Hypergraph-guided Global Mean-field Negotiation For Multiview Evidential Classification

Shaobo Hu[1,2,*], **Shiliang Sun**[3,4,*], **Nan Zhang**[1,2,5,†]
[1]College of Computer Science and Artificial Intelligence, Wenzhou University, China
[2]Wenzhou Institute of Data Research, Wenzhou University, China
[3]State Key Laboratory of Submarine Geoscience, the School of Automation and Intelligent Sensing, Shanghai Jiao Tong University, China
[4]Key Laboratory of System Control and Information Processing, Ministry of Education of China
[5]Wenzhou Turing Institute for Advanced Study in Artificial Intelligence, China
24451352015@stu.wzu.edu.cn, shiliangsun@gmail.com, nzhang@wzu.edu.cn

Abstract

Multiview learning integrates complementary information from diverse observations to enhance model performance, where to quantify predictive uncertainty evidential deep learning has recently gained significant attention. However, existing methods generally overlook the critical importance of high-order correlation structures in multiview features, and traditional evidential combination rules frequently suffer from issues such as order dependence and irrational belief assignments when handling highly conflictive multiview features. To address these issues, we propose the Hypergraph-guided Global Mean-field Negotiation (HGMN) method for multiview evidential classification. Specifically, HGMN first leverages hypergraph convolutional networks to capture high-order topological correlations within each view. Subsequently, instead of relying on pairwise or hierarchical fusion strategies, HGMN introduces a global mean-field negotiation mechanism that enables multiple views to dynamically reach a global consensus, effectively isolating dissenting noise. Finally, HGMN incorporates a multi-objective collaborative optimization strategy that enhances decision robustness and trustworthiness in complex open-world environments. Extensive experimental results on eight public datasets demonstrate that our method significantly outperforms state-of-the-art baselines in terms of accuracy and robustness.

1 Introduction

Multiview learning takes advantage of complementary and consistent information from different views of the same object to construct more robust and comprehensive representations [1], which has given rise to numerous applications in fields such as computer vision [2], natural language processing [3] and beyond [4–6]. However, enhancing model accuracy alone is no longer sufficient to meet the demands of today's society. Instead, such models are expected to not only excel in predictive performance but also demonstrate reliability in decision-making processes, uncertainty estimation, robustness, and human interpretability. In particular, trustworthy multiview learning has become a research hotspot, especially in safety-critical domains [7, 8]. Several recent studies have further explored this direction [9–11]. For example, trusted multiview classification [12] extended evidence deep learning [13] to multiview scenes by using evidence fusion techniques, which replaced the softmax output with Dirichlet distributions to quantify uncertainty based on subjective logic [14] and

*Equal contribution.
†Corresponding author: Nan Zhang.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

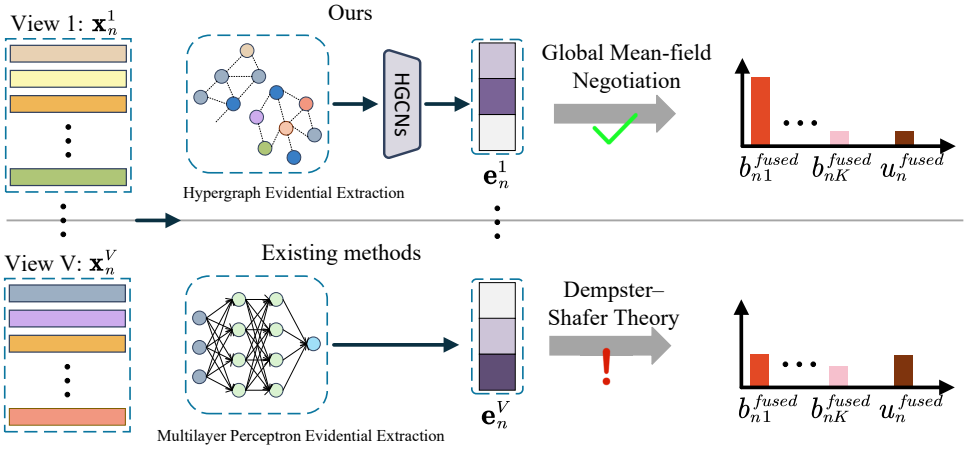


Figure 1: The issues of existing multiview evidential classification and our solutions. Existing methods typically employ multilayer perceptron for evidence extraction, whereas we adopt a hypergraph-based approach to obtain higher-quality initial evidence. In the multiview fusion stage, traditional methods rely on Dempster–Shafer theory, which suffers from limitations such as high sensitivity to conflicting data, sequential dependence, and pairwise combination rules. To overcome these limitations, we introduce a global mean-field negotiation mechanism.

Dempster-Shafer theory (DST) at each view. Meanwhile, Liu et al. [15] proposed a trusted multiview deep learning method that models each view's output as an opinion, accumulates evidence to reduce uncertainty, and enforces view consistency via opinion entropy.

However, DST-based fusion inherently suffers from two critical limitations, i.e., its sequential pairwise combination rules scale poorly with an increasing number of samples and are highly sensitive to conflicting evidence [16]. To overcome these issues, recent work has developed more robust fusion mechanisms tailored for conflicting multiview data [17, 18]. For example, Xu et al. [17] used conflicting opinion aggregation and a consistency loss to fuse view opinions. Similarly, Hu et al. [19] employed contrastive learning to align views and jointly learned consistent and complementary evidence for decision fusion. Nevertheless, as illustrated in Fig. 1, these methods still overlook the critical importance of high-order correlation structures within and across views, leading to unreliable evidence extraction. For example, in cancer multimodal diagnosis, focusing solely on isolated sample features may produce misleading evidence due to device artifacts in one view or information conflicts across views. Without exploiting hypergraph topological relationships across samples and views, a model struggles to capture high-order consistency and conflict signals, resulting in distorted uncertainty estimates and overconfident predictions [20, 21].

To address these challenges, we propose Hypergraph-guided Global Mean-field Negotiation (HGMN), a robust multiview evidential classification method that operates in two sequential stages. In the feature extraction stage, we first construct a joint local-global hypergraph and employ hypergraph convolutional networks to capture higher-order manifold structures and complex interaction information among samples, thus producing high-quality initial evidence. In the evidential fusion stage, we introduce a global mean-field negotiation mechanism that dynamically evaluates the topological consistency of each view to discount unreliable inputs and then places all views under a unified global constraint to directly compute the mainstream opinion. Conflicting views that deviate from the mainstream have their belief mass dynamically converted into uncertainty, achieving efficient conflict adaptation and global consensus alignment. Finally, we design a multi-objective optimization strategy: supervised contrastive learning enhances inter-class discriminability on multiview features, uncertainty-guided Jensen–Shannon consistency learning encourages safe cross-view consensus, and an evidential variance penalty forces high uncertainty under severe noise or semantic conflicts. In this way, HGMN fundamentally improves decision robustness and trustworthiness in complex open-world scenarios. The core contributions of this paper are summarized as follows:

1. We propose a novel multiview evidential learning framework (HGMN) that leverages a joint local-global hypergraph and hypergraph convolutional networks to extract high-quality evidence from higher-order manifold structures.

2. We design an adaptive global mean-field negotiation mechanism that overcomes the sequential dependency and sensitivity to conflicting evidence in traditional DST and its combination rules, thus addressing the challenges posed by conflicting multiview data.
3. We perform extensive experiments to evaluate HGMN, with results showing its superior performance against state-of-the-art methods in terms of accuracy, reliability, and robustness.

2 Related Work

Hypergraph Neural Networks. Hypergraph neural networks handle high-order relational data by leveraging hyperedges to model complex multi-node interactions [22]. Feng et al. [23] designed hyperedge convolution for representation learning on hypergraphs, aiming to capture high-order information. Yadati et al. [24] trained graph convolutional networks on hypergraphs via linear edge approximation, which enables efficient modeling of high-order relationships. In order to improve adaptability, several dynamic mechanisms were introduced, including generalized message passing [25], continuous diffusion [26], ordinary differential equations [27], partial differential equations [28], and cellular sheaf theory [29]. Meanwhile, hypergraph neural networks have demonstrated application potential in various domains, including traffic detection [30] and medical segmentation [31]. Existing research has made significant progress in the framework construction, dynamic optimization, and application expansion of hypergraph neural networks, providing abundant tools and methods for processing complex high-order relational information. Recent work has explored the use of hypergraphs in multiview learning, including multiview multilabel classification [32], multiview clustering with hyper-Laplacian regularization [33], and superpixel-level HSI clustering [34]. To the best of our knowledge, hypergraph convolution has not yet been introduced into the trusted multiview classification domain. Therefore, we construct a joint local-global hypergraph and introduce hypergraph convolutional networks to process high-order relational information, thereby generating high-quality initial evidence in the feature extraction stage.

Trusted Multiview Learning. Multiview learning aims to integrate complementary information across different views, yet traditional methods often ignore decision uncertainty, limiting their use in safety-critical applications [35–37].To address this limitation, subsequent work has focused on handling view conflicts and heterogeneity. Huang et al. [38] identified over-reliance on conflict-sensitive DST and proposed selective Markov random fields for robust fusion. Shi et al. [39] developed hierarchical opinion aggregation to mitigate within-view noise. Liu et al. [40] dynamically decoupled consistent and complementary evidence to handle semantic ambiguity. Beyond explicit conflicts, robustness against low-quality data, label noise, and adversarial attacks has also been explored. Zhou et al. [41] designed a confidence-aware fusion mechanism for corrupted views. Xu et al. [42] constructed a noise correlation matrix to detect label noise and model uncertainty. Yue et al. [43] developed an evidence inconsistency metric with belief ensemble for adversarial defense, while Dong et al. [44] extended recognition to unknown classes for open-world settings. However, these methods that rely on DST's sequential pairwise combination rules still suffer from poor scalability and high sensitivity to conflicting evidence. To address this issue, we introduce a global mean-field negotiation mechanism that places all views under a unified global constraint to directly compute the mainstream opinion, thereby achieving efficient conflict adaptation and global consensus alignment.

3 Methodology

As illustrated in Fig. 2, the proposed HGMN method comprises two stages, i.e., feature extraction and evidential fusion. This section provides a comprehensive overview of HGMN.

3.1 Problem Definition

In a typical multiview classification task, let $\mathcal{D} = \{(\mathcal{X}_n, \mathbf{y}_n)\}_{n=1}^{N}$ be a dataset consisting of N samples belonging to K mutually exclusive classes. Each sample is observed from V distinct views, denoted as the set $\mathcal{X}_n = \{\mathbf{x}_n^1, \mathbf{x}_n^2, \ldots, \mathbf{x}_n^V\}$, where $\mathbf{x}_n^v \in \mathbb{R}^{d_v}$ is the feature vector of the n-th sample in the v-th view with d_v dimensions. The ground-truth label is represented as a one-hot vector $\mathbf{y}_n \in \{0,1\}^K$. By stacking the feature vectors of all N samples, we obtain the overall feature matrix for the v-th view, denoted as $\mathbf{X}^{(v)} \in \mathbb{R}^{N \times d_v}$.

