ROBUST MULTI-VIEW LEARNING VIA HYBRID GEOMETRIC RELIABILITY UNDER TEST-TIME NOISY CORRESPONDENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

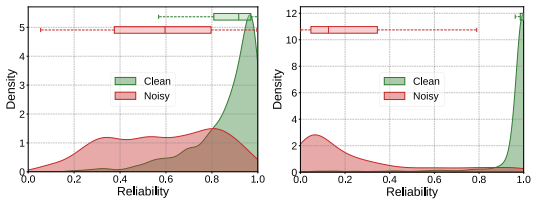
The effectiveness of multi-view learning hinges on correct correspondences across different views. However, data acquisition and transmission failures in practical deployment may cause certain views of a sample to be mismatched, leading to test-time noise correspondence that severely undermines multi-view fusion. Existing methods predominantly rely on view-wise statistical information for reliability assessment, while largely overlooking the intrinsic geometric structure of multi-view features. To address these limitations, we propose a robust multi-view learning framework based on **H**ybrid **G**eometric **R**eliability, termed **HGR**. First, Noisy Correspondence Augmented Learning simulates diverse view mismatch patterns during training and learns view reliability through explicit supervision. Second, Hybrid Geometric Learning constructs Euclidean and Hyperbolic Gramian matrices to capture the high-order joint geometric structure of multiple views, and performs reliability evaluation based on cross-view structural relationships. Furthermore, the Geometric Consistency Principle is introduced, where the Gramian volume contrastive loss and the view consistency loss are devised. The two losses reinforce the geometric coherence of correctly matched samples while amplifying the structural distortion of mismatched ones. Extensive experiments on multiple multi-view benchmark datasets demonstrate that the proposed HGR achieves superior performance and robustness over the state-of-the-art methods.

1 INTRODUCTION

A vast amount of data naturally exhibits multi-view (MV) characteristics (Fang et al., 2023; Huang et al., 2021; Qin et al., 2025) in real-world applications, where a single entity can be described from distinct modalities or sensors. Multi-view learning (MVL) (Li et al., 2018; Fu et al., 2020; Berahmand et al., 2025) emerges as a paradigm that harnesses heterogeneous sources of information, aiming to explore cross-view complementarity and learn more discriminative data representations. However, the effective utilization of multi-view information hinges on the correct correspondences across different views. Cross-view noisy correspondence (NC) (Han et al., 2023a; Huang et al., 2024a; He et al., 2024; Li et al., 2025; 2026) severely compromises semantic consistency, thereby significantly undermining the reliability of MVL.

During the deployment phase, practical challenges including data acquisition and transmission failures induce the problem of test-time NC (Li et al., 2026; He et al., 2026), where partial views of a given sample are no longer drawn from the original instance, but are replaced by features from other samples, resulting in view mismatches. In such cases, certain views carry semantics that conflict with the true labels for the sample, and naively aggregating all view information can severely mislead the predictions of models. To address this issue, some studies (Sun et al., 2024; He et al., 2024) employed contrastive learning to enforce consistency among aligned samples while pushing mismatched ones apart. Others (Han et al., 2023b; Liang et al., 2025) leveraged evidential learning to quantify view-wise uncertainty and performed logit fusion based on the estimated uncertainty. A very recent work (He et al., 2026) further predicted view reliability by jointly considering cross-view prediction consistency and intra-view prediction stability.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



(a) BML (b) HGR

Figure 1: Reliability density comparison between BML and HGR on NUSWIDEOBJ dataset. BML relies on view-specific and pairwise view statistics to assess view reliability, its predicted reliability scores for clean and noisy views overlap considerably, exhibiting low discriminability, while HGR, based on hybrid Gram geometry, clearly distinguishes clean views from noisy views.

Although the above methods can alleviate the performance degradation caused by test-time NC to some extent, most of them overlook the intrinsic geometric structure embedded in MV data. One of the fundamental objectives (Zhao et al., 2017) of MVL is to learn representations that conform to the underlying manifold structure of the data. When cross-view correspondences remain intact, the MV embeddings of the same samples are expected to exhibit strong geometric consistency in the latent space. Once the correspondences are disrupted by view misaligned, the geometric constraints across views are substantially violated. Existing methods (Sun et al., 2024; He et al., 2024; Han et al., 2023b; Liang et al., 2025; He et al., 2026) fail to fully exploit this geometric prior, mostly relying on statistical metrics that are independent of geometric structure to evaluate view reliability, leading to degraded discriminative capability under high noise ratios. Fig. 1 (a) exhibits the reliability density visualization of the state-of-the-art baseline BML on NUSWIDEOBJ dataset, where the test NC ratio is set to 50%. BML relies on view-specific information and pairwise view statistics to assess view reliability. However, its predicted reliability scores for clean and noisy views overlap substantially, and the median of predicted scores for noisy views is excessively large, even exceeding 0.5, resulting in very limited discriminability. In contrast, explicitly modeling the relationships among different views from a geometric perspective directly characterizes the structural position of each view relative to others, which facilitates robust view reliability assessment under extreme noise conditions.

Motivated by the deficiencies discussed above, we introduce a robust MVL framework based on **H**ybrid **G**eometric **R**eliability (HGR), which aims to to tackle the challenges of test-time NC. We first introduce Noisy Correspondence Augmented Learning (NCAL), which proactively simulates diverse view mismatch patterns during training and introduces view-level explicit supervisory signals to guide the model in identifying incorrectly corresponded views. Subsequently, we propose Hybrid Geometric Learning (HGL), which constructs sample-wise Gramian matrices (GM) in Euclidean and Hyperbolic spaces. These two geometries provide complementary perspectives on the inter-view relationship structure, serving as crucial cues for view reliability estimation. Fig. 1 (b) shows that the proposed HGR can clearly distinguishes clean views from noisy views. Furthermore, we establish the geometric consistency principle consisting of the Gramian volume (GV) contrastive loss and the view consistency loss, which reinforces the geometric consistency of aligned samples while amplifying the structural distortion of mismatched ones. Fig. 2 presents the framework of HGR. Generally, the contributions of this paper are summarized as follows:

- We present a robust MVL framework based on hybrid geometric reliability. By incorporating noisy correspondence augmentation and hybrid geometric learning, the model can explicitly learn view-specific reliability, which facilitates adaptive suppression and integration of views that are corrupted at test time.
- We propose Gramian Geometry-Driven Reliability Evaluation, which exploits Euclidean and Hyperbolic GM to capture MV high-order geometric structure. By encoding inter-view structural relationships into reliability-aware features, it enables view reliability assessment from joint MV structure instead of isolated statistical metrics.
- We propose Geometric Consistency Principle, based on which the GV contrastive loss and the view consistency loss are designed. The two losses aim to reinforce the geometric coherence of clean samples while amplifying the structural distortion of mismatched ones.
- To verify the effectiveness of the proposed method, we conduct experiments on multiple challenging MV benchmark datasets. Experimental results demonstrate that the proposed HGR consistently achieves superior performance across various noise levels compared to the SOTA algorithms.

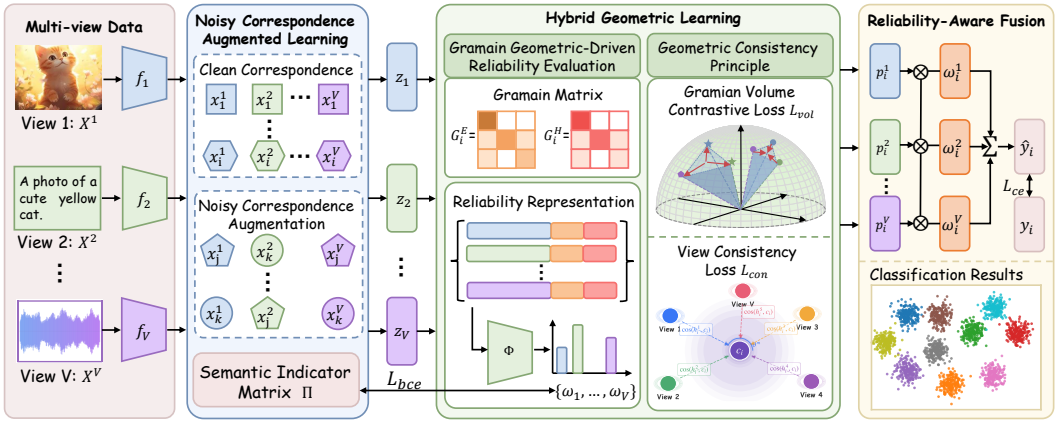


Figure 2: Overview of the proposed HGR. HGR comprises Noisy Correspondence Augmented Learning, which simulates diverse view-level NC and learns view reliability through explicit supervision, and Hybrid Geometric Learning, which exploits Euclidean and hyperbolic Gramian geometry to capture high-order inter-view relationships for reliability estimation.

2 METHODOLOGY

2.1 PRELIMINARIES

Given a set of MV features $\mathcal{X} = \{\mathbf{X}^v\}_{v=1}^V$ with V views, where $\mathbf{X}^v \in \mathbb{R}^{N \times d^v}$ contains N samples with d^v dimensions. For classification tasks, each sample $\{\mathbf{x}_i^v\}_{v=1}^V$ is associated with a corresponding label $y_i \in \{1, 2, ..., C\}$, where C denotes the number of classes. Then, a MV dataset is denoted as $\mathcal{D} = \{\mathcal{X}, Y\}$, where $Y = \{y_i\}_{i=1}^N$. For each view, a view-specific neural network is employed, consisting of an encoder $f^v(\cdot)$ and a classifier $g^v(\cdot)$. $f^v(\cdot)$ is used to encode the raw features into the embedding space while $g^v(\cdot)$ is used to project the embedding into the label space. Following the definition in (He et al., 2026) , the test-time NC is formulated as follows.

Definition 1 (**Test-Time Noisy Correspondence**). *Given a test multi-view dataset $\mathcal{D}_{test}$, a sample $\mathbf{x}_i$ is considered to be affected by test-time NC if at least one view is mismatched. Let k_i denote the number of corrupted views for the i-th sample, which reflects the severity of NC and is defined as*

$$k_i = \sum_{v=1}^{V} \mathbb{I}(\mathcal{T}(\mathbf{x}_i^v) \neq y_i), \tag{1}$$

where $\mathcal{T}(\cdot)$ denotes the ground truth for a sample in a specific view, $\mathbb{I}(\cdot)$ denotes an indicator function. Furthermore, To avoid excessive misalignment that may introduce semantic ambiguity, k_i is constrained to satisfy

$$1 \leq k_i \leq \left\lfloor \frac{V}{2} \right\rfloor. \tag{2}$$

2.2 NOISY CORRESPONDENCE AUGMENTED LEARNING

In the scenario of test-time NC, cross-view semantic alignments for the same samples collapse. Observations from certain views may originate from other samples, introducing semantic inconsistency with the labels of current samples. Since standard MV training data assume correct correspondences, models trained on clean data cannot effectively identify or suppress misaligned views. To this end, this section introduces the NCAL module, which actively simulates the NC during training, constructing diverse misalignment patterns and providing explicit view-level reliability supervision.

