A Dynamic Learning Strategy for Dempster-Shafer Theory with Applications in Classification and Enhancement

Linlin Fan[1], Xingyu Liu[1], Mingliang Zhou[1]*, Xuekai Wei[1], Weizhi Xian[2], Jielu Yan[1], Weijia Jia[3]
[1]School of Computer Science, Chongqing University,
[2]Chongqing Research Institute of HIT, Harbin Institute of Technology,
[3]BNU-UIC Institute of Artificial Intelligence and Future Networks, Beijing Normal University
linlinfan@stu.cqu.edu.cn,xingyuliu@stu.cqu.edu.cn,mingliangzhou@cqu.edu.cn,
xuekaiwei2-c@my.cityu.edu.hk,wasxxwz@163.com, yanjielu@cqu.edu.cn,
jiawj@bnu.edu.cn

Abstract

Effective modelling of uncertain information is crucial for quantifying uncertainty. Dempster–Shafer evidence (DSE) theory is a widely recognized approach for handling uncertain information. However, current methods often neglect the inherent a priori information within data during modelling, and imbalanced data lead to insufficient attention to key information in the model. To address these limitations, this paper presents a dynamic learning strategy based on nonuniform splitting mechanism and Hilbert space mapping. First, the framework uses a nonuniform splitting mechanism to dynamically adjust the weights of data subsets and combines the diffusion factor to effectively incorporate the data a priori information, thereby flexibly addressing uncertainty and conflict. Second, the conflict in the information fusion process is reduced by Hilbert space mapping. Experimental results on multiple tasks show that the proposed method significantly outperforms state-of-the-art methods and effectively improves the performance of classification and low-light image enhancement (LLIE) tasks. The code is available at https://anonymous.4open.science/r/Third-ED16.

1 Introduction

Uncertain information is inevitably encountered in the process of modelling data-based complex systems via deep learning. Effective modelling and processing of uncertain information is an important technique that plays a key role in the decision-making process and improves the ultimate decision-making level. Currently, many methods have been proposed to solve this problem, such as evidence-theoretic methods [57] and fuzzy logic methods [10]. and deep learning-based methods [44]. These techniques have been applied in several fields, such as graph clustering [49], classification [56], and target detection [32]. However, how to effectively handle uncertain information from different sources and combine them effectively while avoiding mutual conflicts is still a challenging problem.

In machine learning, classification is the task of making predictions about new sample categories by learning the features of samples from known categories [26]. In the classification task, the reliability of the classifier plays an important role. However, owing to the uncertainty of the data itself and the possible conflict or redundancy of information from different sources, effectively and rationally addressing information from different sources to improve classification accuracy has become an urgent problem. Currently, many classical machine learning methods have been proposed. Denoeux

*Corresponding author: Mingliang Zhou.

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

 https://doi.org/10.52202/085713-0409

et al. [15] viewed the neighbors of unclassified samples as evidence for the hypothesis, with support as a function of distance between the vectors, and used Dempster's combination (DC) rule to combine the evidence for the classification task. Freund *et al.* [19] proposed the classification method of an alternating decision tree, which solved the problem that the original classifier is complicated and difficult to understand. Chang *et al.* [5] proposed a support vector machine (SVM) and an SVM with radial basis functions to implement classification problems. Xu *et al.* [60] utilized a normality test and normality transformation to address nonnormal data for classification tasks. Hu *et al.* [23] investigated Bayesian and mutual information classifiers and applied them in a classification task. Xu *et al.* [61] extended the classical probabilistic calibration approach to an evidence-theoretic framework when dealing with different sources of information to address the problem that a single probability measure may not adequately express uncertainty when modelling the calibration step. Xiao *et al.* [58] proposed weighted belief–jensen–shannon divergence for decision-making improvement on the basis of Dempster–Shafer evidence (DSE) theory. Although the above methods can address information from different sources, the importance of the information and the measurement of discrepancies are still not comprehensively considered. These shortcomings are reflected mainly in the following aspects:

- In DSE theory, the current method does not consider the a priori information brought by the data itself, nor can it dynamically adjust the tendency of splitting on the basis of the importance between different subsets in the process of splitting, which means that the uncertainty information cannot be well handled.

- When fusing multiple basic belief assignments (BBAs), different evidence sources conflict due to uncertainty or data distribution differences. Direct use of DC rules may lead to conflict amplification or even produce unreasonable results. Moreover, traditional methods are usually based on Euclidean space or simple statistical metrics to calculate the differences between evidence, which makes it difficult to capture the nonlinear characteristics of complex data distributions.

- In practical tasks, data distributions are often unbalanced, which can lead to insufficient focus on key information in the model. In addition, current methods are unable to customize the degree of attention according to different data, and traditional training methods treat all features equally, which may lead to insufficient learning of key regions.

To address the above problems, we propose a dynamic learning strategy based on nonuniform splitting mechanism and Hilbert space mapping. First, the framework dynamically adjusts the weights of different subsets through the nonuniform splitting mechanism and uses the a priori information of the data in combination with the diffusion factor to flexibly address uncertainties and conflicts. Second, the data are mapped into the Hilbert space for computation to alleviate the information conflict problem that may occur during the information fusion process. Third, a targeted training strategy is proposed to enhance the model's ability to learn important features and regions, which achieves results in both classification and enhancement tasks. In summary, the main contributions of our work are as follows:

- We propose a nonuniform splitting mechanism. This mechanism can be dynamically adjusted according to the importance between different subsets, giving more weight to some subsets and less weight to others. The a priori information of the data can be utilized by introducing a diffusion factor, and this splitting mechanism, which is based on a priori information, can be more flexible in addressing uncertainty and conflicting information.

- We propose a scheme for fusing different BBAs. This scheme maps the data into Hilbert space for computation, which is more responsive to the differences in the true distributions of complex data. A specific way to compute the differences before fusing different BBAs is used to reduce the conflicts between different information.

- We propose an effective targeted training strategy (TTS). This strategy enhances the model's ability to learn specific information and regions. Higher weights are assigned to important features to increase attention, thus improving the overall performance of the task. Accuracy is improved in classification tasks, and data imbalance is alleviated in low-light image enhancement (LLIE) tasks.

