sum up, the overall loss function for a specific instance $\{\mathbf{x}_n^v\}_{v=1}^V$ can be calculated as:

$$L = L_{acc}(\boldsymbol{\alpha}_n) + \beta \sum_{v=1}^{V} L_{acc}(\boldsymbol{\alpha}_n^v) + \gamma L_{con}. \tag{17}$$

The model optimization is elaborated in Algorithm 1 (Technical Appendix).

Discussion and Analyses

In this subsection, we theoretically analyze the advantages of ECML, especially the conflictive opinion aggregation for the conflictive multi-view data. The following propositions provide the theoretical analysis to support the conclusions. The proof is shown in the Technical Appendix.

Proposition 1 The conflictive opinion aggregation $\boldsymbol{w}^{A\underline{\Diamond}B} = \boldsymbol{w}^A \underline{\Diamond} \boldsymbol{w}^B$ is equivalent to averaging the view-specific evidences $\boldsymbol{e}^{A\underline{\Diamond}B} = \frac{1}{2}(\boldsymbol{e}^A + \boldsymbol{e}^B)$.

Based on this proposition, in the multi-view fusion stage, we establish a simple and effective average pooling fusion layer, $g(\cdot)$, to realize conflictive opinion aggregation.

Proposition 2 For the conflictive opinion aggregation, after aggregating a new opinion into the original opinion, if the uncertain mass of the new opinion is smaller than the original uncertain mass, the uncertain mass of the aggregated opinion would be smaller than the original one; conversely, it would be larger.

An important characteristic of most existing trust multi-view learning methods (Han et al. 2021; Jung et al. 2022; Liu et al. 2022; Xu et al. 2022; Liu et al. 2023; Zhang et al. 2023c) is "After integrating another opinion into the original opinion, the obtained uncertainty mass will be reduced." We argue that this is unreasonable since: 1) when integrating a reliable perspective, the fusion process should ideally reduce the overall uncertainty; 2) when incorporating an unreliable or conflicting perspective, the fusion should increase the uncertainty. Furthermore, existing methods often overlook the possibility of conflicts between opinions gathered from different views. These conflicts may arise due to misaligned data or variations in the model's performance across different views.

To illustrate this issue, let's consider the scenario of two observers, A and B, observing colored balls drawn from a box. The observers could be seem as sensors to collect multi-view data. Different kinds of observers (normal or color

blind) can produce conflictive multi-view data. The balls can be one of four colors: black, white, red, or green. Observer B is color blind, specifically having difficulty distinguishing between red and green balls while being able to differentiate between other color combinations. On the other hand, Observer A has perfect color vision and can usually identify the correct color when a ball is selected. Consequently, when a red ball is chosen, observer A typically identifies it as red, while observer B may perceive it as green. This disagreement between A and B leads to conflicting opinions regarding the same object.

Assuming that it is initially unknown whether one of the observers is color blind, their opinions are considered equally reliable. However, the existing fusion methods would erroneously reduce the uncertainty after combining the opinions of both observers. But in this case, we should treat the information from each perspective equally, given the possibility of conflicting opinions, and not automatically reduce the uncertainty. In summary, existing fusion methods should take into account the potential conflicts between opinions collected from different views and treat each perspective's information equally, rather than assuming a reduction in uncertainty based solely on the fusion process.

Experiments

In this section, we evaluate ECML on 6 real-world multi-view datasets. Furthermore, we also analyze the conflictive degree and uncertainty on conflictive multi-view data.

Experimental Setup

Datasets. **HandWritten**[3] comprises 2000 instances of handwritten numerals ranging from '0' to '9', with 200 patterns per class. It is represented using six feature sets. **CUB**[4]consists of 11788 instances associated with text descriptions of 200 different categories of birds. In this study, we focus on the first 10 categories and extract image features using GoogleNet and corresponding text features using doc2vec. **HMDB**[5] is a large-scale human action recognition dataset containing 6718 instances from 51 action categories. We extract the HOG and MBH features as multiple views for this dataset. **Scene15**[6] includes 4485 images from 15 indoor and outdoor scene categories. We extract three types of features GIST, PHOG, and LBP. **Caltech101**[7] comprises 8677 images from 101 classes. We select the first 10 categories and extract two deep features (views) using DECAF and VGG19 models. **PIE**[8] contains 680 instances belonging to 68 classes. We extract intensity, LBP, and Gabor as 3 views. Table 1 summarizes a summary of the datasets.