NC Construction. To simulate NC at test time, in each training epoch, a random subset of samples is selected with probability $p \in (0, 1]$ for corruption. The subset is denoted as $\mathcal{S} \subseteq \{1, 2, ..., N\}$ with $|\mathcal{S}| = pN$. For each selected sample, a random portion of its views are kept intact while the rest are replaced by views from other samples via cross-sample permutation:

$$\hat{\mathbf{x}}_i^v \leftarrow \mathbf{x}_j^v, i \in \mathcal{S}. \tag{3}$$

After view-wise permutation, we obtain the new MV dataset $\hat{\mathcal{D}}$ and further determine whether a view is genuinely noisily correlated by comparing its post-permutation features with the original label of the sample, which is formulated as

$$\pi_i^v = \begin{cases} 0, & \text{if } \mathcal{T}(\hat{\mathbf{x}}_i^v) \neq y_i, \\ 1, & \text{if } \mathcal{T}(\hat{\mathbf{x}}_i^v) = y_i. \end{cases} \tag{4}$$

π_i^v does not strictly indicate whether the view preserves the original sample-level correspondence, but rather measures whether the view remains semantically consistent with the current sample at the class level. For each sample, a view-wise semantic indicator vector $\boldsymbol{\pi}_i = [\pi_i^1, ..., \pi_i^V]$ is defined, and the collection of these vectors forms the global indicator matrix $\Pi = [\boldsymbol{\pi}_i; ...; \boldsymbol{\pi}_N] \in \mathbb{R}^{N \times V}$. Notably, the construction of NC is also subject to the constraint in Eq. 2, that is

$$1 \leq \sum_{v=1}^{V} (1 - \pi_i^v) \leq \left\lfloor \frac{V}{2} \right\rfloor. \tag{5}$$

View Reliability Assessment. For the NC augmented MV dataset $\hat{\mathcal{D}}$, each view is first passed through its corresponding view-specific encoder $f(\cdot)$ to extract latent representations. Subsequently, a view-specific classifier $g(\cdot)$ is applied to obtain the classification logits. For the v-th view of the i-th sample, we have

$$\mathbf{z}_i^v = f^v(\hat{\mathbf{x}}_i^v), \mathbf{p}_i^v = g^v(\mathbf{z}_i^v), \tag{6}$$

where $\mathbf{z}_i^v$ denotes the latent representation and $\mathbf{p}_i^v$ denotes the predicted logit. Based on the late fusion strategy, $\{\mathbf{p}_i^v\}_{v=1}^V$ are fused into a unified logit. However, under test-time NC, different views contribute to the final prediction with varying degrees of trustworthiness. For views that preserve correct semantic correspondences, their predictions should be assigned greater influence, whereas the influence of misaligned views should be suppressed. To this end, each view of every sample is assigned a reliability weight ω_i^v, the logit fusion is written as

$$\mathbf{p}_i = \sum_{v=1}^{V} \omega_i^v \mathbf{p}_i^v. \tag{7}$$

Thus, the cross-entropy (CE) loss is used to minimize the distance between the unified logit and the ground truth within a training batch:

$$\mathcal{L}_{ce} = -\frac{1}{B} \sum_{i=1}^{B} y_i \log \tilde{\mathbf{p}}_i, \tag{8}$$

where B denotes the number of samples in a training batch, $\tilde{\mathbf{p}}_i = \mathrm{softmax}(\mathbf{p}_i)$ is the normalized predicted label. Under the supervision of $\mathcal{L}_{ce}$, the model can acquire basic classification capability.

How to assess the reliability of sample-wise view features remains a critical challenge. A straightforward approach is to directly rely on the view features themselves with supervision from the semantic indicator matrix to determine reliability. First, an assessment network $\Phi(\cdot)$ is introduced to produce the reliability scores for all views of every sample:

$$\omega_i^v = \sigma[\Phi(\hat{\mathbf{x}}_i^v)], \tag{9}$$

where $\sigma[\cdot]$ is the sigmoid function that projects the weights into $(0, 1)$. Thus, the binary CE loss is employed to minimize the distance between π_i^v and ω_i^v, the supervised loss is written as:

$$\mathcal{L}_{bce} = -\frac{1}{BV} \sum_{i=1}^{B} \sum_{v=1}^{V} \mathrm{BCE}(\pi_i^v, \omega_i^v) = -\frac{1}{BV} \sum_{i=1}^{B} \sum_{v=1}^{V} \left[(1 - \pi_i^v) \log(1 - \omega_i^v) + \pi_i^v \log \omega_i^v \right]. \tag{10}$$

In addition to the view-specific features, BML (He et al., 2026) further incorporates view-specific prediction entropy and cross-view prediction consistency as complementary cues. However, this approach primarily relies on prediction-level statistics, making it difficult to fully characterize the more fundamental relational structures among view features. In the next section, we elaborate on how the geometric structure of views is leveraged to guide the reliability assessment.

2.3 HYBRID GEOMETRIC LEARNING

Relying solely on view-specific and pairwise statistics fails to characterize the overall structures of multiple views, which limits the reliability estimation capability. To address this issue, this section introduces HGL, which integrates GM from both Euclidean and Hyperbolic spaces to capture cross-view relationships from two complementary aspects. Furthermore, the Geometric Consistency Principle is proposed to amplify the structural discrepancy between aligned and misaligned samples.

Gramian Geometry-Driven Reliability Evaluation. Conventional methods employ pairwise statistics to characterize view relationships, which cannot capture the joint coupling among more than two views. When multiple views are simultaneously mismatched, such methods are prone to misjudgment due to accidental similarities. In contrast, GM can encode the holistic high-order geometric structure of multiple views, which enables the identification of structural distortions induced by mismatches across multiple views. The definition of Euclidean GM (EGM) and Euclidean GV (EGV) are introduced as follows.

Definition 2 (**EGM and EGV**). *Given a set of feature vectors $\mathbf{X} = \{\mathbf{x}_1; \mathbf{x}_2; ...; \mathbf{x}_n\} \in \mathbb{R}^{n\times d}$, where n and d denote the number of samples and the feature dimension, respectively, the EGM $\mathbf{G}^E \in \mathbb{R}^{n\times n}$ is defined as the matrix of pairwise inner products:*

$$\mathbf{G}^E = [\langle \mathbf{x}_i, \mathbf{x}_j \rangle_E]_{(i,j)\in\{n\}}, \tag{11}$$

where each element is computed by $\langle \mathbf{x}_i, \mathbf{x}_j \rangle_E = \mathbf{x}_i \mathbf{x}_j^\top$. Further, the EGV is defined as

$$V^E = \sqrt{\det(\mathbf{G}^E)}. \tag{12}$$

V^E measures the volume of the parallelotope spanned by the given vectors, and therefore characterizes the overall geometric configuration of the vector set.

The EGM effectively captures high-order relationships among MV features via joint geometric structure, but feature normalization based on L_2-norm makes it prone to variance collapse, where the matrix degenerates toward an identity matrix and discriminative structural variance vanishes. In contrast, Hyperbolic geometry has exponential volume growth and provides greater capacity to preserve geometric variation and capture high-order structural distortions that Euclidean space tends to flatten. However, Hyperbolic Gram alone focuses on fine-grained differences and lacks the cross-class discriminative stability of Euclidean space. We therefore exploit Gramain geometry in both spaces to build the view reliability representation. The Lorentz model of Hyperbolic space, Hyperbolic GM (HGM) and Hyperbolic GV (HGV) are defined as follows.

Definition 3 (**Lorentz Model of Hyperbolic Space**). *Let $\mathcal{H}^d$ denote the d-dimensional Lorentz model of Hyperbolic space, which is defined as the upper sheet of a two-sheeted hyperboloid embedded in (d+1)-dimensional Minkowski space and has constant sectional curvature -1:*

$$\mathcal{H}^d = \{\mathbf{h} = (t, \mathbf{s}) \in \mathbb{R}^+ \times \mathbb{R}^d : -t^2 + ||\mathbf{s}||^2 = -1\}, \tag{13}$$

where $t \in \mathbb{R}^+$ denotes the temporal component and $\mathbf{s} \in \mathbb{R}^d$ denotes the spatial component. The Lorentzian inner product between two samples $\mathbf{h}_i = (t_i, \mathbf{s}_i)$ and $\mathbf{h}_j = (t_j, \mathbf{s}_j)$ is defined as

$$\langle \mathbf{h}_i, \mathbf{h}_j \rangle_L = -t_i t_j + \langle \mathbf{s}_i, \mathbf{s}_j \rangle_E. \tag{14}$$

Accordingly, each sample $\mathbf{h} = (t, \mathbf{s}) \in \mathcal{H}^d$ satisfies

$$t = \sqrt{(1 + ||\mathbf{s}||^2)}. \tag{15}$$

According to **Definition** 3, a convenient mapping from an arbitrary Euclidean vector $\mathbf{x}$ into Hyperbolic space by the Lorentz model is written as:

$$\varphi(\mathbf{x}) = (\sqrt{(1 + ||\mathbf{x}||^2)}, \mathbf{x}). \tag{16}$$

Based on **Definitions** 2 and 3, we can derive the definitions of the HGM and HGV.

Definition 4 (**HGM and HGV**). *Given a set of Euclidean feature vectors $\mathbf{X} = \{\mathbf{x}_1; \mathbf{x}_2; ...; \mathbf{x}_n\} \in \mathbb{R}^{n\times d}$, each sample is mapped into projected into the Hyperbolic space through the Lorentz model: $\mathbf{h}_i = (\sqrt{1 + ||\mathbf{x}_i||^2}, \mathbf{x}_i)$. Then, the HGM is defined as*

$$\mathbf{G}^H = [\langle \mathbf{h}_i, \mathbf{h}_j \rangle_L]_{(i,j)\in\{n\}} = \left[\langle \mathbf{x}_i, \mathbf{x}_j \rangle_E - \sqrt{(1 + ||\mathbf{x}_i||^2)}\sqrt{(1 + ||\mathbf{x}_j||^2)}\right]_{(i,j)\in\{n\}}. \tag{17}$$

The corresponding HGV is computed by

$$V_H = \sqrt{\det(\mathbf{G}^H)}. \tag{18}$$

Assessing view reliability from raw view representations or pairwise statistics is inherently limited, it reflects only single-view semantics or local pairwise interactions, without modeling the joint high-order structure of all views. To address this, we propose the following Gramain geometry-driven reliability representation construction method.

Proposition 1 (**Gramian Geometry-Driven Reliability Representation**). *Given a set of MV features for the i-th sample, $\mathbf{z}_i = \{\mathbf{z}_i^1; \mathbf{z}_i^2; ...; \mathbf{z}_i^V\}$, its EGM $\mathbf{G}_i^E \in \mathbb{R}^{V\times V}$ and HGM $\mathbf{G}_i^H \in \mathbb{R}^{V\times V}$ can be constructed by **Definitions** 2 and 4, respectively. For the v-th view, its geometric reliability representation is defined by concatenating the v-th rows of EGM and HGM:*

$$\mathbf{r}_i^v = \left[\mathbf{z}_i^v || \mathbf{G}_i^E(v,:) || \mathbf{G}_i^H(v,:)\right]. \tag{19}$$

The reliability representation $\mathbf{r}_i^v$ characterizes the relational structure of the v-th view with respect to all views from complementary Euclidean and Hyperbolic geometric perspectives, and therefore provides a geometry-aware basis for estimating its view-specific reliability.

Based on the geometry-aware reliability representation $\mathbf{r}_i^v$, the reliability score of the v-view of the i-th sample is obtained by the assessment network:

$$\omega_i^v = \sigma[\Phi(\mathbf{r}_i^v)]. \tag{20}$$

Geometric Consistency Principle. Relying solely on the classification loss is insufficient to constrain the MV feature space, we further regularize the feature space from the perspective of geometric consistency. For samples with aligned cross-view correspondence, different views are expected to exhibit a consistent geometric structure in the latent space. Conversely, when certain views are corrupted by misalignment, the cross-view feature relationships are disrupted, resulting in a more dispersed holistic feature structure. Since the GV reflects the holistic spanning extent of feature vectors in a geometric space, we minimize the GV of view-aligned samples while maximizing that of view-misaligned samples.

Specifically, for the i-th sample $\mathbf{z}_i = \{\mathbf{z}_i^1, \mathbf{z}_i^2, ..., \mathbf{z}_i^V\}$, its EGV V_i^E and HGV V_i^H are computed based on **Definitions** 2 and 4. To integrate the complementary strengths of Euclidean and Hyperbolic geometries, we propose an adaptive hybrid geometry fusion strategy:

$$V_i = \beta V_i^E + (1-\beta)V_i^H, \tag{21}$$

where $\beta \in (0,1)$ is a learnable weight. In a training batch, the GV-induced regularization loss is written as:

$$\mathcal{L}_{vol} = -\frac{1}{B_a}\sum_{i=1}^{B_a} \log \frac{\exp(-V_i/\tau)}{\exp(-V_i/\tau) + \sum_{j=1}^{B_u}\exp(-V_j/\tau)}, \tag{22}$$

where B_a denotes the number of aligned samples, B_u denotes the number of unaligned samples, τ denotes the temperature parameter.

