Trusted Multi-view Learning with Label Noise

Cai Xu, Yilin Zhang, Ziyu Guan*, Wei Zhao
School of Computer Science and Technology, Xidian University
{cxu@, ylzhang_3@stu., zyguan@, ywzhao@mail.}xidian.edu.cn

Abstract

Multi-view learning methods often focus on improving decision accuracy while neglecting the decision uncertainty, which significantly restricts their applications in safety-critical applications. To address this issue, researchers propose trusted multi-view methods that learn the class distribution for each instance, enabling the estimation of classification probabilities and uncertainty. However, these methods heavily rely on high-quality ground-truth labels. This motivates us to delve into a new generalized trusted multi-view learning problem: how to develop a reliable multi-view learning model under the guidance of noisy labels? We propose a trusted multi-view noise refining method to solve this problem. We first construct view-opinions using evidential deep neural networks, which consist of belief mass vectors and uncertainty estimates. Subsequently, we design view-specific noise correlation matrices that transform the original opinions into noisy opinions aligned with the noisy labels. Considering label noises originating from low-quality data features and easily-confused classes, we ensure that the diagonal elements of these matrices are inversely proportional to the uncertainty, while incorporating class relations into the off-diagonal elements. Finally, we aggregate the noisy opinions and employ a generalized maximum likelihood loss on the aggregated opinion for model training, guided by the noisy labels. We empirically compare TMNR with state-of-the-art trusted multi-view learning and label noise learning baselines on 5 publicly available datasets. Experiment results show that TMNR outperforms baseline methods on accuracy, reliability and robustness. The code and appendix are released at https://github.com/YilinZhang107/TMNR.

1 Introduction

Multi-view data is widely present in various real-world scenarios. For instance, in the field of healthcare, a patient's comprehensive condition can be reflected through multiple types of examinations; social media applications often include multi-modal contents such as textual and visual reviews [Liu *et al.*, 2024]. Multi-view learning synthesizes both consistency and complementary information to obtain a more comprehensive understanding of the data. It has generated significant and wide-ranging influence across multiple research areas, including classification [Chen *et al.*, 2024; Wang *et al.*, 2022], clustering [Xu *et al.*, 2023; Huang *et al.*, 2023; Wen *et al.*, 2022], recommendation systems [Lin *et al.*, 2023; Nikzad-Khasmakhi *et al.*, 2021], and large language models [Min *et al.*, 2023].

Most existing multi-view learning methods focus on improving decision accuracy while neglecting the decision uncertainty. This significantly limits the application of multi-view learning in safety-critical scenes, such as healthcare. Recently, Han *et al.* propose a pioneering work [Han *et al.*, 2020], Trusted Multi-view Classification (TMC), to solve this problem. TMC calculates and aggregates the evidences of all views from the original data features. It then utilizes these evidences to parameterize the class distribution, which could be used to estimate the class probabilities and uncertainty. To train the entire model, TMC requires the estimated class probabilities to be consistent with the ground-truth labels. Following this line, researchers propose novel evidence aggregation methods, aiming to enhance the reliability and robustness in the presence of feature noise [Gan *et al.*, 2021; Qin *et al.*, 2022], conflictive views [Xu *et al.*, 2024] and incomplete views [Xie *et al.*, 2023].

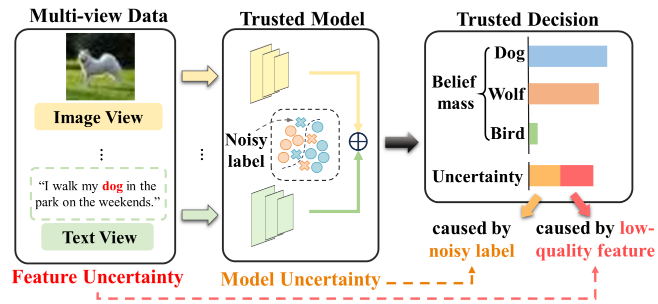

Figure 1: The generalized trusted multi-view learning problem: the model should recognize the feature and model uncertainty caused by low-quality feature and noisy labels, respectively.

*Corresponding author

Regretfully, these trusted multi-view learning methods consistently rely on high-quality ground-truth labels. The labeling task is time-consuming and expensive especially when dealing with large scale datasets, such as user generated multi-modal contents in social media applications. This motivates us to delve into a new *Generalized Trusted Multi-view Learning (GTML) problem: how to develop a reliable multi-view learning model under the guidance of noisy labels? This problem encompasses two key objectives: 1) detecting and refining the noisy labels during the training stage; 2) recognizing the model's uncertainty caused by noisy labels.* For example, instances belonging to classes like "dog" and "wolf" might exhibit similarities and are prone to being mislabelled. Consequently, the model should exhibit higher decision uncertainty in such cases. An intuitive analogy is an intern animal researcher (model) may not make high-confidence decisions for all animals (instances), but is aware of the cases where a definitive decision is challenging.

In this paper, we propose an Trusted Multi-view Noise Refining (TMNR) method for the GTML problem. We consider the label noises arising from two sources: low-quality data features, such as blurred and incomplete features, and easily-confused classes, such as classes "dog" and "wolf". Our objective is to leverage multi-view consistent information for noise detection. To achieve this, 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. We then model the view-specific distributions of class probabilities using the Dirichlet distribution, parameterized with view-specific evidence. These distributions allow us to construct opinions, which consist of belief mass vectors and uncertainty estimates. We design view-specific noise correlation matrices to transform the original opinions to the noisy opinions, which aligns with the noisy labels. Considering that low-quality data features are prone to mislabeling, we require the diagonal elements of the noise correlation matrices to be inversely proportional to the uncertainty. Additionally, we incorporate class relations into the off-diagonal elements. For instance, the elements corresponding to "dog" and "wolf" should have larger values since these two classes are easily mislabelled. Next, we aggregate the noisy opinions to obtain the common evidence. Finally, we employ a generalized maximum likelihood loss on the common evidence, guided by the noisy labels, for model training.

The main contributions of this work are summarized as follows: 1) we propose the generalized trusted multi-view learning problem, which necessitates the model's ability to make reliable decisions despite the presence of noisy guidance; 2) we propose the TMNR method to tackle this problem. TMNR mitigates the negative impact of noisy labels through two key strategies: leveraging multi-view consistent information for detecting and refining noisy labels, and assigning higher decision uncertainty to instances belonging to easily mislabelled classes; 3) we empirically compare TMNR with state-of-the-art trusted multi-view learning and label noise learning baselines on 5 publicly available datasets. Experiment results show that TMNR outperforms baseline methods on accuracy, reliability and robustness.