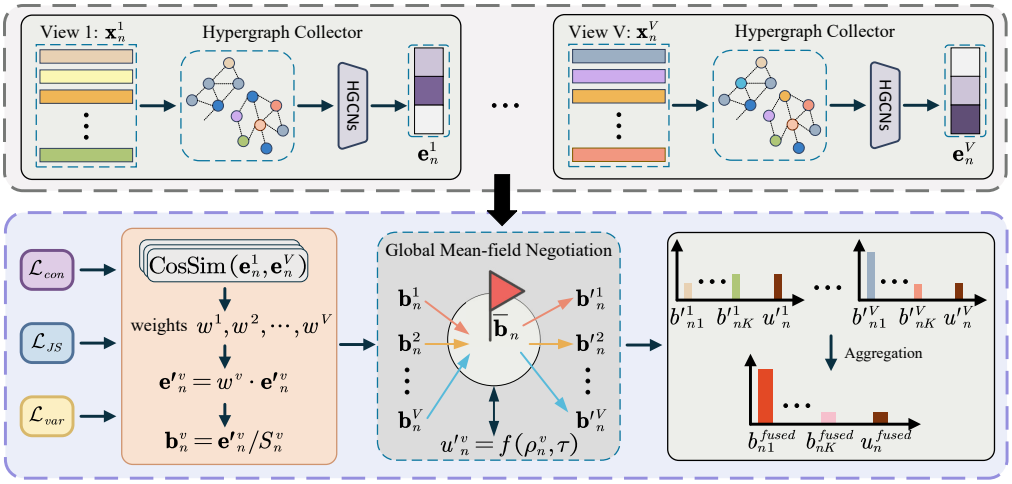


Figure 2: Illustration of the proposed HGMN framework. First, a joint local-global hypergraph is constructed to generate initial high-quality evidence. Subsequently, a global mean-field negotiation mechanism fuses opinions by first discounting unreliable ones according to topological consistency to obtain $\mathbf{e'}_n^v$, then computing the mean-field opinion under a unified global constraint while dynamically converting the belief mass of conflicting opinions into uncertainty $u'^v_n = f(\rho_n^v, \tau) = u_n^v + \tau(1 - \rho_n^v)(1 - u_n^v)$, thereby yielding $\mathbf{b'}_n^v$ and performing average fusion for decision making. Finally, a multi-objective optimization strategy is adopted, including supervised contrastive learning $\mathcal{L}_{con}$, uncertainty-guided weighted JS consistency learning $\mathcal{L}_{JS}$, and an evidential variance penalty $\mathcal{L}_{var}$.

3.2 Hybrid Hypergraph Evidential Extraction

Traditional multiview extractors (e.g., standard multilayer perceptron) fundamentally struggle to capture the complex, high-order semantic correlations inherent in multiview data, often leading to severe structural information decay. To address this limitation, we propose hybrid hypergraph convolutional networks designed to extract reliable subjective evidence from each view.

For a mini-batch feature matrix $\mathbf{X}^v \in \mathbb{R}^{B \times d_v}$ in the v-th view, we construct a joint hypergraph that simultaneously encodes fine-grained local geometry and global semantic attractors. Specifically, after L_2 normalization to obtain $\tilde{\mathbf{X}}^v$, we compute the dense similarity matrix $\mathbf{S}^v = \tilde{\mathbf{X}}^v(\tilde{\mathbf{X}}^v)^\top$. We instantiate local hyperedges by linking each sample to its m-nearest neighbors, yielding the local incidence matrix $\mathbf{H}_{local}^v \in \{0,1\}^{B \times B}$. To mitigate overfitting to noisy local neighborhoods and to anchor the global data distribution, we further sample C distinct nodes as global semantic attractors. Assigning samples to their nearest attractors generates the global incidence matrix $\mathbf{H}_{global}^v \in \{0,1\}^{B \times C}$ where $\mathbf{H}_{global}^v[i,j] = 1$ if sample i is assigned to centroid j. The hybrid topological structure $\mathbf{H}^v$ is derived by concatenating two matrices:

$$\mathbf{H}^v = [\mathbf{H}_{local}^v \parallel \mathbf{H}_{global}^v] \in \mathbb{R}^{B \times (B+C)}. \tag{1}$$

For spectral convolution on the hypergraph, the hypergraph transition matrix $\mathbf{G}^v$ is formulated as:

$$\mathbf{G}^v = \mathbf{D}_q^{-1/2}\mathbf{H}^v\mathbf{W}\mathbf{D}_r^{-1}(\mathbf{H}^v)^\top\mathbf{D}_q^{-1/2}, \tag{2}$$

where $\mathbf{W}$ is the hyperedge weights, and $\mathbf{D}_q$ and $\mathbf{D}_r$ represent the diagonal degree matrices of vertices and hyperedges, respectively. The high-order representation at the $(l+1)$-th layer is subsequently updated via $\mathbf{Z}^{(l+1)} = \sigma(\mathbf{G}^v\mathbf{Z}^{(l)}\mathbf{\Theta}^{(l)})$. To map these topological features into subjective logic without over-smoothing, we employ a two-layer hypergraph block coupled with a linear residual shortcut. A softplus activation strictly enforces the non-negativity constraint required for evidential parameters:

$$\mathbf{E}^v = Softplus\Big(HGCN(\mathbf{X}^v, \mathbf{G}^v) + Shortcut(\mathbf{X}^v)\Big). \tag{3}$$

where $HGCN(\cdot)$ denotes the two-layer hypergraph block, $Shortcut(\cdot)$ denotes the linear residual shortcut, $Softplus(\cdot)$ denotes the softplus activation function (shortcut connections are introduced

to preserve instance-specific discriminative information and mitigate over-smoothing, while Softplus is adopted to enforce non-negativity of evidence parameters with smooth gradients, thereby avoiding the gradient-dead zone of ReLU that would otherwise prevent recovery of zeroed dimensions). For the n-th sample, the extracted evidence $\mathbf{e}_n^v = (e_{n1}^v, \ldots, e_{nK}^v)$ serves as a subjective prior. To ensure these priors are sufficiently discriminative before entering the negotiation phase, we formalize the representation learning objective using information theory.

Assumption 1 (Semantic Latent Consensus) *Although multiview observations $\{\mathbf{x}^1, \ldots, \mathbf{x}^V\}$ are perturbed by heterogeneous and conditionally independent noise, they inherently share a task-relevant latent semantic variable mapped to the ground-truth Y.*

Based on Assumption 1, the optimal topological extraction should maximize the mutual information between the evidence and the true labels. However, computing the exact mutual information in high-dimensional continuous space is computationally intractable. Therefore, minimizing the supervised contrastive loss $\mathcal{L}_{con}$ in the evidence space strictly maximizes a tractable lower bound on the mutual information between the extracted topological representations and the latent semantics [45]. We impose the supervised contrastive regularization directly on the evidence space:

$$\mathcal{L}_{con} = -\frac{1}{2N} \sum_{i}^{V} \sum_{i \neq j} \sum_{n=1}^{N} \log \frac{\exp(sim(e_n^i, e_n^j)/\tau)}{\sum_{n \neq m, w=i,j} \exp(sim(e_n^i, e_m^w)/\tau)}, \tag{4}$$

where $sim\,(\cdot, \cdot)$ denotes the cosine similarity. This physically compresses the intra-class entropy while expanding the inter-class margins, ensuring a mathematically compact and separable manifold for robust evidential reasoning.

3.3 Global Mean-field Negotiation Evidential Fusion

Traditional evidential fusion heavily relies on the sequential DST rule, which is computationally intractable and perilously sensitive to highly conflicting views. To circumvent this, we propose a global mean-field negotiation mechanism to adaptively calibrate and align multiview beliefs. The mean-field approximation originates from statistical mechanics and was later formalized as a variational inference technique. The core idea of replacing complex multi-body interactions with an averaged effective field acting on each entity independently originates from statistical mechanics, where it was developed to solve interacting spin systems such as the Ising model [46]. This same principle was later formalized as a general-purpose variational inference technique: minimizing the KL divergence between the true joint distribution and a fully-factorized surrogate yields a set of coupled self-consistency equations, each solved as an independent optimization [47].

Topological Discounting. Before participating in the global mean-field, anomalous views must be strictly restrained. We measure the topological consensus score c_n^v by evaluating the cross-view agreement of the L_2-normalized evidence vectors $\tilde{\mathbf{e}}_n^v$:

$$c_n^v = \sum_{j=1}^{V} (\tilde{\mathbf{e}}_n^v)^\top \tilde{\mathbf{e}}_n^j, \tag{5}$$

where $\tilde{\mathbf{e}}_n^v = \mathbf{e}_n^v / \|\mathbf{e}_n^v\|_2$. The topological consensus score c_n^v is relatively normalized to yield a dynamic discounting weight $w_n^v = c_n^v / (\max_j c_n^j + \epsilon)$, where $\epsilon = 10^{-8}$. Consequently, the discounted evidence is updated as $\mathbf{e'}_n^v = w_n^v \mathbf{e}_n^v$. Based on subjective logic [14], the view-specific Dirichlet distribution is parameterized by $\boldsymbol{\alpha}_n^v = \mathbf{e'}_n^v + \mathbf{1}$. The Dirichlet strength S_n^v, the probability belief mass $\mathbf{b}_n^v$, and the epistemic uncertainty u_n^v are formulated as:

$$S_n^v = \sum_{k=1}^{K} \alpha_{nk}^v, \quad \mathbf{b}_n^v = \frac{\boldsymbol{\alpha}_n^v - \mathbf{1}}{S_n^v}, \quad u_n^v = 1 - \sum_{k=1}^{K} b_{nk}^v = \frac{K}{S_n^v}\,. \tag{6}$$

Cognitive Boundary Calibration. We define the global mean-field as the expectation of all subjective beliefs, i.e., $\bar{\mathbf{b}}_n = \frac{1}{V} \sum_{v=1}^{V} \mathbf{b}_n^v$. The consensus degree of each view is quantitatively assessed via the inner product $\rho_n^v = (\mathbf{b}_n^v)^\top \bar{\mathbf{b}}_n$.

Crucially, rather than blindly averaging conflicting opinions, we actively penalize views that drastically deviate from the mean-field. Guided by a learnable negotiation intensity $\tau \in [0, 1]$, the discordant belief mass is dynamically converted into penalized uncertainty ${u'}_n^v$:

$$ {u'}_n^v = u_n^v + \tau(1 - \rho_n^v)(1 - u_n^v). \tag{7} $$

When a corrupted view heavily conflicts with the consensus ($\rho_n^v \to 0$), the model enforces a severe penalty driving ${u'}_n^v \to 1$ (the global mean-field acts as a consensus anchor, while the learnable τ controls the negotiation stiffness—lower τ tolerates dissent, higher τ enforces strict alignment with the mainstream opinion. When all views provide entirely contradictory evidence, driving epistemic uncertainty ${u'}_n^v \to 1$ is a desirable property, not a flaw. In safety-critical evidential learning, outputting "I don't know" under total conflict is far more trustworthy than forcing a confidently erroneous prediction). Concurrently, the raw beliefs are forced to align with the mean-field $\tilde{\mathbf{b}}_n^v = \mathbf{b}_n^v + \tau \bar{\mathbf{b}}_n \odot (\mathbf{1} - |\mathbf{b}_n^v - \bar{\mathbf{b}}_n|)$. The aligned beliefs ${\mathbf{b}'}_n^v$ are then strictly redistributed to satisfy the residual belief budget M_n^v:

$$ {\mathbf{b}'}_n^v = M_n^v \frac{\tilde{\mathbf{b}}_n^v}{\|\tilde{\mathbf{b}}_n^v\|_1} = (1 - {u'}_n^v) \frac{\tilde{\mathbf{b}}_n^v}{\|\tilde{\mathbf{b}}_n^v\|_1}. \tag{8} $$

Ultimately, the mean-pooled global consensus is mapped back to the fused Dirichlet parameters $\boldsymbol{\alpha}_n^{fused}$ for downstream decision-making:

$$ \mathbf{b}_n^{fused} = \frac{1}{V} \sum_{v=1}^{V} {\mathbf{b}'}_n^v, \quad u_n^{fused} = \frac{1}{V} \sum_{v=1}^{V} {u'}_n^v, \quad \boldsymbol{\alpha}_n^{fused} = \mathbf{b}_n^{fused} \frac{K}{u_n^{fused}} + \mathbf{1}. \tag{9} $$

Dynamic Belief Alignment. To drive the global negotiation, individual view distributions must converge toward the consensus. To prevent clean views from catastrophically aligning with anomalous noise, we define an uncertainty-guided weighted alignment metric. Let $\mathbf{p}_n^i$ and $\mathbf{p}_n^j$ be the probability distributions of two views. The uncertainty-guided weighted JS divergence scales the standard distribution distance by the joint epistemic reliability $(1 - u_n^i)(1 - u_n^j)$:

$$ \mathcal{L}_{JS} = \frac{2}{NV(V-1)} \sum_{n=1}^{N} \sum_{i<j}^{V} (1 - u_n^i)(1 - u_n^j) \mathrm{JS}(\mathbf{p}_n^i \parallel \mathbf{p}_n^j). \tag{10} $$

The gradient flow from severely corrupted views ($u_n \to 1$) is mathematically truncated, ensuring a unidirectional and robust cross-view optimal transport toward the clean consensus.

