Trusted Multi-view Learning for Long-tailed Classification

Chuanqing Tang[1,2], Yifei Shi[1], Guanghao Lin[1], Lei Xing[3], Long Shi[1,2*]

[1]School of Computing and Artificial Intelligence, Southwestern University of Finance and Economics, Chengdu, China
[2]Artificial Intelligence and Digital Finance Key Laboratory of Sichuan Province, Chengdu, China
[3]Xi'an Jiaotong University, Xi'an, China
223081200012@smail.swufe.edu.cn, 42211176@smail.swufe.edu.cn, 42211096@smail.swufe.edu.cn,
xingl@xjtu.edu.cn, shilong@swufe.edu.cn

Abstract

Class imbalance has been extensively studied in single-view scenarios; however, addressing this challenge in multi-view contexts remains an open problem, with even scarcer research focusing on trustworthy solutions. In this paper, we tackle a particularly challenging class imbalance problem in multi-view scenarios: long-tailed classification. We propose TMLC, a Trusted Multi-view Long-tailed Classification framework, which makes contributions on two critical aspects: opinion aggregation and pseudo-data generation. Specifically, inspired by Social Identity Theory, we design a group consensus opinion aggregation mechanism that guides decision-making toward the direction favored by the majority of the group. In terms of pseudo-data generation, we introduce a novel distance metric to adapt SMOTE for multi-view scenarios and develop an uncertainty-guided data generation module that produces high-quality pseudo-data, effectively mitigating the adverse effects of class imbalance. Extensive experiments on long-tailed multi-view datasets demonstrate that our model is capable of achieving superior performance.

Introduction

Class imbalance refers to a situation in which the distribution of classes in a dataset is non-uniform, characterized by significant or even extreme disparities (Kang et al. 2020). This phenomenon, widely observed in real-world scenarios such as fraud detection (Zhang et al. 2021) and medical image analysis (Ju et al. 2021), often arises due to limitations or biases in the data collection process. In the machine learning community, class imbalance is recognized as a critical research challenge, as models trained on imbalanced data tend to favor the majority class while neglecting the minority class (Cao et al. 2019; Tan et al. 2020). This can result in suboptimal performance, particularly for those predictions where accurate classification of the minority class is essential.

Research on class imbalance mainly follows three technical routes: re-sampling (Liu, Wu, and Zhou 2008), cost-sensitive learning (Hsieh et al. 2021), and data augmentation (Liu et al. 2020). Re-sampling methods aim to balance class distributions by either over-sampling minority classes or under-sampling majority classes, with two well-known techniques being SMOTE (Chawla et al. 2002) and Tomek Links (430 1976). In contrast to re-sampling, which focuses on data generation, cost-sensitive learning addresses class imbalance by assigning higher misclassification costs to minority classes through weighted loss functions (Cui et al. 2019; Ren et al. 2020) or specialized algorithms (Cao et al. 2019; Khan et al. 2019). This approach offers flexibility but often requires careful tuning of cost parameters. Data augmentation, on the other hand, highlights the diversity of minority class samples using techniques such as generative models (Ji and Liu 2024) or feature-space transformations (Wang, Ramanan, and Hebert 2017). It is worth noting that, although both re-sampling and data augmentation methods essentially fall under pseudo-data generation approaches, they differ significantly in their mechanisms: re-sampling focuses on adjusting data distribution, whereas data augmentation emphasizes enriching data diversity.

The aforementioned studies on class imbalance have been conducted in single-view scenarios, where decision-making relies on a single modality of data. However, real-world applications often involve multi-view scenarios (Shi et al. 2024a,b), in which multiple data sources are available for analysis. For instance, in medical diagnosis (Ren et al. 2024), multi-view data may include X-ray images, MRI scans, and other modalities, each offering unique and complementary insights into patient conditions. To the best of our knowledge, addressing class imbalance in such multi-view settings remains an open and under-explored challenge, which demands innovative solutions and deeper investigation.

In this paper, we address a particularly challenging class imbalance problem—long-tailed distribution class imbalance—in multi-view scenarios. *We aim to answer two critical questions: (1) How to tackle the challenge of multi-view long-tailed classification, and (2) how to ensure reliable decision-making?* To this end, we propose TMLC, a Trusted Multi-View Long-Tailed Classification Framework , as shown in Fig. 1. TMLC is designed to achieve trustworthy classification for multi-view long-tailed data by leveraging a deep evidence learning-inspired paradigm (Sensoy, Kaplan, and Kandemir 2018) and adopting a oversampling-based scheme. Specifically, we make significant advancements in opinion aggregation and pseudo-data generation,

*Corresponding author.

which respectively enhance the reliability of the trustworthy framework and improve the effectiveness of the oversampling process. The main contributions of this work are as follows:

- Motivated from Social Identity Theory (Hogg 2016), we design a group consensus opinion aggregation mechanism that guides decision-making toward the direction favored by the majority of the group, thereby preventing negative individual opinions from dominating the final outcome.
- To address the limitation of SMOTE, which cannot be directly applied to multi-view settings, we introduce a novel distance metric based on joint subjective evidence and propose an uncertainty-based sample generation scheme that enables the production of high-quality pseudo-data.
- Extensive experiments on long-tailed multi-view datasets demonstrated that TMLC outperforms serveral state-of-art baselines.

Related Work

Long-tailed Learning. Traditional long-tailed learning methods include re-sampling (Han, Wang, and Mao 2005; Liu, Wu, and Zhou 2008), cost-sensitive learning (Cui et al. 2019; Cao et al. 2019; Wang et al. 2021) and data augmentation (Chu et al. 2020; Kim, Jeong, and Shin 2020; Yin et al. 2019). As one of the most widely used techniques, Re-sampling rebalances classes by adjusting the number of samples per class. SMOTE (Chawla et al. 2002) enhances the representation of minority classes by generating synthetic samples. Decoupling (Kang et al. 2019) identifies effective sampling strategies for standard model training. DCL (Wang et al. 2019) and LOCE (Feng, Zhong, and Huang 2021) propose adaptive sampling strategies. Balanced Meta-Softmax (Ren et al. 2020) employs meta-learning to estimate optimal sampling rates for different classes. SimCal (Wang et al. 2020) uses a novel two-level class-balanced sampling strategy that combines image-level and instance-level. However, these methods do not fully explore the uncertainty of samples during the sampling phase.

