# Multi-view Learning via Trusted Pairwise Entity Energy

**Yalan Qin, Guorui Feng *, Xinpeng Zhang**

School of Communication and Information Engineering, Shanghai University, Shanghai, China
{ylqin, grfeng, xzhang}@shu.edu.cn

## Abstract

Learning on multi-view data is a fundamental task, which integrates the information from different views to improve the final performance. It is also a basic task for learning on the long-tailed data in real applications, followed by the downstream tasks, i.e., classification. The existing works for trusted classification on multi-view data or long-tailed data usually aim to improve the final performance and dynamically consider the confidence of prediction for the data which is crucial in cost-sensitive domains. However, these methods pay few attentions to the pairwise trusted problem which considers the trusted pairs instead of trusted annotated data points. Besides, the problem of classification on long-tailed multi-view data has never been studied so far. In this work, we focus on the pairwise trusted problem on long-tailed multi-view classification and give a general framework, which considers the trusted pairs instead of trusted annotated data points. We then construct a specific example under the general framework and introduce a novel Enhanced Normal-Inverse Gamma distribution (ENIG). ENIG is a joint probabilistic distribution built on Dirichlet distribution and NIG. A novel combination rule based on ENIG for long-tailed multi-view data is also given, which adaptively integrates the long-tailed data from different views to achieve a consensus one at the level of evidence and effectively produces a trusted long-tailed multi-view classification result. Our method is robust and able to be dynamically aware of the uncertainty for the long-tailed data from each view. The accurate uncertainty can be induced by the proposed learning framework, leading to both robustness and reliability for classification on long-tailed multi-view data. Experimental results on different long-tailed multi-view datasets demonstrate the effectiveness of our method in terms of accuracy, robustness and reliability.

## Introduction

In real-world applications, data are usually long-tailed distributed over various categories (Wu et al. 2020) and associated with multiple types of features or different modalities. The long-tailed classification task is a challenging task since it needs to deal with the few-shot learning problem for the tail classes. Besides, the over-all imbalance for class deviates the models to extremely focus on the head classes, leading to unpromising results on tail classes. Exploiting the information from different views is a long-standing goal to improve the learning performance in machine learning. Multi-view learning methods have achieved great success in recent years (Qin, Pu, and Wu 2024b; Qin et al. 2023c; Qin, Pu, and Wu 2024a; Qin et al. 2025d, 2022b, 2023b, 2025e; Qin, Feng, and Zhang 2025b; Qin et al. 2025f; Qin and Qian 2024; Qin et al. 2024a,b, 2025c; Qin, Feng, and Zhang 2025a; Qin et al. 2025a, 2022a, 2023d; Li et al. 2023a,b; Wang, Zhang, and Zhou 2025a,b; Wang et al. 2025; Liu et al. 2024, 2023a, 2025, 2023b, 2022b,a), which are different from the methods for singe view (Qin, Wu, and Feng 2021; Qin et al. 2022c, 2023e,a; Pu et al. 2023; Qin et al. 2025b) and typically depend on building complex models. Although accurate results can be obtained by these methods, they usually inevitably produce unreliable predictions, especially for views which are not well represented. It limits their deployment in real applications which are safety-critical, *i.e.*, medical diagosis.

Existing methods for long-tailed classification have been presented based on different strategies (Corbière et al. 2019; Lin et al. 2020). To achieve robust predictions, the redundant ensemble reduces the model variance to obtain the state-of-the-art performance. Some methods (Wu et al. 2020; Lin et al. 2017) assign larger importance to tail data points, which mainly rebalances the training for different classes. However, these methods usually inevitably produce unreliable prediction and their deployments in failure-sensitive applications are limited. They uniformly assign experts to all classes by assuming that all classifiers are trained on all data points, which often needs excessive computational cost. Trustworthy Long-tailed Classification (TLC) (Li et al. 2022) simultaneously performs uncertainty estimation and classification task in a multi-expert framework to identify hard data points. It is observed that the single-view long-tailed data is used for TLC, which lacks flexibility since the data with multiple views are common in real scenarios.

It has been shown effective that using the data with multiple views to learn a shared representation in various downstream tasks. For multi-view learning, traditional approaches usually assign a fixed weight or assume equal values for different views (Tsai et al. 2019). They are under the assumption that the importance or quality of all views are basically stable for all data points. Multi-view learning meth-

*Corresponding author

ods (Andrew et al. 2013) based on Canonical Correlation Analysis (CCA) are representative ones, which essentially maximize the correlation between different views to seek for a shared representation. Recently, some methods (Lin et al. 2023) built on contrastive learning have been presented and achieved good performance. However, the quality of a view usually varies for different data points and the designed methods should be aware of this for adaption. Therefore, we are expected to simultaneously obtain the classification result and show the confidence of the given decision. That is, we should provide the accurate uncertainty for the prediction result of each data point or individual view of each data point based on the model.

Estimating the uncertainty provides a way to obtain trusted prediction (Amini et al. 2020; Han et al. 2023). Without uncertainty estimation, the decisions made by used models are untrusted since they are easily influenced by the limited training data or noise. The untrusted models are vulnerable to be attacked, leading to wrong decisions and unbearable cost in critical domains. Then, it is desired to quantify the uncertainty in the learning systems. Algorithms based on uncertainty can be classified into two different categories,*i.e.*, non-Bayesian and Bayesian. Various Bayesian methods including Markov Chain Monte Carlo (MCMC) (Hernández-Lobato and Adams 2015), Laplacian approximation (MacKay 1992) and variational techniques (Blundell et al. 2015) have been proposed, which learn a distribution over the weight to characterize uncertainty. Compared with standard neural networks, these approaches are computationally expensive since it is a challenge to model the distribution of weights. Recently, plenty of non-Bayesian approaches including evidential deep learning (Sensoy, Kaplan, and Kandemir 2018), deep ensemble (Lakshminarayanan, Pritzel, and Blundell 2017) and deterministic uncertainty estimation (van Amersfoort et al. 2020) have been developed. Despite significant progress, most existing methods usually focus on the trusted annotated data points and pay few attentions to the pairwise trusted problem. Besides, these approaches are presented to estimate the uncertainty on the data with single view, which fail to fuse different views led by the uncertainty for improving both reliability and classification performance for long-tailed multi-view data.