The rest of the paper is organized as follows. In section 2, related work is briefly introduced. section 3 describes the proposed method. section 4 describes the experiments and analysis of the results. Finally, section 5 provides a discussion and conclusion.

2 Related Work

2.1 Modelling of uncertain information

In classification tasks, information uncertainty often leads the classifier to make incorrect decisions. To address this problem effectively, many methods based on statistical and distance metrics have been proposed with the aim of improving classification performance by means of different mathematical models. Cover and Hart [12] proposed a method based on a distance metric function, which is inferred by selecting nearest neighbor samples. Cortes and Vapnik [11] effectively differentiate between different data distributions by constructing a separating hyperplane that can correctly divide the training dataset and has a maximum geometric interval. Xanthopoulos *et al.* [55] proposed discriminant analysis on the basis of a statistical approach that uses the grouping information of known samples and their corresponding multivariate variable characteristics to infer the group to which new samples belong. Quinlan *et al.* [41] designed a model that is based on a tree structure, where each internal node represents a judgment on a feature and each branch represents a possible output of the judgment. However, the above methods may have difficulty making accurate inferences because of their own limitations when dealing with data with conflicting or redundant information, thus affecting inference efficiency.

2.2 Evidence theory-based modelling of uncertain information

Uncertain information can be classified into either the empty set $\emptyset$ or the whole set Ω on the basis of the modelling of the uncertain information, thus addressing the information uncertainty. Zhao *et al.* [71] obtained the final classification results by evaluating the reliability of single and multiple sources through independent and combined reliability assessments, respectively. Liu *et al.* [38] combined the inferred results from multiple models and used the average as the final output. Jousselme *et al.* [30] introduced the distance calculation method of the similarity measure to generate more reliable inference results. Zhang *et al.* [69] proposed a multisource information fusion algorithm based on belief χ^2 scatter for inference tasks in complex data scenarios. However, owing to the uncertainty of the data itself and the possible conflicts or redundant information from different sources, effectively combining information from different sources by taking their importance into account remains a challenging research problem.

2.3 Deep learning-based modelling of uncertain information

Deep learning models often face the problem of decreased prediction reliability due to uncertain information. Chen *et al.* [8] proposed a radial basis function network learning algorithm based on orthogonal least squares to solve the underperformance problem caused by randomly selecting the centroid method to improve the performance of the task network. Castro *et al.* [4] investigated the problem of biased results due to data imbalance based on a multilayer perceptron (MLP) neural network and statistically improved the classification performance of the MLP. Sensoy *et al.* [45] used the prediction output from the network as subjective information for modelling, which in turn served as data support for the deterministic neural network to accomplish the subsequent classification task. Zaidi *et al.* [67] proposed two methods for automatically building a collection of different network architectures that can weigh the advantages of different structures well and use architectural variations as a source of diversity. However, these methods rely mostly on specific assumptions, and when there are multiple sources of conflicting contradictions in the input data, these models lack an effective dynamic processing framework, which may lead to bias in the final results.

 https://doi.org/10.52202/085713-0409

3 Method

3.1 Motivation and overview

In DSE theory, when dealing with highly conflicting information, existing methods cannot adequately consider the lack of precision due to ambiguity or uncertainty in BBAs. First, the existing allocation methods based on the splitting idea assume that the allocation is based on the premise of uniform splitting, which assumes that all subsets are equally proportioned to distribute the quality and cannot be dynamically adjusted according to the quality among different subsets. Thus, the differences between different subsets cannot be adequately considered. Therefore, a nonuniform splitting mechanism is introduced to give more weight to some subsets and less weight to others. This splitting mechanism, which is based on a priori information, namely the inherent characteristics of the data's own structure and the initial evidence distribution, can handle uncertainties and conflicting information more flesxibly. The mechanism is adjustable to assign different split weights to different subsets through the diffusion coefficient. Second, conflicts arise from different evidence sources due to uncertainty or data distribution differences. To avoid the occurrence of the conflict phenomenon in the process of fusing different BBAs, high-order dynamic maximum mean difference (HODMMD) is proposed, which maps the data into the Hilbert space for computation and is more responsive to the differences in the real distributions of complex data. A specific way to calculate the difference before fusing different BBAs is used to reduce the conflict between different information. Third, in practical applications, the imbalance problem inherent in the data leads to insufficient attention of the model to key information. Therefore, an effective TTS is needed to assign higher weights to important features and increase attention, thus improving the overall performance of the task.

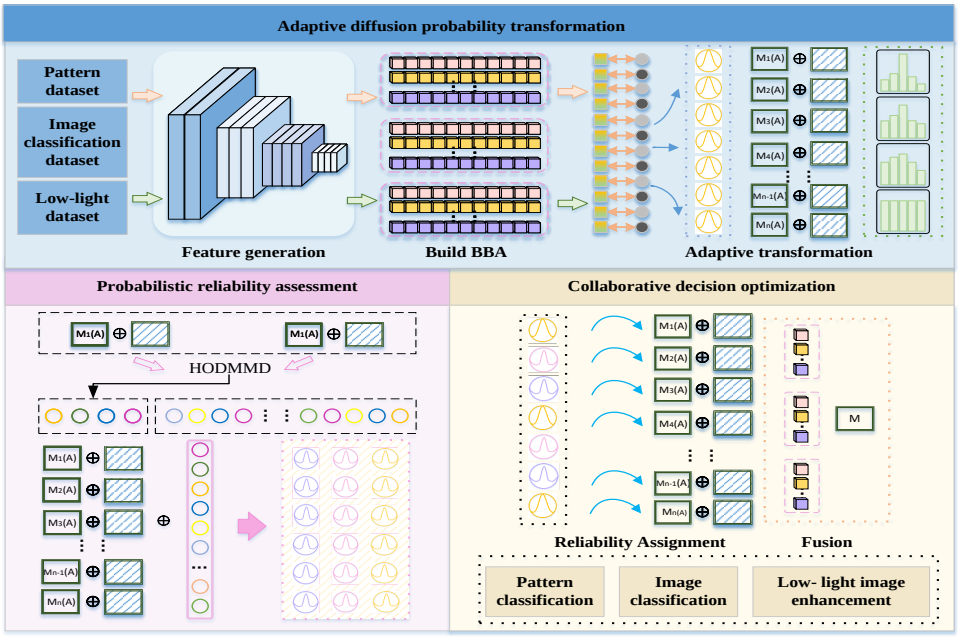

Figure 1: Overview of our targeted training strategy (TTS).