However, GV primarily characterizes the holistic geometric structure formed by MV features as a whole, without explicitly specifying the direction toward a shared semantic representation. Thus, constraining only the overall volume still lacks an explicit view-level alignment signal. To this end, we leverage fully aligned samples to construct direct cross-view consistency supervision. Specifically, for the i-th fully aligned sample, we obtain the consistent representation by

$$\mathbf{c}_i = \frac{1}{V}\sum_{v=1}^{V}\mathbf{z}_i^v. \tag{23}$$

By maximizing the cosine similarity between each view-specific and the consistent representations, we enforce the semantic consistency of different views in the latent space. The corresponding loss is defined as

$$\mathcal{L}_{con} = \frac{1}{B_a V}\sum_{i=1}^{B_a}\sum_{v=1}^{V}[1-\cos(\mathbf{z}_i^v, \mathbf{c}_i)]. \tag{24}$$

Different from $\mathcal{L}_{vol}$, $\mathcal{L}_{con}$ directly constrains the consensus relationships of fully aligned samples, providing explicit alignment guidance for the feature space.

Table 1: Comparison of Classification performance (ACC±STD) under different noise rates. The best results are highlighted in **bold**, and the second-best results are underlined.

Noise	Method	GRAZ02	Caltech8677	flower17	MITIndoor	100Leaves	Wikipedia
50%	TMC	59.22±2.22	72.23±0.90	50.11±2.51	50.57±1.88	79.69±1.55	47.99±3.39
	UIMC	58.01±3.02	74.52±0.84	63.79±2.12	55.40±0.84	78.09±1.37	48.13±2.42
	ECML	55.86±2.82	70.03±1.01	35.33±4.36	48.23±2.18	77.69±1.84	47.99±3.39
	CCML	60.84±2.28	71.22±1.18	53.79±3.15	50.07±1.91	78.88±1.46	46.55±3.15
	MAMC	64.86±2.22	59.84±1.38	51.69±2.91	28.31±2.17	73.62±2.11	49.21±3.00
	ETF	62.09±1.10	74.57±1.10	71.87±1.65	57.75±0.91	78.72±1.58	45.61±3.53
	TMCEK	54.19±2.95	71.73±0.82	35.11±6.69	48.40±1.11	77.62±2.05	49.06±3.31
	FUML	29.56±5.07	74.56±0.89	62.98±6.08	47.75±1.57	94.62±1.50	35.61±6.11
	RML	59.12±2.80	67.83±0.74	65.18±2.01	54.97±1.00	86.22±2.00	24.10±2.02
	BML	67.23±1.95	84.26±0.66	84.63±2.04	55.24±1.46	96.16±0.84	53.38±0.71
	HGR	**68.58±2.91**	**85.73±0.51**	**86.47±1.34**	**59.89±1.42**	**96.75±0.42**	**54.68±0.64**
100%	TMC	51.93±2.53	49.76±1.73	37.72±2.52	34.34±1.54	61.78±2.94	36.62±4.08
	UIMC	52.06±2.52	54.12±1.17	49.04±2.38	37.96±0.88	60.78±3.31	36.26±4.61
	ECML	46.39±2.39	47.04±1.11	27.68±3.74	32.14±1.70	59.91±3.41	36.33±3.98
	CCML	52.57±2.08	48.16±1.34	42.83±3.64	34.53±1.73	60.09±3.04	34.89±4.38
	MAMC	58.78±1.49	48.85±1.01	39.26±3.43	16.08±1.10	58.81±2.42	35.76±3.70
	ETF	54.86±2.10	53.19±1.27	58.23±3.06	38.97±1.15	59.53±2.27	34.24±3.93
	TMCEK	45.61±1.71	49.56±2.02	26.88±6.70	32.78±0.86	59.91±3.52	36.76±4.14
	FUML	28.89±4.14	66.25±0.99	49.74±6.28	32.58±0.80	89.84±1.61	26.98±4.79
	RML	51.22±2.31	48.41±1.37	51.91±3.37	37.09±0.76	76.44±2.94	18.20±2.16
	BML	59.06±1.34	77.21±0.64	76.91±2.89	42.89±1.16	93.56±0.73	35.68±3.17
	HGR	**61.49±2.34**	**80.37±1.02**	**79.04±1.53**	**49.77±1.56**	**94.13±0.69**	**37.70±2.44**

2.4 Training and Inference

Training. In the training phase, we adopt an end-to-end training paradigm, and the overall objective loss function is summarized as

$$\mathcal{L} = \mathcal{L}_{ce} + \lambda_b \mathcal{L}_{bce} + \lambda_v \mathcal{L}_{vol} + \lambda_c \mathcal{L}_{con}, \tag{25}$$

where λ_b, λ_v, and λ_c denote the three trade-off parameters that balance the contributions of the four loss terms. By minimizing the overall loss $\mathcal{L}$, the view-specific encoders, classifiers, and the assessment network can be optimized.

Inference. In the inference phase, for each test sample $\mathbf{x}_i = \{\mathbf{x}_i^1, \mathbf{x}_i^2, ..., \mathbf{x}_i^V\}$, we compute its latent representations $\{\mathbf{z}_i^v\}_{v=1}^V$, logit outputs $\{\mathbf{p}_i^v\}_{v=1}^V$, EGM $\mathbf{G}_i^E$, and HGM $\mathbf{G}_i^H$. Then, the reliability representations $\{\mathbf{r}_i^v = [\mathbf{z}_i^v || \mathbf{G}_i^E(v,:) || \mathbf{G}_i^H(v,:)]\}_{v=1}^V$ are obtained by concatenation, which are further passed into the assessment network to generate the reliability scores $\{\omega_i^v\}_{v=1}^V$. Finally, the predicted label for the i-th sample is computed by

$$\hat{y}_i = \arg \max_{c \in \{1,...,C\}} \left[\sum_{v=1}^{V} \omega_i^v \mathbf{p}_i^v \right]_c. \tag{26}$$

3 Experiments

3.1 Datasets and Compared Methods

To verify the effectiveness of the proposed HGR, we conduct experiments on 12 widely-used multi-view benchmark datasets, including **GRAZ02** (Opelt et al., 2006), **Caltech8677** (Fei-Fei et al., 2004), **flower17** (Nilsback & Zisserman, 2006), **MITIndoor** (Quattoni & Torralba, 2009), **100Leaves** (Wang et al., 2019b), **Wikipedia** (Rasiwasia et al., 2010), **LandUse** (Yang & Newsam, 2010), **MNIST10k** (Asuncion et al., 2007), **CCV** (Jiang et al., 2011), **NUSWIDEOBJ** (Chua et al., 2009), **AWA** (Lampert et al., 2013), **YouTubeFace** (Wolf et al., 2011).

To demonstrate the superiority of the proposed HGR, we compare it against ten representative MVL approaches, including trustworthy methods **TMC** (Han et al., 2021), **UIMC** (Xie et al., 2023),

Table 2: Comparison of Classification performance (ACC±STD) under different noise rates. The best results are highlighted in **bold**, and the second-best results are underlined.

Noise	Method	LandUse	MNIST10k	CCV	NUSWIDEOBJ	AWA	YouTubeFace
50%	TMC	39.62±1.97	80.11±0.53	38.92±0.75	34.51±0.28	28.99±0.41	63.48±0.27
	UIMC	42.43±1.12	80.66±0.48	39.68±0.97	35.62±0.67	28.66±0.43	68.72±0.23
	ECML	37.79±1.93	78.92±0.64	36.37±1.32	30.73±0.41	26.05±0.48	45.79±0.87
	CCML	34.90±1.82	79.78±0.69	37.25±0.98	32.48±0.56	24.15±0.58	35.81±0.48
	MAMC	62.67±2.21	95.70±0.38	46.27±1.50	38.70±0.60	21.84±0.59	76.45±0.44
	ETF	38.98±1.51	82.74±0.51	41.67±1.04	38.26±0.45	36.94±0.36	68.97±0.24
	TMCEK	38.86±2.35	78.84±0.73	37.70±0.68	31.27±0.56	26.63±0.53	46.57±0.93
	FUML	68.55±1.80	92.96±0.57	42.21±2.60	42.58±0.39	31.42±0.52	74.10±0.41
	RML	66.10±2.33	92.36±0.49	41.39±1.01	37.51±0.46	32.56±0.49	73.79±0.33
	BML	75.62±1.29	96.43±0.19	51.73±1.10	45.99±0.44	40.52±0.31	84.12±0.32
	HGR	**76.86±1.24**	**96.93±0.15**	**53.49±0.87**	**47.57±0.22**	**43.15±0.27**	**86.53±0.25**
100%	TMC	33.50±1.43	71.53±0.62	33.12±1.35	29.56±0.43	21.38±0.53	51.64±0.31
	UIMC	34.90±1.55	69.92±0.80	34.04±1.16	29.86±0.45	21.53±0.21	55.61±0.39
	ECML	32.48±1.82	69.18±1.00	31.99±0.89	27.44±0.39	19.16±0.40	38.99±0.60
	CCML	29.55±1.75	71.19±0.67	30.66±1.02	27.21±0.32	16.77±0.57	31.43±0.31
	MAMC	52.62±2.15	94.63±0.52	37.91±1.66	31.47±0.32	11.45±0.53	66.27±0.62
	ETF	31.52±1.05	76.93±0.88	33.30±1.45	31.56±0.29	25.92±0.51	55.57±0.45
	TMCEK	33.02±1.42	69.87±0.89	32.18±0.83	27.40±0.37	19.40±0.31	39.46±0.62
	FUML	63.38±1.91	92.20±0.40	36.55±2.90	38.14±0.53	24.35±0.39	65.10±0.56
	RML	55.31±2.37	89.12±1.06	35.01±0.95	31.19±0.38	23.91±0.39	60.99±0.44
	BML	69.02±1.50	95.71±0.26	44.16±0.97	39.98±0.53	32.22±0.51	78.42±0.26
	HGR	**72.43±1.29**	**96.48±0.24**	**46.83±0.80**	**43.15±0.46**	**37.22±0.44**	**83.39±0.18**

ECML (Xu et al., 2024), **CCML** (Liu et al., 2024b), **ETF** (Lu et al., 2025), **TMCEK** (Liang et al., 2025), **FUML** (Duan et al., 2025), alignment methods **MAMC** (Lin et al., 2025), **RML** (Xu et al., 2025), and reliability eatimation mathod **BML** (He et al., 2026).

3.2 Experimental Results

The experimental results on twelve MV datasets are reported in Tables 1 and 2, where the noise ratios are set to 50% and 100%. It can be seen that he proposed HGR achieves the best classification performance under both types of NC scenarios across all datasets, demonstrating its superiority over the SOTA baselines. BML also introduces a view reliability assessment mechanism and achieves the second-best results. However, it estimates reliability solely based on single and pairwise view information, ignoring the overall MV geometric structure, and therefore performs worse than HGR. In contrast, baseline methods that rely on uncertainty measurement and cross-view alignment perform poorly under NC scenarios, most of which focus on feature-level noise. When cross-view mismatch occurs, the original alignment objective is disturbed by erroneous signals, and uncertainty estimation is also easily misled by mismatched samples, ultimately leading to a marked degradation in classification performance.

3.3 Ablation Study

To verify the effectiveness of each loss component, we conduct ablation studies on $\mathcal{L}_{bce}$, $\mathcal{L}_{vol}$, and $\mathcal{L}_{con}$, the results are reported in Table 3. When all three loss terms are removed and only the classification loss $\mathcal{L}_{ce}$ is retained, the model achieves the worst performance. Introducing any one of the three loss terms leads to a significant improvement in the experimental results, demonstrating their effectiveness of improving the performance. When any two loss terms are employed, performance generally improves further. Notably, equipping the model with $\mathcal{L}_{bce}$ and $\mathcal{L}_{vol}$ yields the second-best results on most test datasets, while achieving the best results requires the additional inclusion of $\mathcal{L}_{con}$, highlighting the critical role of the $\mathcal{L}_{con}$ in guiding the optimization direction.

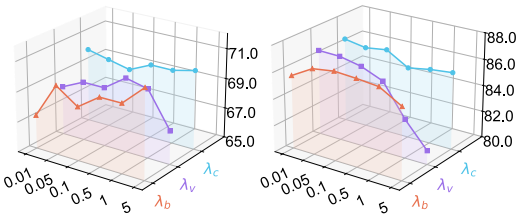
3.4 Parameter Sensitivity Analysis

To investigate the impact of λ_b, λ_v, and λ_c, we perform the parameter sensitivity analysis in Fig. 3. We can see that a small λ_b weakens the ability of identifying mismatched views and degrades the

Table 3: Ablation study of three loss items, where the test NC ratio is set to 0.5.