In this paper, we focus on the pairwise trusted problem for the long-tailed multi-view classification and present a general framework, which is able to consider the trusted pairs rather than trusted annotated data points. It is an important and challenging problem since trusted pairwise relation between different representations of the same data point usually exists and helps lead to a desired classification result. The main contributions in this work are:

- We propose the pairwise trusted problem for long-tailed multi-view classification and give a general framework, which considers the trusted pairs instead of trusted annotated data points. This problem is pervasive but ignored in real world applications and we give a novel insight for long-tailed multi-view classification based on the uncertainty guided by pairwise pairs for the long-tailed data from each view. It is able to provide trusted decisions and effectively fuse the long-tailed data from different views at the evidence level under the general framework.
- We provide a specific example under the general framework and introduce Enhanced Normal-Inverse Gamma distribution (ENIG), which is a joint probabilistic distribution based on Dirichlet distribution and NIG for trusted long-tailed multi-view classification. We also give the combination rule of ENIG to integrate the long-tailed data from different views, which can measure the uncertainties of the long-tailed multi-view data.
- We perform extensive experiments on different long-tailed multi-view datasets, which demonstrates the effectiveness, reliability and robustness of our method compared with the representative approaches in terms of different metrics.

# Formulation

## General framework

Given trusted representation matrices $C_1 \in R^{n\times d_1}$ and $C_2 \in R^{n\times d_2}$ of the same dataset $X \in R^{n\times d}$, we propose to define the trusted energy function for long-tailed multi-view classification based on the trusted pairwise term $c_{ij}^{12}$ between $c_i^1$ and $c_j^2$. Here, $c_i^1$ and $c_j^2$ are trusted representations of $i$-th and $j$-th data point in $C_1$ and $C_2$, respectively. For a trusted pairwise entity $U \in R^{n\times n}$, we adopt $E(U)$ to define the trusted energy function regarding $U$. The value of $E(U)$ should be relatively large when $U$ is a correct trusted pairwise entity. Therefore, it turns to the maximizing problem of $E(U)$ in determining the correctness of $U$, which is formulated as

$$E(U) = \sum_{c_i^1, c_j^2 \in U} \pi(c_i^1, c_j^2), \tag{1}$$

where $\pi(\ )$ denotes the trusted pairwise potential. This kind of potential is defined in the following.

**Definition 1.** (Trusted Pairwise Potential) The trusted pairwise potential of each pair between $c_i^1$ and $c_j^2$ is defined as

$$\pi(c_i^1, c_j^2) = c_{ij}^{12}. \tag{2}$$

Note that the trusted pairwise potential $c_{ij}^{12}$ corresponds to the joint probabilistic distribution of $i$-th and $j$-th data point in $C_1$ and $C_2$. We adopt the vector $y$ as the indicator with $y_{c_i^1} = 1$ when $c_i^1$ leads to a correct output of downstream task or zero otherwise. Considering the influences brought by the number of columns in $U$, we average the trusted energy function $E(U)$ regarding the number of columns in $U$, formulated as:

$$\begin{aligned} \widehat{E(U)} &= \frac{1}{n^2} \sum_{c_i^1, c_j^2 \in U} \pi(c_i^1, c_j^2) y_{c_i^1} y_{c_j^2} \\ &= \sum_{c_i^1, c_j^2 \in U} \pi(c_i^1, c_j^2) \frac{y_{c_i^1}}{n} \frac{y_{c_j^2}}{n}, \end{aligned} \tag{3}$$

where $z = \frac{y}{n}$ and $\sum_i z_i = 1$. For the trusted pairwise entity $U$ with $n$ columns, the corresponding values of these columns are unknown. Since the problem of determining $z$ is NP hard, we relax $z$ in the continuous range $[0, \xi]$, where $0 < \xi \leq 1$. The relaxation is able to simultaneously simplify the computation complexity and endow $z$ with intuitive probabilistic interpretation. Then the above optimized problem is defined as:

$$\max \widehat{E(U)},\ s.t.\ z \geq 0,\ \sum_i z_i = 1,\ z_i \in [0, \xi]. \tag{4}$$

We then seek for the optimal solution to the above optimization problem by introducing Lagrangian multipliers $\lambda$, $\beta_1, \cdots, \beta_n$ and $\gamma_1, \cdots, \gamma_n$, where $\lambda \geq 0,\ \beta_i \geq 0$ and $\gamma_i \geq 0$. For simplicity, we use $g(z)$ to represent $\widehat{E(U)}$ and the corresponding Lagrangian function can be written as:

$$\begin{aligned} L(\lambda, \beta, \gamma, z) = g(z) + \lambda(1 - \sum_{i=1}^{n} z_i) \\ + \sum_{i=1}^{n} \beta_i z_i + \sum_{i=1}^{n} \gamma_i (\xi - z_i). \end{aligned} \tag{5}$$

We alternately update the variables $\lambda,\ \beta,\ \gamma$ and $z$ to minimize the above Lagrangian function. During the optimization process, the local optimal $z^*$ should satisfy the KKT condition, written as:

$$\begin{cases} \dfrac{\partial g(z_i^*)}{z_i^*} - \lambda + \beta_i - \gamma_i = 0, \\ \sum\limits_{i=1}^{n} (\xi - z_i^*)\gamma_i = 0, \\ \sum\limits_{i=1}^{n} z_i^* \beta_i = 0, \end{cases} \tag{6}$$

where $i = 1, \cdots, n$. We then define the following reward to explore the first-order necessary condition for the local optimal $z^*$.

**Definition 2.** (Trusted Pairwise Reward) The reward of trusted pairwise energy at column $t$ in $U$, represented by $R^{U_t}(z)$, is defined by

$$R^{U_t}(z) = \sum_{j \neq t} \pi(c_t^1, c_j^2) z_{c_j^2}. \tag{7}$$

Based on the definition of trusted pairwise reward, the value of $z_t^*$ is divided into three ranges, *i.e.*, $z_t^* = 0$, $z_t^* \in (0, \xi)$ and $z_t^* = \xi$. Then the above objective is rewritten as

$$R^{U_t}(z) \begin{cases} \leq \xi,\ z_t^* = 0, \\ = \xi,\ z_t^* \in (0, \xi), \\ \geq \xi,\ z_t^* = \xi. \end{cases} \tag{8}$$

By combining $z_t^* = 0$ and $z_t^* \in (0, \xi)$, we can achieve the set of components $z_t^* \in [0, \xi)$. The set of components $z_t^* \in (0, \xi]$ is achieved by integrating $z_t^* \in (0, \xi)$ and $z_t^* = \xi$. Based on these two combinations, we have the theorem in the following.

**Theorem 1.** If $z^*$ is the optimal solution with $z_i^* \in [0, \xi)$ and $z_j^* \in (0, \xi]$, then $R^{U_t}(z_i^*) \leq R^{U_t}(z_j^*)$ is satisfied. Otherwise, $z^*$ is not the optimal solution if $R^{U_t}(z_i^*) > R^{U_t}(z_j^*)$.