Given a set of data I, after feature extraction, different BBAs $\widehat{m_1}, \widehat{m_2}, \cdots, \widehat{m_n}$ are obtained via adaptive diffusion probability transformation (ADPT) (Tadpt):

$$\widehat{m_1}, \widehat{m_2}, \cdots, \widehat{m_n} = \mathrm{Tadpt}(I) \tag{1}$$

After that, the reliability of the different BBAs is calculated in the probabilistic reliability assessment (PRA) (Tpra) stage:

$$\dot{\phi_i} = \mathrm{Tpra}(\widehat{m_i}, \widehat{m_j}) \quad (1 \leq i, j \leq n) \tag{2}$$

These reliabilities are utilized as discount factors to perform collaborative decision optimization (CDO) (Tcdo) operations on the BBAs, which are fused to obtain the final decision result:

$$\widehat{m} = \mathrm{Tcdo}(\dot{\phi_1}, \dot{\phi_2}, \cdots, \dot{\phi_i}, \cdots \dot{\phi_n}; \quad \widehat{m_1}, \widehat{m_2}, \cdots, \widehat{m_i}, \cdots, \widehat{m_n}) \tag{3}$$

where $\dot{\phi}_1, \dot{\phi}_2, \cdots, \dot{\phi}_i, \cdots \dot{\phi}_n$ are discount factors. The final decision result is obtained after fusion.

3.2 Adaptive diffusion probability transformation

On the basis of the definition of DSE theory, the proposed ADPT model is as follows:

$$\dot{m}_g^\tau(A_i) = \sum_{A_i \subseteq A_j} D(A_i, A_j) \frac{\tau^{|A_j|-|A_i|}}{(\tau+1)^{|A_j|} - \tau^{|A_j|}} \dot{m}_g^{\tau-1}(A_j) \tag{4}$$

where $A_i, A_j \in 2^\Omega \backslash \{\emptyset\}$, τ is the number of iterative splits. $\dot{m}_g^\tau$ is τ order of $\dot{m}_g$, and $\dot{m}_g^0$ is the initial basic belief assignment (BBA). $D(A_i, A_j) = \left(\frac{|A_i|}{|A_j|}\right)^{\frac{1}{\tau}}$ is a diffusion function used to control the quality of the distribution from subset A_i to A_j, which regulates the process in which the mass function is transferred from a fuzzy hypothesis A_j to a more specific hypothesis A_i, effectively quantifying and integrating prior information. When τ tends to ∞, the result of $D(A_i, A_j)$ tends to 1, and Equation (4) degenerates into a uniform distribution [25], which is no longer affected by the importance of different regions. $|\cdot|$ is the number of elements contained, $\Omega = \{\theta_1, \theta_2, \theta_3, \ldots, \theta_N\}$ is the frame of discernment, and $2^\Omega = \{\emptyset, \{\theta_1\}, \{\theta_2\}, \ldots, \{\theta_N\}, \{\theta_1 \cup \theta_2\}, \ldots, \Omega\}$ is the power set. After normalization processing,

$$\widehat{m}_g^\tau(A_i) = \sum_{A_i \in 2^\Omega} \frac{\dot{m}_g^\tau(A_i)}{\sum_{A_i \in 2^\Omega} \dot{m}_g^\tau(A_i)} \tag{5}$$

3.3 Probabilistic reliability assessment and its properties

When different quality functions are obtained, if they are simply combined according to the DC rule [48], the integration between different BBAs may be hindered because of conflicting information. For this reason, the idea of discounting techniques [46] is introduced. Defined HODMMD:

$$\text{HODMMD}^\tau(\widehat{m}_1, \widehat{m}_2) = \left\| \sum_{i=1}^{N} \left(\widehat{m}_{g_1}^\tau(A_i) - \widehat{m}_{g_2}^\tau(A_i) \right) \right\|_{\mathcal{H}} \tag{6}$$

where $\mathcal{H}$ is the regenerated kernel Hilbert space and the kernel function is computed via a Gaussian kernel function. The properties of the HODMMD are as follows.

Property 1. When $\tau \to \infty$, the HODMMD is equivalent to measuring the difference between the pignistic probability transformations Bet$P$ with $\widehat{m}_1$ and $\widehat{m}_2$ of the maximum mean difference.

$$\lim_{\tau \to \infty} \text{HODMMD}^\tau(\widehat{m}_1, \widehat{m}_2) = \left\| \sum_{i=1}^{N} (\text{Bet}P_1(\theta_i) - \text{Bet}P_2(\theta_i)) \right\|_{\mathcal{H}} \tag{7}$$

Property 2. When $\widehat{m}_1$ and $\widehat{m}_2$ degenerate into probability distributions, that is, $U = (u_1, u_2, \ldots, u_N)$ and $V = (v_1, v_2, \ldots, v_N)$, the proposed HODMMD degenerates into a maximum mean difference.

$$\text{HODMMD}^\tau(\widehat{m}_1, \widehat{m}_2) = \text{MMD}(\widehat{m}_1, \widehat{m}_2) \tag{8}$$

Property 3. HODMMD$^\tau(\widehat{m}_1, \widehat{m}_2)$ and HODMMD$^\tau(\widehat{m}_2, \widehat{m}_1)$ are equivalent.

$$\text{HODMMD}^\tau(\widehat{m}_1, \widehat{m}_2) = \text{HODMMD}^\tau(\widehat{m}_2, \widehat{m}_1) \tag{9}$$

Property 4. When $\widehat{m}_1 = \widehat{m}_2$, the value of HODMMD is always equal to 0.

$$\text{HODMMD}^\tau(\widehat{m}_1, \widehat{m}_2) = 0 \tag{10}$$

Proofs of these properties can be found in the technical appendix.