2 Related Work

2.1 Deep Multi-view Fusion

Multi-view fusion has demonstrated its superior performance in various tasks by effectively combining information from multiple sources or modalities [Liang *et al.*, 2021; Zhou *et al.*, 2023]. According to the fusion strategy, existing deep multi-view fusion methods can be roughly classified into feature fusion [Hu *et al.*, 2024; Xu *et al.*, 2022; Liu *et al.*, 2023] and decision fusion [Jillani *et al.*, 2020; Liu *et al.*, 2022]. A major challenge of feature fusion methods is each view might exhibit different types and different levels of noise at different points in time. Trust decision fusion methods solve this making view-specific trust decisions to obtain the view-specific reliabilities, then assigning large weights to these views with high reliability in the multi-view fusion stage. Following this line, Xie *et al.* [Xie *et al.*, 2023] tackle the challenge of incomplete multi-view classification through a two-stage approach, involving completion and evidential fusion. Xu *et al.* [Xu *et al.*, 2024] focus on making trust decisions for instances that exhibit conflicting information across multiple views. They propose an effective strategy for aggregating conflicting opinions and theoretically prove this strategy can exactly model the relation of multi-view common and view-specific reliabilities. However, it should be noted that these trusted multi-view learning methods heavily rely on high-quality ground-truth labels, which may not always be available or reliable in real-world scenarios. This limitation motivates for us to delve into the problem of GTML, which aims to learn a reliable multi-view learning model under the guidance of noisy labels.

2.2 Label-Noise Learning

In real-world scenarios, the process of labeling data can be error-prone, subjective, or expensive, leading to noisy labels. Label-noise learning refers to the problem of learning from training data that contains noisy labels. In multi-classification tasks, label-noise can be categorized as Class-Conditional Noise (CCN) and Instance-Dependent Noise (IDN). CCN occurs when the label corruption process is independent of the data features, and instances in a class are assigned to other classes with a fixed probability. Dealing with CCN noise often involves correcting losses by estimating an overall category transfer probability matrix [Patrini *et al.*, 2017; Hendrycks *et al.*, 2018]. IDN refers to instances being mislabeled based on their class and features. In this work, we focus on IDN as it closely resembles real-world noise. The main challenge lies in approximating the complex and high-dimensional instance-dependent transfer matrix. Several approaches have been proposed to address this challenge. For instance, Cheng *et al.* [Cheng *et al.*, 2020] proposes an instance-dependent sample sieve method that enables model to process clean and corrupted samples individually. Cheng *et al.* [Cheng *et al.*, 2022] effectively reduce the complexity of the instance-dependent matrix by streaming embedding. Berthon *et al.* [Berthon *et al.*, 2021] approximate the transfer distributions of each instance using confidence scores. However, the confidence depends on the pre-trained model and may not reliable. The proposed TMNR reliably bootstraps

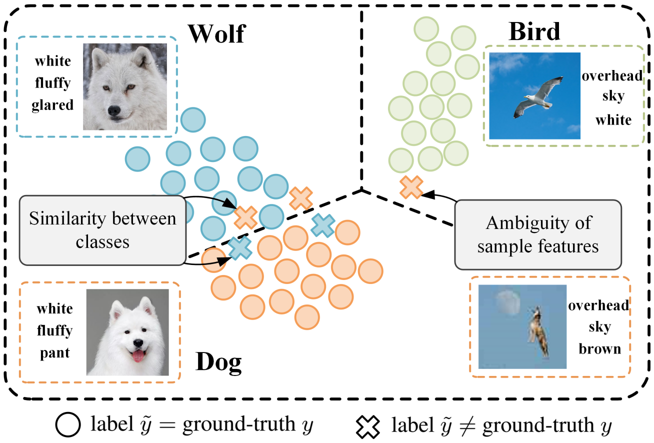

Figure 2: Illustration of the label-noise. Each color represents a ground-truth category y.

the correlation matrix based on multi-view opinions, leading to superior performance. In addition, TMNR not only refines the noise in the labels but also recognizes the model's uncertainty caused by noisy labels.

3 The Method

In this section, we first define the generalized trusted multi-view learning problem, then present Trusted Multi-view Noise Refining (TMNR) in detail, together with its implementation.

3.1 Notations and Problem Statement

We use $\{\boldsymbol{x}_n^v \in \mathbb{R}^{d_v}\}_{v=1}^V$ to denote the feature of the n-th instance, which contains V views, d_v denotes the dimension of the v-th view. $y_n \in \{1, ..., K\}$ denotes the ground-truth category, where K is the number of all categories. In the generalized trusted multi-view classification problem, the labels of some data instances contain noise as shown in Figure 2. Therefore, we utilize $\{\tilde{y}_n \in \{1, ..., K\}\}_{n=1}^N$ as the set of noisy labels that may have been corrupted

The objective is to learn a trusted classification model according to noisy training instances $\{\{\boldsymbol{x}_n^v\}_{v=1}^V, \tilde{y}_n\}_{n=1}^{N_{train}}$. For the instances of the test sets, the model should predict the category $\{y_n\}$ and uncertainty $\{u_n\}$, which can quantify the uncertainty caused by low-quality feature and noisy labels.

3.2 Trusted Multi-view Noise Refining Pipeline

As shown in Figure 3, We first construct view-specific opinions using evidential DNNs $\{f^v(\cdot)\}_{v=1}^V$. To account for the presence of label noise, the view-specific noise correlation matrices $\{T^v\}_{v=1}^V$ transform the original opinions into noisy opinions aligned with the noisy labels. Finally, we aggregates the noisy opinions and trains the whole model by the noisy labels. Details regarding each component will be elaborated as below.

View-specific Evidence Learning

In this subsection, we introduce the evidence theory to quantify uncertainty. Traditional multi-classification neural networks usually use a Softmax activation function to obtain the probability distribution of the categories. However, this provides only a single-point estimate of the predictive distribution, which can lead to overconfident results even if the predictions are incorrect. This limitation affects the reliability of the results. To address this problem, EDL [Sensoy *et al.*, 2018] introduces the evidential framework of subjective logic [Jøsang, 2016]. It converts traditional DNNs into evidential neural networks by making a small change in the activation function to get a non-negative output (e.g.,ReLU) to extract the amount of support (called evidence) for each class.

In this framework, the parameter $\boldsymbol{\alpha}$ of the Dirichlet distribution $Dir(\boldsymbol{p}|\boldsymbol{\alpha})$ is associated with the belief distribution in the framework of evidence theory, where $\boldsymbol{p}$ is a simplex representing the probability of class assignment. We collect evidence, $\{\boldsymbol{e}_n^v\}$ by view-specific evidential DNNs $\{f^v(\cdot)\}_{v=1}^V$. The corresponding Dirichlet distribution parameter is $\boldsymbol{\alpha}^v = \boldsymbol{e}^v + 1 = [\alpha_1^v, \cdots, \alpha_K^v]^T$. After obtaining the distribution parameter, we can calculate the subjective opinion $\mathcal{O}^v = (\boldsymbol{b}^v, u^v)$ of the view including the quality of beliefs $\boldsymbol{b}^v$ and the quality of uncertainty u, where $\boldsymbol{b}^v = (\boldsymbol{\alpha}^v - 1)/S^v = \boldsymbol{e}^v/S^v$, $u^v = K/S^v$, and $S^v = \sum_{k=1}^K \alpha_k^v$ is the Dirichlet intensity.