**Proof.** For the optimal $z^*$, $R^{U_t}(z_i^*) \leq R^{U_t}(z_j^*)$ with $z_t^* \in [0, \xi)$ and $z_t^* \in (0, \xi]$ is necessary conditions according to the above formulations. Therefore, $R^{U_t}(z_i^*) \leq R^{U_t}(z_j^*)$ with $z_t^* \in [0, \xi)$ and $z_t^* \in (0, \xi]$ achieved by the necessary conditions is satisfied if $z^*$ is the optimal solution. The detailed proof ends here.

The above Theorem 1 built on the trusted pairwise energy for long-tailed multi-view classification forms the important theoretical foundation of the proposed general framework in this part.

## Specific example

Within the general framework of trusted pairwise energy, we give a specific example that relies on the joint probabilistic distribution based on Dirichlet distribution and Normal Inverse-Gamma distribution (NIG) (Amini et al. 2020) for trusted long-tailed multi-view classification considering the independence between these two distributions, termed Enhanced Normal-Inverse Gamma distribution (ENIG), where $c_i^1$ and $c_j^2$ in Eq. (1) correspond to the trusted representations extracted by Dirichlet distribution and NIG, respectively. That is, the trusted pairwise term is built on predicting the label and location by Dirichlet distribution and NIG of input data, respectively. The trusted pairwise reward can be computed based on Eq. (7), which satisfies Theorem 1 in the general framework. We also show how to conduct uncertainty estimation with ENIG and give ENIG's combination rule for integrating the long-tailed data from different views.

ENIG is built on the generalization of Bayesian theory for subjective probability, which explicitly considers the source trust and uncertainty. It measures the chances to find the true class labels for prediction by assigning belief masses to the sets of class labels. Besides, ENIG is also guided by the advantage of NIG in evidential regression, which helps produce the trusted pairwise term. For single-view case, we define $\text{ENIG}(\eta, \delta, \alpha, \beta, \gamma)$ as follows:

$$\text{ENIG} = \begin{cases} \frac{1}{B(\eta)} \prod_{k=1}^{K} p_k^{\eta_k - 1} \text{NIG}, & p \in S_K, \\ 0, & \text{else}, \end{cases} \tag{9}$$

with

$$\begin{aligned} &\text{NIG}(\delta, \alpha, \beta, \gamma) = p(\mu, \sigma^2 \mid \delta, \gamma, \alpha, \beta) \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\sqrt{\gamma}}{\sigma\sqrt{2\pi}} (\frac{1}{\sigma^2})^{\alpha+1} \exp(-\frac{2\beta + \gamma(\delta - \mu)^2}{2\sigma^2}), \end{aligned} \tag{10}$$

where $B(\ )$ denotes the beta function, $K$ is the number of classes, $\eta$ is parameter of the Dirichlet distribution $D(p_i \mid \eta_i)$, $p_k \in [0, 1]$ is the unit simplex with $K$ dimensions and $S_K = \left\{ p \mid \sum_{k=1}^{K} p_k = 1 \right\}$. Based on the formulation of Dirichlet distribution (Darbellay and Vajda 2000), we can obtain the belief masses and uncertainty as:

$$b_k = \frac{\eta_k - 1}{S},\ \ u = \frac{K}{S}, \tag{11}$$

where $S = \sum_{k=1}^{K} \eta_k$ denotes the Dirichlet strength. We can use the evidence $e = [e_1, e_2, ..., e_K]$ to measure the support derived from the data, which is obtained from the output of neural networks. The parameters of the Dirichlet distribution (Darbellay and Vajda 2000; Li et al. 2022) is calculated by

$$\eta_k = e_k + 1. \tag{12}$$

Then the belief masses and uncertainty in Eq. (9) can be obtained. Note that the total amount of belief masses and uncertainty is a constant according to Eqs. (9)-(10), which is formulated as:

$$\sum_{k=1}^{K} b_k + u = 1. \tag{13}$$

We can observe that the belief masses will be low when the evidences are insufficient for prediction on all classes. The prediction has a high probability to be wrong when the uncertainty of the output is high. Having introduced ENIG measuring uncertainty and belief masses for single-view long-tailed data, we then study its extension to the long-tailed multi-view data. The combination rule of ENIG integrates the evidences from different views, resulting in a degree of belief represented by an objective termed belief function which considers all the available evidences. Specifically, we combine independent belief masses and uncertainties from $V$ views to achieve a joint one.

**Definition 3.** (Combination of ENIG distributions) Given two ENIG distributions, *i.e.*, $\mathrm{ENIG}(\eta_1, \delta_1, \alpha_1, \beta_1, \gamma_1)$ and $\mathrm{ENIG}(\eta_2, \delta_2, \alpha_2, \beta_2, \gamma_2)$, the combination of these two ENIG distributions is defined as

$$\begin{aligned} &\mathrm{ENIG}(\eta, \delta, \alpha, \beta, \gamma) = \mathrm{ENIG}(\eta_1, \delta_1, \alpha_1, \beta_1, \gamma_1) \\ &\uplus \mathrm{ENIG}(\eta_2, \delta_2, \alpha_2, \beta_2, \gamma_2), \end{aligned} \tag{14}$$

*with*

$$\begin{aligned} &\alpha = \alpha_1 + \alpha_2 + \frac{1}{2}, \;\; \gamma = \gamma_1 + \gamma_2, \\ &\beta = \beta_1 + \beta_2 + \frac{1}{2}\gamma_1(\delta_1 - \delta)^2 + \frac{1}{2}\gamma_2(\delta_2 - \delta)^2, \\ &\eta = \frac{1}{2}(\eta_1 + \eta_2), \;\; \delta = (\gamma_1 + \gamma_2)^{-1}(\gamma_1\delta_1 + \gamma_2\delta_2), \end{aligned} \tag{15}$$

where $\uplus$ corresponds to the multiplication between two distributions. As an extension, the combination of ENIG distributions for total $V$ views is as:

$$\begin{aligned} &\mathrm{ENIG}(\eta, \delta, \alpha, \beta, \gamma) = \mathrm{ENIG}(\eta_1, \delta_1, \alpha_1, \beta_1, \gamma_1) \\ &\uplus \mathrm{ENIG}(\eta_2, \delta_2, \alpha_2, \beta_2, \gamma_2) \\ &\uplus ... \uplus \mathrm{ENIG}(\eta_V, \delta_V, \alpha_V, \beta_V, \gamma_V). \end{aligned} \tag{16}$$

**Definition 4.** (ENIG's combination rule for two independent belief masses) The combination of belief masses $\{b_k\}_{k=1}^{K}$ is calculated from two belief masses $\left\{b_k^1\right\}_{k=1}^{K}$ and $\left\{b_k^2\right\}_{k=1}^{K}$ as (Han et al. 2021):

$$b_k = \frac{1}{1-C}(b_k^1 b_k^2 + b_k^1 u^2 + b_k^2 u^1), \tag{17}$$

where $C = \sum_{i \neq j} b_i^1 b_j^2$ denotes the conflict factor between two belief masses and the scale factor $\frac{1}{1-C}$ is used to perform normalization.