Uncertainty-induced Trusted Learning. Inspired by evidential deep learning (Sensoy, Kaplan, and Kandemir 2018), which serves as a prominent paradigm for uncertainty estimation, Trusted Multi-view Classification (TMC) has recently gained particular attention. TMC leverages the Dirichlet distribution to model class probabilities and integrates Dempster-Shafer theory to achieve reliable classification results (Han et al. 2022). Based on TMC, numerous extensions have been proposed to tackle specific challenges, involving opinion aggregation (Liu et al. 2022; Xu et al. 2024a; Zheng et al. 2023), noisy inputs and labels (Xu et al. 2024b; Shi et al. 2024c), semantic ambiguity (Liu et al. 2024), etc. Recent advancements in trusted learning have expanded its research to areas such as weakly supervised learning (Wang et al. 2024; Hu et al. 2025), robust learning (Zhou et al. 2023a,b; Du et al. 2023; Deng et al. 2025), and multimodal applications (Huang et al. 2025; Li et al. 2025).

Methodology

In this section, we first provide the problem definition and theoretical foundation on evidential deep learning. Subsequently, we detail our key contributions, including the group consensus opinion aggregation mechanism and the uncertainty-based sample generation scheme.

Problem Definition and Preliminaries

Problem Definition A long-tailed multi-view dataset typically contains an imbalanced training set and a balanced test set. For a K classification problem, we define the multi-view data as $\{\boldsymbol{X}^v\}_{v=1}^V$ and the corresponding ground-truth labels as $\{\mathbf{y}_n\}_{n=1}^N$ with V views and N samples, where $\boldsymbol{X}^v = [\boldsymbol{x}_1^v, \boldsymbol{x}_2^v, \cdots, \boldsymbol{x}_N^v] \in \mathbb{R}^{d_v \times N}$ represents the v-th view data with d_v dimension. For each view data $\boldsymbol{X}^v$, the distribution of classes is highly imbalanced, with tail classes having significantly fewer samples compared to head classes. The goal of TMLC is to address the challenges caused by imbalanced class distributions.

Evidential Deep Learning Evidential Deep Learning (EDL) (Sensoy, Kaplan, and Kandemir 2018) is a probabilistic framework to model prediction uncertainty by employing evidential priors and subjective logic (Jøsang 2016). In the EDL framework, the evidences $\boldsymbol{e}^v = [e_1^v, \cdots, e_K^v]$ for the v-th are obtained using Deep Neural Networks (DNNs), where the Softmax layer is replaced by a ReLU activation function.

Subjective logic provides a technical approach to associate the evidence $\boldsymbol{e}^v = [e_1^v, \cdots, e_K^v]$ with the parameters of the Dirichlet distribution $\boldsymbol{\alpha}^v = [\alpha_1^v, \cdots, \alpha_K^v]$. Specifically, each parameter α_k^v of the Dirichlet distribution is derived from e_k^v as $\alpha_k^v = e_k^v + 1$. The belief mass b_k^v and the uncertainty u^v are then calculated as follows:

$$b_k^v = \frac{e_k^v}{S^v} = \frac{\alpha_k^v - 1}{S^v} \quad \text{and} \quad u^v = \frac{K}{S^v}, \tag{1}$$

where $S^v = \sum_{k=1}^K (e_k^v + 1) = \sum_{k=1}^K \alpha_k^v$ denotes the Dirichlet strength. The projection probability is calculated by

$$P_k^v = b_k^v + a_k^v u^v. \tag{2}$$

From Eq. (1), we can obtain an ordered triplet $\omega = (\boldsymbol{b}, u, \boldsymbol{a})$ that reflects the multinomial opinion of an instance $\boldsymbol{x}_n^v$. Since the Dirichlet distribution parameter $\boldsymbol{a}$ is derived from the belief mass $\boldsymbol{b}$, a simplified representation of the multinomial opinion can be expressed as $\mathcal{M} = \{\boldsymbol{b}, u\}$. Once multinomial opinions from multiple views are derived, designing an efficient opinion aggregation strategy becomes crucial for achieving trusted decision-making.

Group Consensus Opinion Aggregation

Social Identity Theory posits that individuals tend to value shared group opinions over personal viewpoints, which helps avoid conflict and fosters stable social environments (Hogg 2016). Inspired by this theory, we argue that in a trusted decision-making process, partial aggregated group opinions should be prioritized over individual opinions. To this end, we propose a novel opinion aggregation strategy,

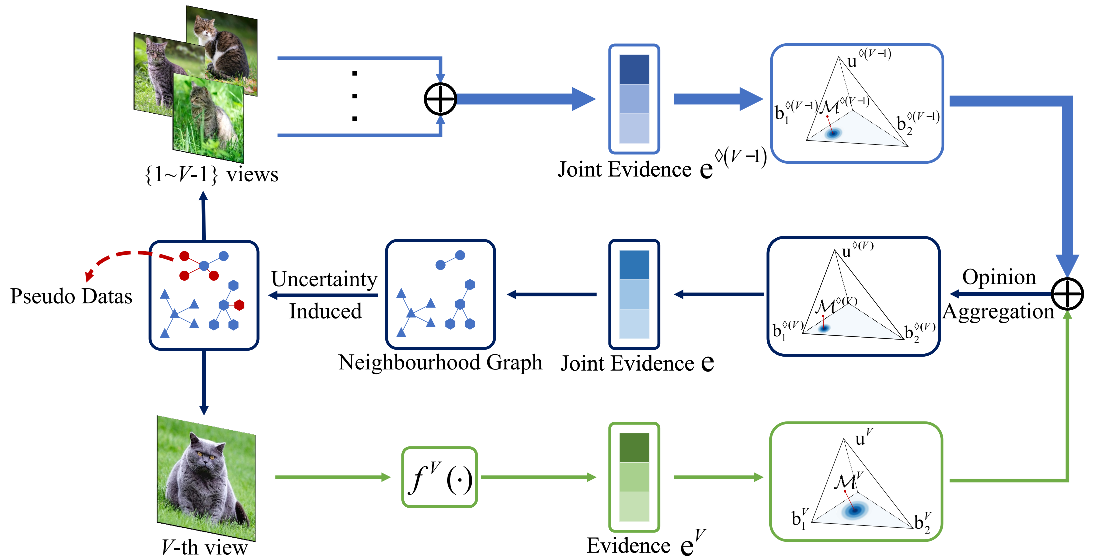

Figure 1: Illustration of TMLC. Assuming that the opinions from the first $V-1$ views have been aggregated, we now focus on combining the resulting joint opinion $\boldsymbol{e}^{\lozenge(V-1)}$ with the opinion $\boldsymbol{e}^V$ from the V-th view. In this process, a larger weight is assigned to $\boldsymbol{e}^{\lozenge(V-1)}$ which respresents the group joint opinion (bold arrow line), while a smaller weight is assigned to the individual opinion $\boldsymbol{e}^V$ (thin arrow line). Subsequently, we propose an uncertainty-based sample generation scheme to ensure the production of reliable pseudo-data for the minority class.

termed Group Consensus Opinion Aggregation, which is detailed below.