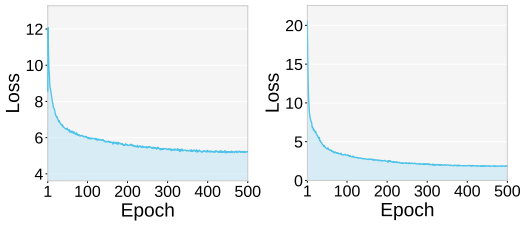
$\mathcal{L}_{bce}$	$\mathcal{L}_{vol}$	$\mathcal{L}_{con}$	GRAZ02	Caltech8677	flower17	MITIndoor	NUSWIDEOBJ	AWA
✗	✗	✗	60.00±1.16	69.90±0.53	70.66±2.35	52.37±1.62	40.82±0.50	33.20±0.51
✓	✗	✗	66.69±1.18	85.58±0.42	85.74±1.57	59.50±2.25	47.20±0.72	42.86±0.19
✗	✓	✗	67.36±1.43	84.90±0.38	85.37±1.58	57.84±1.14	45.77±0.31	40.19±0.32
✗	✗	✓	67.57±2.00	85.56±0.42	85.22±1.37	57.39±0.97	45.47±0.12	40.39±0.58
✓	✓	✗	67.77±2.17	85.65±0.43	85.88±0.79	59.78±1.17	47.38±0.34	42.83±0.29
✓	✗	✓	67.43±1.64	85.61±0.67	85.73±1.63	58.49±1.48	47.42±0.25	42.53±0.56
✗	✓	✓	67.63±2.23	85.54±0.57	85.59±1.20	58.40±1.35	45.92±0.35	40.65±0.49
✓	✓	✓	**68.58±2.91**	**85.73±0.51**	**86.47±1.34**	**59.89±1.42**	**47.57±0.22**	**43.15±0.27**

model performance, while an excessively large λ_v makes geometric regularization dominate training and impairs the main classification optimization. λ_c yields moderate performance fluctuation, but a proper value is required to preserve MV consistency.



(a) GRAZ02 (b) Caltech8677

Figure 3: Parameter sensitivity study about three trade-off hyperparameters.

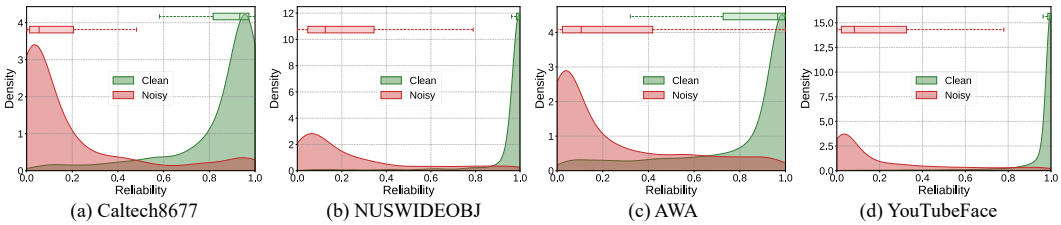


(a) GRAZ02 (b) Caltech8677

Figure 4: The loss curves with increasing training epochs.

3.5 Visualization Analysis

Some visualizations are provided to interpret model behavior. The training loss curves in Fig. 4 show that the total loss drops rapidly and converges stably, demonstrating stable training of the proposed optimization objective. In addition, the density plots of view-wise reliability scores demonstrate that HGR clearly separates clean views from noisy mismatched ones. Clean views yield high reliability scores while noisy views occupy low-score regions.



(a) Caltech8677 (b) NUSWIDEOBJ (c) AWA (d) YouTubeFace

Figure 5: The density visualization of the view reliability score ω_i^v on four datasets, where the test NC ratio is set to 0.5.

4 Conclusion

Test-time NC poses a significant challenge to MVL, severely degrading the performance of model. To address this issue, we propose the HGR framework. Specifically, HGR simulates view mismatch through NVAL, performs geometry-driven view reliability assessment using EGM and HGM. Furthermore, two loss objectives are designed to constrain the view consistency. Experiments on multiple datasets demonstrate the superiority of HGR over the SOTA compared methods.

LLM USE STATEMENT

During the preparation of this manuscript, the authors utilized LLM for manuscript translation, English polishing, and code debugging. All authors have thoroughly reviewed, verified, and taken full responsibility for all contents, experimental results, claims, and references in this paper. No LLM was used to generate core research ideas, model designs, or experimental conclusions.

REFERENCES

Arthur Asuncion, David Newman, et al. Uci machine learning repository, 2007.

Kamal Berahmand, Fatemeh Daneshfar, Maryam Rahmaninia, Maryam Haghighat, and Mahdi Jalili. A comprehensive survey on multi-view classification: Methods, applications, and challenges. *ACM Transactions on Intelligent Systems and Technology*, 16(6):1–34, 2025.

Jie Chen, Jian Pu, Junping Zhang, et al. Gmv: A unified and efficient graph multi-view learning framework. *Advances in Neural Information Processing Systems*, pp. 145359–145382, 2026.

Zhaoliang Chen, Lele Fu, Jie Yao, Wenzhong Guo, Claudia Plant, and Shiping Wang. Learnable graph convolutional network and feature fusion for multi-view learning. *Information Fusion*, 95: 109–119, 2023.

Tat-Seng Chua, Jinhui Tang, Richang Hong, Haojie Li, Zhiping Luo, and Yantao Zheng. Nus-wide: a real-world web image database from national university of singapore. In *Proceedings of the ACM international conference on image and video retrieval*, pp. 1–9, 2009.

Giordano Cicchetti, Eleonora Grassucci, Luigi Sigillo, and Danilo Comminiello. Gramian multimodal representation learning and alignment. In *Proceedings of International Conference on Learning Representations*, pp. 42128–42149, 2025.

Zhuohang Dang, Minnan Luo, Jihong Wang, Chengyou Jia, Haochen Han, Herun Wan, Guang Dai, Xiaojun Chang, and Jingdong Wang. Disentangled noisy correspondence learning. *IEEE Transactions on Image Processing*, 34:2602–2615, 2025.

Zhibin Dong, Siwei Wang, Jiaqi Jin, Xinwang Liu, and En Zhu. Cross-view topology based consistent and complementary information for deep multi-view clustering. In *Proceedings of IEEE/CVF International Conference on Computer Vision*, pp. 19383–19394, 2023.

Zhibin Dong, Shengju Yu, Siwei Wang, Xinwang Liu, and En Zhu. Cross-view graph matching for unsupervised learning with unaligned multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 38(4):2221–2233, 2026.

Siyuan Duan, Yuan Sun, Dezhong Peng, Guiduo Duan, Xi Peng, and Peng Hu. Deep fuzzy multiview learning for reliable classification. In *Proceedings of the International Conference on Machine Learning*, 2025.

Uno Fang, Man Li, Jianxin Li, Longxiang Gao, Tao Jia, and Yanchun Zhang. A comprehensive survey on multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 35 (12):12350–12368, 2023.

Li Fei-Fei, Rob Fergus, and Pietro Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop*, pp. 178–178, 2004.

Lele Fu, Pengfei Lin, Athanasios V Vasilakos, and Shiping Wang. An overview of recent multi-view clustering. *Neurocomputing*, 402:148–161, 2020.

Lele Fu, Zhaoliang Chen, Yongyong Chen, and Shiping Wang. Unified low-rank tensor learning and spectral embedding for multi-view subspace clustering. *IEEE Transactions on Multimedia*, 25:4972–4985, 2022.

Lele Fu, Sheng Huang, Lei Zhang, Jinghua Yang, Zibin Zheng, Chuanfu Zhang, and Chuan Chen. Subspace-contrastive multi-view clustering. *ACM Transactions on Knowledge Discovery from Data*, 18(9):1–35, 2024.

Lele Fu, Bowen Deng, Sheng Huang, Tianchi Liao, Chuanfu Zhang, and Chuan Chen. Learn from global rather than local: Consistent context-aware representation learning for multi-view graph clustering. In *Proceedings of International Joint Conference on Artificial Intelligence*, pp. 5145–5153, 2025.

Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2414–2423, 2016.

Haochen Han, Kaiyao Miao, Qinghua Zheng, and Minnan Luo. Noisy correspondence learning with meta similarity correction. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7517–7526, 2023a.

Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. Trusted multi-view classification. In *Proceedings of the International Conference on Learning Representations*, 2021.

Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. Trusted multi-view classification with dynamic evidential fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):2551–2566, 2023b.

Changhao He, Hongyuan Zhu, Peng Hu, and Xi Peng. Robust variational contrastive learning for partially view-unaligned clustering. In *Proceedings of the ACM International Conference on Multimedia*, pp. 4167–4176, 2024.

Changhao He, Di Xue, Shuxian Li, Yanji Hao, Xi Peng, and Peng Hu. Bootstrapping multi-view learning for test-time noisy correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1627–1638, 2026.

Zhenyu Huang, Joey Tianyi Zhou, Hongyuan Zhu, Changqing Zhang, Jiancheng Lv, and Xi Peng. Deep spectral representation learning from multi-view data. *IEEE Transactions on Image Processing*, 30:5352–5362, 2021.

Zhenyu Huang, Peng Hu, Guocheng Niu, Xinyan Xiao, Jiancheng Lv, and Xi Peng. Learning with noisy correspondence. *International Journal of Computer Vision*, 132(9):3656–3677, 2024a.

Zhenyu Huang, Mouxing Yang, Xinyan Xiao, Peng Hu, and Xi Peng. Noise-robust vision-language pre-training with positive-negative learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(1):338–350, 2024b.

Yu-Gang Jiang, Guangnan Ye, Shih-Fu Chang, Daniel Ellis, and Alexander C Loui. Consumer video understanding: A benchmark database and an evaluation of human and machine performance. In *Proceedings of the ACM International Conference on Multimedia Retrieval*, pp. 1–8, 2011.

Zisen Kong, Dongxia Chang, Yiming Wang, Pengyuan Li, and Yao Zhao. Tensorial multi-view clustering via alternative rank minimization and inter-view alignment. *IEEE Transactions on Knowledge and Data Engineering*, 38(5):3196–3209, 2026.

Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling. Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):453–465, 2013.

Haobin Li, Yijie Lin, Peng Hu, Mouxing Yang, and Xi Peng. Learning with dual-level noisy correspondence for multi-modal entity alignment. In *Proceedings of the International Conference on Learning Representations*, volume 2026, pp. 151700–151731, 2026.

Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *Proceedings of the International Conference on Machine Learning*, pp. 12888–12900, 2022.

Shuxian Li, Changhao He, Xiting Liu, Joey Tianyi Zhou, Xi Peng, and Peng Hu. Learning with noisy triplet correspondence for composed image retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19628–19637, 2025.

Yingming Li, Ming Yang, and Zhongfei Zhang. A survey of multi-view representation learning. *IEEE transactions on knowledge and data engineering*, 31(10):1863–1883, 2018.

Xinyan Liang, Shijie Wang, Yuhua Qian, Qian Guo, Liang Du, Bingbing Jiang, Tingjin Luo, and Feijiang Li. Trusted multi-view classification with expert knowledge constraints. In *Proceedings of the Forty-second International Conference on Machine Learning*, 2025.

Yuena Lin, Yiyuan Wang, Gengyu Lyu, Yongjian Deng, Haichun Cai, Huibin Lin, Haobo Wang, and Zhen Yang. Enhance multi-view classification through multi-scale alignment and expanded boundary. In *Proceedings of the International Conference on Learning Representations*, 2025.

Suyuan Liu, Qing Liao, Siwei Wang, Xinwang Liu, and En Zhu. Robust and consistent anchor graph learning for multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 36 (8):4207–4219, 2024a.

Xiaohao Liu, Xiaobo Xia, See-Kiong Ng, and Tat-Seng Chua. Principled multimodal representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(8):9114–9128, 2026.

Ying Liu, Lihong Liu, Cai Xu, Xiangyu Song, Ziyu Guan, and Wei Zhao. Dynamic evidence decoupling for trusted multi-view learning. In *Proceedings of the ACM International Conference on Multimedia*, pp. 7269–7277, 2024b.

Jueqing Lu, Wray Buntine, Yuanyuan Qi, Joanna Dipnall, Belinda Gabbe, and Lan Du. Navigating conflicting views: Harnessing trust for learning. In *Proceedings of the International Conference on Machine Learning*, 2025.