**Definition 5.** (ENIG's combination rule for two independent uncertainties) The combination of uncertainties $u$ is calculated from two uncertainties $u^1$ and $u^2$ as (Li et al. 2022) in the following manner:

$$u = \frac{1}{1-C} u^1 u^2. \tag{18}$$

Likewise, we can extend ENIG's combination rule for two independent belief masses and uncertainties to total $V$ views as the combination of ENIG distributions. After obtaining the joint belief masses $\{b_k\}_{k=1}^{K}$ and uncertainty $u$, we can induce the parameters for the ENIG and the joint evidence from different views as:

$$S = \frac{K}{u}, \;\; e_k = b_k \times S, \;\; \alpha_k = e_k + 1. \tag{19}$$

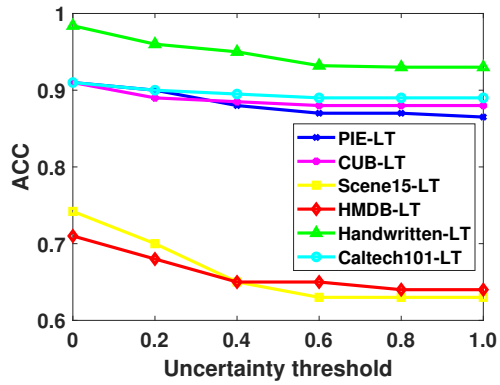
Then, we can achieve the final uncertainty and probability of each class by the parameter $\eta$ and joint evidence $e$.

We then discuss how to obtain evidence for each view by training neural networks, which can be used to achieve belief masses and uncertainty. The classification result is induced by capturing the evidence from the input with neural networks. To ensure that the network produces non-negative values as the evidence, we replace the conventional classifier with an activation function layer, *i.e.*, ReLU. We then achieve the parameters of the Dirichlet distribution by the evidence.

Given the evidence of the $i$-th data point obtained by the evidence network, we can achieve the parameter $\eta_i$ and $D(p_i \mid \eta_i)$, where $p_i$ is the class assignment probability. Then we can have the adjusted cross-entropy loss as:

$$\begin{aligned} L_i^{ace} &= \int\int\int \left[\sum_{j=1}^{K} -y_{ij} \log(p_{ij})\right] .p(y_i \mid \mu, \sigma^2) \\ &\quad .\mathrm{ENIG}(\eta, \delta, \alpha, \beta, \gamma) d\mu d\sigma^2 dp_i \\ &= \sum_{j=1}^{K} y_{ij}(\psi(S_i) - \psi(\eta_{ij})) \\ &\quad - \alpha \log(\Omega) + (\alpha + \frac{1}{2}) \log((y-\delta)^2\gamma + \Omega) \\ &\quad + \log \Upsilon + \frac{1}{2} \log(\frac{\pi}{\gamma}), \end{aligned} \tag{20}$$

where $\Omega = 2\beta(1+\gamma)$, $\psi(.)$ is the digamma function and $\Upsilon = (\frac{\Sigma(\alpha)}{\Sigma(\alpha+\frac{1}{2})})$. The above loss function is able to ensure that the right label of each data point produces more evidence. However, it fails to guarantee that less evidence is generated for wrong labels. Then, we introduce the KL divergence term based on the Dirichlet distribution $D(p_i \mid \eta_i)$




(a)

Figure 1: Classification performance with different uncertainty thresholds.

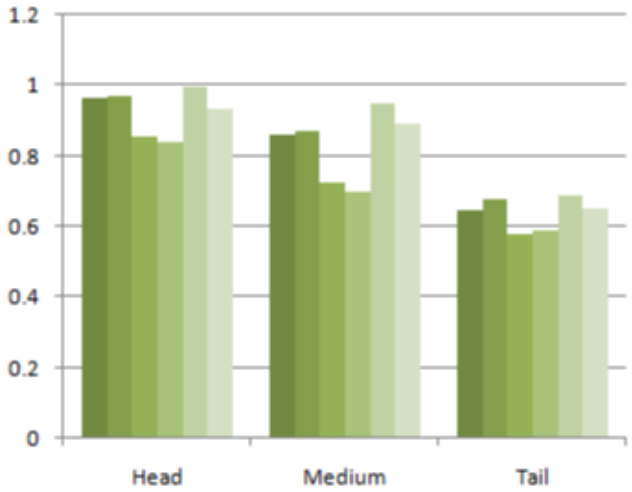
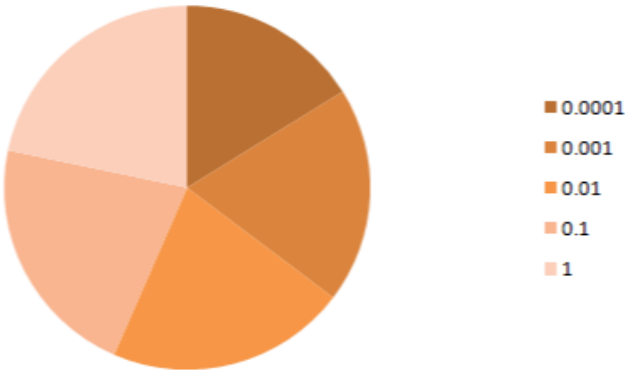



Figure 2: Classification accuracy for head, medium and tail.




(a) ACC

Figure 3: Performance on PIE-LT with different $\lambda_l$.

as:

$$\begin{aligned} L_{kl} &= KL[D(p_i \mid \hat{\eta}_i) \mid\mid D(p_i \mid 1)] \\ &= \log(\frac{\Gamma(\sum_{k=1}^{K} \hat{\eta}_{ik})}{\Gamma(K)\prod_{k=1}^{K}\Gamma(\hat{\eta}_{ik})}) \\ &+ \sum_{k=1}^{K}(\hat{\eta}_{ik}-1)\left[\psi(\hat{\eta}_{ik}) - \psi(\sum_{j=1}^{K}\hat{\eta}_{ik})\right], \end{aligned} \tag{21}$$

where $\Gamma(.)$ is the gamma function and $\hat{\eta}_i = y_i + (1-y_i)\odot\eta_i$ denotes the adjusted parameter of the Dirichlet distribution to avoid the evidence of the groundtruth class to be 0. Given parameter $\eta_i$, the loss of data point $i$ is formulated as:

$$L_i = L_i^{ace} + \lambda_l L_{kl}, \tag{22}$$

where $\lambda_l > 0$ is the parameter for balancing different terms. To simultaneously utilize all views, we use the overall loss $L'$ based on the multi-task strategy as follows:

$$L' = \sum_{i=1}^{n}\sum_{v=1}^{V} L_i^v, \tag{23}$$

where $L'$ is obtained by $L_i^v$. For simplicity, we omit the details here.