Definition 1. ***Group Consensus Opinion Aggregation.*** Without loss of generality, each view is assumed to be equally important. For V opinions $\{\mathcal{M}^1, \mathcal{M}^2, \cdots, \mathcal{M}^V\}$, we aggregate them via

$$\mathcal{M}^{\lozenge(V)} = \underbrace{\underbrace{\underbrace{\mathcal{M}^1 \oplus \mathcal{M}^2}_{\mathcal{M}^{\lozenge(2)}} \oplus \mathcal{M}^3}_{\mathcal{M}^{\lozenge(3)}} \oplus \cdots \oplus \mathcal{M}^{V-1}}_{\mathcal{M}^{\lozenge(V-1)}} \oplus \mathcal{M}^V, \tag{3}$$

where $\mathcal{M}^{\lozenge(V)} = \{\boldsymbol{b}^{\lozenge(V)}, u^{\lozenge(V)}\}$. Regarding the calculation of $\boldsymbol{b}^{\lozenge(V)}$ and $u^{\lozenge(V)}$, we present the aggregation process corresponding to their k-th components:

$$b_k^{\lozenge(V)} = \frac{e_k^{\lozenge(V)}}{S^{\lozenge(V)}} = \frac{\gamma^{\lozenge(V-1)} e_k^{\lozenge(V-1)} + \gamma^V e_k^V}{\gamma^{\lozenge(V-1)} S^{\lozenge(V-1)} + \gamma^V S^V}, \tag{4}$$

and

$$\begin{aligned} u^{\lozenge(V)} &= \frac{K}{\sum_{k=1}^{K}(e_k^{\lozenge(V)} + 1)} \\ &= \frac{K}{\sum_{k=1}^{K}(\gamma^{\lozenge(V-1)} e_k^{\lozenge(V-1)} + \gamma^V e_k^V + 1)}, \end{aligned} \tag{5}$$

with $\gamma^{\lozenge(V-1)}$ and γ^V being

$$\gamma^{\lozenge(V-1)} = \frac{V-1}{V}, \quad \gamma^V = \frac{1}{V}. \tag{6}$$

From Eqs. (4) to (6), we see that the aggregated evidence $e_k^{\lozenge(V-1)}$ for the first $V-1$ views is assigned with a large weight $\gamma^{\lozenge(V-1)}$, whereas the evidence e_k^V to be aggregated is assigned with a small weight γ^V. This indicates that, in the opinion aggregation process, the aggregated group opinion will play a dominant role. The further calculation results for Eq. (4) and (5) are given by

$$b_k^{\lozenge(V)} = \frac{\gamma^{\lozenge(V-1)} b_k^{\lozenge(V-1)} u^V + \gamma^V b_k^V u^{\lozenge(V-1)}}{\gamma^V u^{\lozenge(V-1)} + \gamma^{\lozenge(V-1)} u^V} \tag{7}$$

and

$$u^{\lozenge(V)} = \frac{u^{\lozenge(V-1)} u^V}{\gamma^V u^{\lozenge(V-1)} + \gamma^{\lozenge(V-1)} u^V} \tag{8}$$

The detailed derivation for the above equations and its effectiveness can be found in the supplementary material.

Our proposed aggregation strategy ensures that the group joint opinion is given higher priority than the individual subjective opinion. When dealing with conflicting opinions, our scheme tends to favor the group consensus opinion, thereby enhancing the realibity of decision-making.

Uncertainty-based Sample Generation

A common strategy for addressing the challenge of class imbalance is to generate new samples for the minority class to construct a balanced dataset. Among various techniques, the SMOTE method (Chawla et al. 2002) is a well-established and straightforward approach that yields new samples by linearly interpolating between minority class samples and their nearest neighbors. While traditional SMOTE is effective in single-view scenarios, it is not directly applicable to multi-view settings. The primary challenge lies in the fact that SMOTE relies on neighbor analysis based on Euclidean distance calculations, which may yield inconsistent results across different views due to the inherent diversity among them. Specifically, the neighbor relationships identified in each view can differ significantly, potentially leading to conflicts in the synthetic samples generated from different views.

To overcome the aforementioned drawback, we attempt to introduce a distance metric to adapt SMOTE for multi-view scenarios. We build upon a natural premise: samples belonging to the same class exhibit consistency in their subjective opinions. Consequently, it is reasonable to infer that samples from the same category should also show similar evidence distributions. Based on this assumption, we develop a novel distance metric based on joint evidences.

Definition 2. ***Distance measures based on joint subjective evidence.*** Given any two samples $\{\boldsymbol{x}_i^v\}_{v=1}^V$ and $\{\boldsymbol{x}_j^v\}_{v=1}^V$, along with their corresponding joint opinions $\boldsymbol{e}_i$, $\boldsymbol{e}_j$, the distance between these two samples is defined as:

$$D(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sqrt{(e_{i1} - e_{j1})^2 + \cdots + (e_{iK} - e_{jK})^2}. \quad (9)$$

In Eq. (9), a smaller distance indicates a higher similarity between samples.

Apart from the limitation that the traditional SMOTE method cannot be applied to multi-view settings, another problem is that during the pseudo-data generation process, the weights for combining the center sample with its neighors are randomly selected, which may result in undesirable pseudo samples. To address this, we design an uncertainty-based sample generation scheme that ensures neighbors with low-uncertainty opinion are assigned with higher weights, thereby facilitating the generation of high-quality pseudo-samples. In the following, we present the details of our designed scheme.