Saiyang Na, Feng Jiang, Qifeng Zhou, Wenliang Zhong, Thao M Dang, Yuzhi Guo, Hehuan Ma, Chunyuan Li, Weizhi An, and Junzhou Huang. Hyperbolic gramian volumes for multimodal alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 37756–37765, 2026.

M-E Nilsback and Andrew Zisserman. A visual vocabulary for flower classification. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pp. 1447–1454, 2006.

A. Opelt, A. Pinz, M. Fussenegger, and P. Auer. Generic object recognition with boosting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):416–431, 2006.

Yalan Qin, Xinpeng Zhang, Shui Yu, and Guorui Feng. A survey on representation learning for multi-view data. *Neural Networks*, 181:106842, 2025.

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

Ariadna Quattoni and Antonio Torralba. Recognizing indoor scenes. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pp. 413–420, 2009.

Nikhil Rasiwasia, Jose Costa Pereira, Emanuele Coviello, Gabriel Doyle, Gert RG Lanckriet, Roger Levy, and Nuno Vasconcelos. A new approach to cross-modal multimedia retrieval. In *Proceedings of the ACM International Conference on Multimedia*, pp. 251–260, 2010.

Jongbin Ryu, Dongyoon Han, and Jongwoo Lim. Gramian attention heads are strong yet efficient vision learners. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5818–5828, 2023.

Yuan Sun, Yang Qin, Yongxiang Li, Dezhong Peng, Xi Peng, and Peng Hu. Robust multi-view clustering with noisy correspondence. *IEEE Transactions on Knowledge and Data Engineering*, 36(12):9150–9162, 2024.

Yuan Sun, Yongxiang Li, Zhenwen Ren, Guiduo Duan, Dezhong Peng, and Peng Hu. Roll: Robust noisy pseudo-label learning for multi-view clustering with noisy correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 30732–30741, 2025.

Chuan Tang, Miaomiao Li, Jun Wang, Renxiang Guan, Siwei Wang, Chang Tang, En Zhu, and Xinwang Liu. Align then tensorize: Multi-level consistent anchor graph learning for scalable multi-view clustering. *IEEE Transactions on Image Processing*, 35:7049–7062, 2026.

Hao Wang, Yan Yang, and Bing Liu. Gmc: Graph-based multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 32(6):1116–1129, 2019a.

Hao Wang, Linlin Zong, Bing Liu, Yan Yang, and Wei Zhou. Spectral perturbation meets incomplete multi-view data. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 3677–3683, 2019b.

Lior Wolf, Tal Hassner, and Itay Maoz. Face recognition in unconstrained videos with matched background similarity. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pp. 529–534, 2011.

Jianlong Wu, Zhouchen Lin, and Hongbin Zha. Essential tensor learning for multi-view spectral clustering. *IEEE Transactions on Image Processing*, 28(12):5910–5922, 2019.

Zhihao Wu, Jinyu Cai, Yunhe Zhang, Jielong Lu, Zhaoliang Chen, Shuman Zhuang, and Haishuai Wang. Where graph meets heterogeneity: multi-view collaborative graph experts. *Advances in Neural Information Processing Systems*, pp. 80096–80117, 2026.

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

Cai Xu, Jiajun Si, Ziyu Guan, Wei Zhao, Yue Wu, and Xiyue Gao. Reliable conflictive multi-view learning. In *Proceedings of the AAAI conference on artificial intelligence*, pp. 16129–16137, 2024.

Jie Xu, Na Zhao, Gang Niu, Masashi Sugiyama, and Xiaofeng Zhu. Robust multi-view learning via representation fusion of sample-level attention and alignment of simulated perturbation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4232–4241, 2025.

Yi Yang and Shawn Newsam. Bag-of-visual-words and spatial extensions for land-use classification. In *Proceedings of the SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 270–279, 2010.

Zixiao Yu, Lele Fu, Yongyong Chen, Zhiling Cai, and Guoqing Chao. Hyper-laplacian regularized concept factorization in low-rank tensor space for multi-view clustering. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 9(2):1728–1742, 2024.

Changqing Zhang, Yajie Cui, Zongbo Han, Joey Tianyi Zhou, Huazhu Fu, and Qinghua Hu. Deep partial multi-view learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44 (5):2402–2415, 2020.

Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun. Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, 38:43–54, 2017.

Qiyu Zhong, Yi Shan, Haobo Wang, Zhen Yang, and Gengyu Lyu. Tensorized multi-view multi-label classification via laplace tensor rank. In *Proceedings of the International Conference on Machine Learning*, 2025.

Xin Zou, Ruimeng Liu, Chang Tang, Zhenglai Li, Xinwang Liu, Kunlun He, and Wanqing Li. Learning disentangled representations for generalized multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(9):10489–10503, 2026.

APPENDIX CONTENTS

A Related Work **15**
A.1 Multi-view Learning 15
A.2 Gramian Volume 15
A.3 Noisy Correspondence. 15
B The Details of Datasets and Compared Methods **15**
B.1 Datasets 15
B.2 Compared Methods. 17
C Implementation Details **18**
D Performance under Different Noise Ratios **18**
E Ablation Study of Gramian Geometry **18**
F Parameter Sensitivity Analysis **20**
G Comparison of Reliability Density Visualization **21**
H Convergence Verification **21**

A Related Work

A.1 Multi-view Learning

MVL (Li et al., 2018; Fu et al., 2020; Berahmand et al., 2025) is a modeling paradigm for multi-source data analysis, which enhances the discriminability of sample representations by exploiting the complementarity and consistency across different views. Some representative studies have played pivotal roles in advancing the field of MVL. For example, tensor-based approaches (Wu et al., 2019; Fu et al., 2022; Yu et al., 2024; Zhong et al., 2025; Tang et al., 2026; Kong et al., 2026) captured cross-view consistent structures by leveraging low-rank tensor decomposition. Graph-based approaches (Wang et al., 2019a; Chen et al., 2023; Liu et al., 2024a; Fu et al., 2025; Wu et al., 2026; Dong et al., 2026) exploited a shared manifold structure to regularize the learning of view-consistent representations. Deep learning-based approaches (Zhang et al., 2020; Dong et al., 2023; Fu et al., 2024; Zou et al., 2026; Chen et al., 2026) employed neural networks to model nonlinear relationships among multi-view data and integrated compact representations in a low-dimensional latent space. These methods typically rely on a fundamental assumption that cross-view features of all samples are correctly aligned. However, this assumption is often violated in the presence of cross-view NC, causing existing MVL methods to severely degrade.

A.2 Gramian Volume

GV (Gatys et al., 2016; Ryu et al., 2023) is a metric that characterizes high-order geometric relationships in a representation space. For a set of feature representations, the corresponding GM encodes the pairwise relationships among different feature vectors. In general, a smaller GV indicates stronger dependencies among features and a more compact geometric structure, whereas a larger GV implies that the features are more dispersed and exhibit greater geometric diversity. In contrast to conventional metrics that focus solely on pairwise similarities, GV is capable of holistically characterizing the high-order geometric dependencies among multiple representations. Hence, it is widely adopted in multimodal representation learning (Cicchetti et al., 2025; Na et al., 2026; Liu et al., 2026) to minimize the discrepancy among representations from different modalities. Motivated by this property, we further employ GV to capture the global geometric relationships among multi-view representations, and exploit complementary structural information across different geometric spaces to facilitate view reliability modeling under NC.

A.3 Noisy Correspondence

NC commonly arise in modeling of paired data, such as MVL (Sun et al., 2024; 2025), cross-modal retrieval (Qin et al., 2022; Dang et al., 2025), and vision-language model training (Li et al., 2022; Huang et al., 2024b), where a subset of paired samples are incorrectly aligned. NC cause models to learn erroneous cross-view or cross-modal mappings during training, thereby severely compromising the accuracy and reliability of inference. In MVL under NC, some studies (Huang et al., 2024a; Sun et al., 2024; 2025; He et al., 2026) mitigated the adverse effects of mismatched samples during training and inference through enhanced robustness strategies or reliability estimation. However, they largely overlook the contribution of the global geometric structure of samples. To address this limitation, we propose to assess the reliability of view-specific features using a hybrid Gramian geometry, thereby guiding late prediction fusion.

B The Details of Datasets and Compared Methods

B.1 Datasets

GRAZ02 (Opelt et al., 2006) contains 1,476 object images from 4 classes, represented by 6 visual features, including Speeded-Up Robust Features (SURF), Scale-Invariant Feature Transform (SIFT), Local Binary Pattern (LBP), Wavelet Transform (WT), Pyramid Histogram of Oriented Gradients (Pyramid HOG), and GIST descriptor (GIST). The dimension of each views is 512, 32, 256, 500, 500, and 680, respectively.

Caltech8677 (Fei-Fei et al., 2004) comprises 8,677 object images across 101 categories. The three handcrafted views are Pyramid Histogram of Visual Words (PHOW), LBP, and CENsus TRansform

Algorithm 1 The main steps of the proposed HGR

Input: Training MV dataset $\mathcal{D} = \{\{\mathbf{x}_i^v\}_{v=1}^V, y_i\}_{i=1}^N$, trade-off parameters λ_b, λ_v and λ_c, the augmentation ratio p, the number of training epochs E, the learning rate η.

Output: The predicted labels $\{\hat{y}_i\}_{i=1}^{N_{test}}$.

Training Phase:
for $e = 1 : E$ **do**
  Generate random subset of samples $\mathcal{S}$ for NC construction;
  Construct the NC augmented MV dataset $\hat{\mathcal{D}}$ by Eq. 3;
  for $i = 1 : N$ **do**
    for $v = 1 : V$ **do**
      Obtain the embedding and logit: $\mathbf{z}_i^v = f^v(\hat{\mathbf{x}}_i^v), \mathbf{p}_i^v = g^v(\mathbf{z}_i^v)$;
      Obtain the EGM $\mathbf{G}_i^E$ and HGM $\mathbf{G}_i^H$ by Eqs. 11 and 17;
      Obtain the EGV $\mathbf{V}_i^E$ and HGV $\mathbf{V}_i^H$ by Eqs. 12 and 18;
      Calculate the hybrid GV $\mathbf{V}_i$ by Eq. 21;
      Obtain the reliability representation $\mathbf{r}_i$ by Eq. 19;
      Calculate the reliability score $\omega_i^v = \sigma[\Phi(\mathbf{r}_i^v)]$;
    end for
    Calculate the fused logit $\mathbf{p}_i$ by Eq. 7;
    Calculate the CE loss $\mathcal{L}_{ce}$, the BCE loss $\mathcal{L}_{bce}$, the volume loss $\mathcal{L}_{vol}$, and the consistency loss $\mathcal{L}_{con}$ by Eqs. 8, 10, 24;
    Calculate the overall loss $\mathcal{L} = \mathcal{L}_{ce} + \lambda_b\mathcal{L}_{bce} + \lambda_v\mathcal{L}_{vol} + \lambda_c\mathcal{L}_{con}$;
    for $v = 1 : V$ **do**
      Update the view-specific encoders and classifiers: $f_v^e = f_v^{e-1} - \eta\nabla\mathcal{L}$, $g_v^e = g_v^{e-1} - \eta\nabla\mathcal{L}$;
    end for
  end for
end for
Inference Phase:
for $i = 1 : N_{test}$ **do**
  for $v = 1 : V$ **do**
    Obtain the embedding and logit: $\mathbf{z}_i^v = f^v(\mathbf{x}_i^v), \mathbf{p}_i^v = g^v(\mathbf{z}_i^v)$;
    Obtain the EGM $\mathbf{G}_i^E$ and HGM $\mathbf{G}_i^H$ by Eqs. 11 and 17;
    Obtain the reliability representation $\mathbf{r}_i$ by Eq. 19;
    Calculate the reliability score $\omega_i^v = \sigma[\Phi(\mathbf{r}_i^v)]$;
  end for
  Calculate the fused logit $\mathbf{p}_i$ by Eq. 7 and obtain the predicted label $\hat{y}_i$ by Eq. 26;
end for

hISTogram (CENTRIST), and an additional view is obtained by using Inception V3 to extract deep features. The dimension of each view is 2048, 4800, 3540, and 1240, respectively.

flower17 (Nilsback & Zisserman, 2006) consists of 1,360 images spanning 17 flower categories, represented by seven precomputed distance matrices, each corresponding to a distinct view. The dimension of each view is 1360.