## Experiments

In this section, we perform experiments to investigate the proposed method on different long-tailed multi-view datasets.

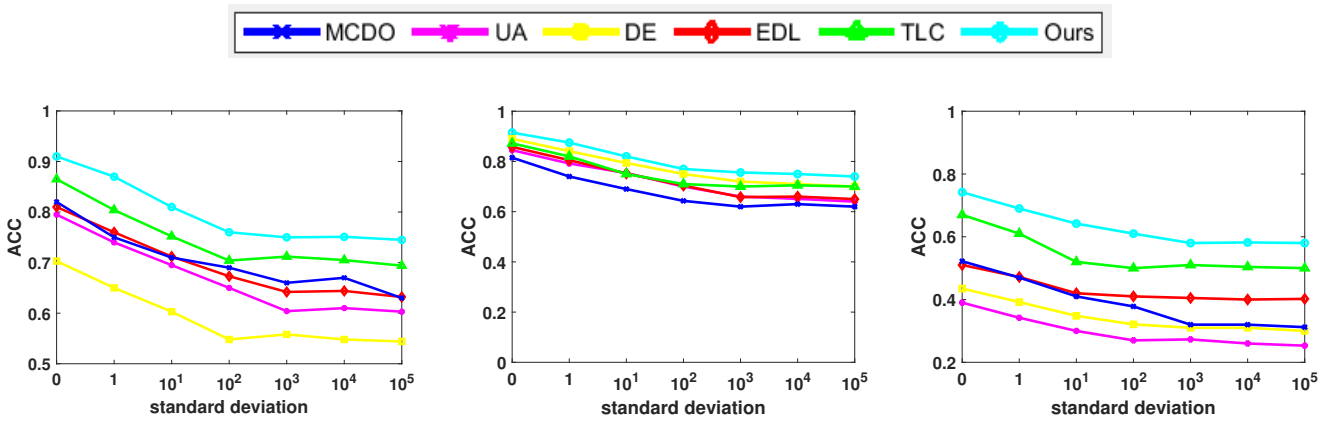
### Experimental setup

**Datasets** We use six long-tailed multi-view datasets including PIE-LT, CUB-LT, Scene15-LT, HMDB-LT, Handwritten-LT and Caltech101-LT. These datasets are obtained from the original datasets (*i.e.*, PIE, CUB, Scene15 (Fei-Fei and Perona 2005), HMDB (Kuehne et al. 2011), Handwritten and Caltech101 (Fei-Fei, Fergus, and Perona 2004)) with the transformation over exponential distributions. We set the imbalance ratio as 50 for these datasets, *i.e.*, the first class has 50 times more training data points than the last class.

**Compared methods** We compare the proposed method with the following six methods: Monte Carlo DropOut (MCDO) (Gal and Ghahramani 2015), Uncertainty-aware Attention (UA) (Heo et al. 2018), Deep Ensemble (DE) (Lakshminarayanan, Pritzel, and Blundell 2017), Evidential Deep Learning (EDL) (Sensoy, Kaplan, and Kandemir 2018), Trusted Long-tailed Classification (TLC) (Li et al. 2022), and Trusted Multi-view Classification (TMC) (Han et al. 2021). In implementation, the backbone for all datasets is ResNet32 (He et al. 2016) and SGD is used to optimize.

### Experimental results

We first compare the proposed method with current classification methods based on uncertainty using the view which performs the best since most existing uncertainty-based classification methods use single-view data as input. We com-




(b) PIE-LT (c) CUB-LT (d) Scene15-LT

Figure 4: Classification performance (ACC) on long-tailed multi-view datasets with different levels of noise.

| PIE-LT | MCDO | UA | DE | EDL | TMC | TLC | Ours |
|---|---|---|---|---|---|---|---|
| ACC | 80.0 | 78.4 | 65.3 | 78.3 | 83.5 | 85.0 | 89.6 |
| AUROC | 90.2 | 85.3 | 87.4 | 87.4 | 86.0 | 89.4 | 92.3 |
| CUB-LT | MCDO | UA | DE | EDL | TMC | TLC | Ours |
| ACC | 80.3 | 82.5 | 87.3 | 82.0 | 82.5 | 84.9 | 89.2 |
| AUROC | 92.0 | 64.3 | 91.7 | 89.4 | 90.0 | 92.5 | 94.7 |
| Scene15-LT | MCDO | UA | DE | EDL | TMC | TLC | Ours |
| ACC | 48.2 | 37.5 | 36.0 | 42.0 | 58.0 | 60.5 | 70.0 |
| AUROC | 87.9 | 82.0 | 67.4 | 86.5 | 89.3 | 90.5 | 92.0 |
| HMDB-LT | MCDO | UA | DE | EDL | TMC | TLC | Ours |
| ACC | 48.7 | 49.7 | 52.9 | 52.7 | 57.5 | 61.5 | 68.4 |
| AUROC | 89.2 | 86.4 | 88.0 | 89.5 | 91.5 | 92.7 | 94.0 |
| Handwritten-LT | MCDO | UA | DE | EDL | TMC | TLC | Ours |
| ACC | 90.2 | 91.5 | 94.7 | 93.0 | 94.0 | 95.2 | 97.0 |
| AUROC | 94.2 | 92.0 | 93.0 | 95.0 | 93.0 | 96.0 | 98.1 |
| Caltech101-LT | MCDO | UA | DE | EDL | TMC | TLC | Ours |
| ACC | 88.9 | 86.7 | 86.9 | 85.7 | 84.5 | 86.2 | 89.0 |
| AUROC | 90.6 | 92.0 | 92.6 | 94.0 | 93.5 | 96.0 | 98.1 |

Table 1: Classification performance (mean )(%) using the view which performs the best.