For a minority class, we randomly select a data sample $\{\boldsymbol{x}_\mathrm{c}^v\}_{v=1}^V$ as the center. The R samples with the shortest distances to the center sample, calculated using Eq. (9), are identified as its neighbors and denoted as $\{\boldsymbol{x}_1^v, \boldsymbol{x}_2^v, \ldots, \boldsymbol{x}_R^v\}_{v=1}^V$. For the v-th view, the new pseudo sample is generated by

$$\boldsymbol{x}_{new}^v = w_0^v \boldsymbol{x}_\mathrm{c}^v + w_1^v \boldsymbol{x}_1^v + \cdots + w_R^v \boldsymbol{x}_R^v, \quad (10)$$

where $\sum_{r=0}^R w_r^v = 1$ and $w_r^v \geq 0$. We expect that the weight w_r^v to be assigned a high value if the aggregated opinion between the corresponding neighbor $\boldsymbol{x}_r^v$ and the center sample $\boldsymbol{x}_c^v$ exhibits low uncertainty, indicating that the relationship between this neighbor and the center sample is highly reliable. To achieve this objective, we begin with integrating the evidence of $\boldsymbol{x}_c^v$ with that of $\boldsymbol{x}_r^v$

$$\boldsymbol{e}_{cr}^v = \frac{1}{2}\boldsymbol{e}_\mathrm{c}^v + \frac{1}{2}\boldsymbol{e}_r^v, \ \ r = 0, 1, \cdots, R \quad (11)$$

With the integrated evidence in Eq. (11) and recalling (1), we can obtain the corresponding opinions and projection probabilities for the neighbors, denoted as $\{\boldsymbol{b}_{cr}^v, u_{cr}^v\}_{r=0}^R$ and $\{\boldsymbol{P}_{cr}^v\}_{r=0}^R$, respectively.

Here, we treat the projection probalibity as the prediction probability. This leads to a straightforward approach for aligning the distribution of the generated pseudo-samples with the true distribution, typically measured by cross-entropy. However, directly using cross entropy can not reflect the reliability of the generated samples. Thus, we introduce the influence of uncertainty and propose a novel criterion, termed Uncertainty Entropy.

Definition 3. ***Uncertainty Entropy.*** Given an opinion $\mathcal{M} = \{\boldsymbol{b}, u\}$ and the true label $\boldsymbol{y}$, the uncertainty entropy is defined as:

$$H = -\exp(u) \sum_{k=1}^{K} y_k \log(b_k + u a_k), \quad (12)$$

where $\exp(u)$ is the penalty coefficient. If the uncertainty $u = 0$, then $\exp(u) = 1$, and the uncertainty has no effect on the entropy. Conversely, as the uncertainty u increases, the entropy value also increases, indicating a greater discrepancy between the predicted classification and the true class.

With uncertainty entropy, we further present the following formulation for calculating w_r^v

$$w_r^v = \frac{\mathcal{F}(H_{cr}^v)}{\mathcal{F}(H_c^v) + \sum_{r=1}^R \mathcal{F}(H_{cr}^v)}, \ r = 0, 1, \cdots, R \quad (13)$$

where $\mathcal{F}(\cdot)$ is a monotonically decreasing function. If the uncertainty is low, from the above discussion, we know that the uncertainty entropy is small. Due to the monotonically decreasing property of $\mathcal{F}(\cdot)$, $\mathcal{F}(H_{cr}^v)$ will be assigned a large value. This implies that our scheme is capable of ensuring that neighbors with low-uncertainty opinions are assigned higher weights. In this paper, we use $\mathcal{F}(x) = x^{-1}$.

Loss Function

In this subsection, we introduce the loss function for taining our model. We first employ an adjusted cross-entropy loss function to ensure that all views generate appropriate non-negative view-specific evidence for classification:

$$\begin{aligned} L_{ace}(\boldsymbol{\alpha}_n) &= \int \left[\sum_{k=1}^{K} -y_{nk} \log p_{nk}\right] \frac{\Pi_{k=1}^K p_{nk}^{\alpha_{nk}-1}}{B(\alpha_n)} d_{P_n} \\ &= \sum_{k=1}^{K} y_{nk}(\psi(S_n) - \psi(\alpha_{nk}), \end{aligned} \quad (14)$$

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

Additionally, to ensure that incorrect classes in each sample produce lower evidence, we incorporate an extra KL divergence term into the loss function

$$\begin{aligned} L_{KL}(\boldsymbol{\alpha}_n) =& KL\left[D(p_n|\tilde{\alpha}_n)||D(p_n|1)\right] \\ &= \log(\frac{\Gamma(\sum_{k=1}^K \tilde{\alpha}_{nk})}{\Gamma(K)\Pi_{k=1}^K \Gamma(\tilde{\alpha}_{nk})}) \\ &+ \sum_{k=1}^{K} (\tilde{\alpha}_{nk} - 1) \left[\psi(\tilde{\alpha}_{nk}) - \psi(\sum_{j=1}^{K} \tilde{\alpha}_{nj})\right], \end{aligned} \quad (15)$$

where $D(p_n|1)$ is the uniform Dirichlet distribution, $\tilde{\boldsymbol{\alpha}}_n = \boldsymbol{y}_n + (1 - \boldsymbol{y}_n) \odot \boldsymbol{\alpha}_n$ is the Dirichlet parameters that reduces the incorrect classes to zero for the n-th instance, and $\Gamma(\cdot)$ is the gamma function. Therefore, given the Dirichlet distribution with parameter $\boldsymbol{\alpha}_n$ for the n-th instance, the loss is:

$$L_{acc}(\boldsymbol{\alpha}_n) = L_{ace}(\boldsymbol{\alpha}_n) + \lambda_t L_K L(\boldsymbol{\alpha}_n), \tag{16}$$

where $\lambda_t \in [0, 1]$ is a balance factor with t being the current training epoch index. The implementation procedures are summarized in Algorithm 1.