MITIndoor (Quattoni & Torralba, 2009) consists of 5,360 images covering 67 classes. In addition to PHOW, LBP, and CENTRIST, a type of deep features based on a pretrained VGG-VD [51] network is extracted. The dimension of each view is 3600, 1770, 1240, 4096, respectively.

100Leaves (Wang et al., 2019b) contains 1,600 leaf images from 100 plant species, represented by three visual views, including Texture Histogram (TH), Fine-Scale Margin (FSM), and Shape Descriptor (SD), each with 64 dimensions.

Wikipedia (Rasiwasia et al., 2010) comprises 693 image-text pairs across 10 semantic categories. The image and text modalities are represented by 128-dimensional SIFT features and 10-dimensional Linear Discriminant Analysis (LDA) features, respectively.

LandUse (Yang & Newsam, 2010) contains 2,100 satellite images from 21 landuse categories, represented by three visual views, including GIST, PHOG, and LBP features with dimensions of 20, 59, and 40, respectively.

MNIST10k (Asuncion et al., 2007) contains 10,000 handwritten digit images from 10 classes, represented by Isomap (ISO), LDA, and Neighborhood Preserving Embedding (NPE). The dimension of each view is 30, 9, 30, respectively.

CCV (Jiang et al., 2011) contains 6,773 video samples from 20 categories and is represented by three views corresponding to SIFT, Space-Time Interest Points (STIP), and Mel-Frequency Cepstral Coefficients (MFCC) features. The dimension of each view is 20.

NUSWIDEOBJ (Chua et al., 2009) contains 30,000 images from 31 object categories, represented by five visual views, including Color Histogram (CH), Color Moments (CM), Color Correlation (CORR), Edge Distribution Histogram (EDH), and Wavelet Texture (WT), with feature dimensions of 65, 226, 145, 74, and 129, respectively.

AWA (Lampert et al., 2013) comprises 30,475 images from 50 animal categories, represented by six visual views, including CH, Local Self-Similarity (LSS), PHOG, SIFT, Color SIFT, and SURF, with feature dimensions of 2,688, 2,000, 252, 2,000, 2,000, and 2,000, respectively.

YouTubeFace (Wolf et al., 2011) contains 101,499 samples from 31 classes, represented by five views with feature dimensions of 64, 512, 64, 647, and 838, respectively.

B.2 Compared Methods

TMC (Han et al., 2021) proposed a trusted multi-view classification paradigm that dynamically assessed per-sample view reliability via uncertainty estimation. By integrating view-specific evidence through a variational Dirichlet and Dempster-Shafer theory, it achieved reliable uncertainty quantification and robustness against noisy or out-of-distribution data.

UIMC (Xie et al., 2023) proposed an uncertainty-induced incomplete multi-view classification model that modeled missing views as distributions conditioned on available views instead of deterministic imputations. It then adopted an evidence-based fusion strategy to adaptively integrate the sampled imputations by their quality, enabling trustworthy classification.

ECML (Xu et al., 2024) introduced a reliable conflictive MVL problem that required models to output both decisions and reliabilities for conflictive instances rather than merely discarding them. To address it, an evidential conflictive MVL method was developed to learn view-specific evidence and aggregate the resulting opinions via a conflictive opinion aggregation strategy.

CCML (Liu et al., 2024b) proposed a consistent and complementary trusted MVL method to address semantic vagueness that suppressed view-specific evidence learning. It constructed view opinions via evidential deep neural networks and dynamically decoupled consistent evidence aligned with ground truth from complementary evidence allowed for vagueness.

ETF (Lu et al., 2025) developed a computational trust-based discounting method that enhanced the evidential multi-view framework by weighting each view according to its instance-wise reliability via a probability-sensitive trust mechanism, relaxing the assumptions of perfect alignment and equal view importance.

TMCEK (Liang et al., 2025) proposed a trusted multi-view classification framework constrained with expert knowledge that enhanced feature-level interpretability and introduced a distribution-aware subjective opinion mechanism for more reliable confidence estimation.

FUML (Duan et al., 2025) proposed a deep fuzzy MVL method that modeled outputs as fuzzy memberships to quantify category credibility and estimated uncertainty in conflicting instances. An entropy-based uncertainty measure and a dual reliable fusion strategy guaranteed the reliability.

MAMC (Lin et al., 2025) proposed a multi-view classification method with a multi-scale alignment module combining instance- and prototype-scale alignment to alleviate feature heterogeneity. A fuzzy-set based expanded class boundary was integrated into prototype-scale alignment to accommodate ambiguous data, reducing boundary ambiguity and improving generalization.

RML (Xu et al., 2025) proposed a robust MVL method that fused heterogeneous views into homogeneous embeddings via a transformer with sample-level attention. A perturbation-based contrastive learning framework further aligned fused representations under imperfect conditions, yielding discriminative and robust representations for downstream tasks.

BML (He et al., 2026) roposed a bootstrapping MVL framework that constructed a controllable noise-augmented training set to close the train-test gap for test-time NC. A lightweight estimator jointly modeled intra-view uncertainty and inter-view discrepancy to produce calibrated reliability weights that suppressed corrupted views during fusion.

Table 4: Details of the used datasets.

Dataset	Samples	Views	Classes	Dimensions
GRAZ02	1,476	6	4	[512, 32, 256, 500, 500, 680]
Caltech8677	8,677	4	101	[2048, 4800, 3540, 1240]
flower17	1,360	7	17	[1360, 1360, 1360, 1360, 1360, 1360, 1360]
MITIndoor	5,360	4	67	[3600, 1770, 1240, 4096]
100Leaves	1,600	3	100	[64, 64, 64]
Wikipedia	693	2	10	[128, 10]
LandUse	2,100	3	21	[20, 59, 40]
MNIST10k	10,000	3	10	[30, 9, 30]
CCV	6,773	3	20	[20, 20, 20]
NUSWIDEOBJ	30,000	5	31	[65, 226, 145, 74, 129]
AWA	30,475	6	50	[2688, 2000, 252, 2000, 2000, 2000]
YouTubeFace	101,499	5	31	[64, 512, 64, 647, 838]

C Implementation Details

Each view is encoded by an independent four-layer fully connected encoder with dimensions [1024, 1024, 1024, 128], mapping the input to a 128-dimensional latent embedding. Each view is equipped with an independent linear classification head, which outputs a predicted label. The Adam optimizer with a learning rate of 0.002 is adopted, the batch size is set as 2048. For the three trade-off parameters in loss function, λ_b is set to 1.0, λ_v and λ_c are tuned in $\{0.01, 0.05, 0.1, 0.5, 1.0\}$. In the training phase, the augment ratio of NC is set to 0.5. Each dataset is split into training and test sets at a ratio of 8:2. At inference time, the NC ratio ranges in [0, 1] with an interval of 0.1. For each dataset, we run experiments with five different random seeds and report the mean and variance. All experiments are implemented in PyTorch and conducted on a single NVIDIA RTX 5090 GPU.

D Performance under Different Noise Ratios

Under the clean setting with zero noise ratio, the proposed HGR achieves SOTA accuracy on most datasets in Table 5, while falling behind the best baseline on three datasets, including Caltech8677, MITIndoor, and CCV. This indicates that several competing baselines can yield competitive results on clean MV datasets owing to their ability of exploring MV complementary information. Nevertheless, the proposed HGR does not suffer from obvious performance drop and maintains decent adaptability for clean MV data. Furthermore, Fig. 6 presents performance curves of all compared methods across varying noise ratios. As the noise ratio increases, most baseline methods exhibit substantial performance decay. Nevertheless, HGR obtains the best results under the vast majority of noise levels. These observations verify that the proposed reliability evaluation built upon Gramian geometry can robustly identify mismatched views over a broad spectrum from clean to highly noisy inputs, preserving strong classification robustness.

E Ablation Study of Gramian Geometry

We conduct ablation experiments to investigate the individual contributions of Euclidean and Hyperbolic Gramian components. Three configurations are compared: (1) The variant with Euclidean

Table 5: Comparison of Classification performance (ACC±STD) without NC. The best results are highlighted in **bold**, and the second-best results are underlined.

Method	GRAZ02	Caltech8677	flower17	MITIndoor	100Leaves	Wikipedia
TMC	67.06±2.40	94.15±0.51	60.77±2.88	65.79±2.47	95.59±0.81	60.86±2.88
UIMC	65.30±2.42	94.82±0.42	76.47±2.06	71.96±1.29	93.94±1.42	59.71±2.51
ECML	63.41±2.43	93.06±0.69	43.86±4.57	61.97±2.50	93.47±1.23	60.86±2.43
CCML	67.74±1.87	93.31±0.84	62.98±3.03	65.58±2.20	96.81±0.65	60.36±2.99
MAMC	71.15±2.68	70.00±1.16	61.25±3.07	39.56±2.49	89.84±1.43	61.30±2.38
ETF	70.78±1.31	**95.78±0.34**	84.67±2.55	**75.41±0.89**	98.44±0.31	57.84±2.45
TMCEK	61.32±3.24	93.98±0.82	42.61±8.62	63.36±1.04	93.53±1.21	61.65±2.41
FUML	30.88±6.46	83.47±0.74	74.56±5.30	61.85±1.57	99.25±0.53	44.82±5.63
RML	66.45±1.10	87.47±1.09	78.97±2.10	72.46±1.16	97.25±0.78	29.78±2.21
BML	72.43±1.17	91.37±0.26	89.71±1.38	66.55±1.85	99.75±0.27	65.47±1.76
HGR	**73.24±1.86**	91.51±0.46	**90.73±0.59**	67.95±0.54	**99.94±0.12**	**68.06±1.74**
Method	LandUse	MNIST10k	CCV	NUSWIDEOBJ	AWA	YouTubeFace
TMC	46.14±1.72	89.04±0.40	44.58±0.77	39.72±0.48	36.91±0.42	75.44±0.34
UIMC	49.90±1.58	91.03±0.48	45.48±1.22	41.27±0.64	35.98±0.43	81.95±0.23
ECML	44.64±1.75	88.43±0.33	42.59±1.03	35.34±0.46	32.92±0.53	52.52±1.15
CCML	41.40±1.96	88.73±0.28	43.62±0.57	38.09±0.55	31.56±0.84	40.36±0.59
MAMC	71.45±1.82	96.43±0.26	54.40±1.46	46.07±0.48	32.24±0.54	87.16±0.24
ETF	46.10±1.71	89.42±0.46	48.40±0.79	45.98±0.41	48.38±0.35	83.01±0.25
TMCEK	45.10±2.17	87.82±0.28	42.90±0.92	35.40±0.56	33.70±0.45	53.69±1.20
FUML	74.76±1.63	93.67±0.59	47.39±2.40	47.17±0.42	38.15±0.50	83.18±0.29
RML	75.48±2.39	95.82±0.38	47.61±0.92	44.18±0.48	41.44±0.44	86.54±0.35
BML	81.43±1.06	96.88±0.16	**59.56±0.99**	51.91±0.40	48.90±0.27	89.72±0.22
HGR	**82.38±0.52**	**97.25±0.19**	59.49±0.82	**52.10±0.42**	**49.14±0.27**	**90.06±0.16**

Gramian geometry (EGG) only, where HGM and HGV are removed. (2) The variant with Hyperbolic Gramian geometry (HGG) only, where EGM and EGV are removed. (3) The full model with both EGM and HGM as well as their corresponding EGV and HGV. Table 6 reports the ablation experiment results, which reveal some notable phenomenon.

On the one hand, the variant with EGG only yields the second-best performance across most datasets. Euclidean space provides basic global discrimination and mitigates partial mismatching view cases. However, limited by variance collapse issues, Euclidean geometry alone cannot sufficiently capture high-order structural distortions induced by view mismatches. On the other hand, the variant with HGG only achieves even worse performance. Although hyperbolic geometry preserves rich structural variance via exponential volume growth, it lacks the cross-class discriminative stability offered by Euclidean geometry. Without Euclidean components, the model tends to overemphasize fine-grained geometric discrepancies and introduces undesired geometric bias, which interferes with the primary classification objective. The best performance is achieved only when both EGG and HGG are equipped. This verifies that HGG is indispensable for the optimal performance, its exponential volume property compensates the drawbacks of Euclidean geometry. The complementary combination of both geometries can fully exploit MV geometric cues and improves the estimation of views.

Table 6: Ablation study of Euclidean and Hyperbolic Gramian Geometry, where the test NC ratio is set to 0.5.