Cognitive Variance Regularization. Standard evidential learning optimizes the expected risk via cross-entropy but succumbs to the overconfidence pitfall under severe distribution shifts. To satisfy the principle of least commitment under extreme data corruption, a subjective logic network must not only minimize the expected classification risk but also strictly bound the second-order variance of the predicted Dirichlet simplex. To this end, we introduce the evidential variance penalty $\mathcal{L}_{var}$:

$$ \mathcal{L}_{var} = \frac{1}{(V+1)N} \sum_{n=1}^{N} \left(Var(\alpha_n^{fused}, S_n^{fused}) + \sum_{v=1}^{V} Var(\alpha_n^v, S_n^v) \right), \tag{11} $$

where $Var(\alpha_n^v, S_n^v) = \sum_{k=1}^{K} \frac{\alpha_{nk}^v (S_n^v - \alpha_{nk}^v)}{(S_n^v)^2 (S_n^v + 1)}$. This explicitly regularizes the false evidence bounds for non-target classes, which inherently guarantees that unresolvable multiview conflicts translate into high epistemic uncertainty.

3.4 Overall Optimization

To empower the framework with end-to-end task awareness, we employ the standard evidential cross-entropy risk coupled with the Kullback-Leibler (KL) divergence annealing term. The classification loss $\mathcal{L}_{acc}$ supervises both the final fused consensus and the individual view-specific predictions to prevent gradient vanishing in early layers:

$$ \mathcal{L}_{acc} = \mathcal{L}_{ace}(\boldsymbol{\alpha}_n^{fused}) + \sum_{v=1}^{V} \mathcal{L}_{ace}(\boldsymbol{\alpha}_n^v), \tag{12} $$

where $\mathcal{L}_{ace}(\boldsymbol{\alpha}_n) = \sum_{k=1}^{K} y_{nk} \left(\psi(S_n) - \psi(\alpha_{nk})\right) + \lambda_t \mathcal{L}_{KL}(\boldsymbol{\alpha}_n)$, and λ_t is the annealing coefficient. The relevant formulas for the KL divergence are provided in Appendix A.

By unifying the topological representation, dynamic negotiation, and classification objectives, the overall joint optimization function of the HGMN framework is formulated as:

$$\mathcal{L}_{total} = \mathcal{L}_{acc} + \lambda \mathcal{L}_{con} + \gamma \mathcal{L}_{JS} + \mu \mathcal{L}_{var}, \tag{13}$$

where λ, γ and μ are hyperparameters. Furthermore, the computational complexity of the HGMN framework is $\mathcal{O}(n)$, with a detailed analysis provided in Appendix B.

3.5 Discussion and Analysis

The following discussion focuses on two design choices in our framework. In the evidential extraction stage, hypergraphs are introduced to capture higher-order evidence relations that pairwise graphs cannot represent, with a detailed discussion of how this design alleviates the noise problem provided in Appendix C. In the evidential fusion stage, a global mean-field negotiation mechanism fuses all evidence sources jointly rather than sequentially, thereby removing the strict order dependence of traditional Dempster–Shafer theory and reducing the fusion complexity from $\mathcal{O}(VK^2)$ to $\mathcal{O}(VK)$. Below, we provide a theoretical justification showing that this mechanism can be interpreted as a variational approximation to the true joint evidence distribution under mild assumptions.

Let $\mathcal{B} = \{b^1, b^2, \ldots, b^V\}$ be the set of beliefs from V views. The exact joint consensus distribution $P(\mathcal{B})$ is intractable under severe conflicts. By mean-field theory in statistical mechanics, we approximate the intractable distribution $P(\mathcal{B})$ with a fully factorized variational distribution $Q(\mathcal{B})$:

$$Q(\mathcal{B}) = \prod_{v=1}^{V} q^v(b^v), \tag{14}$$

where q^v is a single view's distribution. Our objective is to minimize the KL divergence between the approximate distribution Q and the true joint consensus P:

$$Q^* = \arg\min_{Q} KL(Q(\mathcal{B}) || P(\mathcal{B})), \tag{15}$$

By expanding the KL divergence and applying the calculus of variations with respect to q^v, we obtain the optimal update rule for the v-th view as the expectation over all other views $j \neq v$:

$$\ln q^{v*}(b_v) = \mathbb{E}_{Q_{-v}}[\ln P(\mathcal{B})] + C. \tag{16}$$

where C is constant and q^{v*} is the distribution of q^v when it has converged.

Assumption 2 (Global Consensus Vector) *Conditioned on a global consensus vector $\bar{\mathbf{b}}_n$, the views are independent and each follows a distribution whose natural parameter is a linear function of $\bar{\mathbf{b}}_n$.*

As outlined in Assumption 2, the common mean-field assumption for coupled systems implies that the true joint distribution belongs to an exponential family with linearly factorizing sufficient statistics. Consequently, the optimal solution q^{v*} is a distribution centered on the arithmetic mean of the expected beliefs of all views. When all views are treated symmetrically and no prior knowledge distinguishes them, the fixed-point solution satisfies:

$$\mathbb{E}_{q^{v*}}[\mathbf{b}^v] = \frac{1}{V} \sum_{j=1}^{V} \mathbb{E}_{q^{j*}}[\mathbf{b}^j]. \tag{17}$$

Thus, the global consensus mean $\bar{\mathbf{b}} = \frac{1}{V} \sum_{v=1}^{V} \mathbf{b}^v$ emerges as a consistent fixed point. Instead of iterative variational updates, we directly compute this closed-form mean-field context and project each view's belief toward it, yielding a tensor-parallelizable, order-independent approximation. Therefore, our global mean-field negotiation is not a heuristic average but a mathematically grounded variational approximation under the stated assumption, with order independence and linear complexity $\mathcal{O}(VK)$.

4 Experiments

In this section, we verify the effectiveness of the proposed HGMN method from three aspects: (1) comparison with state-of-the-art (SOTA) methods; (2) ablation analysis; and (3) quantitative analysis.

4.1 Experimental Settings

Datasets and Comparison Methods. We conduct experiments on eight multiview datasets (i.e., ALOI [48], Caltech101-20 [49], COIL100 [50], Leaves [51], NUS-WIDE [52], OutdoorScene [53], Scene15 [54] and VGGFace2 [55]) and compare seven state-of-the-art evidential multiview classification methods (i.e., TMC [12], ECML [17], CCML [40], TUNED [38], TMCEK [56], GTMC-HOA [39] and RCMCL [19]). Their detailed descriptions can be found in Appendix D.

Implementation Details. For each dataset, we split the data into 80% for training and 20% for testing. The highly conflicting test set is constructed by applying two corruption operations to the clean test set. Neither operation is applied to the training set. The protocol for each test sample is as follows. 1) Gaussian noise injection: With probability r_{noise}, the feature vectors of the sample in a random subset of views are perturbed by additive Gaussian noise $\varepsilon \sim N(0, \sigma^2)$ with $\sigma = 1.0$; 2) Cross-class feature replacement: With probability $r_{unaligned}$, one randomly selected view of the sample has its feature vector replaced by the corresponding view's feature vector of a sample from a different class. Specifically, we adopt a cyclic class shift, i.e., the replacement feature is taken from the same view of a sample in class $(y + 1)$ mod K. In our experiments, the highly conflicting setting uses $r_{noise} = 0.2$ and $r_{unaligned} = 0.6$. For each method, we perform five independent repeated experiments and report the mean and standard deviation. To ensure fairness, the hyperparameters for the comparison methods were determined based on either the default global settings or the configuration of the first dataset. For details on the experimental compute resources, parameter settings, and hyperparameter analysis of HGMN, see Appendix E and F.3.

4.2 Comparison to Existing SOTA Methods

To validate the effectiveness of our method, we perform comprehensive experiments on multiple datasets under both normal and highly conflicting test sets. Tables 1 and 2 provide the Accuracy of HGMN and contrast methods respectively on the two settings and the Precision rate and F1 score are provided in the Appendix F.1. From Tables 1 and 2, we draw the following conclusions: 1) On the normal test sets, HGMN achieves performance highly competitive with existing SOTA methods. Specifically, it achieves the highest accuracy on COIL100, Leaves, NUS-WIDE, OutdoorScene and Scene15. Although it ranks second ALOI and Caltech101-20, the performance gap with the top-ranked method is very small. These results indicate that incorporating global mean-field negotiation does not impair the model's ability to extract normal semantic features and that HGMN maintains extremely high classification accuracy under ideal clean data conditions; 2) On highly conflicting test sets, HGMN demonstrates strong adaptability. Across all eight datasets, it consistently outperforms all baseline methods by a significant margin, particularly on ALOI, COIL100, Leaves, and OutdoorScene; 3) HGMN achieves minimum improvements of 6.79% in average accuracy compared to all baselines on highly conflicting test sets. This substantial performance margin provides direct and

Table 1: Accuracy (%) of HGMN and comparison methods on normal test sets. The best and the second best results are highlighted by boldface and underlined respectively.

Dataset	TMC	ECML	CCML	TUNED	TMCEK	GTMC-HOA	RCMCL	Ours
ALOI	71.14±0.34	69.32±0.73	61.75±1.27	85.08±1.55	69.61±0.82	**97.56±0.48**	90.69±0.54	97.11±0.41
Caltech101-20	92.05±0.82	92.64±1.10	93.60±0.72	93.72±0.70	92.09±0.87	92.22±1.49	**94.39±0.64**	94.27±0.85
COIL100	96.67±0.68	98.01±0.45	96.47±0.48	97.89±0.34	97.82±0.49	97.03±0.95	98.60±0.19	**98.89±0.25**
Leaves	34.50±3.04	94.12±1.03	96.56±0.20	90.62±1.56	93.87±0.42	97.31±0.42	97.00±0.92	**97.88±0.70**
NUS-WIDE	36.54±0.96	39.37±1.89	45.29±1.88	41.37±1.96	38.62±0.98	43.67±0.50	40.87±1.78	**45.71±0.81**
OutdoorScene	82.42±0.67	84.28±1.13	84.50±0.98	85.13±1.47	85.32±0.76	88.14±0.91	84.65±1.39	**88.77±1.02**
Scene15	61.11±2.09	70.68±1.09	71.17±1.30	73.24±0.62	71.53±0.82	77.37±1.06	73.53±1.10	**78.06±0.33**
VGGFace2	63.36±0.29	60.81±0.41	62.04±0.32	62.02±0.62	62.76±0.31	58.64±0.80	64.39±0.23	**64.62±0.21**

Table 2: Accuracy (%) of HGMN and comparison methods on highly conflicting test sets.