3.4 Collaborative decision optimization

When ADPT is used to obtain multiple BBAs, the reliability of different BBAs needs to be calculated to measure the impact of different sources of evidence. Let $\phi = [\phi_1, \phi_2, ..., \phi_K]$ be the reliability of

 https://doi.org/10.52202/085713-0409

Algorithm 1: A dynamic learning framework based on DSE theory

Input: Training data $X_{training}$ and testing set $X_{testing}$.
Output: Category probabilistic decision results

```
for i = 1 to K do
    Generate values on the basis of the data attribute characteristics of X_training
end
Reliability of different classifiers obtained via decision optimization scheme Eq. (11)
Calculate the discount factor via Eq. (12) and normalize it
for i = 1 to K do
    Obtain the results of K classifiers
    Discounting the different classification results via Eq. (13)
    Fuse different BBAs via Eq. (14)
    Test the kth basic classifiers
end
Use the decision results for subsequent tasks
```

different query patterns or BBAs that satisfy $\phi_k \in [0,1]$ and $\sum_1^K \phi_k = 1$. To measure the reliability of the different BBAs, the HODMMD is used for the calculation:

$$\phi = \arg\min_{\phi} \left(\sum_{l=1}^{L} \text{HODMMD}^{\tau} \left(G^l, \sum_{k=1}^{K} \phi_k \widehat{m}_k^l \right) \right) \quad \text{s.t.} \sum_{k=1}^{K} \phi_k = 1 \tag{11}$$

where l is the index of the different query patterns. $G_l = [G_l(1), G_l(2), ..., G_l(K)]$ is the ground truth. m_k^l is the possibility of the query pattern belonging to the class θ_i. $\widehat{m}_k^l = [\widehat{m}_k^l(1), \widehat{m}_k^l(2), \ldots, \widehat{m}_k^l(\Omega)]$, which also satisfies $m_k^l \in [0,1]$ and $\sum_{k=1}^{2^N-1} m_k^l = 1$. The reliability vector that minimizes the BBA error is calculated via sequential least squares programming (SLSQP). Therefore, the discount factor for the kth BBA is defined as:

$$\dot{\phi_k} = \phi_k / \max\{\phi_1, \phi_2, \ldots, \phi_K\} \tag{12}$$

According to the idea of discounting techniques [46], the discounted BBA is as follows:

$$\begin{cases} \overline{m}_k(X) = \dot{\phi_k} \widehat{m}_k(X), \quad \forall X \subsetneq \Omega \\ \overline{m}_k(\Omega) = \dot{\phi_k} \widehat{m}_r(\Omega) + 1 - \dot{\phi_k} \end{cases} \tag{13}$$

Fusing of different BBAs according to the DC rule:

$$\mathbf{m} = \overline{m}_1 \oplus \overline{m}_2 \oplus \cdots \oplus \overline{m}_R \tag{14}$$

The probabilities of different patterns obtained after fusion are prepared for subsequent classification and LLIE tasks. In LLIE tasks, the image is divided into different image blocks and then input into the proposed decision framework $\mathrm{T_{rans}}$ to obtain the degradation degree representation of different features for each image block. The lower the quality is, the greater the weight assigned to the region. Given low-light image I_{low} and normal light image I_{normal},

$$I_{pro} = \mathrm{T_{rans}}(I_{low}, I_{normal}) \tag{15}$$

where I_{pro} represents the representation of the degree of degradation in different regions of the image. Next, the LLIE task can be formalized as:

$$arg\min_{\xi} Loss(T_{net_\xi}(I_{pro} \cdot I_{low}), I_{normal}) \tag{16}$$

where $Loss$ is the loss function of the original network and where T_{net_ξ} is the network with parameter ξ, $\cdot$ represents the dot product. During this process, the network focuses more intensely on regions of the image that have undergone greater degradation, effectively integrating an understanding of the image content into the learning process. The above process of our method leads to Algorithm 1, which contains three modules: the ADPT, PRA, and CDO of classifiers. These three modules are co-optimized to work together. In this process, initial values are first generated on the basis of the features of the training data. Next, adaptive targeted iterations are performed via ADPT to achieve the generation of BBAs under different classifiers. After that, the reliability of each classifier is calculated as a discount factor using HODMMD. Finally, fusion is performed with the help of a discount factor to obtain the final decision.

4 Experimental Results

4.1 Numerical experiment

To better understand the working mechanism of the ADPT, in this section, a discussion of the proposed ADPT is presented. by several concrete examples.

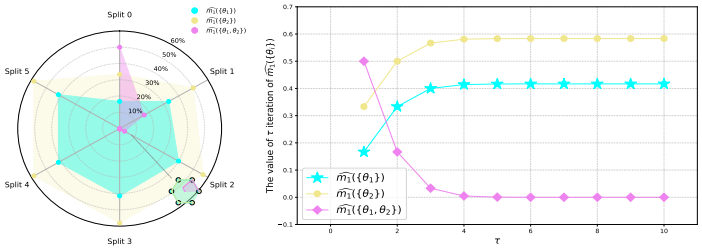

Figure 2: Diagram of the proportion of the mass function increasing with τ in Example 5.1.

Example 5.1 Let a frame of discernment be $\Omega = \{\theta_1, \theta_2\}$. The BBA is as follows:

$$\widehat{m_1} : \widehat{m_1}(\{\theta_1\}) = 1/6, \ \ \widehat{m_1}(\{\theta_2\}) = 1/3, \ \ \widehat{m_1}(\{\theta_1, \theta_2\}) = 1/2.$$

When split once, when $\tau = 1$ is substituted into Equation (4), three scenarios exist: $m_g^1(\{a_1\}) = 0.3$, $m_g^1(\{a_2\}) = 0.5$, and $m_g^1(\{a_1, a_2\}) = 0.2$. When the second, $\tau = 2$, is substituted into Equation (4), there can be three scenarios: $m_g^2(\{a_1\}) = 0.3741$, $m_g^2(\{a_2\}) = 0.5839$, and $m_g^2(\{a_1, a_2\}) = 0.0420$. The process was plotted as a radar chart and a line graph, and the results are shown in Figure 2. The results show that as τ increases, the values of the mass function belonging to $\{\theta_1, \theta_2\}$ gradually shift to the mass functions of $\{\theta_1\}$ and $\{\theta_2\}$. The mass function with a larger set of initial values still maintains a larger weight.