EGG	HGG	GRAZ02	Caltech8677	flower17	MITIndoor	NUSWIDEOBJ	AWA
✗	✓	67.16±1.80	85.20±0.64	86.25±0.95	57.05±1.02	47.00±0.52	41.77±0.27
✓	✗	67.57±1.59	85.40±0.32	86.03±1.23	59.57±0.97	47.53±0.53	42.96±0.18
✓	✓	**68.58±2.91**	**85.73±0.51**	**86.47±1.34**	**59.89±1.42**	**47.57±0.22**	**43.15±0.27**

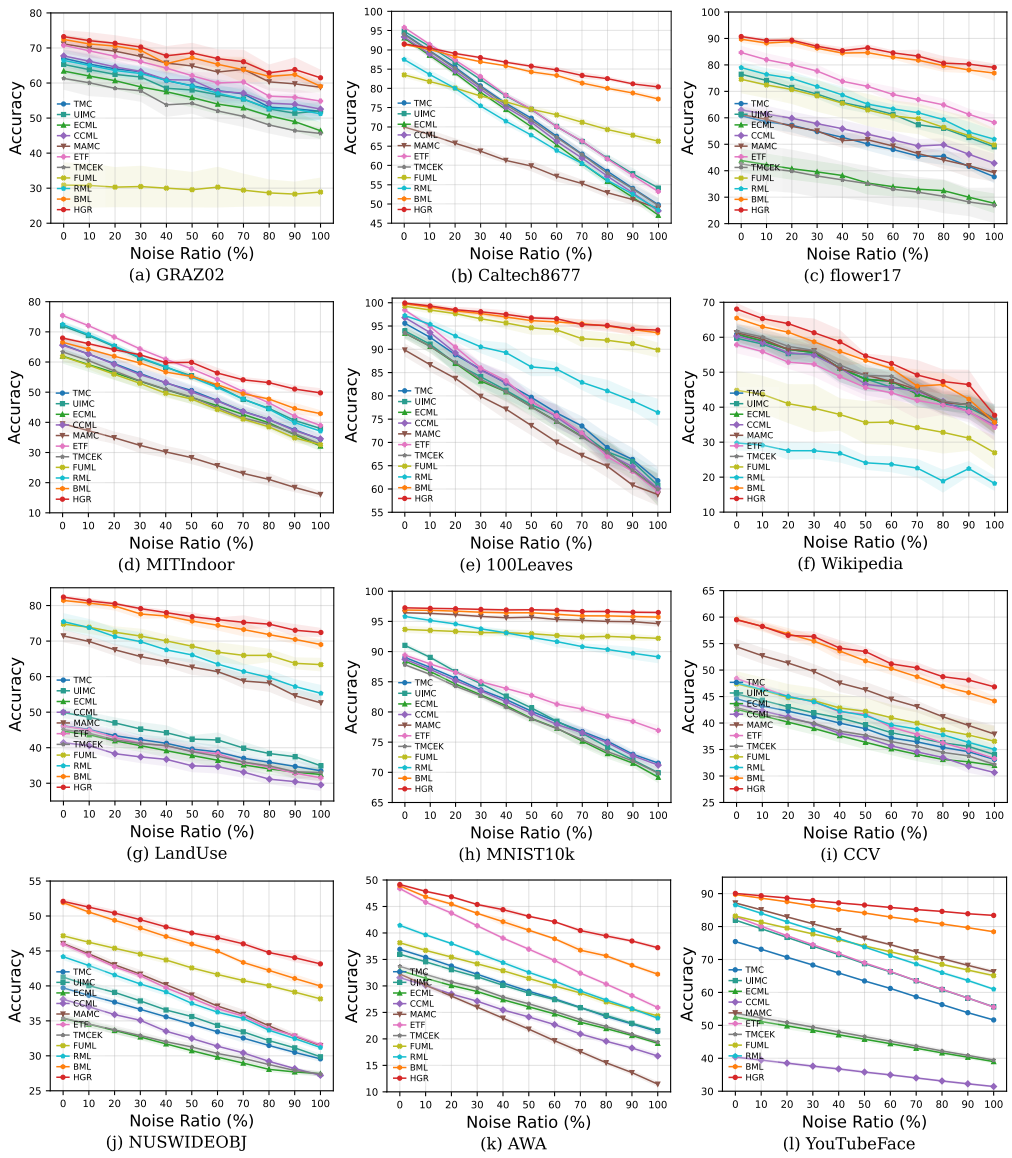
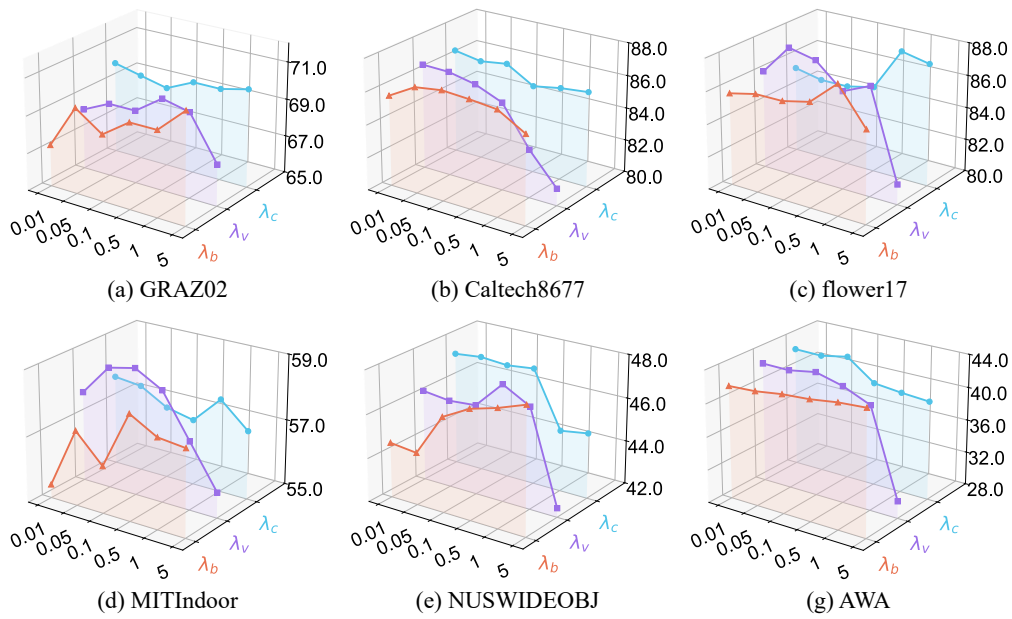


Figure 6: The performance of compared methods under different noise ratios.

F Parameter Sensitivity Analysis

In this section, we perform parameter sensitivity experiments on the three principle trade-off hyperparameters λ_b, λ_v, and λ_c, investigating the influence of different hyperparameter settings on classification performance. Fig. 7 demonstrates that a smaller λ_b leads to a marked decline in model performance. As λ_b controls the weight of the binary classification loss, an insufficient weight weakens the classification supervision and prevents the model from effectively distinguishing view reliability. In contrast, the model performance deteriorates as λ_v and λ_c increase, which correspond to the GV contrastive loss and the view consistency loss, respectively. When these two weights are excessively large, the geometric regularization terms dominate the training process, and the model over-prioritizes MV geometric consistency. This undermines the discriminative representations required by the original classification task, and ultimately leads to accuracy degradation.



(a) GRAZ02 (b) Caltech8677 (c) flower17

(d) MITIndoor (e) NUSWIDEOBJ (g) AWA

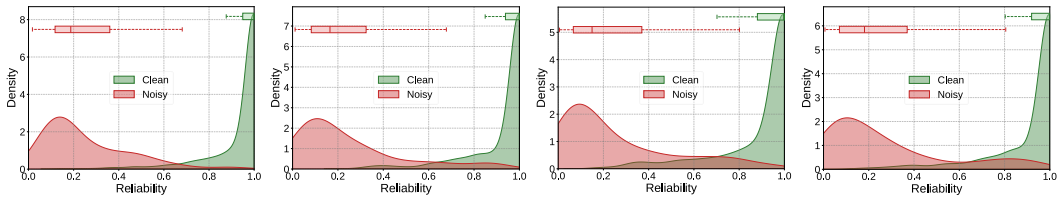
Figure 7: Parameter sensitivity analysis with respect to λ_b, λ_v, and λ_c, where the values are tuned in $\{0.01, 0.05, 0.1, 0.5, 1.0, 5.0\}$.

G Comparison of Reliability Density Visualization

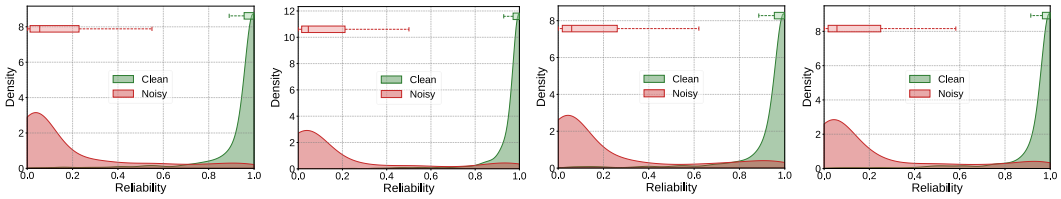
To intuitively compare reliability estimation behaviors, we plot both boxplots and density curves for view-specific reliability scores on the test set in Figs. 8, 9, 10, 11, 12, 13, 14. In the density curves of BML on some datasets, the scores of clean and noisy samples overlap substantially, indicating limited discriminability. Meanwhile, the boxplots show that the median gap between the clean and noisy groups of BML is relatively small. This is because pairwise statistics only capture local interactions, making it difficult to generate highly discriminative reliability scores. For the proposed HGR, the density curves show that clean views cluster in the high-score region and mismatched views cluster in the low-score region, with a clearly stronger separation between the two distributions. Combined with the boxplots, clean views have significantly higher median reliability scores than BML, while mismatched views have significantly lower median than BML. This phenomenon indicates that the reliability assessment mechanism based on hybrid Gramain geometry can achieve superior reliability estimation.

H Convergence Verification

To investigate training stability, we plot loss convergence curves over all twelve MV datasets in Fig. 15. We observe that the total loss decreases rapidly with growing training epochs and gradually stabilizes in later iterations. This observation demonstrates that the proposed loss objective, consisting of classification loss, binary classification loss, GV contrastive loss and view consistency loss, enables effective parameter optimization. After sufficient epochs, the loss exhibits little fluctuation, validating that the proposed loss objective exhibits favorable stability across MV datasets with varying sizes and view configurations.