[3]https://archive.ics.uci.edu/ml/datasets/Multiple+Features

[4]http://www.vision.caltech.edu/visipedia/CUB-200.html

[5]https://serre-lab.clps.brown.edu/resource/hmdb-a-large-human-motion-database

[6]https://doi.org/10.6084/m9.figshare.7007177.v1

[7]http://www.vision.caltech.edu/Image Datasets/Caltech101

[8]http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/Multi-Pie/Home.html

Dataset	Size	K	Dimensionality
HandWritten	2000	10	240/76/216/47/64/6
CUB	11788	10	1024/300
HMDB	6718	51	1000/1000
Scene15	4485	15	20/59/40
Caltech101	8677	101	4096/4096
PIE	680	68	484/256/279

Table 1: Dataset summary.

Compared Methods. The baselines based on feature fusion include: (1) **DCCAE** (Deep Canonically Correlated AutoEncoders) (Wang et al. 2015) is the classical method, which employs autoencoders to seek a common representation. (2) **CPM-Nets** (Cross Partial Multi-view Networks) (Zhang et al. 2022) is a SOTA multi-view feature fusion method, which focuses on learning a versatile representation to handle complex correlations among different views. (3) **DUA-Nets** (Dynamic Uncertainty-Aware Networks) (Geng et al. 2021) is an uncertainty-aware method, which utilizes reversal networks to integrate intrinsic information from different views into a unified representation. The baselines based on decision fusion include: (1) **TMC** (Trusted Multi-view Classification) (Han et al. 2021) is the pioneer uncertainty-aware method, which addresses the uncertainty estimation problem and produces reliable classification results. (2) **TMDL-OA** (Trusted Multi-View Deep Learning with Opinion Aggregation) (Liu et al. 2022) is a SOTA multi-view decision fusion method, which is also based on the evidential DNN and proposes a consistency measure loss to achieve trustworthy learning results.

To create a test set with conflictive instances, we perform the following transformations: (1) For noise views, we introduce Gaussian noise with varying levels of standard deviations σ to a partial percentage of the test instances. (2) For unaligned views, we select a portion of the instances and modify the information of a random view, causing the label corresponding to that view to be misaligned with the true label of the instance. We conduct 10 runs for each method and report the mean values and standard deviations.

Experiment Results

Performance Comparison. Tables 2 and 3 show the classification performance on normal and conflictive test sets, respectively. We obtain that: (1) Even on normal test sets, ECML outperforms all the other baselines. For instance, on the HMDB dataset, ECML achieves an accuracy improvement of approximately 2.64% compared to the second-best (TMDL-OA) model. The reason would be attributed to the incorporation of consistency loss, which enhances the model's learning capability, as validated by the ablation study. (2) When evaluating on conflictive test sets, the accuracy of all the compared methods notably decreases. Nonetheless, thanks to the conflictive opinion aggregation, ECML exhibits an awareness of view-specific conflicts, leading to impressive results across all datasets. This highlights the effectiveness of ECML for both normal and conflictive multi-view data.

Data	DCCAE	CPM-Nets	DUA-Nets	TMC	TMDL-OA	Ours	△%
HandWritten	95.45±0.35	94.55±1.36	98.10±0.32	98.51±0.13	99.25±0.45	**99.40 ± 0.00**	0.15
CUB	85.39±1.36	89.32±0.38	80.13±1.67	90.57±2.96	95.43±0.20	**98.50 ± 2.75**	3.21
HMDB	49.12±1.07	63.32±0.43	62.73±0.23	65.17±2.42	88.20±0.58	**90.84 ± 1.86**	2.99
Scene15	55.03±0.34	67.29±1.01	68.23±0.11	67.71±0.30	75.57±0.02	**76.19 ± 0.12**	0.82
Caltech101	89.56±0.41	90.35±2.12	93.43±0.34	92.80±0.50	94.63±0.04	**95.36 ± 0.38**	0.77
PIE	81.96±1.04	88.53±1.23	90.56±0.47	91.85±0.23	92.33±0.36	**94.71 ± 0.02**	2.57

Table 2: Accuracy (%) on normal test sets. The best and the second best results are highlighted by boldface and underlined respectively. △% denotes the performance improvement of ECML over the best baseline.