Algorithm 1: Implementation pseudocode of TMLC

Input: Multi-view dataset: $\{\boldsymbol{X}^v, \boldsymbol{y}\}_{v=1}^V$
/—Train—/
Output: Network's parameters.
01: **while** not converged **do**
02: **for** $v = 1 : V$ **do**
03: $\mathbf{e}^v \leftarrow$ evidential network output.
04: Calculate opinion $\mathcal{M}^v$ by Eq. (1).
05: **end for**
06: Calculate joint opinion $\mathcal{M}^{\Diamond(V)}$ by Eq. (3).
07: Calculate joint evidence $\boldsymbol{e}$ by Eq. (1).
08: Calculate the joint distribution $\boldsymbol{\alpha} = \boldsymbol{e} + 1$.
09: Calculate the overall loss $\mathcal{L}$ by Eq. (16).
10: Update the networks by gradient descent by $\mathcal{L}$.
11: **end while**
12: Initialize pseudo-data set $\{\}$.
13: **while** minority class k-th not balanced **do**
14: Randomly select a center sample $\{\boldsymbol{x}_c^v\}_{v=1}^V$.
15: Calculate multi-view neighbors by Eq. (9).
16: **for** $v = 1 : V$ **do**
17: Generate pseudo date $\boldsymbol{x}_{new}^v$ by Eq. (10).
18: **end while**
18: Add $\left\{\{\boldsymbol{x}_{new}^v\}_{v=1}^V, y_k = 1\right\}$ to pseudo-data set.
19: **end for**
20: Add pseudo-data set to the training set.
21: Repeat training steps 01-11.
22: **return** networks parameters.
/—Test—/
Output: The final classification prediction and uncertainty.
01: **for** $v = 1 : V$ **do**
02: $\mathbf{e}^v \leftarrow$ evidential network output.
03: Calculate opinion $\mathcal{M}^v$ by Eq. (1).
04: **end for**
05: Calculate joint opinion $\mathcal{M}^{\Diamond(V)}$ by Eq. (3).
06: **return** the decision $\boldsymbol{b}^{\Diamond(V)}$ and uncertainty $u^{\Diamond(V)}$.

Experiment

Experimental Setup

Datasets. HandWritten[1] comprises 2000 instances of handwritten numerals ranging form '0' to '9', with 200 patterns per class. It is represented using six feature sets. **PIE**[2] contains 680 instances belonging to 68 classes. We extract intensity, LBP, and Gabor as 3 views. **Caltech101**[3] com-

[1]https://archive.ics.uci.edu/ml/datasets/Multiple+Features
[2]http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/MultiPie Home.html
[3]http://www.vision.caltech.edu/Image Datasets/Caltech101

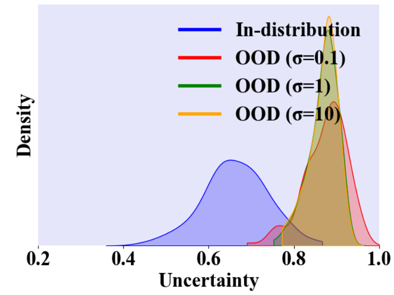

Figure 2: Density of overall uncertainty on the PIE dataset.

prises 8677 images from 101 classes. We select 2386 samples from 20 classes, where each image has six features: Gabor, Wavelet Moments, CENTRIST, HOG, GIST, and LBP. **NUS-WIDE**[4] consists of 269,648 images with 81 concepts. For the top 12 classes, we select 200 images from each class, with a total of 6 different views. **Scene15**[5] includes 4485 images from 15 indoor and outdoor scene categories. We extract three types of features GIST, PHOG, and LBP. **Animal**[6] contains 37,322 images of 50 animal species with six features for each image. We select 11,673 samples contained 20 animal classes and four features for 4 views.

Compared Methods. The compared methods involve long-tail methods, trusted multi-view learning and other multi-view baselines. Specifically, 1) long-tail methods include: **TLC** (Trustworthy Long-tailed Classification) (Li et al. 2022) combines classification and uncertainty estimation in a multi-expert framework with dynamic expert engagement. **H2T** (Head-To-Tail) (Li et al. 2024) augments tail classes by incorporating the diverse semantic information from head classes. 2) Trusted multi-view decision baselines include: **TMC** (Trusted Multi-view Classification) (Han et al. 2022) integrates different views at an evidence level using variational Dirichlet distribution and Dempster-Shafer theory. **ECML** (Evidential Conflictive Multi-view Learning) (Xu et al. 2024a) puts forward a conflictive opinion aggregation model. 3) Multi-view classification baselines include: **DUA-Nets** (Dynamic Uncertainty-Aware Networks) (Geng et al. 2021) is an uncertainty aware method that uses reversal networks to integrate intrinsic information from different views into a unified representation. **DCP** (Dual Contrastive Prediction) (Lin et al. 2022) enables consistency learning and missing view recovery by using an information-theoretical framework. **UIMC** (Uncertainty for Incomplete Multi-View Classification) (Xie et al. 2023) models the uncertainty of missing views through distribution construction and sampling and employs an evidence-based fusion strategy for trustworthy integration of imputed views. **DMVLS** (Dynamic Multi-view Labeling Strategy) (Wan et al. 2024) leverages shared and specific information with two classifiers to prioritize labeling low-confidence samples.

Implementation. See the arXiv paper for details.

[4]https://paperswithcode.com/dataset/nus-wide
[5]https://doi.org/10.6084/m9.figshare.7007177.v1
[6]http://attributes.kyb.tuebingen.mpg.de/

Methods	Datasets					
	HandWritten	PIE	Caltech101	NUS-WIDE	Scene15	Animal
TLC (2022)	84.20±1.05	36.32±0.67	79.04±1.22	19.00±0.61	42.54±1.30	16.54±0.10
H2T (2024)	90.13±0.12	31.54±0.61	83.12±0.22	17.97±0.33	38.59±0.25	15.84±0.10
TMC (2022)	92.65±0.28	46.18±0.92	84.91±0.21	22.44±0.39	32.22±0.50	24.23±0.12
ECML (2024)	91.48±0.34	47.28±0.74	82.87±0.56	22.35±0.27	41.31±0.71	24.31±0.08
DUA-Nets (2021)	91.53±1.86	8.96±1.08	77.60±2.00	20.45±1.72	36.72±1.10	14.89±0.65
DCP (2022)	93.48±1.33	4.56±0.64	86.54±1.15	29.04±2.21	53.56±1.22	18.77±0.63
UIMC (2023)	91.35±0.46	45.88±1.51	79.58±0.74	20.98±0.80	48.02±0.38	20.90±0.41
DMVLS (2024)	81.28±1.05	11.62±1.18	72.03±1.29	21.48±1.19	30.19±0.88	18.63±0.81
Ours-v1	93.40±0.28	48.38±1.03	83.76±0.48	26.31±0.19	45.20±0.41	24.34±0.06
Ours-v2	96.15±0.12	49.26±0.66	88.85±0.81	38.22±0.44	57.69±0.44	30.83±0.15
Ours	**96.25±0.32**	**62.43±0.70**	**89.48±0.24**	**38.96±0.37**	**58.57±0.37**	**31.35±0.20**

Table 1: Accuracy (%) on class-balanced test sets. The best results are highlighted in bold, while the second-best among baselines are underlined.