Dataset	TMC	ECML	CCML	TUNED	TMCEK	GTMC-HOA	RCMCL	Ours
ALOI	59.05±0.91	57.69±0.97	50.60±0.72	69.26±0.60	58.03±1.19	77.82±0.43	77.15±0.81	**89.42±0.40**
Caltech101-20	85.40±1.24	82.55±1.11	86.36±0.85	83.85±0.83	82.51±1.11	83.81±1.74	88.58±0.54	**89.33±1.77**
COIL100	69.94±1.13	69.65±0.95	70.43±0.94	66.54±1.23	70.33±0.81	70.49±1.09	74.64±0.74	**86.03±0.95**
Leaves	26.00±1.33	62.87±1.76	64.62±1.97	59.31±3.01	62.56±1.80	65.00±2.25	65.50±1.66	**76.56±0.91**
NUS-WIDE	31.42±1.16	32.29±1.23	36.00±0.61	33.50±2.33	32.12±1.50	36.33±2.03	36.92±0.98	**38.12±2.55**
OutdoorScene	67.06±0.54	68.92±1.14	69.29±0.80	67.55±0.89	69.78±1.66	70.04±1.38	74.39±1.60	**81.56±1.73**
Scene15	45.64±1.69	53.96±1.49	52.91±1.76	53.76±1.59	54.89±1.56	57.95±1.58	57.32±1.09	**63.83±1.03**
VGGFace2	43.95±0.39	43.88±0.32	45.49±0.69	43.75±0.26	45.14±0.44	40.10±4.03	48.76±0.32	**52.68±0.51**

potent evidence for the superiority of our proposed mechanism. When severe conflicts arise, the DST evidential fusion rules utilized by baseline methods become highly susceptible to being misguided by views that are significantly corrupted. Conversely, through topological consistency-based antecedent discounting and global mean-field alignment, HGMN acutely identifies and actively isolates contaminated views, thereby enabling reliable, robust decisions in highly uncertain open environments. To further demonstrate robustness, we conduct experiments under three escalated configurations, i.e., (0.4, 0.7), (0.6, 0.8), and (0.8, 0.9), alongside the clean baseline (0, 0). HGMN consistently exhibits the slowest accuracy degradation across all eight datasets and still clearly outperforms all baselines under the most extreme setting; the corresponding figure and detailed analysis are provided in Appendix F.2.

4.3 Ablation Analysis

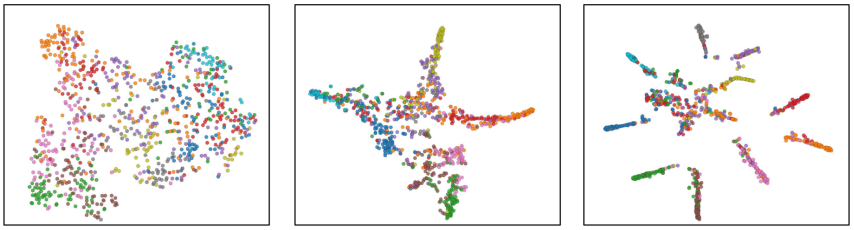
In this subsection, we conducted ablation studies to validate the effectiveness of three key components, i.e., the hypergraph-based evidential extraction framework (S1), the global mean-field negotiation mechanism (S2) and the multi-objective collaborative optimization strategy (S3). Specifically, the hypergraph is replaced by a standard GCN encoder (identical depth, hidden dimensions, and shortcut design) for evidence extraction in HGMN w/o S1, the global mean-field negotiation mechanism is substituted with a simple average fusion strategy in HGMN w/o S2, and the multi-objective optimization term is omitted from the overall loss function, with only $\mathcal{L}_{acc}$ retained in HGMN w/o S3. Furthermore, we performed ablation study on each module within the multi-objective optimization individually, with the experimental results detailed in Appendix F.4. As outlined in Table 3, we draw the following conclusions: 1) GCN can be considered a special case of HGCN under minimal hyperedge cardinality and HGCN performs competitively with or better than GCN across all datasets, confirming that hypergraph topology provides additional modeling capacity without sacrificing performance in the worst case; 2) Replacing the global mean-field negotiation module with a simple evidence averaging method leads to performance degradation, demonstrating that the global mean-field negotiation successfully isolates highly conflicting beliefs and serves as the primary robustness shield; 3) The complete HGMN model achieves the highest accuracy across most of the ablation experiments, demonstrating the effectiveness of the three modules working together to handle highly conflicting multiview classification.

Table 3: Accuracy (%) of HGMN and ablation models on highly conflicting test sets. '✓' and '−' represent the used and not used corresponding item, respectively.

S1	S2	S3	ALOI	Caltech101-20	COIL100	Leaves	NUS-WIDE	OutdoorScene	Scene15	VGGFace2
−	✓	✓	89.31±0.39	88.54±1.26	85.42±0.72	74.38±1.62	37.38±1.70	**82.19±1.55**	63.03±1.15	**52.73±0.44**
✓	−	✓	82.27±0.75	85.31±0.63	73.33±0.70	65.19±2.12	33.29±1.23	72.53±1.46	61.25±1.01	45.61±0.60
✓	✓	−	89.01±0.50	88.58±2.20	85.83±1.00	73.88±2.15	38.08±1.52	79.48±1.09	63.37±1.44	52.39±0.65
✓	✓	✓	**89.42±0.40**	**89.33±1.77**	**86.03±0.95**	**76.56±0.91**	**38.12±2.55**	81.56±1.73	**63.83±1.03**	52.68±0.51

4.4 Quantitative Analysis

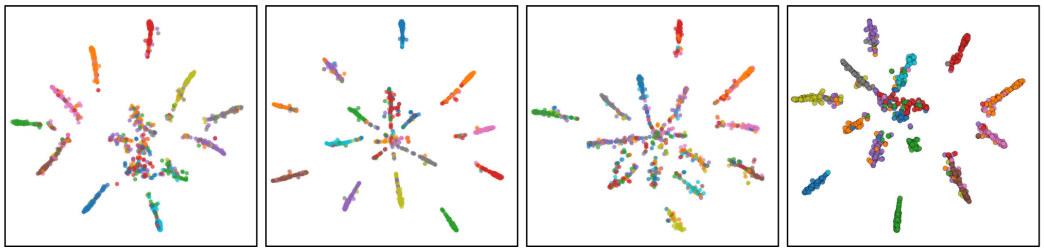
To compare feature representation quality, we first apply t-SNE to visualize the original feature space, the MLP-extracted features, and the hypergraph-extracted features for the first view of Scene15. As shown in Fig. 3, the features extracted by the hypergraph achieve a significant improvement over those extracted by an MLP, with much clearer separation between different categories.



(a) Raw Feature Space (b) MLP Feature Space (c) Hypergraph Feature Space

Figure 3: T-SNE visualization of feature space on the Scene15 dataset: (a) the original feature space, (b) the feature space extracted by MLP, and (c) the feature space extracted by the hypergraph.

Furthermore, we present a t-SNE visualization comparing single-view evidence and mean-field fused evidence on the Scene15 dataset. As illustrated in Fig. 4, after passing through the global mean-field negotiation module, the fused evidence exhibits excellent inter-class separability and intra-class compactness in the embedding space, validating that the mean-field mechanism effectively integrates complementary information from multiple sources.



(a) View 1 (b) View 2 (c) View 3 (d) Multiview Fusion

Figure 4: T-SNE visualization of evidence on the Scene15 dataset. (a) - (c) show the t-SNE results for evidence from three individual views. (d) illustrates the fused evidence after t-SNE.

In order to further demonstrate the advantages of our fusion mechanism. We have conducted a controlled experiment with the fusion mechanism as the sole variable. Both variants use the identical HGCN encoder for evidence extraction, and the only difference is that one employs the standard DST sequential pairwise combination while the other uses our global mean-field negotiation. All hyperparameters are kept identical. The results under the highly conflicting setting are summarized in Table 4. Across all eight datasets, our method consistently outperforms the DST-based fusion variant. On COIL100, DST achieves an Accuracy of 75.65% whereas ours reaches 86.03%; on Leaves, DST achieves an Accuracy of 67.75% whereas ours reaches 76.56%; on ALOI, DST achieves an Accuracy of 83.04% whereas ours reaches 89.42%.

Table 4: Accuracy (%) of HGMN and the DST-based fusion variant on highly conflicting test sets. The best results are highlighted by boldface.

	ALOI	Caltech101-20	COIL100	Leaves	NUS-WIDE	OutdoorScene	Scene15	VGGFace2
DST	83.04±0.68	86.49±0.79	75.65±0.90	67.75±2.17	32.88±2.62	77.17±1.24	60.58±1.64	45.54±0.81
Ours	**89.42±0.40**	**89.33±1.77**	**86.03±0.95**	**76.56±0.91**	**38.12±2.55**	**81.56±1.73**	**63.83±1.03**	**52.68±0.51**

To further demonstrate scalability, HGMN incurs training costs comparable to lightweight baselines and is 5–10× faster than TUNED and RCMCL, while scaling stably on large-scale datasets. Detailed results and analysis are provided in Appendix F.5.

5 Conclusion

In this paper, we address two critical limitations of existing DST-based multiview learning methods: the poor scalability and conflict sensitivity of DST's sequential pairwise combination rules, and the overlook of higher-order correlation structures in multiview features. To this end, we propose a novel robust multiview classification framework, termed HGMN. Specifically, it comprises a hypergraph convolutional network to capture intra-view higher-order correlations, a global mean-field negotiation mechanism to dynamically build consensus and isolate noise, and a multi-objective collaborative optimization strategy to enhance robustness and trustworthiness in open-world settings. However, real open-world scenarios expose models to entirely unseen novel classes. While the current framework can reject highly conflicting noise via elevated epistemic uncertainty, it lacks the explicit capability to isolate and discover unknown classes. In future work, we will incorporate open-set evidential theory to enable dynamic expansion of the probability simplex.

Acknowledgments and Disclosure of Funding

This work was supported by the National Natural Science Foundation of China (No. 62576206), the Zhejiang Provincial Natural Science Foundation of China (No. MS26F020024).

References

[1] F. Wang, Q. Zhu, D. Chang, Q. Gao, J. Han, T. Zhang, R. Hartley, and M. Pollefeys, "Learning-based multi-view stereo: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–20, 2026.

[2] Y. Huang, L. Ji, R. Xiao, and J. Sun, "Multi-view invariance learning for 3d scene graph pre-training via collaborative cross-modal regularization," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2026, pp. 5203–5211.

[3] H. Chen, B. Xu, S. Zhang, H. Liu, J. Lin, and J. Wang, "Vid-LLM: A compact video-based 3d multimodal LLM with reconstruction–reasoning synergy," in *Proceedings of the International Conference on Learning Representations*, 2026, pp. 1–16.

[4] Z. Yang, X. Tao, T. Cai, Y. Tang, H. Xie, L. Li, J. Li, and Q. Li, "A survey on multi-view knowledge graph: generation, fusion, applications and future directions," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2025, pp. 10 788–10 796.

[5] W. Dong, X.-J. Wu, Z. Feng, S. Atito, M. Awais, and J. Kittler, "Probabilistically aligned view-unaligned clustering with adaptive template selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 48, no. 2, pp. 1904–1916, 2026.

[6] H. Li, Y. Lin, P. Hu, M. Yang, and X. Peng, "Community-aware multi-view representation learning with incomplete information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 48, no. 3, pp. 3517–3529, 2026.

[7] X. Liang, P. Fu, Y. Qian, Q. Guo, and G. Liu, "Trusted multi-view classification via evolutionary multi-view fusion," in *Proceedings of the International Conference on Learning Representations*, 2025, pp. 1–14.

[8] C. Gong, Z. gang Su, and T. Denoeux, "Multi-view evidential k-nn classification," *Information Fusion*, vol. 120, p. 103113, 2025.

[9] Z. Zhao, X. Yang, C. Zhao, and C. Zhou, "Tmvf: Trusted multi-view fish behavior recognition with correlative feature and adaptive evidence fusion," *Information Fusion*, vol. 118, p. 102899, 2025.