Data	DCCAE	CPM-Nets	DUA-Nets	TMC	TMDL-OA	Ours	△%
HandWritten	82.85±0.38	83.34±1.07	87.16±0.34	92.76±0.15	93.05±0.05	**94.40 ± 0.05**	1.45
CUB	63.57±1.28	68.82±0.17	60.53±1.17	73.37±2.16	74.43±0.26	**76.50 ± 1.15**	2.78
HMDB	29.62±1.79	42.62±1.43	43.53±0.28	47.17±0.15	67.62±0.28	**70.84 ± 1.19**	4.76
Scene15	25.97±2.86	29.63±1.12	26.18±1.31	42.27±1.61	48.42±1.02	**56.97 ± 0.52**	17.66
Caltech101	60.90±2.32	66.54±2.89	75.19±2.34	90.16±2.50	90.63±2.05	**92.36 ± 1.48**	1.91
PIE	26.89±1.10	53.19±1.17	56.45±1.75	61.65±1.03	68.16±0.34	**84.00 ± 0.14**	23.24

Table 3: Accuracy (%) on conflictive test sets.

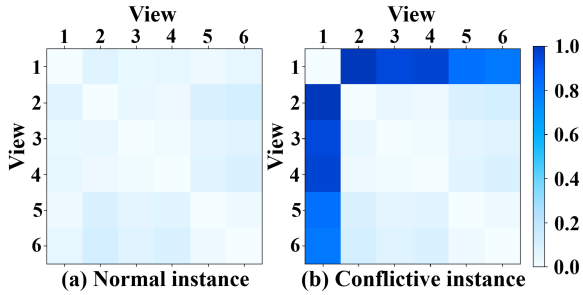

Figure 4: Conflictive degree visualization.

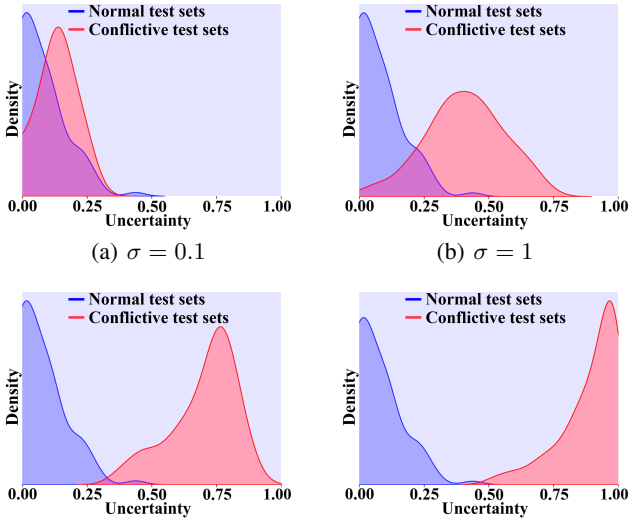

Figure 5: Density of uncertainty.

Conflictive Degree Visualization Fig. 4 shows the conflictive degree on the HandWritten dataset with six views. The left and right parts show the conflictive degree of normal and conflictive instances, respectively. To create conflicts, we modify the content from the first view, causing unaligned content from the other views. The results clearly demonstrate that ECML can effectively capture and quantify conflictive degrees between views. This finding further validates the reliability of ECML.

Uncertainty Estimation To further evaluate the estimated uncertainty, we visualize the distribution of normal and conflictive test sets on the CUB dataset. To construct conflictive test sets, we introduce Gaussian noise with standard deviation $\sigma = 0.1, 1, 5, 10$ to 50% of the test instances. The experimental results are presented in Fig. 5. The results reveal that, when the noise intensity is low ($\sigma = 0.1$), the distribution curve of the conflictive instances closely aligns with that of the normal instances. However, as the noise intensity increases, the uncertainty of the conflictive instances also increases. This finding indicates that the estimated uncertainty is correlated with the quality of the instances, thereby validating the capability of our method in uncertainty estimation.

Conclusion

In this paper, we proposed an Evidential Conflictive Multi-view Learning (ECML) method for the RCML problem. ECML tries to form view-specific opinions consisting of belief mass vector and decision reliability. It further aggregates conflictive opinions by a simple and effective average pooling layer. We theoretically proved it can exactly model the relation of multi-view common and view-specific reliabilities. Furthermore, we also extended our method by minimizing the degree of conflict between opinions to guarantee the consistency of results between different opinions. Experimental results on six real-world datasets confirmed the effectiveness of ECML.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant Nos. 62133012, 61936006, 62103314, 62302370, 62073255, 62303366), the Key Research and Development Program of Shanxi (Program No. 2020ZDLGY04-07), Innovation Capability Support Program of Shanxi (Program No. 2021TD-05), Natural Science Basic Research Program of Shaanxi under Grant No.2023-JC-QN-0648, the Open Project of Anhui Provincial Key Laboratory of Multimodal Cognitive Computation, Anhui University, No. MMC202105.