Datasets	Test sets	Methods	
		ECML	Ours
HandWritten	Normal	97.00	**97.50**
	Conflictive	92.69	**94.80**
PIE	Normal	94.85	**95.59**
	Conflictive	81.62	**83.67**
Caltech101	Normal	**93.88**	93.08
	Conflictive	90.53	**90.99**
NUS-WIDE	Normal	40.54	**46.04**
	Conflictive	37.04	**42.17**
Scene15	Normal	70.12	**71.57**
	Conflictive	61.50	**62.74**
Animal	Normal	58.31	**59.48**
	Conflictive	49.39	**51.05**

Table 2: Accuracy (%) on normal and conflictive test sets. The best results are highlighted in bold.

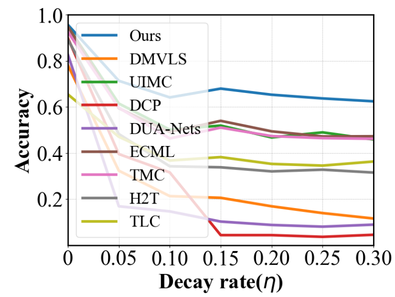

Figure 3: Performance comparison on PIE datasets with varying decay rates.

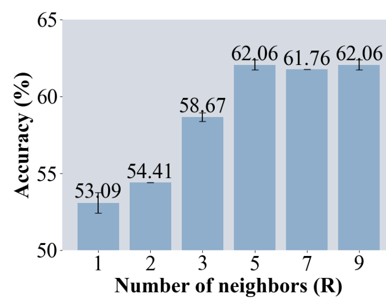

Figure 4: Accuracy comparison with respect to R.

Experiment Results

Accuracy Comparison. Table 1 presents the comparison results of various methods on standard test sets. It is evident that our proposed TMLC method significantly outperforms all baseline methods on the class-balanced test set, particularly on the PIE, NUS-WIDE, and Animal datasets, where it achieves performance improvements exceeding 5%. We attribute this enhancement primarily to our uncertainty-based sample generation scheme, which ensures that the generated pseudo-samples more closely approximate the true sample distribution. These pseudo-samples effectively mitigate the class imbalance issue in multi-view data, thereby improving the model's classification accuracy. To validate the effectiveness of the generated pseudo-samples, we conduct a comparative experiment by excluding the pseudo-sample generation module, resulting in a variant labeled as Ours-v1. Furthermore, to verify the superiority of the weight calculation scheme, we replace the designed weight mechanism with random weights, resulting in a variant labeled as Ours-v2. The results show that Ours-v1 performs worse than the complete model across all tested datasets, demonstrating that the uncertainty-based sample generation scheme plays a crucial role in enhancing the network's classification performance. It is also seen that Ours-v2 exhibits inferior performance relative to the complete model, confirming that the uncertainty-induced weight scheme more effectively integrates neighboring samples from different views, thereby generating synthetic samples that better align with the characteristics of the minority class.

Performance for Conflictive Noise. To validate the superiority of the proposed group consensus opinion aggregation mechanism, we conduct experiments by introducing conflictive noise into the test data (Xu et al. 2024a). We compare our approach with ECML, a state-of-the-art (SOTA) baseline for handling conflictive instances. Table 2 summarizes the comparison results on six datasets under both normal and conflict noise conditions. The results demonstrate that, in both scenarios, the group consensus opinion aggregation mechanism generally outperforms ECML. Specifically, on

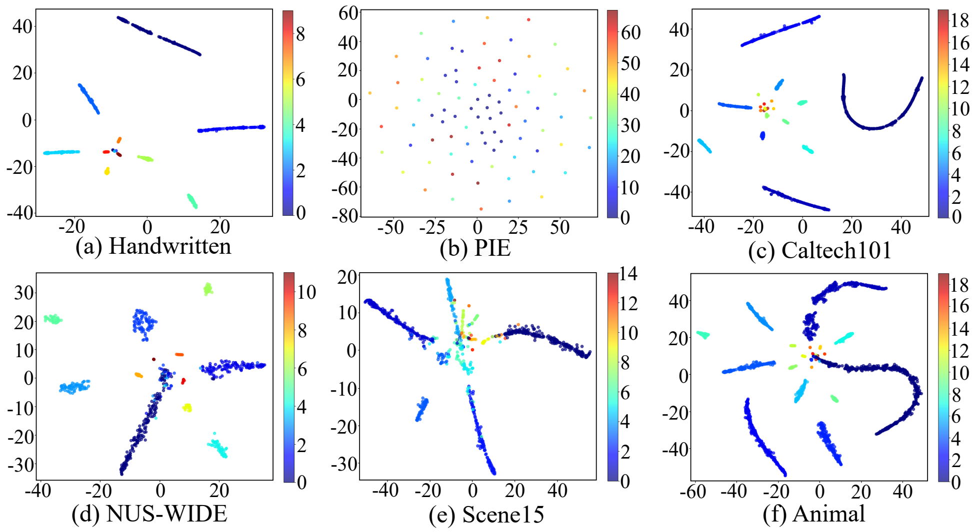

Figure 5: Joint evidence visualization on different datasets.

the NUS-WIDE dataset, our method achieves performance improvements of 5.5% and 5.13% in normal and conflict noise settings, respectively. These findings highlight the effectiveness of our proposed group consensus opinion aggregation in enhancing the aggregation of multi-view opinions, thereby leading to more reliable classification results.