| PIE-LT | MCDO | UA | DE | EDL | TMC | TLC | Ours |
|---|---|---|---|---|---|---|---|
| ACC | 82.0 | 79.5 | 70.3 | 81.0 | 84.5 | 86.5 | 91.0 |
| AUROC | 91.4 | 87.0 | 89.0 | 88.5 | 89.0 | 91.3 | 94.0 |
| CUB-LT | MCDO | UA | DE | EDL | TMC | TLC | Ours |
| ACC | 81.5 | 84.5 | 89.0 | 85.8 | 86.9 | 87.2 | 91.5 |
| AUROC | 93.5 | 69.0 | 93.0 | 91.2 | 92.0 | 94.5 | 96.7 |
| Scene15-LT | MCDO | UA | DE | EDL | TMC | TLC | Ours |
| ACC | 52.2 | 39.0 | 43.5 | 51.0 | 65.2 | 67.0 | 74.2 |
| AUROC | 89.0 | 84.8 | 71.2 | 89.0 | 91.0 | 92.4 | 93.9 |
| HMDB-LT | MCDO | UA | DE | EDL | TMC | TLC | Ours |
| ACC | 53.0 | 52.5 | 58.0 | 57.1 | 62.0 | 64.5 | 71.0 |
| AUROC | 90.0 | 88.0 | 89.5 | 91.5 | 91.5 | 93.0 | 95.9 |
| Handwritten-LT | MCDO | UA | DE | EDL | TMC | TLC | Ours |
| ACC | 92.0 | 93.6 | 98.4 | 95.9 | 94.2 | 96.5 | 98.4 |
| AUROC | 96.8 | 94.0 | 95.6 | 95.7 | 96.0 | 97.1 | 98.5 |
| Caltech101-LT | MCDO | UA | DE | EDL | TMC | TLC | Ours |
| ACC | 89.1 | 88.4 | 87.0 | 87.8 | 85.7 | 87.0 | 91.0 |
| AUROC | 91.4 | 93.5 | 93.0 | 95.2 | 94.6 | 97.1 | 98.5 |

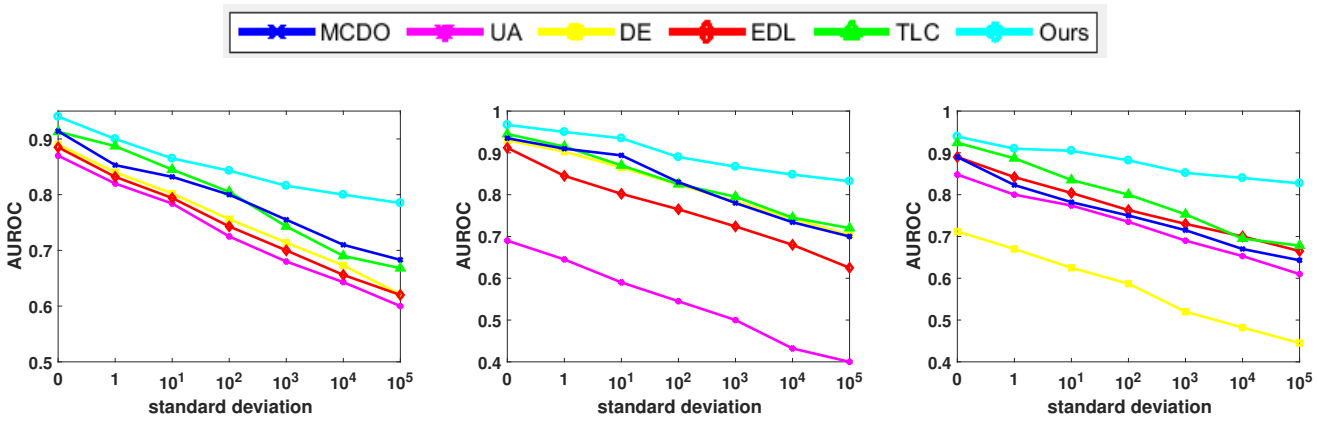
Table 2: Classification performance (mean)(%) using multiple views.

prehensively compare our method with others by reporting the results of all methods in terms of accuracy and AUROC (Hand and Till 2001), which is shown in Table 1. We repeat each experiment for 30 times and record the mean in the table. In all experiments, we set $\lambda_l = 0.1$. According to Table 1, we can observe that our method is able to achieve better classification performance than other methods on all long-tailed multi-view datasets. For example, our method improves about 1.8% compared with the method which performs the second best in terms of accuracy on Handwritten-LT dataset.

To further demonstrate the effectiveness of our method using multiple views, we integrate different views by concatenating the original features from multiple views for all methods. We also repeat each experiment for 30 times and record the mean in the table. We show the classification results using multiple views in Table 2. It can be found that our method can still achieve satisfied results, which shows the effectiveness of our method using multiple views.

## Uncertainty estimation

To perform the uncertainty estimation, we show how the prediction of our method changes with different prediction uncertainty values. Here, we use uncertainty threshold to denote value of uncertainty. According to Fig. 1, we can find that our method is able to provide better predictions based on ACC on all long-tailed multi-view datasets with the decreasing uncertainty thresholds, which demonstrates that the




(b) PIE-LT (c) CUB-LT (d) Scene15-LT

Figure 5: Classification performance (AUROC) on long-tailed multi-view datasets with different levels of noise.

| Dataset | CUB-LT | Scene-LT | HMDB-LT | Handwritten-LT | Caltech-LT |
|---|---|---|---|---|---|
| ACC | 0.840 | 0.680 | 0.625 | 0.905 | 0.810 |
| AUROC | 0.932 | 0.905 | 0.927 | 0.965 | 0.970 |

Table 3: Ablation study on the proposed ENIG.

| $L'$ | $L_i^{ace}$ | $L_{kl}$ | PIE-LT | CUB-LT | Scene15-LT | HMDB-LT |
|---|---|---|---|---|---|---|
| ✓ | | ✓ | 0.790 | 0.825 | 0.627 | 0.650 |
| ✓ | ✓ | | 0.789 | 0.832 | 0.610 | 0.546 |
| ✓ | ✓ | ✓ | **0.910** | **0.915** | **0.742** | **0.710** |

Table 4: Ablation study on different terms in the overall loss.

produced classification result and the corresponding uncertainty of our method are supported by the trusted decisions.

## Ablation study

We show how the proposed ENIG influences the classification performance and the results are listed in Table 3. Note that we randomly choose a view for each dataset to report the classification performance for the ablation study of ENIG. It is easy to conclude that ENIG is beneficial to obtain desirable classification performance.

We also compare different terms in the overall loss $L'$ ($L_i^{ace}$ and $L_{kl}$) on four datasets in terms of accuracy and the results of the full objective are included at the bottom of the table. According to Table 4, we can find that both $L_i^{ace}$ and $L_{kl}$ are both important to achieve satisfied classification performance on these datasets.

We study the accuracy on three class regions (head, medium and tail) for different datasets orderded by experimental setup in Fig. 2. We find that the proposed method achieves satisfied performance on head classes and relatively poor performance for the tail classes. This observation justifies the motivation that studying the long-tailed multi-view classification problem is important from the experimental perspective.