References

Ektefaie, Y.; Dasoulas, G.; Noori, A.; Farhat, M.; and Zitnik, M. 2023. Multimodal learning with graphs. *Nature Machine Intelligence*, 5(4): 340–350.

Fan, G.; Zhang, C.; Wang, K.; and Chen, J. 2023. MV-HAN: A Hybrid Attentive Networks Based Multi-View Learning Model for Large-Scale Contents Recommendation. In *Proceedings of the IEEE/ACM International Conference on Automated Software Engineering*.

Gal, Y.; and Ghahramani, Z. 2016. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In *International Conference on Machine Learning*, volume 48, 1050–1059.

Gawlikowski, J.; Tassi, C. R. N.; Ali, M.; Lee, J.; Humt, M.; Feng, J.; Kruspe, A.; Triebel, R.; Jung, P.; Roscher, R.; et al. 2023. A survey of uncertainty in deep neural networks. *Artificial Intelligence Review*, 1–77.

Geng, Y.; Han, Z.; Zhang, C.; and Hu, Q. 2021. Uncertainty-Aware Multi-View Representation Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(9): 7545–7553.

Han, Z.; Zhang, C.; Fu, H.; and Zhou, J. T. 2021. Trusted Multi-View Classification. In *International Conference on Learning Representations*.

Han, Z.; Zhang, C.; Fu, H.; and Zhou, J. T. 2023. Trusted Multi-View Classification With Dynamic Evidential Fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2): 2551–2566.

Hou, D.; Cong, Y.; Sun, G.; Dong, J.; Li, J.; and Li, K. 2022. Fast Multi-View Outlier Detection via Deep Encoder. *IEEE Transactions on Big Data*, 8(4): 1047–1058.

Huang, S.; Wu, H.; Ren, Y.; Tsang, I.; Xu, Z.; Feng, W.; and Lv, J. 2022. Multi-view Subspace Clustering on Topological Manifold. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 25883–25894.

Huang, Z.; Hu, P.; Zhou, J. T.; Lv, J.; and Peng, X. 2020. Partially View-aligned Clustering. In *Advances in Neural Information Processing Systems*, volume 33, 2892–2902.

Huang, Z.; Ren, Y.; Pu, X.; Huang, S.; Xu, Z.; and He, L. 2023. Self-Supervised Graph Attention Networks for Deep Weighted Multi-View Clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(7): 7936–7943.

Jøsang, A. 2016. *Subjective logic*, volume 3. Springer.

Jung, M. C.; Zhao, H.; Dipnall, J.; Gabbe, B.; and Du, L. 2022. Uncertainty estimation for multi-view data: The power of seeing the whole picture. *Advances in Neural Information Processing Systems*, 35: 6517–6530.

Lakshminarayanan, B.; Pritzel, A.; and Blundell, C. 2017. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. In *Advances in Neural Information Processing Systems*, volume 30.

Lampert, C. H.; and Krömer, O. 2010. Weakly-Paired Maximum Covariance Analysis for Multimodal Dimensionality Reduction and Transfer Learning. In *Computer Vision – ECCV 2010*, 566–579.

Liang, P. P.; Zadeh, A.; and Morency, L.-P. 2022. Foundations and recent trends in multimodal machine learning: Principles, challenges, and open questions. *arXiv preprint arXiv:2209.03430*.

Liu, W.; Chen, Y.; Yue, X.; Zhang, C.; and Xie, S. 2023. Safe Multi-View Deep Classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 8870–8878.

Liu, W.; Yue, X.; Chen, Y.; and Denoeux, T. 2022. Trusted Multi-View Deep Learning with Opinion Aggregation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(7): 7585–7593.

Lyzhov, A.; Molchanova, Y.; Ashukha, A.; Molchanov, D.; and Vetrov, D. 2020. Greedy Policy Search: A Simple Baseline for Learnable Test-Time Augmentation. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 124, 1308–1317.

Marcos Alvarez, A.; Yamada, M.; Kimura, A.; and Iwata, T. 2013. Clustering-Based Anomaly Detection in Multi-View Data. In *Proceedings of the ACM International Conference on Information & Knowledge Management*, 1545–1548.

Mostafazadeh, N.; Brockett, C.; Dolan, W. B.; Galley, M.; Gao, J.; Spithourakis, G.; and Vanderwende, L. 2017. Image-Grounded Conversations: Multimodal Context for Natural Question and Response Generation. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 462–472.

Qin, Y.; Peng, D.; Peng, X.; Wang, X.; and Hu, P. 2022. Deep Evidential Learning with Noisy Correspondence for Cross-Modal Retrieval. In *Proceedings of the ACM International Conference on Multimedia*, 4948–4956.