Example 5.2 Let a frame of discernment be $\Omega = \{\theta_1, \theta_2\}$. The BBAs are as follows:

$$\widehat{m_1} : \widehat{m_1}(\{\theta_1\}) = (1-x)/3, \ \ \widehat{m_1}(\{\theta_2\}) = (1-x)/3, \ \ \widehat{m_1}(\{\theta_1, \theta_2\}) = (2x+1)/3,$$
$$\widehat{m_2} : \widehat{m_2}(\{\theta_1\}) = (1-y)/3, \ \ \widehat{m_2}(\{\theta_2\}) = (1-y)/3, \ \ \widehat{m_2}(\{\theta_1, \theta_2\}) = (2y+1)/3.$$

The variation of the proposed HODMMD is further explored by changing the values of x and y and plotting Figure 3. As τ increases, the difference between $\widehat{m_1}$ and $\widehat{m_2}$ gradually decreases, verifying that the uncertainty information of the BBAs gradually decreases with increasing ADPT.

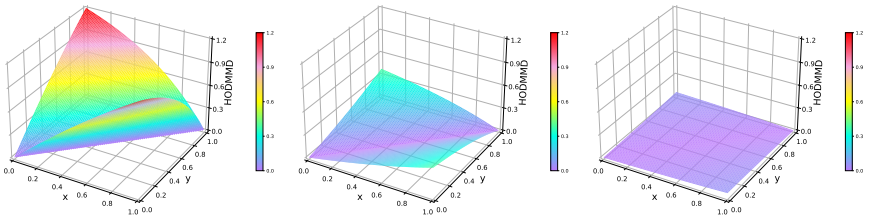

Figure 3: HODMMD between $\widehat{m_1}$ and $\widehat{m_2}$ when τ is 1, 3, 5 in Example 5.2.

4.2 Pattern classification

For each dataset, the same testing method as in [58] was adopted to perform fivefold cross-validation according to a training set-to-test machine ratio of 4:1. The classification accuracy was evaluated, and then all the results were averaged to arrive at the final result, which ensures a fair comparison between the different methods. The experimental results are shown in Table 1. The proposed method achieves the optimal performance except for the Parkinsons dataset because the features of the different samples in this dataset are too close to each other, resulting in poor differentiation of importance when performing ADPT. On the basis of the evaluation metric of classification accuracy, improvements of 2.67%, 11.9%, 1.48%, 1.66% and 0.22% are obtained over those of the advanced DMA method on

https://doi.org/10.52202/085713-0409

the Iris, Heart, Hepatitis, Australian and Segment datasets, respectively. This is because our method better handles the contribution of different features to the classification and better solves the conflict problem when fusing different BBAs.

Table 1: Comparison of the classification accuracy of different methods. The best, second and third results are in the red, green and blue colours.

Dataset	Iris[18]	Heart[27]	Hepatitis[1]	Parkinsons[36]	Australian[42]	Segment[2]	CBench [16]
NaB[23]	94.67%	82.59%	76.76%	68.75%	79.56%	80.22%	67.68%
kNN[12]	95.33%	57.78%	65.71%	83.02%	67.40%	96.93%	99.02%
REPTree[19]	92.00%	70.74%	71.64%	80.94%	80.59%	95.11%	65.45%
SVM[5]	94.67%	83.70%	79.96%	70.13%	80.29%	64.50%	88.48%
SVM-RBF[5]	94.67%	82.96%	76.76%	81.03%	79.86%	81.73%	68.18%
MLP[4]	93.33%	75.19%	74.93%	74.39%	82.32%	95.93%	80.40%
RBFN[8]	92.67%	81.85%	81.32%	82.05%	82.61%	87.58%	83.64%
kNN-DST[15]	95.33%	76.30%	80.57%	78.01%	78.41%	93.37%	94.42%
NDC[60]	94.00%	82.59%	79.40%	70.26%	80.01%	79.61%	43.13%
EvC[61]	94.67%	83.70%	79.88%	81.64%	80.60%	95.90%	76.33%
DMA[58]	96.00%	84.07%	83.04%	75.03%	84.14%	99.74%	100.00%
ours	98.67%	95.97%	84.52%	82.50%	85.80%	99.96%	100.00%

4.3 Image classification

For the image classification task, we examine the performance on the CIFAR-10 and CIFAR-100 datasets, which contain 10 and 100 categories, respectively. To verify the effectiveness of the proposed fusion framework, which is based on a nonuniform splitting mechanism and Hilbert space mapping more comprehensively, we perform feature extraction via ResNet-18 and the CNN framework proposed in [48] before adopting the proposed method for decision making. In this task, prior information is the initial probability distribution obtained by performing feature extraction on the input samples for the CNN and ResNet-18, with the last linear layer removed. The comparison results of the different methods are shown in Table 2. Our method achieves better results, which once again validates the effectiveness of the proposed method in addressing uncertain information.

Table 2: Comparison of classification accuracy on the CIFAR-10 dataset[31] and CIFAR-100 dataset[31]. The best, second and third results are in the red, green and blue colours.