[10] X. Wang, S. Duan, Q. Li, G. Duan, Y. Sun, and D. Peng, "Reliable disentanglement multi-view learning against view adversarial attacks," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2025, pp. 6469–6477.

[11] J. Zhang, Y. Yang, M. Shang, L. Guo, D. Zhang, L. Du, and for the Alzheimer's Disease Neuroimaging Initiative, "Toward trustworthy multi-view representation with fine-grained explainability embeddings," *IEEE Transactions on Medical Imaging*, vol. 45, no. 2, pp. 704–719, 2026.

[12] Z. Han, C. Zhang, H. Fu, and J. T. Zhou, "Trusted multi-view classification," in *Proceedings of the International Conference on Learning Representations*, 2021, pp. 1–16.

[13] M. Sensoy, L. Kaplan, and M. Kandemir, "Evidential deep learning to quantify classification uncertainty," in *Advances in Neural Information Processing Systems*, 2018, pp. 3183–3193.

[14] A. Jsang, *Subjective Logic: A Formalism for Reasoning Under Uncertainty (Artificial Intelligence: Foundations, Theory, and Algorithms Series)*. Springer, 2018.

[15] W. Liu, X. Yue, Y. Chen, and T. Denoeux, "Trusted multi-view deep learning with opinion aggregation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 7, 2022, pp. 7585–7593.

[16] H. Huang, Z. Liu, X. Han, X. Yang, and L. Liu, "A belief logarithmic similarity measure based on dempster-shafer theory and its application in multi-source data fusion," *Journal of Intelligent & Fuzzy Systems*, vol. 45, no. 3, pp. 4935–4947, 2023.

[17] C. Xu, J. Si, Z. Guan, W. Zhao, Y. Wu, and X. Gao, "Reliable conflictive multi-view learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, pp. 16 129–16 137.

[18] S. Duan, Y. Sun, D. Peng, G. Duan, X. Peng, and P. Hu, "Deep fuzzy multi-view learning for reliable classification," in *Proceedings of the International Conference on Machine Learning*, vol. 267, 2025, pp. 14 735–14 758.

[19] S. Hu, H. Huang, N. Zhang, and S. Sun, "Robust trusted conflictive multiview collaborative contrastive learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 48, no. 6, pp. 7161–7176, 2026.

[20] S. Xu, Y. Sun, X. Li, S. Duan, Z. Ren, Z. Liu, and D. Peng, "Noisy label calibration for multi-view classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025.

[21] C. Tang, Y. Shi, G. Lin, L. Xing, and L. Shi, "Trusted multi-view learning for long-tailed classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 40, no. 30, 2026, pp. 25 814–25 822.

[22] S. Kim, S. Y. Lee, Y. Gao, A. Antelmi, M. Polato, and K. Shin, "A survey on hypergraph neural networks: An in-depth and step-by-step guide," in *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2024, pp. 1–11.

[23] Y. Feng, H. You, Z. Zhang, R. Ji, and Y. Gao, "Hypergraph neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, pp. 3558–3565.

[24] N. Yadati, M. Nimishakavi, P. Yadav, V. Nitin, A. Louis, and P. Talukdar, "Hypergcn: A new method for training graph convolutional networks on hypergraphs," in *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[25] J. Huang and J. Yang, "Unignn: a unified framework for graph and hypergraph neural networks," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2021, pp. 2563–2569.

[26] P. Wang, S. Yang, Y. Liu, Z. Wang, and P. Li, "Equivariant hypergraph diffusion neural operators," in *Proceedings of the International Conference on Learning Representations*, 2023, pp. 1–15.

[27] J. Yan, Y. Feng, S. Ying, and Y. Gao, "Hypergraph dynamic system," in *Proceedings of the International Conference on Learning Representations*, 2024, pp. 1—12.

[28] Y. Gao, J. Yan, Y. Feng, X. Han, S. Ying, Z. Wu, and H. Hu, "Hgnnv2: Stable hypergraph neural networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1—17, 2026.

[29] I. Duta, G. Cassarà, F. Silvestri, and P. Lió, "Sheaf hypergraph networks," in *Advances in Neural Information Processing Systems*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., vol. 36, 2023, pp. 12 087–12 099.

[30] X. Li, T. Zhang, J. Wang, Z. Han, N. Wang, S. Fan, H. Du, J. Kang, J. Liu, and D. Niyato, "Protect ntn-iot security by malicious traffic detection: A multidimensional hypergraph learning approach," *IEEE Internet of Things Journal*, vol. 13, no. 3, pp. 3594—3607, 2026.

[31] Q. Guo, Y. Wang, Y. Zhang, H. Qi, Y. Hu, and Y. Jiang, "Hyper-bts: Brain tumor segmentation based on hypergraph guidance," *Pattern Recognition*, vol. 169, p. 111926, 2026.

[32] Y. Shan, L. Gao, Y. Lin, Z. Yang, G. Lyu, and H. Han, "Hypergraph-based multi-view multi-label classification via adaptive high-order semantic fusion," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 40, no. 30, 2026, pp. 25 269–25 277.

[33] Z. Gu *et al.*, "Hypergraph-enhanced contrastive learning for multi-view clustering with hyper-laplacian regularization," in *Advances in Neural Information Processing Systems*, vol. 38, 2025, pp. 54 600–54 620.

[34] S. Yang, K. Cui, Y. Jia, Y. Ding, R. Guan, and Y. Ding, "Superpixel-level hypergraph regularized multi-view clustering for hyperspectral images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2026.

[35] Z. Han, C. Zhang, H. Fu, and J. T. Zhou, "Trusted multi-view classification with dynamic evidential fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 2551–2566, 2023.

[36] X. Liang, S. Li, Q. Guo, Y. Qian, B. Jiang, T. Luo, and L. Du, "Improving evolutionary multi-view classification via eliminating individual fitness bias," in *Advances in Neural Information Processing Systems*, vol. 38, 2025, pp. 143 388–143 412.

[37] C. Zhang, X. Zhu, Z. Wang, Y. Zhong, W. Sheng, W. Ding, and B. Jiang, "Discriminative multi-view fusion via adaptive regression," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 8, no. 6, pp. 3821–3833, 2024.

[38] H. Huang, C. Qin, Z. Liu, K. Ma, J. Chen, H. Fang, C. Ban, H. Sun, and Z. He, "Trusted unified feature-neighborhood dynamics for multi-view classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025, pp. 17 413–17 421.

[39] L. Shi, C. Tang, H. Deng, C. Xu, L. Xing, and B. Chen, "Generalized trusted multi-view classification framework with hierarchical opinion aggregation," *IEEE Transactions on Multimedia*, pp. 1–14, 2026.

[40] Y. Liu, L. Liu, C. Xu, X. Song, Z. Guan, and W. Zhao, "Dynamic evidence decoupling for trusted multi-view learning," in *Proceedings of the ACM International Conference on Multimedia*, 2024, pp. 7269–7277.

[41] H. Zhou, Z. Xue, Y. Liu, B. Li, J. Du, M. Liang, and Y. Qi, "Calm: An enhanced encoding and confidence evaluating framework for trustworthy multi-view learning," in *Proceedings of the ACM International Conference on Multimedia*, 2023, pp. 3108—3116.

[42] C. Xu, Y. Zhang, Z. Guan, and W. Zhao, "Trusted multi-view learning with label noise," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2024, pp. 5263–5271.

[43] X. Yue, Z. Dong, Y. Chen, and S. Xie, "Evidential dissonance measure in robust multi-view classification to resist adversarial attack," *Information Fusion*, vol. 113, p. 102605, 2025.

[44] Z. Dong, X. Yue, Y. Chen, and Y. Zhou, "Trusted open-world multi-view classification with dynamic opinion aggregation," in *Proceedings of the ACM International Conference on Multimedia*, 2025, pp. 1181–1189.

[45] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," in *Advances in Neural Information Processing Systems*, vol. 33, 2020, pp. 18 661–18 673.

[46] K. Huang, *Statistical mechanics*. John Wiley & Sons, 2008.

[47] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American Statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[48] G.-Y. Zhang, D. Huang, and C.-D. Wang, "Facilitated low-rank multi-view subspace clustering," *Knowledge-Based Systems*, vol. 260, p. 110141, 2023.

[49] X. Li, H. Zhang, R. Wang, and F. Nie, "Multiview clustering: A scalable and parameter-free bipartite graph fusion method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 330–344, 2022.

[50] X. Hu, X. Liu, W. Pedrycz, Q. Liao, Y. Shen, Y. Li, and S. Wang, "Multi-view fuzzy classification with subspace clustering and information granules," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 11, pp. 11 642–11 655, 2023.

[51] J. Yang and C.-T. Lin, "Multi-view adjacency-constrained hierarchical clustering," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 7, no. 4, pp. 1126–1138, 2023.

[52] B. Wang, Y. Xiao, Z. Li, X. Wang, X. Chen, and D. Fang, "Robust self-weighted multi-view projection clustering," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, pp. 6110–6117.

[53] S.-G. Fang, D. Huang, X.-S. Cai, C.-D. Wang, C. He, and Y. Tang, "Efficient multi-view clustering via unified and discrete bipartite graph learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 8, pp. 11 436–11 447, 2024.

[54] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2004, pp. 178–178.

[55] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *Proceedings of the IEEE International Conference on Automatic Face Gesture Recognition*, 2018, pp. 67–74.

[56] X. Liang, S. Wang, Y. Qian, Q. Guo, L. Du, B. Jiang, T. Luo, and F. Li, "Trusted multi-view classification with expert knowledge constraints," in *Proceedings of the International Conference on Machine Learning*, 2025, pp. 37 409–37 426.

A KL Divergence Loss Function

The Kullback-Leibler (KL) divergence loss function $\mathcal{L}_{KL}\left(\alpha_n^v\right)$ is defined as follows:

$$\begin{aligned}\mathcal{L}_{KL}\left(\alpha_n^v\right) &= KL\left[D\left(p_n^v|\tilde{\alpha}_n^v\right) \parallel D\left(p_n^v|\mathbf{1}\right)\right] \\ &= \log\left(\frac{\Gamma\left(\sum_{k=1}^K \tilde{\alpha}_{nk}^v\right)}{\Gamma(K)\prod_{k=1}^K \Gamma\left(\tilde{\alpha}_{nk}^v\right)}\right) + \sum_{k=1}^K \left(\tilde{\alpha}_{nk}^v - 1\right)\left[\psi\left(\tilde{\alpha}_{nk}^v\right) - \psi\left(\sum_{j=1}^K \tilde{\alpha}_{nj}^v\right)\right],\end{aligned} \tag{18}$$

where α_n^v are the predicted Dirichlet parameters for the v-th view of the n-th sample and $\tilde{\alpha}_n^v = y_n + (1 - y_n) \odot \alpha_n^v$ are the Dirichlet parameters after removal of the non-misleading evidence. $D\left(p_n^v|\mathbf{1}\right)$ is the uniform Dirichlet distribution, $\Gamma(\cdot)$ is the gamma function.

B Computational Complexity Analysis

In this section, we provide a computational complexity analysis of the proposed HGMN framework. Let V be the number of views, B be the mini-batch size, d be the feature dimension, K be the number of classes, m be the number of nearest neighbors used in the m-NN search, and C be the number of global semantic attractors.

Training Time Complexity. The training procedure of HGMN primarily consists of three stages for each mini-batch, i.e., hypergraph construction and convolution, evidence generation and mean-field negotiation, and multi-objective loss computation.