Evidential Noise Forward Correction

In the GTML problem, we expect to train the evidence network so that its output is a clean evidence distribution about the input. To minimise the negative impact of IDN in the training dataset, we modify the outputs of the DNNs with an additional structure to adjust the loss of each training sample before updating the parameters of the DNNs. This makes the optimisation process immune to label noise, called evidential noise forward correction. This structure should be removed when predicting test data.

For each view $\{\boldsymbol{x}^v\}_{v=1}^V$ of a specific instance, we construct view-specific noise correlation matrix to model the noise process:

$$\boldsymbol{T}^v = [t_{kj}^v]_{k,j=1}^K \in [0,1]^{K \times K}, \tag{1}$$

where $t_{kj}^v := P(\tilde{y} = j | y = k, \boldsymbol{x}^v)$ and $\sum_{j=1}^K t_{kj}^v = 1$.

To predict probability distributions, the noise class posterior probability is obtained by calculating the correlation matrix:

$$\begin{aligned} P(\tilde{y} = j | \boldsymbol{x}^v) &= \sum_{k=1}^K P(\tilde{y} = j, y = k | \boldsymbol{x}^v) \\ &= \sum_{k=1}^K t_{kj}^v P(y = k | \boldsymbol{x}^v). \end{aligned} \tag{2}$$

Based on the evidence theory described in the previous subsection and considering the constraints within $\boldsymbol{T}^v$ itself, we convert the transfer of predicted probabilities of instances into a transfer of the extracted support at the evidence level using the following equation:

$$\tilde{e}_j^v = \sum_{k=1}^K t_{kj}^v \, e_k^v, \tag{3}$$

where $\tilde{e}_j^v / e_k^v$ denotes the j-/k-th element in the noise/clean class-posterior evidence quantities $\tilde{\boldsymbol{e}}^v / \boldsymbol{e}^v$.

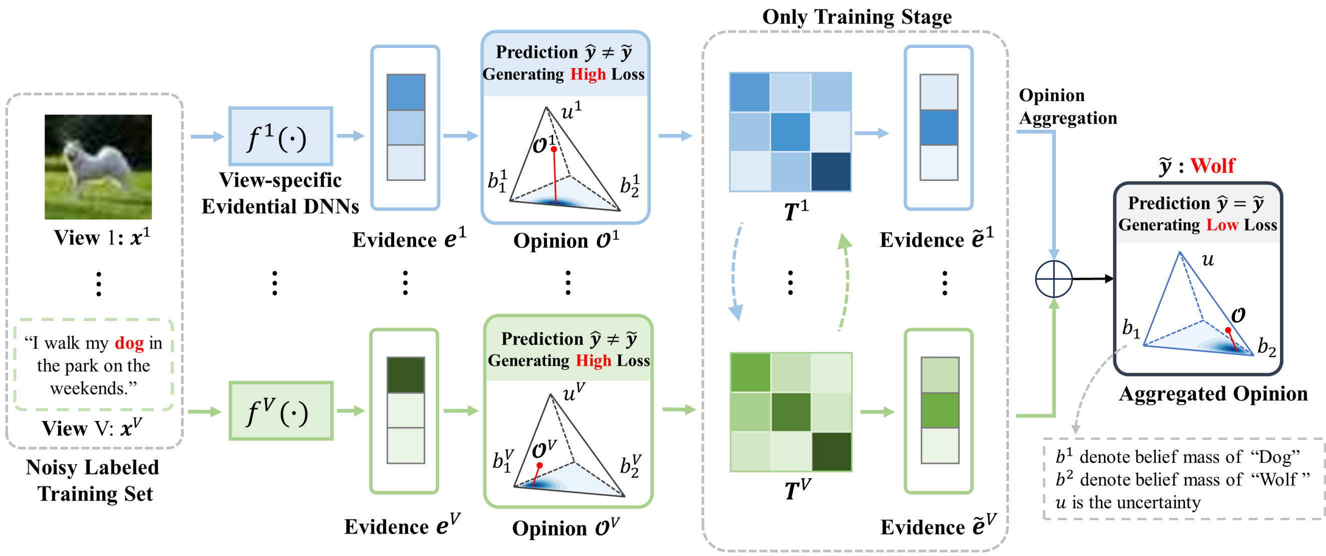

Figure 3: Illustration of TMNR. We first construct view-specific opinions using evidential DNNs $\{f^v(\cdot)\}_{v=1}^V$. Subsequently, the view-specific noise correlation matrices $\{\boldsymbol{T}^v\}_{v=1}^V$ transform the original opinions into noisy opinions aligned with the noisy labels. Finally, we aggregates the noisy opinions and trains the whole model by the noisy labels.

Therefore, the clean posterior $\boldsymbol{e}^v$ obtained from the prediction of each view is transferred to the noisy posterior $\tilde{\boldsymbol{e}}^v$. The parameter $\tilde{\boldsymbol{\alpha}}$ in the noisy Dirichlet distribution is computed in order to align the supervised labels $\tilde{y}$ that contain the noise. The entire evidence vector could be calculabed by:

$$\tilde{\boldsymbol{e}}^v = \boldsymbol{T}^{v\top} \boldsymbol{e}^v, \quad \tilde{\boldsymbol{\alpha}}^v = \tilde{\boldsymbol{e}}^v + 1. \tag{4}$$

Trust Evidential Multi-view Fusion

After obtaining opinions from multiple views, we consider dynamically integrating them based on uncertainty to produce a combined opinion. We achieve this via the Dempster's combination rule [Jøsang, 2016]. We take the clean opinion aggregation as example. Given clean opinions of two views of the same instance, i.e., $\mathcal{O}^1 = (\boldsymbol{b}^1, u^1)$ and $\mathcal{O}^2 = (\boldsymbol{b}^2, u^2)$, the computation to obtain the aggregated opinion $\mathcal{O} = \mathcal{O}^1 \diamond \mathcal{O}^2 = (\boldsymbol{b}, u)$ is defined as follows:

$$b_k = \frac{1}{1-C}\left(b_k^1 b_k^2 + b_k^1 u^2 + b_k^2 u^1\right), u = \frac{1}{1-C} u^1 u^2, \tag{5}$$

where $C = \sum_{i \neq j} b_i^1 b_j^2$.