## Parameter selection

In this part, we study how to choose the proper value of the parameter $\lambda$. We select it in the range of $[0.0001, 0.001, 0.01, 0.1, 1]$ for simplicity and find that better performance can be obtained when $\lambda = 0.1$ according to Fig. 3 on PIE-LT dataset in terms of ACC.

## Robustness study

We also study the robustness of different methods using multiple views by adding Gaussian noise with multiple levels of standard deviations $(\sigma_t)$ to half of the total views. The comparison results based on accuracy are listed in Figs. 4-5. We can observe that our method is able to achieve competitive results when the data has no noise, *i.e.*, ours is robust on Scene15-LT for the ablation study on ENIG. Furthermore, the proposed method can be aware of the noise for specific view and achieve encouraging results on all long-tailed multi-view datasets, which can be explained by the fact that our method is benefited from the fusion based on uncertainty.

# Conclusion

In this paper, we propose pairwise trusted problem on long-tailed multi-view classification and give a general framework, which considers the trusted pairs instead of trusted annotated data points. We construct a specific example under the general framework and show a new trusted classification method based on ENIG for long-tailed multi-view data, which can integrate different views at the level of evidence and produce a trust classification result. We can induce the accurate uncertainty with the unified learning framework, leading to both robustness and reliability of classification problem for long-tailed multi-view data. Experimental results on several datasets show that our method is effective compared with different representative methods based on accuracy, robustness and reliability.

## Acknowledgments

This work was supported by Eastern Talent Plan Leading Project under Grant BJKJ2024011 and National Natural Science Foundation of China (62402303).

## References

Amini, A.; Schwarting, W.; Soleimany, A.; and Rus, D. 2020. Deep Evidential Regression. In *Advances in Neural Information Processing Systems*, 14927–14937.

Andrew, G.; Arora, R.; Bilmes, J. A.; and Livescu, K. 2013. Deep Canonical Correlation Analysis. In *Proceedings of the International Conference on Machine Learning*, 1247–1255.

Blundell, C.; Cornebise, J.; Kavukcuoglu, K.; and Wierstra, D. 2015. Weight Uncertainty in Neural Network. In *Proceedings of the 32nd International Conference on Machine Learning, ICML*, volume 37, 1613–1622.

Corbière, C.; Thome, N.; Bar-Hen, A.; Cord, M.; and Pérez, P. 2019. Addressing Failure Prediction by Learning Model Confidence. In *Advances in Neural Information Processing Systems*, 2898–2909.

Darbellay, G. A.; and Vajda, I. 2000. Entropy expressions for multivariate continuous distributions. *IEEE Trans. Inf. Theory*, 46(2): 709–712.

Fei-Fei, L.; Fergus, R.; and Perona, P. 2004. Learning Generative Visual Models from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops*, 178.

Fei-Fei, L.; and Perona, P. 2005. A Bayesian Hierarchical Model for Learning Natural Scene Categories. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 524–531.

Gal, Y.; and Ghahramani, Z. 2015. Bayesian Convolutional Neural Networks with Bernoulli Approximate Variational Inference. *CoRR*, abs/1506.02158.

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

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

Hand, D. J.; and Till, R. J. 2001. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Mach. Learn.*, 45(2): 171–186.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Heo, J.; Lee, H.; Kim, S.; Lee, J.; Kim, K. J.; Yang, E.; and Hwang, S. J. 2018. Uncertainty-Aware Attention for Reliable Interpretation and Prediction. In *Advances in Neural Information Processing Systems*, 917–926.

Hernández-Lobato, J. M.; and Adams, R. P. 2015. Probabilistic Backpropagation for Scalable Learning of Bayesian Neural Networks. In *Proceedings of the International Conference on Machine Learning*, 1861–1869.

Kuehne, H.; Jhuang, H.; Garrote, E.; Poggio, T. A.; and Gool, L. V. 2011. HMDB: A large video database for human motion recognition. In *IEEE International Conference on Computer Vision*, 2556–2563.

Lakshminarayanan, B.; Pritzel, A.; and Blundell, C. 2017. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. In *Advances in Neural Information Processing Systems*, 6402–6413.

Li, B.; Han, Z.; Li, H.; Fu, H.; and Zhang, C. 2022. Trustworthy Long-Tailed Classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 6960–6969.

Li, X.; Ren, Z.; Sun, Q.; and Xu, Z. 2023a. Auto-weighted Tensor Schatten p-Norm for Robust Multi-view Graph Clustering. *Pattern Recognit.*, 134: 109083.

Li, X.; Sun, Y.; Sun, Q.; and Ren, Z. 2023b. Consensus Cluster Center Guided Latent Multi-Kernel Clustering. *IEEE Trans. Circuits Syst. Video Technol.*, 33(6): 2864–2876.

Lin, T.; Goyal, P.; Girshick, R. B.; He, K.; and Dollár, P. 2017. Focal Loss for Dense Object Detection. In *IEEE International Conference on Computer Vision*, 2999–3007.

Lin, T.; Goyal, P.; Girshick, R. B.; He, K.; and Dollár, P. 2020. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42(2): 318–327.

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

Liu, C.; Wen, J.; Wu, Z.; Luo, X.; Huang, C.; and Xu, Y. 2024. Information Recovery-Driven Deep Incomplete Multiview Clustering Network. *IEEE Transactions on Neural Networks and Learning Systems*, 35(11): 15442–15452.

Liu, C.; Wen, J.; Xu, Y.; Zhang, B.; Nie, L.; and Zhang, M. 2025. Reliable Representation Learning for Incomplete Multi-View Missing Multi-Label Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6): 4940–4956.

Liu, C.; Wu, Z.; Wen, J.; Xu, Y.; and Huang, C. 2023a. Localized Sparse Incomplete Multi-View Clustering. *IEEE Transactions on Multimedia*, 25: 5539–5551.

Liu, J.; Liu, X.; Xiong, J.; Liao, Q.; Zhou, S.; Wang, S.; and Yang, Y. 2022a. Optimal Neighborhood Multiple Kernel Clustering With Adaptive Local Kernels. *IEEE Trans. Knowl. Data Eng.*, 34(6): 2872–2885.

Liu, J.; Liu, X.; Yang, Y.; Guo, X.; Kloft, M.; and He, L. 2022b. Multiview Subspace Clustering via Co-Training Robust Data Representation. *IEEE Trans. Neural Networks Learn. Syst.*, 33(10): 5177–5189.

Liu, J.; Liu, X.; Yang, Y.; Liao, Q.; and Xia, Y. 2023b. Contrastive Multi-View Kernel Learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(8): 9552–9566.