Uncertainty Estimation. In order to assess the uncertainty of the model predictions, we visualize the distribution of both in-distribution and out-of-distribution samples. Specifically, the original samples and the samples with added Gaussian noise are considered as in-distribution and out-of-distribution samples, respectively. The experiment is conducted on the PIE dataset, where Gaussian noise with different standard deviations ($\sigma = 0.1, \sigma = 1, \text{and } \sigma = 10$) is added to the test samples, as shown in Fig. 2. The results indicate that as the noise intensity increases, the uncertainty of the data also increases, and the distribution curve of the out-of-distribution samples deviates more significantly from that of the in-distribution samples. This suggests that the proposed group consensus opinion aggregation strategy effectively captures the uncertainty, thereby improving the overall reliability of the model.

Performance Evaluation under Different Decay Rates. To demonstrate the superiority of our TMLC model in noisy environments, we conduct experiments by introducing varying levels of decay rates into the training data, with the decay rate ranging over $[0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30]$. Fig. 3 illustrates the performance comparison of all methods on the PIE dataset under different decay rates. It is evident that, compared to the baseline methods, our TMLC model consistently outperforms them across all tested datasets at each decay rate, further validating the effectiveness of the proposed pseudo-data generation scheme.

Parameter Selection. Since the model requires a predefined number of neighbors, R, we conduct experiments with different values of $R = [1, 2, 3, 5, 7, 9]$ to investigate its impact. As shown in Fig. 4, as the number of neighbors increases, the generated multi-view pseudo-samples incorporate more relevant sample information, which enhances the diversity of the samples and improves the classification performance. However, beyond a certain value of R, the classification performance approaches stability. This is because the sample knowledge has already been sufficiently integrated, and further increasing the number of neighbors does not provide additional useful information, resulting in negligible improvements in classification performance.

Joint Evidence Visualization. To validate the rationale of the newly defined evidence-based distance metric in the uncertainty-based sample generation scheme, Fig. 5 visualizes the evidence obtained from the aggregation mechanism on six multi-view datasets. Our observations are as follows: (1) Evidence points belonging to the same class are clustered closely together, while those from different classes are distinctly separated. This demonstrates that the evidence-based distance effectively captures class-specific information and accurately constructs neighbor relationships. (2) As the degree of class imbalance increases, the limited number of training samples results in reduced evidence and heightened uncertainty. Consequently, these classes tend to form neighbor relationships with other challenging-to-classify classes.

Conclusion

In this paper, we propose a trusted multi-view learning framework for long-tailed classification, namely TMLC . TMLC makes two significant contributions. First, inspired by Social Identity Theory, we design a group consensus opinion aggregation mechanism that aligns decision-making with the collective judgment of the majority. Second, we develop an innovative pseudo-data generation strategy that integrates a novel evidence-based distance metric and an uncertainty-guided module to produce high-quality synthetic samples. Extensive experiments demonstrate the superior performance of TMLC.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant No.62201475 and No.62406241, Sichuan Science and Technology Program under Grant No.2024NSFSC1436.

References

1976. Two Modifications of CNN. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-6(11): 769–772.

Cao, K.; Wei, C.; Gaidon, A.; Arechiga, N.; and Ma, T. 2019. Learning imbalanced datasets with label-distribution-aware margin loss. *Advances in neural information processing systems*, 32.

Chawla, N. V.; Bowyer, K. W.; Hall, L. O.; and Kegelmeyer, W. P. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16: 321–357.

Chu, P.; Bian, X.; Liu, S.; and Ling, H. 2020. Feature space augmentation for long-tailed data. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16*, 694–710. Springer.

Cui, Y.; Jia, M.; Lin, T.-Y.; Song, Y.; and Belongie, S. 2019. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9268–9277.

Deng, H.; Pan, N.; Tang, C.; and Shi, L. 2025. Trustworthy data recovery for incomplete multi-view learning. *Signal Processing*, 110146.

Du, S.; Fang, Z.; Lan, S.; Tan, Y.; Günther, M.; Wang, S.; and Guo, W. 2023. Bridging Trustworthiness and Open-World Learning: An Exploratory Neural Approach for Enhancing Interpretability, Generalization, and Robustness. In *Proceedings of the 31st ACM International Conference on Multimedia*, 8719–8729.

Feng, C.; Zhong, Y.; and Huang, W. 2021. Exploring classification equilibrium in long-tailed object detection. In *Proceedings of the IEEE/CVF International conference on computer vision*, 3417–3426.

Geng, Y.; Han, Z.; Zhang, C.; and Hu, Q. 2021. Uncertainty-aware multi-view representation learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 7545–7553.

Han, H.; Wang, W.-Y.; and Mao, B.-H. 2005. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In *International conference on intelligent computing*, 878–887. Springer.

Han, Z.; Zhang, C.; Fu, H.; and Zhou, J. T. 2022. Trusted multi-view classification with dynamic evidential fusion. *IEEE transactions on pattern analysis and machine intelligence*, 45(2): 2551–2566.

Hogg, M. A. 2016. *Social identity theory*. Springer.

Hsieh, T.-I.; Robb, E.; Chen, H.-T.; and Huang, J.-B. 2021. Droploss for long-tail instance segmentation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 1549–1557.

Hu, S.; Tian, B.; Liu, W.; and Ye, Y. 2025. Self-supervised Trusted Contrastive Multi-view Clustering with Uncertainty Refined. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 17305–17313.

Huang, C.; Shi, Y.; Wong, W.; Liu, C.; Wang, W.; Wang, Z.; and Wen, J. 2025. Multi-view Evidential Learning-based Medical Image Segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 17386–17394.

Ji, P.; and Liu, J. 2024. Tltscore: Towards long-tail effects in text-to-visual evaluation with generative foundation models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5302–5313.

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

Ju, L.; Wang, X.; Wang, L.; Liu, T.; Zhao, X.; Drummond, T.; Mahapatra, D.; and Ge, Z. 2021. Relational subsets knowledge distillation for long-tailed retinal diseases recognition. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*, 3–12. Springer.

Kang, B.; Li, Y.; Xie, S.; Yuan, Z.; and Feng, J. 2020. Exploring balanced feature spaces for representation learning. In *International conference on learning representations*.

Kang, B.; Xie, S.; Rohrbach, M.; Yan, Z.; Gordo, A.; Feng, J.; and Kalantidis, Y. 2019. Decoupling representation and classifier for long-tailed recognition. *arXiv preprint arXiv:1910.09217*.

Khan, S.; Hayat, M.; Zamir, S. W.; Shen, J.; and Shao, L. 2019. Striking the right balance with uncertainty. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 103–112.