For a set of opinions from multiple views $\{\mathcal{O}^v\}_{v=1}^V$, the joint multi-view subjective opinion is obtained by $\mathcal{O} = \mathcal{O}^1 \diamond \mathcal{O}^2 \diamond \cdots \mathcal{O}^V$. The corresponding parameters of the $Dir(\boldsymbol{p}|\boldsymbol{\alpha})$ are obtained with $S = K/u$, $\alpha_k = b_k \times S + 1$ and the final probability could be estimated via $p_k = \alpha_k / S$.

3.3 Loss Function

In this section, we explore the optimization of the parameter set $\{\boldsymbol{\theta}, \boldsymbol{\omega}\}$ in the evidence extraction network $f(\cdot; \boldsymbol{\theta})$ and the correlation matrices $\{\{\boldsymbol{T}_n^v\}_{v=1}^V\}_{n=1}^{N_{train}}$.

Classification Loss

We capture view-specific evidence from a single view $\boldsymbol{x}^v$ of a sample. The vector $\boldsymbol{e}^v = f^v(\boldsymbol{x}^v)$ denotes the clean class-posterior evidence obtained from the corresponding view network prediction. This evidence undergoes correction through Eq. (4) to yield the noisy class-posterior evidence, denoted as $\tilde{\boldsymbol{e}}^v$, along with the associated Dirichlet parameter $\tilde{\boldsymbol{\alpha}}^v$. Then in the constructed inference framework, the evidence multi-classification loss $\mathcal{L}$ containing the classification loss $\mathcal{L}_{ace}$ and the Kullback-Leibler (KL) divergence term $\mathcal{L}_{KL}$ is defined, where the classification loss $\mathcal{L}_{ace}$ is obtained by adjusting for the conventional cross-entropy loss, i.e., the generalized maximum likelihood loss, as follows:

$$\mathcal{L}_{ace}(\tilde{\boldsymbol{\alpha}}^v) = \sum_{k=1}^{K} \tilde{y}_k \left(\psi\left(\tilde{S}^v\right) - \psi\left(\tilde{\alpha}_k^v\right)\right), \tag{6}$$

where $\psi(\cdot)$ is the digamma function, $\tilde{\alpha}_j^v$ denote the j-th element in $\tilde{\boldsymbol{\alpha}}^v$.

Eq. (6) does not ensure that the wrong classes in each sample produce lower evidence, and we would like it to be reduced to zero. Thus the KL divergence term is expressed as:

$$\mathcal{L}_{KL}(\tilde{\boldsymbol{\alpha}}^v) = KL\left[D(\boldsymbol{p}^v|\bar{\boldsymbol{\alpha}}^v) \parallel D(\boldsymbol{p}^v|\mathbf{1})\right],$$

where $\bar{\boldsymbol{\alpha}}^v = \tilde{\mathbf{y}} + (1 - \tilde{\mathbf{y}}) \odot \tilde{\boldsymbol{\alpha}}^v$ is the Dirichlet parameter adjusted to remove non-misleading evidence. $\tilde{\mathbf{y}}$ denotes the noisy label $\tilde{y}$ in the form of a one-hot vector, and $\Gamma(\cdot)$ is the gamma function. Thus for a given view's Dirichlet parameter $\tilde{\boldsymbol{\alpha}}^v$, the view-specific loss is

$$\mathcal{L}(\tilde{\boldsymbol{\alpha}}^v) = \mathcal{L}_{ace}(\tilde{\boldsymbol{\alpha}}^v) + \lambda \mathcal{L}_{KL}(\tilde{\boldsymbol{\alpha}}^v), \tag{7}$$

where $\lambda \in [0, 1]$ is a changeable parameter, and we gradually increase its value during training to avoid premature convergence of misclassified instances to a uniform distribution.

Uncertainty-Guided Correlation Loss

Optimizing the view-specific correlation matrix, being a function of the high-dimensional input space, presents challenges without any underlying assumptions. In the context of the IDN problem, it is important to recognize that the probability of a sample being mislabeled depends not only on its category but also on its features. When the features contain noise or are difficult to discern, the likelihood of mislabeling increases significantly. As highlighted earlier, the uncertainty provided by the evidence theory has proven effective in assessing the quality of sample features. Therefore, it is natural for us to combine the uncertainty estimation with the IDN problem, leveraging its potential to enhance the overall performance.

In our work, we do not directly reduce the complexity of correlation matrix by simplifying it. Instead, we propose an assumption that "the higher the uncertainty of the model on the decision, the higher the probability that the sample label is noisy". Based on this assumption, a mild constraint is imposed on the correlation matrix to effectively reduce the degrees of freedom of its linear system. Specifically, based on the obtained Dirichlet parameters $\tilde{\boldsymbol{\alpha}}^v$ with its corresponding opinion uncertainty u^v. We impose different constraints on various parts of the correlation matrix $\boldsymbol{T}^v$, aiming to encourage it to transfer evidence for instances with higher uncertainty and uncover potential labeling-related patterns.

Diagonal elements. Since the diagonal element $\{t_{kk}^v\}_{k=1}^K$ in the $\boldsymbol{T}^v$ corresponds to the probability that the labelled category is equal to its true category. Meanwhile the confidence we obtain from subjective opinions is only relevant for the diagonal elements corresponding to their labelled category $\tilde{y}$, for $\{t_{kk}^v\}_{k=1,k\neq\tilde{y}}^K$, u^v no longer provides any direct information. Therefore, we simply make the other diagonal elements close to the confidence mean of the corresponding class of samples from the current batch. It can be expressed as:

$$\mathcal{M}_{\mathcal{D}}(\tilde{\boldsymbol{\alpha}}^v) = \sum_{k=1}^{K} \mathcal{M}_{\mathcal{D}k}(\tilde{\boldsymbol{\alpha}}^v), \tag{8}$$

$$\mathcal{M}_{\mathcal{D}k}(\tilde{\boldsymbol{\alpha}}^v) = \begin{cases} [(1-u^v) - t_{kk}^v]^2, & \text{if } k = \tilde{y}, \\ [(1-\bar{u}_k^v) - t_{kk}^v]^2, & \text{if } k \neq \tilde{y}, \end{cases} \tag{9}$$

where $u^v = K/\sum_{k=1}^K \tilde{\alpha}_k^v$, $\bar{u}_k^v$ is the average of the u^v of all samples with label $\tilde{y} = i$ in the current batch.