1. For hypergraph construction and convolution, constructing the intra-batch m-NN graph by computing pairwise cosine similarities requires $\mathcal{O}(VB^2d)$. Assigning B samples to C global attractors takes $\mathcal{O}(VBCd)$. Crucially, during hypergraph convolution, explicitly computing the dense transition matrix $\mathbf{G}$ would incur a prohibitive cost of $\mathcal{O}(B^2)$. To ensure scalability, we implement the convolution using sparse matrix multiplication associativity, i.e., $\mathbf{H}(\mathbf{W}(\mathbf{H}^\top\mathbf{Z}))$. Since the incidence matrix $\mathbf{H}$ contains at most $B(m+1)$ non-zero entries, the convolution complexity is strictly bounded by $\mathcal{O}(VB(m+C)d)$.
2. For evidence generation and mean-field negotiation, mapping the higher-order features to Dirichlet parameters takes $\mathcal{O}(VBK)$. The tensor-parallelized mean-field negotiation mechanism computes topological consensus scores, calibrates the global mean-field, and dynamically updates the uncertainty u. Unlike traditional DST, which suffers from $\mathcal{O}(VK^2)$ or even exponential $\mathcal{O}(2^K)$, our mean-field protocol operates in strict linear time $\mathcal{O}(VBK)$.
3. For multi-objective loss computation, the supervised contrastive loss $\mathcal{L}_{\text{con}}$ calculates pairwise similarities among $V \cdot B$ evidential vectors, requiring $\mathcal{O}(V^2B^2K)$. The JS divergence $\mathcal{L}_{\text{JS}}$, variance penalty $\mathcal{L}_{\text{var}}$, and classification loss $\mathcal{L}_{\text{acc}}$ are element-wise operations bounded by linear time $\mathcal{O}(V^2BK)$ and $\mathcal{O}(VBK)$.

Overall, the worst-case training time complexity per mini-batch is dominated by $\mathcal{O}(VB^2(d+VK))$. Since the batch size B is a small constant in practice, the entire training process maintains exceptional efficiency. Considering a total of N training samples, T iterations, and a mini-batch size B, the overall complexity of the entire training process is $\mathcal{O}(VNB(d+VK)T)$, which is linear with respect to the number of training samples, i.e., $\mathcal{O}(N)$.

Inference Time Complexity. A paramount advantage of HGMN is its late-stage modular decoupling. During the inference phase for an unseen sample, the model completely bypasses the computationally expensive intra-batch m-NN graph construction and the pairwise contrastive loss computation. The raw multiview features are directly mapped to subjective evidence, followed by the lightweight mean-field negotiation. Thus, the inference time complexity for a single sample decreases dramatically to strictly $\mathcal{O}(V(d+K))$. This ultra-low latency enables HGMN to be effortlessly deployed in high-throughput, real-time open environments.

Space Complexity. During the training phase, the peak GPU memory footprint predominantly stems from the multiview feature representations, the hypergraph incidence matrices, and the intermediate dense computational graphs generated during multi-objective optimization. Specifically, storing

the feature representations requires $\mathcal{O}(VBd)$. The highly sparse hypergraph incidence matrices $\mathbf{H}$ occupy merely $\mathcal{O}(V \cdot B \cdot (m + C))$. However, computing $\mathcal{L}_{con}$ and the m-NN similarities necessitates the construction of temporary dense matrices, occupying $\mathcal{O}(V^2B^2)$. Consequently, the overall training space complexity is bounded by $\mathcal{O}(VBd+V^2B^2+VBK)$. Owing to B being a local constant (i.e., independent of the total number of training samples N) and the intrinsic topological sparsity of the hypergraph, HGMN effectively circumvents the Out-of-Memory (OOM) bottlenecks typically encountered by traditional fully-connected graph neural networks.

In order to make a more vivid comparison of the computational overhead, we report the total training time (seconds) in Table 5. HGMN incurs training costs comparable to lightweight baselines (e.g., TMC, ECML) and is 5 to 10 times faster than TUNED and RCMCL, while maintaining stable scaling on large-scale datasets such as VGGFace2 and ALOI. This confirms that our hypergraph construction and mean-field negotiation introduce negligible computational overhead.

Table 5: Training time (seconds) comparison across all datasets.

Dataset	TMC	ECML	CCML	TUNED	TMCEK	GTMC-HOA	RCMCL	Ours
ALOI	325.54	312.43	300.39	1498.10	469.72	573.64	11933.33	473.28
Caltech101-20	142.31	135.45	130.18	883.40	204.77	210.14	1185.14	199.63
COIL100	154.49	170.54	148.14	830.20	252.71	287.98	5530.03	254.16
Leaves	38.48	35.6	32.35	253.68	52.4	72.12	1319.6	51.08
NUS-WIDE	80.83	91	75.46	853.36	143.89	147.13	679.17	146.62
OutdoorScene	95.4	91.83	84.75	636.99	131.42	157.25	479.35	146.5
Scene15	87.31	103.59	88.51	923.05	145.68	163.12	713.32	156.16
VGGFace2	1434.16	1417.73	1320.68	14813.45	1992.12	2147.31	19731.44	1939.79

C Additional Discussion

We use hypergraphs for the following reason. Firstly, hypergraphs are inherently suitable for group-level noise correction. Under noise, corrupted samples rarely occur in isolation. Instead, they form localized clusters where semantically similar samples within the same class are affected simultaneously. While a standard pairwise graph corrects such noise incrementally by propagating information between two nodes at a time, a hypergraph connects multiple nodes via a single hyperedge and aggregates their collective signals. This allows the model to identify and discount entire corrupted clusters at once, rather than adjusting them neighbor by neighbor. Secondly, our specific hypergraph construction provides dual safeguards against noise. Local kNN hyperedges capture fine-grained manifold structure to preserve local neighborhood integrity, while global semantic attractors anchor the overall data distribution. Together, they prevent the model from overfitting to noisy local patterns.

The noise robustness of our hypergraph stems from two distinct levels. Firstly, during construction, the m-nearest-neighbor selection with $L2$ normalization implicitly suppresses noisy feature dimensions by favoring locally consistent samples and rejecting outliers. Secondly, during spectral convolution, the influence of a corrupted node is diluted by the collective signal of other nodes within the same hyperedge, which inherently reduces the impact of view-specific noise injected into individual samples. The pairwise graph is vulnerable to noise contamination due to its lack of collective correction, which results in the propagation of corrupted signals directly to its neighbors.

D Datasets and Comparison Methods

As shown in Table 6, we perform experiments on eight public datasets, including ALOI, Caltech101-20, COIL100, Leaves, NUS-WIDE, OutdoorScene, Scene15 and VGGFace2. Among these datasets, the number of views varies from 3 to 6, the number of categories varies from 8 to 101, and the number of samples varies from 1600 to 34027. The details of them are summarized as follows.

- ALOI [48], a subset of the Amsterdam Library of Object Images, comprises 10800 images of 100 small objects taken under diverse illumination and rotation angles.
- Caltech101-20 [49] provides a commonly used subset of 20 image categories specifically built for object recognition tasks.

- COIL100 [50] collects 7200 images of 100 categories. Three heterogeneous features, including isometric projection, linear discriminant analysis, and neighborhood preserving embedding, are used as multiple views.
- Leaves [51] contains sixteen leaf samples per species, covering a total of one hundred plant species.
- NUS-WIDE [52] subset consists of 12 categories of animal images selected from the NUS-WIDE-OBJ dataset, with the first 200 images chosen for each category.
- OutdoorScene [53] dataset contains 2688 outdoor scene images, which are divided into 8 groups.
- Scene15 [54] consists of 4,485 images spanning 15 indoor and outdoor scene categories, with the number of images per category ranging from 200 to 400.
- VGGFace2 [55] is a large-scale face recognition dataset introduced by the Visual Geometry Group at the University of Oxford. All images were collected from Google Image Search and cover large variations in pose, age, illumination, ethnicity, and profession.

We compare HGMN with seven state-of-the-art multiview learning methods, summarized as follows.

- Trusted Multiview Classification (TMC) [12] models the class probability distribution using a Dirichlet distribution and incorporates DST to dynamically fuse multiview information at the evidence level. By training the network with an improved loss function, it obtains evidence from each view and estimates uncertainty, thereby achieving highly reliable and robust multiview classification.
- Evidential Conflictive Multiview Learning (ECML) [17] proposes a conflicting opinion aggregation strategy to fuse multiview opinions and introduces a conflict degree loss to ensure view consistency during training, thereby achieving accurate decision-making and reliable uncertainty estimation on conflicting multiview data.
- Consistent and Complementary-aware trusted Multiview Learning (CCML) [40] dynamically decouples the consistent evidence and complementary evidence from each view. It imposes a strict alignment constraint with the ground-truth labels on the consistent evidence, while only requiring the complementary evidence to reflect the true class probabilities, thereby allowing semantic ambiguity.
- Trusted Unified feature NEighborhood Dynamics (TUNED) [38] extracts the local feature-neighborhood structure of each view using a view-specific deep neural network and a graph convolutional network constructed on a compatibility adjacency network . A selective Markov random field adaptively handles cross-view neighborhood dependencies, suppresses low-consistency views and aggregates the local and global evidence.
- Trusted Multiview Classification Constrained with Expert Knowledge (TMCEK) [56] employs Gabor functions embedded with expert knowledge as convolutional kernels during the feature extraction stage to process multiview signals in both the time domain and the time-frequency domain, thereby enhancing feature interpretability. In the decision-making stage, it introduces a distribution-aware subjective opinion mechanism, where uncertainty is quantified by combining the total evidence mass with the evidence distribution characterized by the Gini coefficient.

Table 6: Dataset summary

Dataset	#Samples	#Categories	#Dimensionality
ALOI	10800	100	77/13/64/125
Caltech101-20	2386	20	48/40/254/1984/512/928
COIL100	7200	100	30/99/30
Leaves	1600	100	64/64/64
NUS-WIDE	2400	12	64/144/73/128/225
OutdoorScene	2688	8	512/432/256/48
Scene15	4485	15	20/59/40
VGGFace2	34027	50	944/576/512/640

- Generalized Trusted Multiview Classification framework with Hierarchical Opinion Aggregation (GTMC-HOA) [39] extracts common information from each view by minimizing a confusion adversarial loss and captures view-specific information through orthogonal constraints. It then performs intra-view aggregation based on DST to reduce evidence uncertainty and improve view quality. In the inter-view aggregation stage, an evidence level attention mechanism is designed to capture the importance weights of different views and a conflicting opinion aggregation strategy is adopted to fuse multiview evidence.
- Robust Trusted Conflictive Multiview Collaborative Contrastive Learning (RCMCL) [19] enhances inter-view consistency using discordance-based evidential contrastive learning and it introduces vacuity dynamically weighted complementary evidence, combines category-level contrastive learning to separate consistent and complementary evidence and finally fuses both types of evidence with an evidence gradient penalty to make joint decisions.

E Experimental Settings

Compute Resources. In our experiments, all methods are implemented using PyTorch 2.5.1. The computing environment uses Windows 11 with Python 3.9.21 and CUDA 11.8. Two different GPU setups are employed depending on the dataset type: for normal test datasets, we use an NVIDIA GeForce RTX 4060 Ti (8 GB VRAM) with an Intel i7-12700KF CPU and 32 GB of RAM; for highly conflicting test datasets, we use an NVIDIA GeForce RTX 3090 (24 GB VRAM) with an Intel i7-13700K CPU and 128 GB of RAM.

Parameter Settings. For hypergraph parameters (i.e., C, m), the search ranges were determined as follows. The number of global semantic attractors C is searched over $\{4, 8, 12, 16, 20, 24, 28\}$ to cover the class-count spectrum of eight datasets (i.e., 5 to 100 classes) without over-fragmenting the mini-batch. The number of kNN neighbors m is searched over $\{5, 10, 15, 20, 25\}$ to balance hypergraph sparsity and local manifold capture, satisfying $m < B$ across all batch sizes. For standard deep learning parameters (i.e., λ, γ, μ, B), the search ranges were determined as follows. The contrastive weight λ and JS weight γ share the same range [0.1, 2.0], following the common practice that loss weights in multi-objective optimization are typically explored on a log-spaced grid centered at 1.0. The variance penalty μ uses a smaller range [0.01, 0.1] since its magnitude is inherently smaller than the primary loss terms. The batch size B share the same range [32, 400] covers the practical range bounded by GPU memory and the minimum requirement for meaningful kNN graph construction ($B > m$). Fig. 5 of Appendix E.1 demonstrates that HGMN is robust to hyperparameter variations and does not rely on aggressive per-dataset tuning to achieve strong performance.