	Methods	Architecture	Accuracy		Methods	Architecture	Accuracy
CIFAR-10	DIR-Net [40] IJCV'2023	ResNet-18	92.80%	CIFAR-100	Dspike[34] NeurIPS'2021	ResNet-19	73.12%
	MST[51] ICCV'2023	ResNet-18	93.20 %		GLIF[66] NeurIPS' 2022	ResNet-19	77.05%
	ReSTE[54] ICCV'2023	ResNet-18	92.63 %		Diet-SNN[43] TNNLS'2021	VGG-16	69.67%
	ADMM[9] TIP'2023	ResNet-18	95.40 %		PASNN[17] KBS'2023	ResNet-14	72.63%
	SML[14] ICML'2023	ResNet-19	95.12 %		MPBN [22] ICCV'2023	ResNet-19	74.40%
	UDSP[20] CVPR'2024	ResNet-56	93.78%		MS-ResNet[24] ICCV'2023	MS-ResNet18	75.39%
	BiPer [50] CVPR'2024	ResNet-20	93.75%		BKDSNN[63] ECCV'2024	ResNet-19	74.95%
		ResNet-18	93.75%				
		VGG-small	92.11%				
	TAB[28] ICIL'2024	VGG-9	93.41%		TAB[28] ICIL'2024	VGG-11	76.31%
	APL[70] TPAMI'2023	ResNet-18	96.00%		APL [70] TPAMI'2024	ResNet-18	78.90%
	ESNN[47] EAAI'2025	VGG-16	93.55%		ESNN[47] EAAI'2025	VGG-16	76.55%
	ours	CNN	95.67%		ours	CNN	79.71%
		ResNet-18	95.61%			ResNet-18	79.78%

4.4 Low-light image enhancement

In LLIE, dark areas may contain critical information that is often difficult for models to adequately learn and focus on [35]. Traditional training strategies, which apply uniform processing across the entire image, are inherently limited in their ability to specifically enhance model learning for these critical regions. This often results in suboptimal detail recovery in the target areas during enhancement.

Therefore, a TTS is proposed in conjunction with the proposed method, which explicitly guides the model to increase the level of attention to low-quality regions. The method is a plug-and-play module. In this task, prior information refers to the degree of degradation of different regions in the image. We used the LLIE network of the last few years as a baseline network, and the results are shown in Table 3. A consistent improvement in performance can be seen after using the TTS. To obtain a more intuitive sense of the enhancement, we present the results of the baseline method and our method, and the results are shown in Figure 4. Our method effectively improves the quality of images. For the LOL-v1 dataset, our method enhances the texture details on the glass; for the LOL-v2-real dataset, our method recovers the text on the wall more clearly; for the LOL-v2-real dataset, our method recovers the color of the flower petals closer to the ground truth. This finding verifies the validity of the proposed method as well as the idea and further shows that the proposed ADPT can characterize the data well. As seen through the above experiments, applying different learning strategies to different regions can effectively enhance the model's focus on the target region. By adaptively adjusting the learning weights of different regions, the model's ability to focus on low-quality regions is enhanced, the data distribution imbalance problem is alleviated, and the overall performance of the model is improved. To the best of our knowledge, this is the first time that DSE theory has been introduced into LLIE.

More experiments can be found in the technical appendices.

Table 3: Quantitative comparison of LOL-v1 [53], LOL-v2-real [65] and LOL-v2-syn [65].

Methods	LOL-v1[53]		LOL-v2-real[65]		LOL-v2-syn[65]	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
MIRNet[68] TPAMI'2020	24.14	0.830	20.02	0.820	21.94	0.876
FIDE[59] CVPR'2020	18.27	0.665	16.85	0.678	15.20	0.612
ZeroDCE[21] CVPR'2020	16.76	0.560	18.06	0.577	17.76	0.816
Sparse[53] TIP'2021	17.20	0.640	20.06	0.816	22.05	0.905
DRBN[64] TIP'2021	20.13	0.830	20.29	0.831	23.22	0.927
RUAS[37] CVPR'2021	18.23	0.720	18.37	0.723	16.55	0.652
ZeroDCE++[33] TPAMI'2021	16.11	0.530	18.06	0.577	18.03	0.825
SCI[39] CVPR'2022	14.78	0.525	16.19	0.522	16.67	0.811
Restormer[29] TCSVT'2023	22.43	0.823	19.94	0.827	21.41	0.830
SNR[62] CVPR'2022	24.61	0.842	21.48	0.849	24.14	0.928
SNR-TTS	**24.94**(+0.33)	**0.854**(+0.012)	**22.00**(+0.52)	0.846(-0.003)	**24.32**(+0.18)	**0.929**(+0.001)
LLFlow-L[52] AAAI'2022	24.99	0.870	25.31	0.805	25.88	0.908
LLFlow-L-TTS	**26.70**(+1.71)	0.860(-0.01)	**26.97**(+1.66)	**0.865**(+0.06)	**26.09**(+0.21)	0.906(-0.002)
LLFlow-S[52] AAAI'2022	24.06	0.860	26.80	0.860	25.30	0.877
LLFlow-S-TTS	**26.28**(+2.22)	0.848(-0.012)	**28.15**(+1.35)	**0.866**(+0.006)	**25.33**(+0.03)	**0.880**(+0.003)
Retinexformer[3] ICCV'2023	25.16	0.845	22.80	0.840	25.67	0.930
Retinexformer-TTS	**26.14**(+0.98)	**0.849**(+0.004)	**23.01**(+0.21)	**0.843**(+0.003)	**26.04**(+0.37)	**0.942**(+0.012)

5 Conclusion

This paper proposes a dynamic learning strategy based on nonuniform splitting mechanism and Hilbert space mapping, which is based on DSE theory, as an efficient method for processing uncertain information. The current method cannot dynamically adjust the tendency of splitting on the basis of the importance between different subsets in the process of splitting, and directly using the DC rule in fusion will produce unreasonable results owing to the conflict problem. First, the nonuniform splitting mechanism proposed in this paper takes the data's inherent a priori information into account and thus can handle uncertainty and conflict information more flexibly while improving the accuracy of the task. Second, mapping the data into the Hilbert space for computation is more responsive to the differences in the true distributions of complex data, thus providing an effective strategy for the subsequent fusion of different BBAs. We conducted experiments on multiple tasks as well as multiple publicly available datasets. The experimental results show that our method significantly outperforms existing machine learning methods and deep learning methods. To the best of our knowledge, we are the first to introduce DSE theory to LLIE and provide effective performance enhancement for LLIE tasks.