Kim, J.; Jeong, J.; and Shin, J. 2020. M2m: Imbalanced classification via major-to-minor translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 13896–13905.

Li, B.; Han, Z.; Li, H.; Fu, H.; and Zhang, C. 2022. Trustworthy long-tailed classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6970–6979.

Li, M.; Zhikai, H.; Lu, Y.; Lan, W.; Cheung, Y.-m.; and Huang, H. 2024. Feature fusion from head to tail for long-tailed visual recognition. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 13581–13589.

Li, Y.; Zhen, L.; Sun, Y.; Peng, D.; Peng, X.; and Hu, P. 2025. Deep Evidential Hashing for Trustworthy Cross-Modal Retrieval. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 18566–18574.

Lin, Y.; Gou, Y.; Liu, X.; Bai, J.; Lv, J.; and Peng, X. 2022. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 4447–4461.

Liu, J.; Sun, Y.; Han, C.; Dou, Z.; and Li, W. 2020. Deep representation learning on long-tailed data: A learnable embedding augmentation perspective. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2970–2979.

Liu, W.; Yue, X.; Chen, Y.; and Denoeux, T. 2022. Trusted multi-view deep learning with opinion aggregation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 7585–7593.

Liu, X.-Y.; Wu, J.; and Zhou, Z.-H. 2008. Exploratory undersampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(2): 539–550.

Liu, Y.; Liu, L.; Xu, C.; Song, X.; Guan, Z.; and Zhao, W. 2024. Dynamic evidence decoupling for trusted multi-view learning. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 7269–7277.

Ren, J.; Yu, C.; Ma, X.; Zhao, H.; Yi, S.; et al. 2020. Balanced meta-softmax for long-tailed visual recognition. *Advances in neural information processing systems*, 33: 4175–4186.

Ren, Y.; Pu, J.; Yang, Z.; Xu, J.; Li, G.; Pu, X.; Yu, P. S.; and He, L. 2024. Deep clustering: A comprehensive survey. *IEEE transactions on neural networks and learning systems*, 36(4): 5858–5878.

Sensoy, M.; Kaplan, L.; and Kandemir, M. 2018. Evidential deep learning to quantify classification uncertainty. *Advances in neural information processing systems*, 31.

Shi, L.; Cao, L.; Wang, J.; and Chen, B. 2024a. Enhanced Latent Multi-view Subspace Clustering. *IEEE Transactions on Circuits and Systems for Video Technology*.

Shi, L.; Cao, L.; Ye, Y.; Zhao, Y.; and Chen, B. 2024b. Tensor-based graph learning with consistency and specificity for multi-view clustering. *arXiv preprint arXiv:2403.18393*.

Shi, L.; Tang, C.; Deng, H.; Xu, C.; Xing, L.; and Chen, B. 2024c. Generalized Trusted Multi-view Classification Framework with Hierarchical Opinion Aggregation. *arXiv preprint arXiv:2411.03713*.

Tan, J.; Wang, C.; Li, B.; Li, Q.; Ouyang, W.; Yin, C.; and Yan, J. 2020. Equalization loss for long-tailed object recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11662–11671.

Wan, X.; Liu, J.; Liu, X.; Wen, Y.; Yu, H.; Wang, S.; Yu, S.; Wan, T.; Wang, J.; and Zhu, E. 2024. Decouple then classify: A dynamic multi-view labeling strategy with shared and specific information. In *Forty-first International Conference on Machine Learning*.

Wang, T.; Li, Y.; Kang, B.; Li, J.; Liew, J.; Tang, S.; Hoi, S.; and Feng, J. 2020. The devil is in classification: A simple framework for long-tail instance segmentation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16*, 728–744. Springer.

Wang, T.; Zhu, Y.; Zhao, C.; Zeng, W.; Wang, J.; and Tang, M. 2021. Adaptive class suppression loss for long-tail object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3103–3112.

Wang, X.; Wang, Y.; Wang, Y.; Huang, A.; and Liu, J. 2024. Trusted semi-supervised multi-view classification with contrastive learning. *IEEE Transactions on Multimedia*.

Wang, Y.; Gan, W.; Yang, J.; Wu, W.; and Yan, J. 2019. Dynamic curriculum learning for imbalanced data classification. In *Proceedings of the IEEE/CVF international conference on computer vision*, 5017–5026.

Wang, Y.-X.; Ramanan, D.; and Hebert, M. 2017. Learning to model the tail. *Advances in neural information processing systems*, 30.

Xie, M.; Han, Z.; Zhang, C.; Bai, Y.; and Hu, Q. 2023. Exploring and exploiting uncertainty for incomplete multi-view classification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 19873–19882.

Xu, C.; Si, J.; Guan, Z.; Zhao, W.; Wu, Y.; and Gao, X. 2024a. Reliable conflictive multi-view learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 16129–16137.

Xu, C.; Zhang, Y.; Guan, Z.; and Zhao, W. 2024b. Trusted Multi-view Learning with Label Noise. *arXiv preprint arXiv:2404.11944*.

Yin, X.; Yu, X.; Sohn, K.; Liu, X.; and Chandraker, M. 2019. Feature transfer learning for face recognition with under-represented data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5704–5713.

Zhang, G.; Wu, J.; Yang, J.; Beheshti, A.; Xue, S.; Zhou, C.; and Sheng, Q. Z. 2021. Fraudre: Fraud detection dual-resistant to graph inconsistency and imbalance. In *2021 IEEE international conference on data mining (ICDM)*, 867–876. IEEE.

Zheng, X.; Tang, C.; Wan, Z.; Hu, C.; and Zhang, W. 2023. Multi-level confidence learning for trustworthy multimodal classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 11381–11389.

Zhou, H.; Xue, Z.; Liu, Y.; Li, B.; Du, J.; and Liang, M. 2023a. Rtmc: a rubost trusted multi-view classification framework. In *2023 IEEE international conference on multimedia and expo (ICME)*, 576–581. IEEE.

Zhou, H.; Xue, Z.; Liu, Y.; Li, B.; Du, J.; Liang, M.; and Qi, Y. 2023b. CALM: An Enhanced Encoding and Confidence Evaluating Framework for Trustworthy Multi-view Learning. In *Proceedings of the 31st ACM International Conference on Multimedia*, 3108–3116.