F Further Experiments

F.1 Performance Comparison

We have extended the evaluation to include Precision and F1 score across all eight datasets under both normal and highly conflicting settings. As shown in Tables 7 and 8, HGMN achieves the best performance across all three metrics. Under the highly conflicting setting, HGMN outperforms the strongest baseline (i.e., RCMCL) by over 10 percentage points across Accuracy, Precision, and F1 score on datasets such as COIL100, ALOI, and Leaves, with consistent margins on the remaining datasets. These results confirm that HGMN's robustness is metric-agnostic and not an artifact of a single evaluation criterion.

F.2 Robustness Analysis

In order to further demonstrate the robustness of our method. We have accordingly conducted experiments under three escalated configurations, i.e. (0.4, 0.7), (0.6, 0.8), and (0.8, 0.9), alongside the clean baseline (0, 0). The test-set corruption protocol is as follows: (1) Gaussian noise with a standard deviation of $\sigma = 1.0$ is injected into random views of selected samples with probability r_{noise}; (2) cross-class feature replacement replaces one view's feature with the same view's feature from the next class with probability $r_{unaligned}$. As shown in Fig. 5, HGMN consistently exhibits the slowest accuracy degradation across all eight datasets. On COIL100, HGMN drops from 98.89% to

Table 7: Performance comparisons between HGMN and comparison methods on normal test sets. The best and the second best results are highlighted by boldface and underlined respectively.

Dataset	TMC	ECML	CCML	TUNED	TMCEK	GTMC-HOA	RCMCL	ours
				Accuracy(%)				
ALOI	71.14±0.34	69.32±0.73	61.75±1.27	85.08±1.55	69.61±0.82	**97.56±0.48**	90.69±0.54	97.11±0.41
Cal101-20	92.05±0.82	92.64±1.10	93.60±0.72	93.72±0.70	92.09±0.87	92.22±1.49	**94.39±0.64**	94.27±0.85
COIL100	96.67±0.68	98.01±0.45	96.47±0.48	97.89±0.34	97.82±0.49	97.03±0.95	98.60±0.19	**98.89±0.25**
Leaves	34.50±3.04	94.12±1.03	96.56±0.20	90.62±1.56	93.87±0.42	97.31±0.42	97.00±0.92	**97.88±0.70**
NUS-WIDE	36.54±0.96	39.37±1.89	45.29±1.88	41.37±1.96	38.62±0.98	43.67±0.50	40.87±1.78	**45.71±0.81**
OutdoorScene	82.42±0.67	84.28±1.13	84.50±0.98	85.13±1.47	85.32±0.76	88.14±0.91	84.65±1.39	**88.77±1.02**
Scene15	61.11±2.09	70.68±1.09	71.17±1.30	73.24±0.62	71.53±0.82	77.37±1.06	73.53±1.10	**78.06±0.33**
VGGFace2	63.36±0.29	60.81±0.41	62.04±0.32	62.02±0.62	62.76±0.31	58.64±0.80	64.39±0.23	**64.62±0.21**
				Precision(%)				
ALOI	69.54±1.70	65.92±0.70	56.97±0.90	87.91±0.95	67.79±0.96	**97.61±0.45**	91.65±0.37	97.12±0.40
Cal101-20	89.80±0.77	89.19±1.49	92.13±1.38	**92.39±1.22**	90.61±1.82	88.89±2.28	91.96±1.92	91.57±1.70
COIL100	96.63±0.71	97.96±0.46	96.70±0.42	97.88±0.40	97.82±0.49	97.21±0.82	98.67±0.21	**98.86±0.28**
Leaves	29.25±2.85	94.13±0.90	96.64±0.31	90.77±0.66	93.81±1.29	97.13±0.40	96.94±1.22	**97.96±0.80**
NUS-WIDE	35.30±0.92	38.17±2.20	44.53±2.33	41.50±2.92	37.79±1.32	43.68±0.76	39.49±2.01	**44.56±0.99**
OutdoorScene	82.79±0.56	84.67±1.27	84.74±1.16	85.49±1.72	85.67±0.99	88.24±0.83	84.76±1.31	**88.95±1.11**
Scene15	64.47±4.45	71.20±1.90	71.67±1.10	73.61±1.56	71.90±2.03	76.72±0.76	73.53±1.08	**77.31±0.44**
VGGFace2	63.96±0.29	62.95±0.25	**65.50±0.65**	64.80±0.44	65.36±0.22	63.27±0.88	65.30±0.23	65.24±0.25
				F1 score(%)				
ALOI	66.40±0.23	64.15±0.72	54.15±1.05	84.91±1.35	64.57±0.81	**97.53±0.46**	90.32±0.58	97.03±0.42
Cal101-20	82.60±1.79	83.95±2.59	86.37±1.38	86.02±1.62	82.54±2.15	83.33±3.09	**87.55±2.25**	87.28±2.51
COIL100	96.38±0.79	97.89±0.47	96.19±0.41	97.78±0.39	97.69±0.52	96.97±0.93	98.47±0.25	**98.82±0.29**
Leaves	28.07±2.20	93.02±1.01	96.30±0.36	88.94±1.15	92.75±1.04	96.74±0.47	96.48±1.36	**97.60±0.81**
NUS-WIDE	33.23±0.46	37.61±1.95	43.97±2.09	39.68±2.37	37.01±0.80	43.15±0.69	39.18±2.09	**44.44±0.78**
OutdoorScene	82.60±0.65	84.38±1.07	84.49±1.10	85.18±1.50	85.32±0.79	88.22±0.73	84.68±1.20	**88.90±0.98**
Scene15	55.37±2.68	67.67±1.66	68.26±1.59	70.25±0.81	68.37±1.46	76.35±0.80	70.79±0.55	**76.53±0.57**
VGGFace2	63.14±0.26	60.81±0.50	62.33±0.47	61.88±0.48	62.90±0.31	60.01±0.88	64.16±0.17	**64.41±0.19**

Table 8: Performance comparisons between HGMN and comparison methods on highly conflicting test sets.

Dataset	TMC	ECML	CCML	TUNED	TMCEK	GTMC-HOA	RCMCL	ours
				Accuracy(%)				
ALOI	59.05±0.91	57.69±0.97	50.60±0.72	69.26±0.60	58.03±1.19	77.82±0.43	77.15±0.81	**89.42±0.40**
Cal101-20	85.40±1.24	82.55±1.11	86.36±0.85	83.85±0.83	82.51±1.11	83.81±1.74	88.58±0.54	**89.33±1.77**
COIL100	69.94±1.13	69.65±0.95	70.43±0.94	66.54±1.23	70.33±0.81	70.49±1.09	74.64±0.74	**86.03±0.95**
Leaves	26.00±1.33	62.87±1.76	64.62±1.97	59.31±3.01	62.56±1.80	65.00±2.25	65.50±1.66	**76.56±0.91**
NUS-WIDE	31.42±1.16	32.29±1.23	36.00±0.61	33.50±2.33	32.12±1.50	36.33±2.03	36.92±0.98	**38.12±2.55**
OutdoorScene	67.06±0.54	68.92±1.14	69.29±0.80	67.55±0.89	69.78±1.66	70.04±1.38	74.39±1.60	**81.56±1.73**
Scene15	45.64±1.69	53.96±1.49	52.91±1.76	53.76±1.59	54.89±1.56	57.95±1.58	57.32±1.09	**63.83±1.03**
VGGFace2	43.95±0.39	43.88±0.32	45.49±0.69	43.75±0.26	45.14±0.44	40.10±4.03	48.76±0.32	**52.68±0.51**
				Precision(%)				
ALOI	59.70±2.04	54.74±1.20	46.44±0.91	73.71±0.58	55.77±1.66	79.36±0.62	80.28±1.20	**90.02±0.40**
Cal101-20	75.76±3.46	74.85±1.76	79.74±2.70	78.64±2.34	76.02±3.35	75.51±2.43	80.56±2.38	**82.47±3.00**
COIL100	72.51±0.82	72.13±0.91	72.47±0.73	71.12±1.18	72.65±0.67	69.45±1.74	76.86±0.45	**87.85±0.85**
Leaves	21.65±1.44	63.42±2.92	65.06±2.57	60.32±2.48	63.16±4.11	65.57±3.30	66.07±2.36	**76.73±0.57**
NUS-WIDE	30.90±2.07	30.44±1.15	35.25±0.77	33.41±3.39	29.78±2.27	35.71±2.34	36.20±2.05	**37.16±2.65**
OutdoorScene	67.45±0.41	69.02±1.46	69.54±0.84	68.01±1.23	70.37±2.05	71.08±1.62	74.65±1.47	**82.14±1.55**
Scene15	47.43±2.59	53.75±1.62	52.18±2.13	55.65±2.22	54.81±2.74	59.04±1.21	57.17±0.74	**62.90±0.89**
VGGFace2	44.23±0.51	44.88±0.43	46.79±0.48	45.08±0.17	46.39±0.17	43.15±4.02	49.16±0.22	**52.97±0.61**
				F1 score(%)				
ALOI	54.83±0.87	52.65±0.87	43.40±0.66	68.64±0.60	53.27±1.06	77.79±0.48	76.62±0.76	**89.37±0.41**
Cal101-20	70.77±2.79	69.54±1.76	73.17±1.87	71.16±1.32	68.12±2.37	67.62±1.62	75.79±2.04	**76.65±3.33**
COIL100	69.36±1.29	68.64±1.06	69.44±0.99	65.14±1.65	69.18±1.05	68.05±1.09	73.77±0.79	**85.43±0.98**
Leaves	20.45±0.51	59.26±2.39	60.78±2.02	55.46±2.92	58.87±2.75	61.07±2.39	61.49±2.42	**73.48±1.32**
NUS-WIDE	28.19±1.57	30.22±1.15	34.69±0.60	31.38±2.53	29.89±2.00	35.31±2.34	34.84±1.46	**36.68±2.49**
OutdoorScene	66.73±0.62	68.62±1.35	69.00±0.87	67.34±1.11	69.32±1.83	70.06±1.46	74.12±1.50	**81.78±1.71**
Scene15	40.90±1.63	50.74±1.41	49.95±2.01	51.66±1.72	51.58±1.71	57.05±1.51	54.35±0.78	**62.01±0.65**
VGGFace2	43.39±0.45	43.43±0.38	44.90±0.75	43.14±0.28	44.66±0.45	40.15±4.27	48.25±0.35	**52.27±0.54**

65.72% at the most extreme level, while the best baseline RCMCL falls to 52.04%. Even under the extreme setting (0.8, 0.9), HGMN still maintains a clear lead over all baselines.