Non-diagonal elements. The constraints on the diagonal elements could be regarded as guiding the probability of the sample being mislabeled. The probability of being mislabeled as another category is influenced by the inherent relationship between the different categories. For example, "dog" is more likely to be labeled as "wolf" than as "plane". Considering that samples in the same class can be easily labeled as the same error class, the transfer probabilities of their non-diagonal elements should be close. In addition, since the labeled information may contain noise, we aim to eliminate the misleading of the error samples in the same class. To solve this problem, we construct the affinity matrix $\{\boldsymbol{S}^v\}_{v=1}^V$ for each viewpoint and calculate this loss with only the k most similar samples in the same class, and for the n-th sample:

$$\mathcal{M}_{\mathcal{O}}(\tilde{\boldsymbol{\alpha}}_n^v) = \sum_{m=1}^{N} s_{nm}^v \|\bar{\boldsymbol{T}}_n^v - \bar{\boldsymbol{T}}_m^v\|^2, \tag{10}$$

Algorithm 1 TMNR algorithm

/*Training*/
Input: Noisy training dataset, hyperparameter β, γ
Output: Parameters of model

Initialize the parameters of the evidence neural network.
Initialise all correlation matrices $\boldsymbol{T}$ as unit matrices.
while not converged **do**
  for $v = 1 : V$ **do**
    Obtain clean evidence $\boldsymbol{e}_n^v$ with $f^v(\boldsymbol{x}_n^v; \boldsymbol{\theta})$;
    Obtain $\tilde{\boldsymbol{e}}_n^v$ and $\tilde{\boldsymbol{\alpha}}_n^v$ through Eq. (4);
  end for
  Aggregation to obtain $\tilde{\boldsymbol{\alpha}}_n$ by Eq. (5);
  Calculate overall loss with Eq. (14);
  Update the parameters;
  Correct $\boldsymbol{T}$ to satisfy Eq. (1).
end while

/*Test*/
Calculate the clean joint $Dir(\boldsymbol{p}|\boldsymbol{\alpha})$ and the corresponding uncertainty u by $f(\cdot;\boldsymbol{\theta})$

$$s_{nm}^v = \begin{cases} e^{-\frac{\|\boldsymbol{x}_n^v - \boldsymbol{x}_m^v\|^2}{\sigma^2}}, & \text{if } \boldsymbol{x}_m^v \in \mathcal{N}(\boldsymbol{x}_n^v, k) \text{ and } \tilde{y}_n = \tilde{y}_m, \\ 0, & \text{else}, \end{cases} \tag{11}$$

where s_{nm}^v denotes the (n,m)-th element in the affinity matrix $\boldsymbol{S}^v$ for the v-th view, which measures the similarity between $\boldsymbol{x}_n^v$ and $\boldsymbol{x}_m^v$. $\mathcal{N}(\boldsymbol{x}_n^v, k)$ indicates the k-nearest neighbours of the view $\boldsymbol{x}_n^v$. $\bar{\boldsymbol{T}}_i^v$ is the matrix after zeroing the diagonal elements of $\boldsymbol{T}_i^v$. Thus, the overall regularization term for the inter-sample uncertainty bootstrap is expressed as:

$$\mathcal{M}(\tilde{\boldsymbol{\alpha}}^v) = \mathcal{M}_{\mathcal{D}}(\tilde{\boldsymbol{\alpha}}^v) + \mathcal{M}_{\mathcal{O}}(\tilde{\boldsymbol{\alpha}}^v) \tag{12}$$

Inter-view consistency. In multi-view learning, each view represents different dimensional features of the same instance. In our approach, we leverage the consistency principle of these views to ensure the overall coherence of the correlation matrix across all views. The consistency loss is denoted as:

$$\mathcal{L}_{con} = \frac{1}{V} \sum_{v=1}^{V} \left(\sum_{k=1}^K \sum_{j=1}^K |t_{kj}^v - \bar{t}_{kj}| \right), \tag{13}$$

where $\bar{t}_{kj} = (\sum_{v=1}^V t_{kj}^v)/\, V$.

Overall Loss

To sum up, for a multi-view instance $\{\boldsymbol{x}^v\}_{v=1}^V$, to ensure that view-speicfic and aggregated opinions receive supervised guidance. We use a multitasking strategy and bootstrap the correlation matrix according to the designed regularization term:

$$\mathcal{L}_{all} = \mathcal{L}(\tilde{\boldsymbol{\alpha}}) + \sum_{v=1}^{V} \left[\mathcal{L}\left(\tilde{\boldsymbol{\alpha}}^v\right) + \beta \left(\mathcal{M}(\tilde{\boldsymbol{\alpha}}^v) \right) \right] + \gamma \mathcal{L}_{con}, \tag{14}$$

where β and γ are hyperparameters that balances the adjusted cross-entropy loss with the uncertainty bootstrap regularisation and the inter-view consistency loss. $\tilde{\boldsymbol{\alpha}}$ is obtained by aggregating multiple noise class-posterior parameters $\{\tilde{\boldsymbol{\alpha}}^v\}_{v=1}^V$.

The overall procedure is summarized in Algorithm 1.

4 Experiments

In this section, we test the effectiveness of the proposed method on 5 real-world multi-view datasets with different proportions of labelling noise added. In addition, we also verify the ability of the model to handle low-quality features and noisy labels.

4.1 Experimental Setup

Datasets. **UCI**[1] contains features for handwritten numerals ('0'-'9'). The average of pixels in 240 windows, 47 Zernike moments, and 6 morphological features are used as 3 views. **PIE**[2] consists of 680 face images from 68 experimenters. We extracted 3 views from it: intensity, LBP and Gabor. **BBC**[3] includes 685 documents from BBC News that can be categorised into 5 categories and are depicted by 4 views. **Caltech101**[4] contains 8677 images from 101 categories, extracting features as different views with 6 different methods: Gabor, Wavelet Moments, CENTRIST, HOG, GIST, and LBP. we chose the first 20 categories. **Leaves100**[5] consists of 1600 leaf samples from 100 plant species. We extracted shape descriptors, fine-scale edges, and texture histograms as 3 views.

Compared methods. (1) **Sing-view uncertainty aware methods** contain MCDO [Gal and Ghahramani, 2016] measuring uncertainty by using dropout sampling in both training and inference phases. IEDL [Deng *et al.*, 2023] is the SOTA method that involving evidential deep learning and Fisher's information matrix. (2) **Label noise refining methods** contain: FC [Patrini *et al.*, 2017] corrects the loss function by a CCN transition matrix. ILFC [Berthon *et al.*, 2021] explored IDN transition matrix by training a naive model on a subset. (3) **Multi-view feature fusion methods** contain: DCCAE [Wang *et al.*, 2015] train the autoencoder to obtain a common representation between the two views. DCP [Lin *et al.*, 2022] is the SOTA method that obtain a consistent representation through dual contrastive loss and dual prediction loss. (4) **Multi-view decision fusion methods** contain: ETMC [Han *et al.*, 2022] estimates uncertainty based on EDL and dynamically fuses the views accordingly to obtain reliable results. ECML [Xu *et al.*, 2024] is the SOTA method that propose a new opinion aggregation strategy. We summarize baseline methods in Table 1. For the single-view baselines, we concatenate feature vectors of different views.