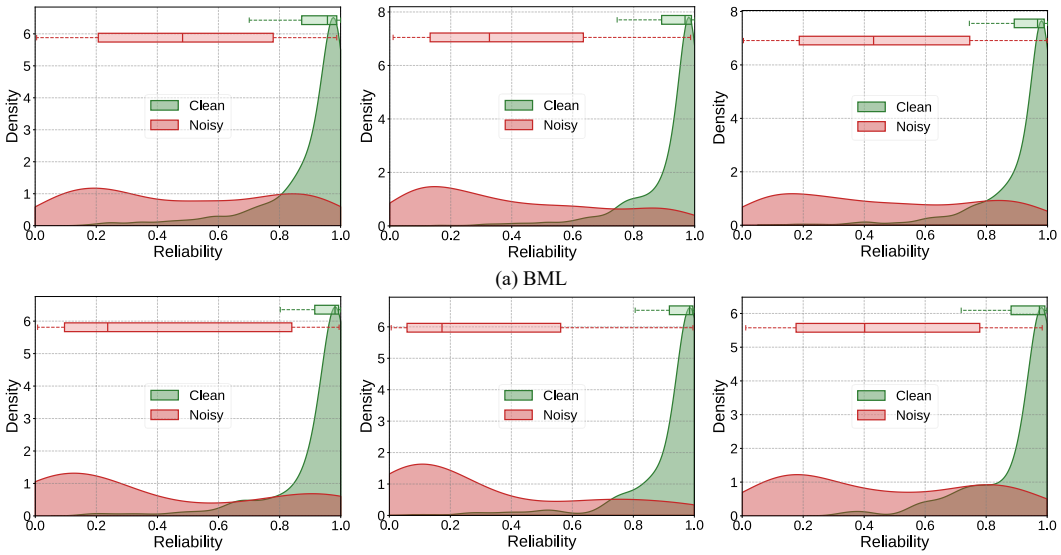


(a) BML



(b) HGR

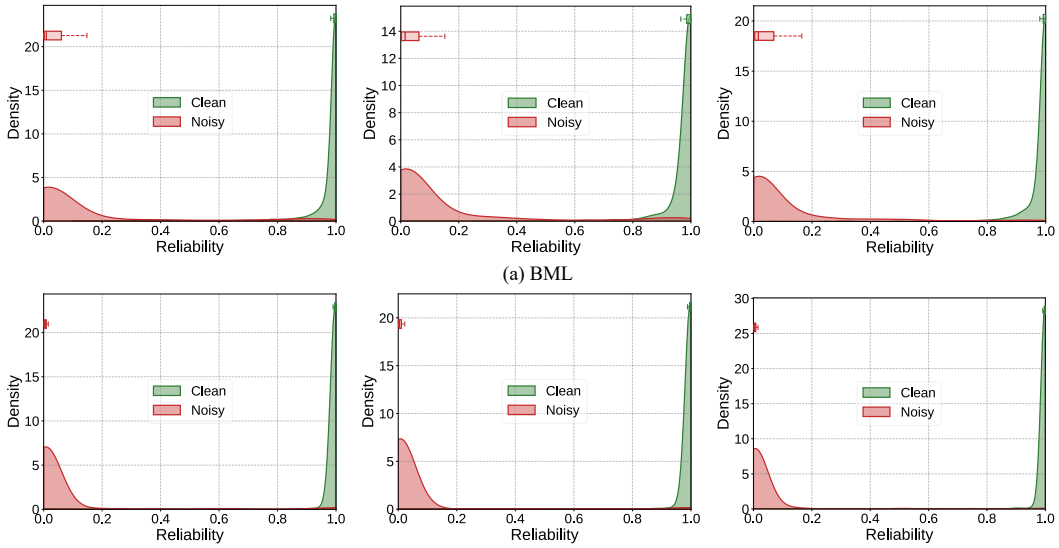
Figure 8: Reliability density comparison between BML and HGR on Caltech8677 dataset, where the test NC ratio is set to 0.5.



(a) BML

(b) HGR

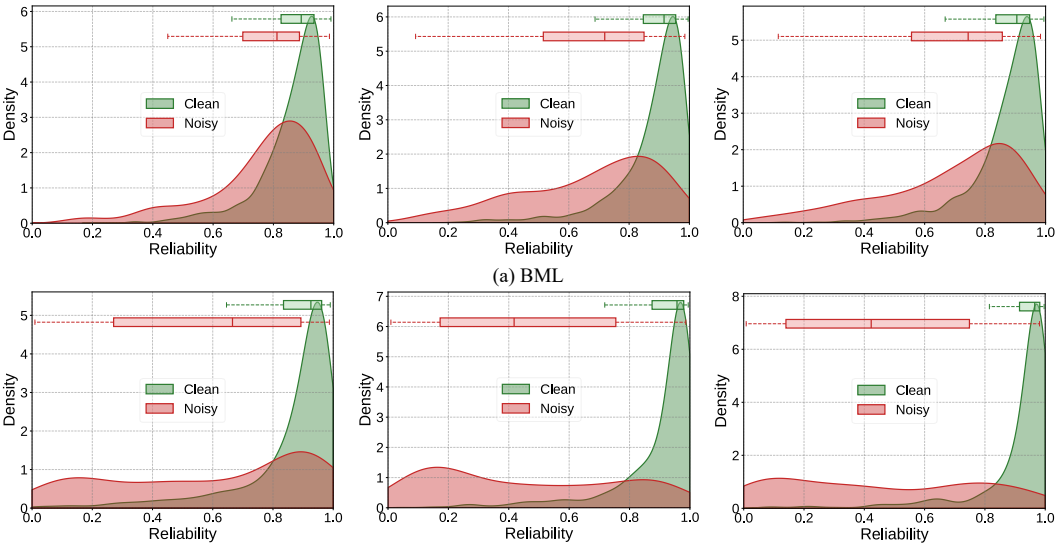
Figure 9: Reliability density comparison between BML and HGR on LandUse dataset, where the test NC ratio is set to 0.5.



(a) BML

(b) HGR

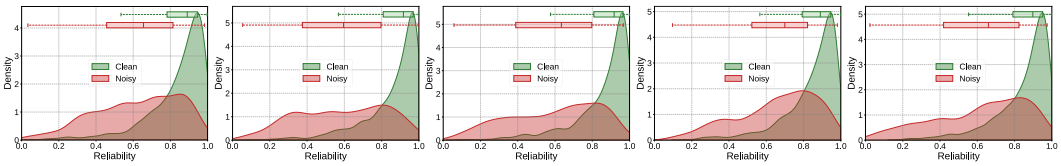
Figure 10: Reliability density comparison between BML and HGR on MNIST10k dataset, where the test NC ratio is set to 0.5.



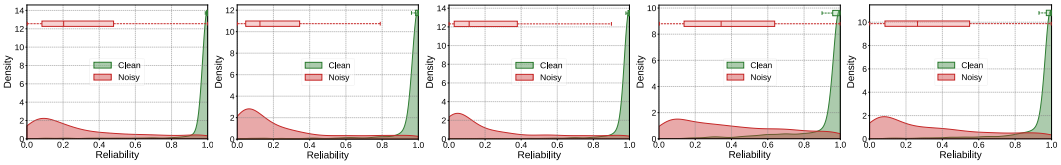
(a) BML

(b) HGR

Figure 11: Reliability density comparison between BML and HGR on CCV dataset, where the test NC ratio is set to 0.5.

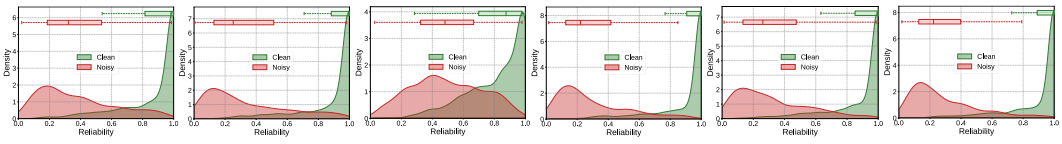


(a) BML

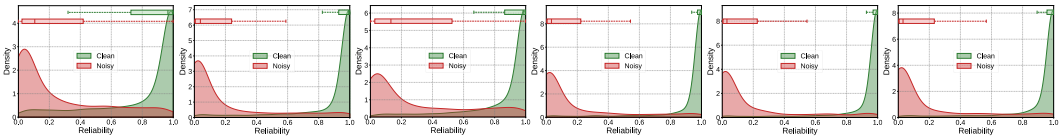


(b) HGR

Figure 12: Reliability density comparison between BML and HGR on NUSWIDEOBJ dataset, where the test NC ratio is set to 0.5.

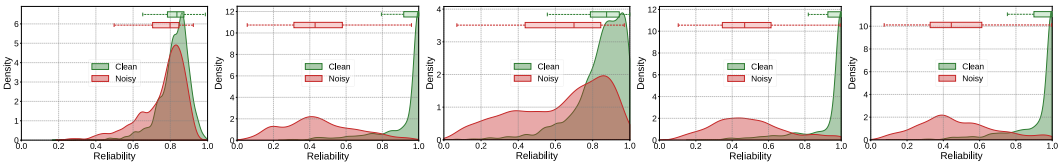


(a) BML

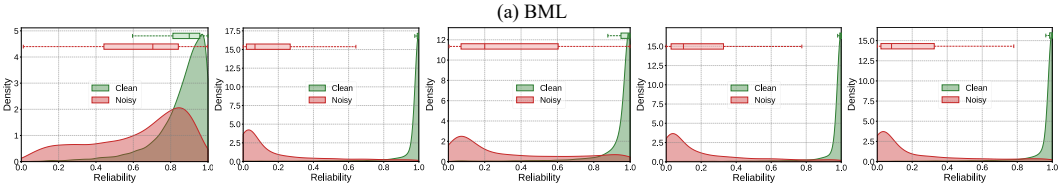


(b) HGR

Figure 13: Reliability density comparison between BML and HGR on AWA dataset, where the test NC ratio is set to 0.5.

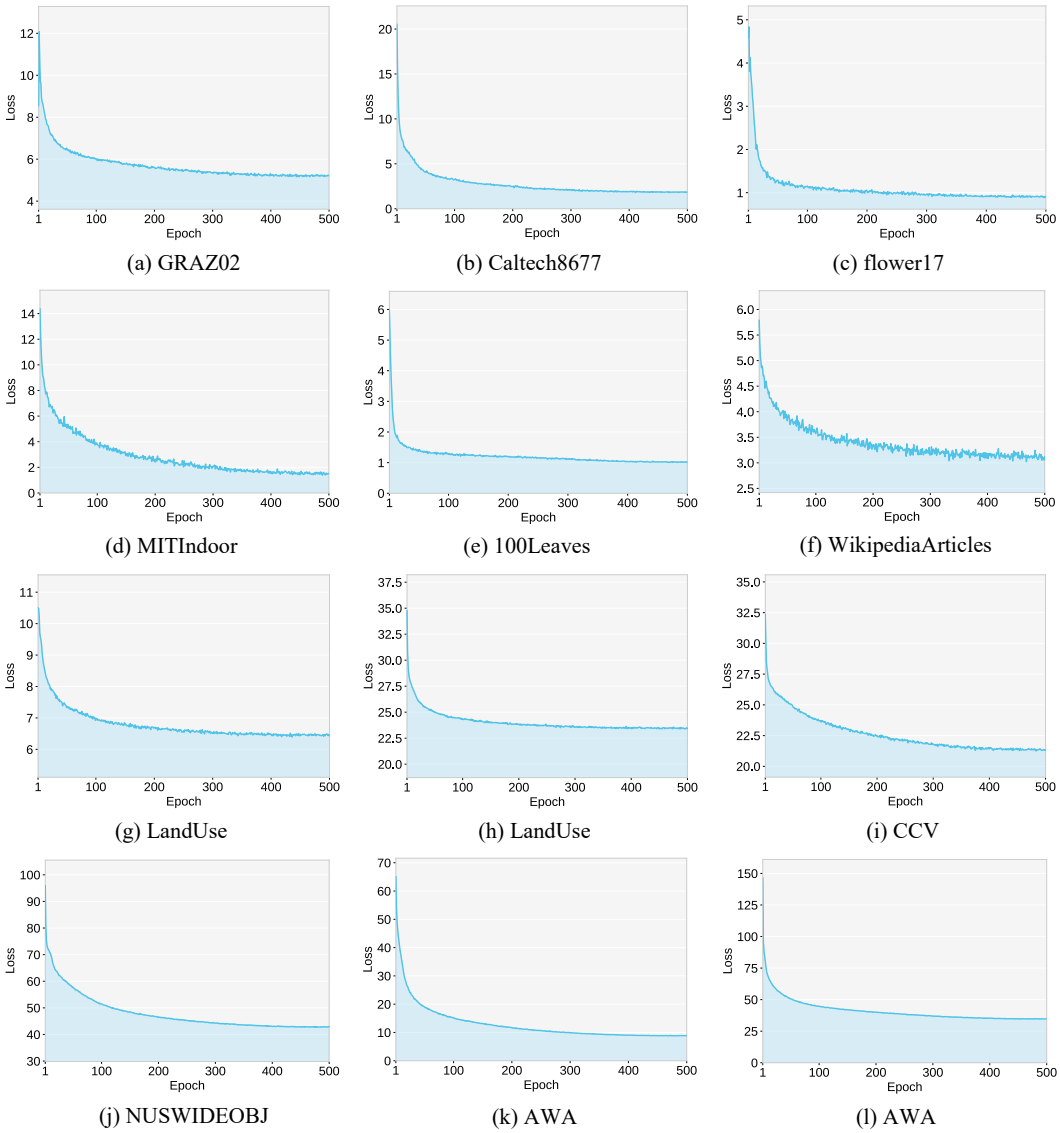


(a) BML



(b) HGR

Figure 14: Reliability density comparison between BML and HGR on YouTubeFace dataset, where the test NC ratio is set to 0.5.



(a) GRAZ02 (b) Caltech8677 (c) flower17

(d) MITIndoor (e) 100Leaves (f) WikipediaArticles

(g) LandUse (h) LandUse (i) CCV

(j) NUSWIDEOBJ (k) AWA (l) AWA

Figure 15: The loss curves with increasing training epochs on twelve MV datasets.