Sensoy, M.; Kaplan, L.; and Kandemir, M. 2018. Evidential Deep Learning to Quantify Classification Uncertainty. In *Advances in Neural Information Processing Systems*, volume 31.

Tan, Y.; Kong, C.; Yu, L.; Li, P.; Chen, C.; Zheng, X.; Hertzberg, V. S.; and Yang, C. 2022. 4sdrug: Symptom-based set-to-set small and safe drug recommendation. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 3970–3980.

Wang, W.; Arora, R.; Livescu, K.; and Bilmes, J. 2015. On Deep Multi-View Representation Learning. In *Proceedings of the International Conference on International Conference on Machine Learning - Volume 37*, 1083–1092.

Wang, X.; Peng, D.; Hu, P.; and Sang, Y. 2019. Adversarial correlated autoencoder for unsupervised multi-view representation learning. *Knowledge-Based Systems*, 168: 109–120.

Wen, J.; Liu, C.; Deng, S.; Liu, Y.; Fei, L.; Yan, K.; and Xu, Y. 2023a. Deep Double Incomplete Multi-View Multi-Label Learning With Incomplete Labels and Missing Views. *IEEE Transactions on Neural Networks and Learning Systems*, 1–13.

Wen, J.; Liu, C.; Xu, G.; Wu, Z.; Huang, C.; Fei, L.; and Xu, Y. 2023b. Highly Confident Local Structure Based Consensus Graph Learning for Incomplete Multi-View Clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15712–15721.

Wen, J.; Zhang, Z.; Fei, L.; Zhang, B.; Xu, Y.; Zhang, Z.; and Li, J. 2022. A survey on incomplete multiview clustering. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(2): 1136–1149.

Wen, Y.; Wang, S.; Liao, Q.; Liang, W.; Liang, K.; Wan, X.; and Liu, X. 2023c. Unpaired Multi-View Graph Clustering With Cross-View Structure Matching. *IEEE Transactions on Neural Networks and Learning Systems*, 1–15.

Xu, C.; Guan, Z.; Zhao, W.; Wu, H.; Niu, Y.; and Ling, B. 2019. Adversarial incomplete multi-view clustering. In *IJCAI*, volume 7, 3933–3939.

Xu, C.; Zhao, W.; Zhao, J.; Guan, Z.; Song, X.; and Li, J. 2022. Uncertainty-aware multiview deep learning for internet of things applications. *IEEE Transactions on Industrial Informatics*, 19(2): 1456–1466.

Yang, M.; Li, Y.; Huang, Z.; Liu, Z.; Hu, P.; and Peng, X. 2021. Partially View-Aligned Representation Learning With Noise-Robust Contrastive Loss. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1134–1143.

Zhang, C.; Cui, Y.; Han, Z.; Zhou, J. T.; Fu, H.; and Hu, Q. 2022. Deep Partial Multi-View Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5): 2402–2415.

Zhang, C.; Lou, Z.; Zhou, Q.; and Hu, S. 2023a. Multi-View Clustering via Triplex Information Maximization. *IEEE Transactions on Image Processing*, 32: 4299–4313.

Zhang, P.; Wang, S.; Li, L.; Zhang, C.; Liu, X.; Zhu, E.; Liu, Z.; Zhou, L.; and Luo, L. 2023b. Let the Data Choose: Flexible and Diverse Anchor Graph Fusion for Scalable Multi-View Clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 11262–11269.

Zhang, Q.; Wu, H.; Zhang, C.; Hu, Q.; Fu, H.; Zhou, J. T.; and Peng, X. 2023c. Provable Dynamic Fusion for Low-Quality Multimodal Data. *arXiv preprint arXiv:2306.02050*.

Zhang, T.; Liu, X.; Gong, L.; Wang, S.; Niu, X.; and Shen, L. 2021. Late fusion multiple kernel clustering with local kernel alignment maximization. *IEEE Transactions on Multimedia*.

Zhang, X.; Chen, M.; Mu, J.; and Zong, L. 2023d. Adaptive View-Aligned and Feature Augmentation Network for Partially View-Aligned Clustering. In *Advances in Knowledge Discovery and Data Mining*, 223–235.

Zhao, H.; Liu, H.; Ding, Z.; and Fu, Y. 2018. Consensus Regularized Multi-View Outlier Detection. *IEEE Transactions on Image Processing*, 27(1): 236–248.

Zhao, S.; Wen, J.; Fei, L.; and Zhang, B. 2023. Tensorized Incomplete Multi-View Clustering with Intrinsic Graph Completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 11327–11335.