Implementation details. We implement all methods on PyTorch 1.13 framework. In our model, the view-specific evidence extracted by fully connected networks with a ReLU layer. The correlation matrices are initially set as unit matrix. We utilize the Adam optimizer with a learning rate of $1e^{-3}$ and l_2-norm regularization set to $1e^{-5}$. In all datasets, 20% of the instances are split as the test set. We run 5 times for each method to report the mean values and standard deviations. We follow [Cheng *et al.*, 2020] to generate the instance-dependent label noise training sets.

Methods	Trusted	Multi-view	Noise Refining
MCDO	✔	✘	✘
IEDL	✔	✘	✘
FC	✘	✘	✔
ILFC	✘	✘	✔
DCCAE	✘	✔	✘
DCP	✘	✔	✘
ETMC	✔	✔	✘
ECML	✔	✔	✘
TMNR	✔	✔	✔

Table 1: Summary of the methods. ✔ denotes the corresponding information is used.

4.2 Experimental Results

Performance comparison. The comparison between TMNR and baselines on clean and noisy datasets are shown in Table 2. We can observe the following points: (1) On the clean training dataset, TMNR achieves performance comparable to state-of-the-art methods. This finding indicates that the noise forward correction module has minimal negtive impact on the model's performance. (2) The performance of multi-view feature fusion methods degrade clearly with the noise ratio increase. The reason is the feature fusion would badly affected by noisy labels. (3) On the noisy training dataset, especially with high noise ratio, TMNR significantly outperforms all baseline. Such performance is a powerful evidence that our proposed method effectively reduces the effect of noisy labels through forward correction. We would further verify this in ablation study and uncertainty evaluation experiments.

Model uncertainty evaluation. In real-world datasets, various categories have varying probabilities of being labeled incorrectly. If we can identify the classes that are more likely to be labeled incorrectly during the labeling process, we can apply specialized processing to address these classes, such as involving experts in secondary labeling. As incorrect labeling leads to increased model uncertainty, our model can effectively identify classes that contain noise by assessing their predicted uncertainty.

To observe significant results, we intentionally flipped the labels of samples belonging to classes '0' and '1', as well as classes '8' and '9', within the UCI dataset during training. Subsequently, predictions were made on the test samples, and the average uncertainty for each category was calculated. The results, depicted in Figure 4(a), demonstrate a notable increase in uncertainty for the categories where the labels were corrupted. Figure 4(b) presents a heat map displaying the mean values of all trained correlation matrix parameters. The results clearly illustrate that the model's structure captures the probability of changes in inter-class evidence.

Correlation matrix evaluation. We analyze the sensitivity of hyperparameter β on all datasets containing 30% noise. The results is shown in Figure 5. It is evident that the sensitivity of the parameter β varies across different datasets,

[1] http://archive.ics.uci.edu/dataset/72/multiple+features

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

[3] http://mlg.ucd.ie/datasets/segment.html

[4] https://github.com/yeqinglee/mvdata

[5] https://archive.ics.uci.edu/dataset/241/one+hundred+plant+species+leaves+data+set

Datasets	Noise	MCDO	IEDL	FC	ILFC	Methods DCCAE	DCP	ETMC	ECML	TMNR
UCI	0%	<u>97.50±0.72</u>	**97.70±0.46**	97.05±0.19	95.50±0.76	85.75±0.32	96.15±0.49	96.20±0.73	97.05±0.48	96.90±0.65
	10%	95.50±0.61	95.25±1.25	**96.20±0.58**	95.45±0.82	85.50±0.70	95.55±0.54	95.50±0.33	95.85±0.19	<u>95.95±0.37</u>
	20%	95.40±0.82	<u>95.50±0.75</u>	95.35±1.06	95.15±0.46	85.00±0.37	95.25±0.87	95.35±0.49	95.35±0.66	**95.90±0.84**
	30%	92.30±0.80	92.50±1,74	91.90±1.21	<u>93.85±1.51</u>	84.50±0.98	92.40±1.24	93.15±1.49	92.85±1.42	**94.00±1.46**
	40%	90.15±1.44	90.30±1.99	89.90±1.58	90.95±1.87	84.50±0.78	90.30±1.08	91.65±1.54	<u>92.35±1.37</u>	**94.65±1.35**
	50%	83.65±1.98	85.85±1.17	83.55±2.93	84.85±1.23	81.75±1.08	83.75±1.64	84.10±1.83	<u>86.15±1.62</u>	**88.90±0.49**
PIE	0%	89.71±3.15	**90.85±3.31**	71.03±5.02	73.28±3.27	53.38±1.82	87.24±2.48	87.06±2.89	<u>89.83±2.66</u>	89.53± 1.89
	10%	77.50±3.37	<u>85.74±4.50</u>	55.00±3.87	66.44±4.12	47.06±0.76	82.70±3.21	83.38±1.28	83.09±3.75	**86.47±1.97**
	20%	64.71±1.40	<u>80.29±3.58</u>	44.85±4.05	61.23±3.21	47.79±1.87	75.76±1.84	77.21±1.04	76.47±1.54	**83.24±1.70**
	30%	55.44±3.28	69.44±2.73	33.97±1.94	58.98±4.10	36.03±0.67	65.38±2.49	63.97±1.86	<u>70.44±4.04</u>	**73.29±2.08**
	40%	46.32±2.03	<u>65.29±4.57</u>	32.31±5.29	51.84±5.29	33.82±1.19	58.46±2.96	61.76±2.33	63.97±4.53	**71.91±2.33**
	50%	38.68±3.50	53.00±2.88	25.44±2.57	44.02±4.23	30.88±1.38	50.85±2.82	51.32±3.40	<u>55.53±2.30</u>	**59.85±2.89**
BBC	0%	93.31±1.79	92.60±2.04	92.12±3.01	92.56±1.87	92.03±2.48	93.20±1.92	**93.58±1.42**	91.82±1.93	<u>93.51±1.35</u>
	10%	89.34±1.88	**90.38±1.81**	89.41±1.57	88.28±1.34	88.24±1.54	88.33±1.41	89.93±1.56	88.18±1.17	<u>90.07±1.53</u>
	20%	86.01±2.86	86.93±1.87	85.11±2.47	86.01±2.02	85.41±2.97	86.17±2.38	<u>86.86±2.73</u>	85.11±2.87	**87.45±2.86**
	30%	74.45±4.05	<u>81.07±3.56</u>	73.87±5.66	77.23±2.98	77.94±2.36	77.80±2.49	80.73±2.47	74.16±1.27	**82.04±2.98**
	40%	69.64±2.72	71.06±1.07	70.51±1.76	72.67±3.88	71.09±2.51	70.87±3.01	<u>72.85±3.36</u>	72.85±3.96	**75.91±3.44**
	50%	56.64±3.59	<u>59.04±4.52</u>	56.93±2.44	57.29±3.90	56.27±3.87	56.24±3.76	57.23±4.30	58.83±3.05	**63.88±4.43**
Caltech101	0%	71.38±5.06	<u>92.35±1.46</u>	64.64±5.57	85.24±1.90	88.03±0.75	91.93±1.39	**92.59±0.86**	91.13±1.61	91.84±1.08
	10%	68.66±5.02	91.05±1.26	57.28±3.46	81.92±2.39	86.19±0.98	90.74±1.22	90.34±1.32	**91.38±1.59**	<u>91.09±0.59</u>
	20%	58.41±2.43	86.82±2.22	42.34±5.38	77.74±2.80	83.47±1.28	87.02±2.31	<u>87.74±1.24</u>	87.12±0.96	**87.78±1.26**
	30%	54.60±3.58	83.01±0.90	52.97±2.54	73.26±2.11	82.85±0.93	84.98±1.01	86.07±1.10	<u>86.11±0.49</u>	**86.82±0.83**
	40%	48.12±5.99	71.92±1.95	46.57±4.77	70.17±3.98	75.94±1.89	76.90±1.55	<u>78.91±0.90</u>	77.82±0.68	**81.59±0.96**
	50%	43.89±6.41	59.04±1.84	35.98±4.78	63.28±3.41	61.09±1.46	65.29±2.52	68.79±2.02	<u>68.91±1.96</u>	**72.89±1.97**
Leaves100	0%	66.12±5.05	72.12±0.71	64.06±4.40	66.27±3.11	63.50±0.75	<u>73.40±2.18</u>	70.62±3.46	73.00±1.73	**73.75±2.55**
	10%	63.81±2.87	65.38±2.18	62.00±3.38	63.02±2.94	62.19±0.97	66.16±2.47	66.31±3.16	<u>68.31±1.45</u>	**68.88±2.17**
	20%	60.06±3.23	<u>63.38±1.67</u>	59.63±2.04	59.61±3.20	61.25±1.24	61.30±3.98	59.31±3.16	62.56±4.67	**64.19±2.03**
	30%	55.31±4.19	57.37±3.44	52.62±3.47	54.75±3.59	58.13±0.76	57.44±2.80	58.75±1.95	<u>59.94±5.61</u>	**60.94±2.04**
	40%	47.69±2.26	52.88±1.39	44.31±2.59	48.57±2.16	<u>55.13±1.87</u>	52.59±1.97	51.69±2.27	54.69±3.89	**57.75±2.12**
	50%	40.88±3.80	48.31±2.11	36.44±3.12	40.35±3.82	50.31±1.89	49.67±3.24	<u>51.81±1.76</u>	50.44±3.10	**55.63±2.91**