 https://doi.org/10.52202/085713-0409

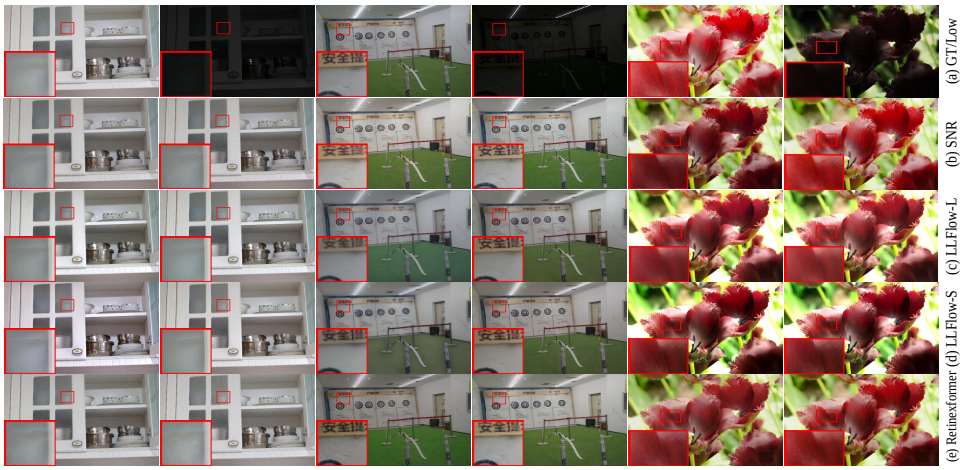

Figure 4: Quantitative comparison of LOL-v1 [53] (1st–2nd columns), LOL-v2-real [65] (3rd–4th columns), and LOL-v2-syn [65] (5th–6th columns). Even columns correspond to the results of adding the TTS module.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 62176027, in part by Chongqing Talent under Grant cstc2024ycjh-bgzxm0082, in part by Chongqing New Yin Cai (YC) Project under Grant CSTB2024YCJH-KYXM0126, in part by the General Program of the Natural Science Foundation of Chongqing under Grant CSTB2024NSCQ-MSX0479, and in part by Chongqing Postdoctoral Foundation Special Support Program under Grant 2023CQBSHTB3119.

References

[1] Bache , K. & Lichman , M. (1983) Hepatitis. *UCI Machine Learning Repository*

[2] Bache , K. & Lichman , M. (1990) Image segmentation. *UCI Machine Learning Repository*

[3] Cai , Y., Bian , H., Lin , J., Wang , H., Timofte , R., & Zhang , Y. (2023) Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* pages 12504–12513.

[4] Castro , C. L. & Braga , A. P. (2013) Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data. *IEEE transactions on neural networks and learning systems* **24**(6):888–899.

[5] Chang , C.-C. & Lin , C.-J. (2011) Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)* **2**(3):1–27.

[6] Chen , C., Chen , Q., Xu , J., & Koltun , V. (2018) Learning to see in the dark pages 3291–3300.

[7] Chen , C., Chen , Q., Do , M. N., & Koltun , V. (2019) Seeing motion in the dark pages 3185–3194.

[8] Chen , S., Cowan , C., & Grant , P. (1991) Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on Neural Networks* **2**(2):302–309.

[9] Chiou , C.-Y., Lee , K.-T., Huang , C.-R., & others (2023) Admm-srnet: Alternating direction method of multipliers based sparse representation network for one-class classification. *IEEE Transactions on Image Processing* **32**:2843–2856.

[10] Constance , L., Driessen , A., Deutschmann , N., & Martinez , M. R. (2022) Fuzzy logic for biological networks as ml regression: Scaling to single-cell datasets with autograd. In *NeurIPS 2022 Workshop on Learning Meaningful Representations of Life*

[11] Cortes , C. & Vapnik , V. (1995) Support-vector networks. *Machine learning* **20**:273–297.

[12] Cover , T. & Hart , P. (1967) Nearest neighbor pattern classification. *IEEE transactions on information theory* **13**(1):21–27.

[13] Dempster , A. P. (2008) Upper and lower probabilities induced by a multivalued mapping. *Classic works of the Dempster-Shafer theory of belief functions* pages 57–72.

[14] Deng , S., Lin , H., Li , Y., & Gu , S. (2023) Surrogate module learning: Reduce the gradient error accumulation in training spiking neural networks. In *International Conference on Machine Learning* pages 7645–7657. PMLR.

[15] Denoeux , T. (1995) A k-nearest neighbor classification rule based on dempster-shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics* **25**(5):804–813.

[16] Deterding , N. M. & Robinson , T. (1988) Connectionist bench (vowel recognition - deterding data). *UCI Machine Learning Repository*

[17] Ding , Y., Zuo , L., Yang , K., Chen , Z., Hu , J., & Xiahou , T. (2023) An improved probabilistic spiking neural network with enhanced discriminative ability. *Knowledge-Based Systems* **280**:111024.

[18] Fisher , R. A. (1936) Iris. *UCI Machine Learning Repository*

[19] Freund , Y. & Mason , L. (1999) The alternating decision tree learning algorithm page 124–133.

[20] Gao , S., Zhang , Y., Huang , F., & Huang , H. (2024) Bilevelpruning: Unified dynamic and static channel pruning for convolutional neural networks. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* pages 16090–16100.

[21] Guo , C., Li , C., Guo , J., Loy , C. C., Hou , J., Kwong , S., & Cong , R. (2020) Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* pages 1780–1789.

[22] Guo , Y., Zhang , Y., Chen , Y., Peng , W., Liu , X., Zhang , L., Huang , X., & Ma , Z. (2023) Membrane potential batch normalization for spiking neural networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* pages 19420–19430.

[23] Hu , B.-G. (2013) What are the differences between bayesian classifiers and mutual-information classifiers? *IEEE transactions on neural networks and learning systems* **25**(2):249–264.

[24] Hu , Y., Deng , L., Wu , Y., Yao , M., & Li , G. (2024) Advancing spiking neural networks toward deep residual learning. *IEEE transactions on neural networks and learning systems* **36**(2):2353–2367.