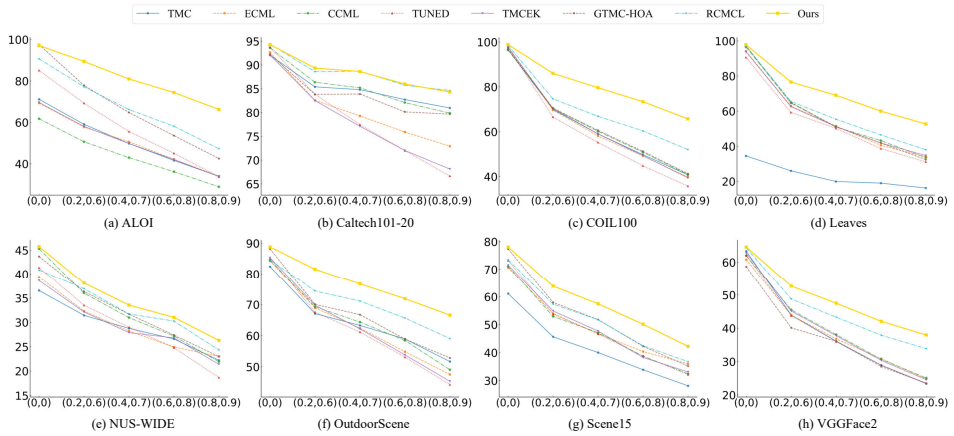


Figure 5: Accuracy (%) of HGMN and comparison methods across varying conflict ratios.

F.3 Parameter Sensitivity Analysis

To validate the stability of the HGMN framework under various configurations, we conduct a comprehensive parameter sensitivity analysis across all 8 benchmark datasets. The evaluation encompasses 6 critical hyperparameters grouped into three dimensions: multi-objective optimization weights (λ, γ, μ), training batch-size (B) and hypergraph construction parameters (m, C). The quantitative results are illustrated in Fig. 6. We draw the following conclusions: 1) λ, γ and μ are the supervised contrastive learning parameter, the uncertainty-guided weighted JS consistency learning parameter and the evidential variance penalty parameter respectively. The accuracy curves for all datasets remain almost perfectly horizontal. The intrinsic global mean-field negotiation mechanism adaptively balances the gradient flows among views, rendering the joint optimization highly stable; 2) For the more complex VGGFace2 dataset, the accuracy demonstrates a consistent upward trajectory as B increases from 32 to 400. This phenomenon rigorously aligns with the theoretical consensus in supervised contrastive learning: larger batch sizes implicitly provide a richer and more diverse set of negative samples for $\mathcal{L}_{con}$, which compels the underlying feature encoder to learn a more discriminative and compact feature manifold; 3) m and C the local kNN parameter and the global clustering parameter. The experiments indicate that even when the local and global connectivities vary multi-fold, the classification performance exhibits no significant fluctuations. This substantiates that our "local-global" dual-stage hypergraph construction is structurally robust, consistently capturing high-order manifold smoothness without being overly sensitive to specific topological configurations. HGMN exhibits a striking insensitivity to most hyperparameters. This extensive parameter robustness endows the proposed framework with immense practical value and high reliability for deployment in real-world open environments.

F.4 Ablation Study of Multi-Objective Optimization

To validate the necessity and synergistic effects of the individual loss functions within our proposed multi-objective joint optimization framework, we conducted a rigorous ablation study on the supervised contrastive learning $\mathcal{L}_{con}$, the uncertainty-guided weighted JS consistency learning $\mathcal{L}_{JS}$, and the evidential variance penalty $\mathcal{L}_{var}$ across all eight datasets. As shown in Table 9, drawn the following conclusions: 1) The full HGMN model integrating all three loss functions achieves the highest classification accuracy on almost all datasets. This compellingly demonstrates that the joint optimization strategy—spanning from bottom-level feature manifold constraints to top-level cognitive boundary regularization—produces a highly complementary and synergistic effect; 2) Removing $\mathcal{L}_{con}$ consistently degrades performance across all datasets, indicating that lacking low-level supervised contrastive constraints leads to insufficient discriminability in the latent space. Similarly, discarding $\mathcal{L}_{JS}$ results in noticeable performance attrition, corroborating that forcing view-specific distributions

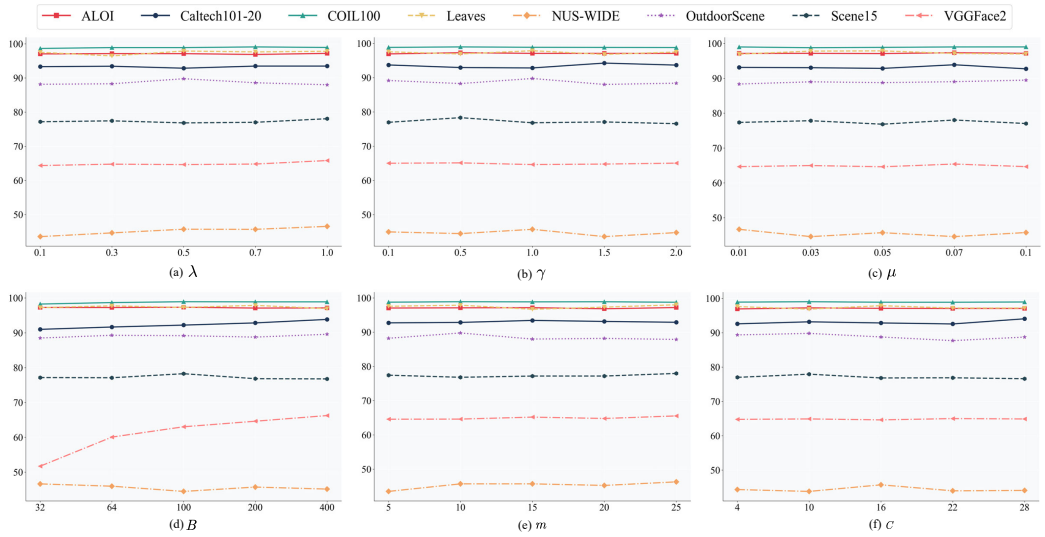


Figure 6: The line chart of accuracy varying with different values of λ, γ, μ, B, m and C.

to align with the global mean-field consensus is indispensable for resolving cross-view semantic conflicts; 3) Eliminating $\mathcal{L}_{var}$ causes a performance drop on seven out of eight datasets, proving that the evidential variance penalty effectively suppresses the generation of false evidence and prevents overconfidence in open environments. On the highly challenging fine-grained dataset VGGFace2, removing $\mathcal{L}_{var}$ yields a marginal performance rebound. In extremely difficult classification tasks, a stringent evidential variance constraint may slightly restrict the model's fitting capacity. Nevertheless, the substantial regularization benefits are provided by $\mathcal{L}_{var}$ in the vast majority of scenarios.

Table 9: Performance comparisons of HGMN and ablation models of multi-objective optimization on highly conflicting test sets. '✓' and '−' represent the used and not used corresponding item, respectively.

$\mathcal{L}_{con}$	$\mathcal{L}_{JS}$	$\mathcal{L}_{var}$	ALOI	Caltech101-20	COIL100	Leaves	NUS-WIDE	OutdoorScene	Scene15	VGGFace2
−	✓	✓	89.01±0.67	88.95±1.10	85.86±1.03	74.50±2.31	37.54±1.94	79.81±1.47	63.14±1.15	52.38±0.12
✓	−	✓	89.41±0.39	89.16±1.63	85.92±0.94	74.88±1.06	38.10±1.11	79.85±1.93	63.68±0.78	53.05±0.53
✓	✓	−	89.30±0.31	88.74±1.07	85.96±1.09	75.38±1.52	38.00±1.32	79.55±0.93	63.55±0.68	**53.08±0.48**
✓	✓	✓	**89.42±0.40**	**89.33±1.77**	**86.03±0.95**	**76.56±0.91**	**38.12±2.55**	**81.56±1.73**	**63.83±1.03**	52.68±0.51

On the highly challenging fine-grained dataset VGGFace2, after removing the evidence variance penalty, the accuracy actually showed a very slight rebound (increasing slightly from 52.68% to 53.08%). From this phenomenon, we derive the following insights: (1) The rebound manifests specifically on VGGFace2 due to its fine-grained nature. Unlike the other seven coarse-grained benchmarks with distinct semantic boundaries, VGGFace2 is a fine-grained dataset featuring subtle inter-class distinctions and massive intra-class variations caused by pose, illumination, and expression changes. Distinguishing similar identities under these conditions requires capturing sharp, sample-specific evidence patterns; (2) L_{var} penalizes high-variance evidence distributions to enforce smooth consensus across views, which can inadvertently smooth out the subtle evidence spikes essential for fine-grained separation. For coarse-grained datasets, this suppression of erratic evidence is purely beneficial. For fine-grained recognition, however, it can inadvertently smooth out the subtle evidence spikes essential for separating highly similar classes, creating a mild fitting limitation; (3) The variance regularization should be adaptively down-scaled for fine-grained tasks. When L_{var} is removed, the model regains flexibility to fit these intricate boundaries, yielding the observed rebound. This suggests that for fine-grained tasks, variance regularization should be adaptively down-scaled (e.g., via a curriculum scheduler).

F.5 Scalability Analysis

To give readers a more intuitive sense of the paper's highly efficient scalability regarding deep learning training dynamics, as shown in Fig. 7, HGMN incurs training costs comparable to lightweight baselines (e.g., TMC, ECML) and is 5 to 10 times faster than TUNED and RCMCL, while maintaining stable scaling on large-scale datasets such as VGGFace2 and ALOI. This confirms that our hypergraph construction and mean-field negotiation introduce negligible computational overhead.

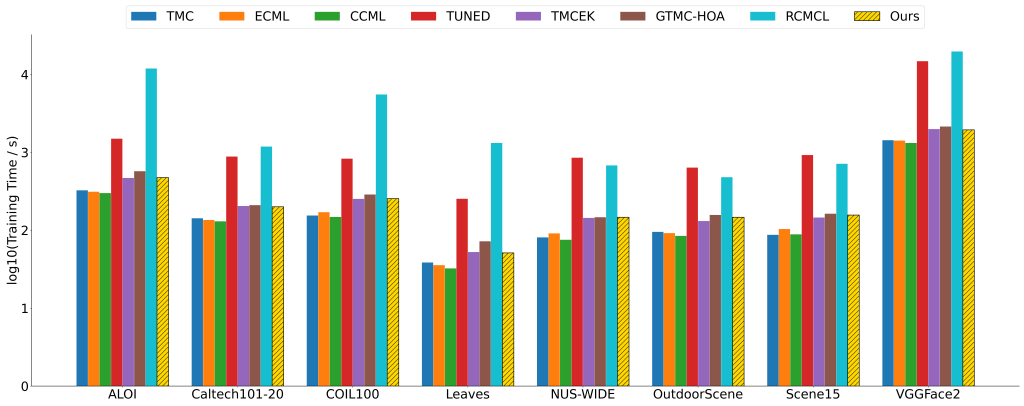


Figure 7: Training time (seconds) comparison across all datasets.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly describe the algorithm implementation, theoretical contributions, and comprehensive experiments presented in this paper, consistent with the subsequent Sections 3 and 4.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: This paper elaborates on the limitation of the current framework's inability to isolate and discover unknown classes in Section 5, and mentions that integrating open-set evidential theory will be considered in future work to enable dynamic expansion of the probability simplex for simultaneous robust classification and novel class discovery.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: This paper provides detailed explanations for all the formulas involved.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The network architecture, datasets, model parameters, and training approaches involved in the experimental implementation are described in detail in both the main text and the Appendix.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: We commit to sharing our code upon the acceptance of the paper.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: All implementation details were provided in Appendix D, Appendix **??**, Appendix E and section 4.1.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: In this paper, to avoid the randomness caused by data partitioning and network initialization, we adopt a five independent repeated experiments across all datasets. The experimental results are presented in the form of mean values and standard deviations. Details can be found in section 4.1.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: This paper provides detailed explanations of the types of computing devices, memory, and storage information used in the experiments, see Appendix E, but does not include relevant computation time. This is because the focus of this paper is on addressing highly conflicting multiview scenarios rather than on acceleration the process.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have checked the NeurIPS code of ethics for our paper.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: There is no societal impact of the work performed.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The assets (datasets, code, models) used in the paper are open source, and our use follows the relevant protocols.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.

- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.