Table 2: Classification accuracy(%) of TMNR and baseline methods on the datasets with different proportions of Instance-Dependent Noise. The 'Noise' column shows the percentage of noisy labelled instances, where 0% denote clean datasets. The best and the second best results are highlighted by **boldface** and <u>underlined</u> respectively.

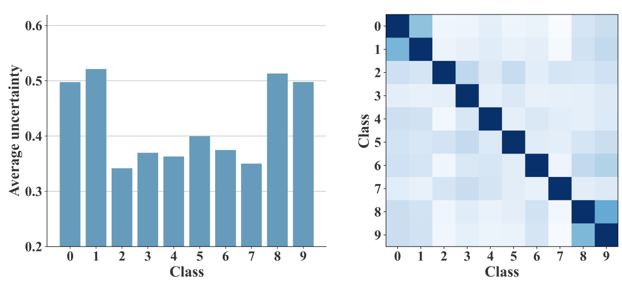

Figure 4: Visualization of the average uncertainty of each category and the correlation matrices.

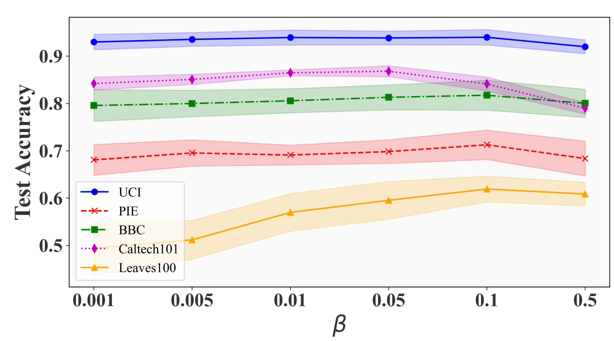

Figure 5: Correlation matrix evaluation. Classification accuracy when adjusting β on all datasets with 30% noise rate.

yet optimal performance is consistently achieved within the range of 0.05 to 0.1. This observation validates the effectiveness of the regularization applied to the diagonal elements of the correlation matrices. Taking this into consideration, we have determined the appropriate value of β for the remaining evaluation experiments.

5 Conclusion

In this paper, we introduced a TMNR method for addressing the generalized trusted multi-view learning problem. TMNR leverages evidential deep neural networks to learn view-specific belief mass vectors and uncertainty estimates. We further designed view-specific noise correlation matrices to effectively correlate the original opinions with the noisy opinions. By aggregating the noisy opinions and training the entire model using the noisy labels, we achieved robust model training. Experimental results on five real-world datasets validated the effectiveness of TMNR, demonstrating its superiority compared to state-of-the-art baseline methods.

Acknowledgments

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

References

[Berthon *et al.*, 2021] Antonin Berthon, Bo Han, Gang Niu, Tongliang Liu, and Masashi Sugiyama. Confidence scores make instance-dependent label-noise learning possible. In *Proceedings of the 38th International Conference on Machine Learning*, pages 825–836, 2021.

[Chen *et al.*, 2024] Changrui Chen, Jungong Han, and Kurt Debattista. Virtual category learning: A semi-supervised learning method for dense prediction with extremely limited labels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[Cheng *et al.*, 2020] Hao Cheng, Zhaowei Zhu, Xingyu Li, Yifei Gong, Xing Sun, and Yang Liu. Learning with instance-dependent label noise: A sample sieve approach. In *International Conference on Learning Representations*, 2020.

[Cheng *et al.*, 2022] De Cheng, Tongliang Liu, Yixiong Ning, Nannan Wang, Bo Han, Gang Niu, Xinbo Gao, and Masashi Sugiyama. Instance-dependent label-noise learning with manifold-regularized transition matrix estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16630–16639, 2022.