MacKay, D. J. C. 1992. A Practical Bayesian Framework for Backpropagation Networks. *Neural Comput.*, 4(3): 448–472.

Pu, N.; Zhong, Z.; Ji, X.; and Sebe, N. 2023. Federated Generalized Category Discovery. *CoRR*, abs/2305.14107.

Qin, Y.; Feng, G.; Ren, Y.; and Zhang, X. 2023a. Block-Diagonal Guided Symmetric Nonnegative Matrix Factorization. *IEEE Trans. Knowl. Data Eng.*, 35(3): 2313–2325.

Qin, Y.; Feng, G.; Ren, Y.; and Zhang, X. 2023b. Consistency-Induced Multiview Subspace Clustering. *IEEE Trans. Cybern.*, 53(2): 832–844.

Qin, Y.; Feng, G.; and Zhang, X. 2025a. Fast Incomplete Multi-view Clustering by Flexible Anchor Learning. In *Forty-second International Conference on Machine Learning*, 1–1.

Qin, Y.; Feng, G.; and Zhang, X. 2025b. Scalable One-Pass Incomplete Multi-View Clustering by Aligning Anchors. In *AAAI Conference on Artificial Intelligence*, 20042–20050.

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

Qin, Y.; Pu, N.; Feng, G.; and Sebe, N. 2025a. Robust Consensus Anchor Learning for Efficient Multi-view Subspace Clustering. In *Forty-second International Conference on Machine Learning*, 1–1.

Qin, Y.; Pu, N.; Sebe, N.; and Feng, G. 2025b. Latent Space Learning-Based Ensemble Clustering. *IEEE Trans. Image Process.*, 34: 1259–1270.

Qin, Y.; Pu, N.; and Wu, H. 2024a. EDMC: Efficient Multi-View Clustering via Cluster and Instance Space Learning. *IEEE Trans. Multim.*, 26: 5273–5283.

Qin, Y.; Pu, N.; and Wu, H. 2024b. Elastic Multi-View Subspace Clustering With Pairwise and High-Order Correlations. *IEEE Trans. Knowl. Data Eng.*, 36(2): 556–568.

Qin, Y.; Pu, N.; Wu, H.; and Fan, Z. 2025c. Flexible Multi-view Clustering with Dynamic Views Generation. In *Proceedings of the 33rd ACM International Conference on Multimedia*, 1072–1081.

Qin, Y.; Pu, N.; Wu, H.; and Sebe, N. 2025d. Discriminative Anchor Learning for Efficient Multi-View Clustering. *IEEE Trans. Multim.*, 27: 1386–1396.

Qin, Y.; Pu, N.; Wu, H.; and Sebe, N. 2025e. Margin-aware Noise-robust Contrastive Learning for Partially View-aligned Problem. *ACM Trans. Knowl. Discov. Data*, 19(1): 26:1–26:20.

Qin, Y.; and Qian, L. 2024. Fast Elastic-Net Multi-view Clustering: A Geometric Interpretation Perspective. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 10164–10172.

Qin, Y.; Qin, C.; Zhang, X.; and Feng, G. 2024a. Dual Consensus Anchor Learning for Fast Multi-View Clustering. *IEEE Trans. Image Process.*, 33: 5298–5311.

Qin, Y.; Qin, C.; Zhang, X.; Qi, D.; and Feng, G. 2023c. NIM-Nets: Noise-Aware Incomplete Multi-View Learning Networks. *IEEE Trans. Image Process.*, 32: 175–189.

Qin, Y.; Sun, Y.; Peng, D.; Zhou, J. T.; Peng, X.; and Hu, P. 2023d. Cross-modal Active Complementary Learning with Self-refining Correspondence. In *Advances in Neural Information Processing Systems*.

Qin, Y.; Tang, Z.; Wu, H.; and Feng, G. 2024b. Flexible Tensor Learning for Multi-View Clustering With Markov Chain. *IEEE Trans. Knowl. Data Eng.*, 36(4): 1552–1565.

Qin, Y.; Wu, H.; and Feng, G. 2021. Structured subspace learning-induced symmetric nonnegative matrix factorization. *Signal Process.*, 186: 108115.

Qin, Y.; Wu, H.; Zhang, X.; and Feng, G. 2022b. Semi-Supervised Structured Subspace Learning for Multi-View Clustering. *IEEE Trans. Image Process.*, 31: 1–14.

Qin, Y.; Wu, H.; Zhao, J.; and Feng, G. 2022c. Enforced block diagonal subspace clustering with closed form solution. *Pattern Recognit.*, 130: 108791.

Qin, Y.; Zhang, X.; Shen, L.; and Feng, G. 2023e. Maximum Block Energy Guided Robust Subspace Clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(2): 2652–2659.

Qin, Y.; Zhang, X.; Yu, S.; and Feng, G. 2025f. A survey on representation learning for multi-view data. *Neural Networks*, 181: 106842.

Sensoy, M.; Kaplan, L. M.; and Kandemir, M. 2018. Evidential Deep Learning to Quantify Classification Uncertainty. In *Advances in Neural Information Processing Systems*, 3183–3193.

Tsai, Y. H.; Bai, S.; Liang, P. P.; Kolter, J. Z.; Morency, L.; and Salakhutdinov, R. 2019. Multimodal Transformer for Unaligned Multimodal Language Sequences. In *Proceedings of the Conference of the Association for Computational Linguistics*, 6558–6569.

van Amersfoort, J.; Smith, L.; Teh, Y. W.; and Gal, Y. 2020. Uncertainty Estimation Using a Single Deep Deterministic Neural Network. In *Proceedings of the International Conference on Machine Learning*, volume 119, 9690–9700.

Wang, X.; Zhang, Y.; Zhang, J.; and Zhou, Y. 2025. Incomplete Multiview Clustering using Discriminative Feature Recovery and Tensorized Matrix Factorization. *IEEE Transactions on Circuits and Systems for Video Technology*, 1–1.

Wang, X.; Zhang, Y.; and Zhou, Y. 2025a. Bidirectional Probabilistic Multi-graph Learning and Decomposition for Multi-view Clustering. *IEEE Transactions on Image Processing*, 1–1.

Wang, X.; Zhang, Y.; and Zhou, Y. 2025b. Pseudo-Supervision Affinity Propagation for Efficient and Scalable Multiview Clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 1–12.

Wu, T.; Huang, Q.; Liu, Z.; Wang, Y.; and Lin, D. 2020. Distribution-Balanced Loss for Multi-label Classification in Long-Tailed Datasets. In *Computer Vision - ECCV*, 162–178.