[Deng *et al.*, 2023] Danruo Deng, Guangyong Chen, Yang Yu, Furui Liu, and Pheng-Ann Heng. Uncertainty estimation by fisher information-based evidential deep learning. In *Proceedings of the 40th International Conference on Machine Learning*, pages 7596–7616, 2023.

[Gal and Ghahramani, 2016] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Proceedings of The 33rd International Conference on Machine Learning*, page 1050–1059, 2016.

[Gan *et al.*, 2021] Jiangzhang Gan, Ziwen Peng, Xiaofeng Zhu, Rongyao Hu, Junbo Ma, and Guorong Wu. Brain functional connectivity analysis based on multi-graph fusion. *Medical image analysis*, 71:102057, 2021.

[Han *et al.*, 2020] Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. Trusted multi-view classification. In *International Conference on Learning Representations*, 2020.

[Han *et al.*, 2022] 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.

[Hendrycks *et al.*, 2018] Dan Hendrycks, Mantas Mazeika, Duncan Wilson, and Kevin Gimpel. Using trusted data to train deep networks on labels corrupted by severe noise. *Advances in neural information processing systems*, 31, 2018.

[Hu *et al.*, 2024] Shizhe Hu, Chengkun Zhang, Guoliang Zou, Zhengzheng Lou, and Yangdong Ye. Deep multiview clustering by pseudo-label guided contrastive learning and dual correlation learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[Huang *et al.*, 2023] Shudong Huang, Yixi Liu, Ivor W Tsang, Zenglin Xu, and Jiancheng Lv. Multi-view subspace clustering by joint measuring of consistency and diversity. *IEEE Transactions on Knowledge and Data Engineering*, 35(8):8270–8281, 2023.

[Jillani *et al.*, 2020] Rashad Jillani, Syed Fawad Hussain, and Hari Kalva. Multi-view clustering for fast intra mode decision in hevc. In *2020 IEEE International Conference on Consumer Electronics (ICCE)*, 2020.

[Jøsang, 2016] Audun Jøsang. *Subjective logic: A formalism for reasoning under uncertainty*. Springer, 2016.

[Liang *et al.*, 2021] Xinyan Liang, Qian Guo, Yuhua Qian, Weiping Ding, and Qingfu Zhang. Evolutionary deep fusion method and its application in chemical structure recognition. *IEEE Transactions on Evolutionary Computation*, 25(5):883–893, 2021.

[Lin *et al.*, 2022] 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.

[Lin *et al.*, 2023] Zhenghong Lin, Yanchao Tan, Yunfei Zhan, Weiming Liu, Fan Wang, Chaochao Chen, Shiping Wang, and Carl Yang. Contrastive intra-and inter-modality generation for enhancing incomplete multimedia recommendation. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 6234–6242, 2023.

[Liu *et al.*, 2022] Wei Liu, Xiaodong Yue, Yufei Chen, and Thierry Denoeux. Trusted multi-view deep learning with opinion aggregation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 7585–7593, 2022.

[Liu *et al.*, 2023] Chengliang Liu, Jie Wen, Zhihao Wu, Xiaoling Luo, Chao Huang, and Yong Xu. Information recovery-driven deep incomplete multiview clustering network. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[Liu *et al.*, 2024] Haoran Liu, Ying Ma, Ming Yan, Yingke Chen, Dezhong Peng, and Xu Wang. Dida: Disambiguated domain alignment for cross-domain retrieval with partial labels. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 3612–3620, 2024.

[Min *et al.*, 2023] Bonan Min, Hayley Ross, Elior Sulem, Amir Pouran Ben Veyseh, Thien Huu Nguyen, Oscar

Sainz, Eneko Agirre, Ilana Heintz, and Dan Roth. Recent advances in natural language processing via large pretrained language models: A survey. *ACM Computing Surveys*, 56(2):1–40, 2023.

[Nikzad-Khasmakhi *et al.*, 2021] Narjes Nikzad-Khasmakhi, Mohammad Ali Balafar, M Reza Feizi-Derakhshi, and Cina Motamed. Berters: Multimodal representation learning for expert recommendation system with transformers and graph embeddings. *Chaos, Solitons & Fractals*, 151:111260, 2021.

[Patrini *et al.*, 2017] Giorgio Patrini, Alessandro Rozza, Aditya Krishna Menon, Richard Nock, and Lizhen Qu. Making deep neural networks robust to label noise: A loss correction approach. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2233–2241, 2017.

[Qin *et al.*, 2022] Yang Qin, Dezhong Peng, Xi Peng, Xu Wang, and Peng Hu. Deep evidential learning with noisy correspondence for cross-modal retrieval. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 4948–4956, 2022.

[Sensoy *et al.*, 2018] Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. *Advances in neural information processing systems*, 31, 2018.

[Wang *et al.*, 2015] Weiran Wang, Raman Arora, Karen Livescu, and Jeff Bilmes. On deep multi-view representation learning. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 1083–1092, 2015.

[Wang *et al.*, 2022] Xu Wang, Peng Hu, Pei Liu, and Dezhong Peng. Deep semisupervised class- and correlation-collapsed cross-view learning. *IEEE Transactions on Cybernetics*, 52(3):1588–1601, 2022.

[Wen *et al.*, 2022] Jie Wen, Zheng Zhang, Lunke Fei, Bob Zhang, Yong Xu, Zhao Zhang, and Jinxing Li. A survey on incomplete multiview clustering. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(2):1136–1149, 2022.

[Xie *et al.*, 2023] Mengyao Xie, Zongbo Han, Changqing Zhang, Yichen Bai, and Qinghua Hu. Exploring and exploiting uncertainty for incomplete multi-view classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19873–19882, 2023.

[Xu *et al.*, 2022] Lei Xu, Hui Wu, Chunming He, Jun Wang, Changqing Zhang, Feiping Nie, and Lei Chen. Multimodal sequence learning for alzheimer's disease progression prediction with incomplete variable-length longitudinal data. *Medical Image Analysis*, 82:102643, 2022.

[Xu *et al.*, 2023] Jie Xu, Yazhou Ren, Xiaoshuang Shi, Heng Tao Shen, and Xiaofeng Zhu. Untie: Clustering analysis with disentanglement in multi-view information fusion. *Information Fusion*, 100:101937, 2023.

[Xu *et al.*, 2024] Cai Xu, Jiajun Si, Ziyu Guan, Wei Zhao, Yue Wu, and Xiyue Gao. Reliable conflictive multi-view learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 16129–16137, 2024.

[Zhou *et al.*, 2023] Hai Zhou, Zhe Xue, Ying Liu, Boang Li, Junping Du, Meiyu Liang, and Yuankai Qi. Calm: An enhanced encoding and confidence evaluating framework for trustworthy multi-view learning. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 3108–3116, 2023.