[25] Huang , Y., Xiao , F., Cao , Z., & Lin , C.-T. (2023) Higher order fractal belief rényi divergence with its applications in pattern classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*

[26] Huang , Y., Hechen , Z., Zhou , M., Li , Z., & Kwong , S. (2025) An attention-locating algorithm for eliminating background effects in fine-grained visual classification. *IEEE Transactions on Circuits and Systems for Video Technology*

[27] Janosi , S. W. P. M. & Detrano , R. (1989) Heart disease. *UCI Machine Learning Repository*

[28] Jiang , H., Zoonekynd , V., De Masi , G., Gu , B., & Xiong , H. (2024) Tab: Temporal accumulated batch normalization in spiking neural networks. In *The Twelfth International Conference on Learning Representations*

[29] Jiang , N., Lin , J., Zhang , T., Zheng , H., & Zhao , T. (2023) Low-light image enhancement via stage-transformer-guided network. *IEEE Transactions on Circuits and Systems for Video Technology* **33**(8): 3701–3712.

[30] Jousselme , A.-L., Grenier , D., & Bossé , É. (2001) A new distance between two bodies of evidence. *Information fusion* **2**(2):91–101.

[31] Krizhevsky , A., Hinton , G., & others (2009) Learning multiple layers of features from tiny images

[32] Lee , H. & Kwon , H. (2019) Dbf: Dynamic belief fusion for combining multiple object detectors. *IEEE transactions on pattern analysis and machine intelligence* **43**(5):1499–1514.

[33] Li , C., Guo , C., & Loy , C. C. (2021.) Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**(8):4225–4238.

 https://doi.org/10.52202/085713-0409

[34] Li , Y., Guo , Y., Zhang , S., Deng , S., Hai , Y., & Gu , S. (2021.) Differentiable spike: Rethinking gradient-descent for training spiking neural networks. *Advances in neural information processing systems* **34**: 23426–23439.

[35] Li , Y., Wei , X., Liao , X., Zhao , Y., Jia , F., Zhuang , X., & Zhou , M. (2024) A deep retinex-based low-light enhancement network fusing rich intrinsic prior information. *ACM Transactions on Multimedia Computing, Communications and Applications* **20**(11):1–23.

[36] Little , M. (2007) Parkinsons. *UCI machine learning repository*

[37] Liu , R., Ma , L., Zhang , J., Fan , X., & Luo , Z. (2021) Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* pages 10556–10565.

[38] Liu , Z., Pan , Q., Dezert , J., Han , J.-W., & He , Y. (2017) Classifier fusion with contextual reliability evaluation. *IEEE transactions on cybernetics* **48**(5):1605–1618.

[39] Ma , L., Ma , T., Liu , R., Fan , X., & Luo , Z. (2022) Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* pages 5637–5646.

[40] Qin , H., Zhang , X., Gong , R., Ding , Y., Xu , Y., & Liu , X. (2023) Distribution-sensitive information retention for accurate binary neural network. *International Journal of Computer Vision* **131**(1):26–47.

[41] Quinlan , J. R. (1996) Learning decision tree classifiers. *ACM Computing Surveys (CSUR)* **28**(1):71–72.

[42] Quinlan , R. (1987) Statlog (australian credit approval). *UCI Machine Learning Repository*

[43] Rathi , N. & Roy , K. (2021) Diet-snn: A low-latency spiking neural network with direct input encoding and leakage and threshold optimization. *IEEE Transactions on Neural Networks and Learning Systems* **34**(6): 3174–3182.

[44] Sanchez , T., Krawczuk , I., Sun , Z., & Cevher , V. (2020) Uncertainty-driven adaptive sampling via gans. In *NeurIPS 2020 workshop on deep learning and inverse problems*

[45] Sensoy , M., Kaplan , L., & Kandemir , M. (2018) Evidential deep learning to quantify classification uncertainty. *Advances in neural information processing systems* **31**.

[46] Shafer , G. (1976) *A mathematical theory of evidence*, *42*. *42*: Princeton university press.

[47] Tang , X., Chen , T., Cheng , Q., Shen , H., Duan , S., & Wang , L. (2025) Spatio-temporal channel attention and membrane potential modulation for efficient spiking neural network. *Engineering Applications of Artificial Intelligence* **148**:110131.

[48] Tong , Z., Xu , P., & Denoeux , T. (2021) An evidential classifier based on dempster-shafer theory and deep learning. *Neurocomputing* **450**:275–293.

[49] Trivedi , P., Heimann , M., Anirudh , R., Koutra , D., & Thiagarajan , J. J. (2023) Estimating epistemic uncertainty of graph neural networks using stochastic centering. In *NeurIPS 2023 Workshop: New Frontiers in Graph Learning*

[50] Vargas , E., Correa , C. V., Hinojosa , C., & Arguello , H. (2024) Biper: Binary neural networks using a periodic function. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* pages 5684–5693.

[51] Vo , Q. H., Tran , L.-T., Bae , S.-H., Kim , L.-W., & Hong , C. S. (2023) Mst-compression: Compressing and accelerating binary neural networks with minimum spanning tree. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* pages 6091–6100.

[52] Wang , Y., Wan , R., Yang , W., Li , H., Chau , L.-P., & Kot , A. (2022) Low-light image enhancement with normalizing flow. In *Proceedings of the AAAI conference on artificial intelligence 36*, pp. 2604–2612.

[53] Wei , C., Wang , W., Yang , W., & Liu , J. (2018) Deep retinex decomposition for low-light enhancement. *IN BMVC*

[54] Wu , X.-M., Zheng , D., Liu , Z., & Zheng , W.-S. (2023) Estimator meets equilibrium perspective: A rectified straight through estimator for binary neural networks training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* pages 17055–17064.

[55] Xanthopoulos , P., Pardalos , P. M., Trafalis , T. B., Xanthopoulos , P., Pardalos , P. M., & Trafalis , T. B. (2013) Linear discriminant analysis. *Robust data mining* pages 27–33.

[56] Xia , T., Han , J., Qendro , L., Dang , T., & Mascolo , C. (2022) Hybrid-edl: Improving evidential deep learning for uncertainty quantification on imbalanced data. In *Workshop on Trustworthy and Socially Responsible Machine Learning, NeurIPS 2022*

[57] Xiao , F. (2023) Quantum x-entropy in generalized quantum evidence theory. *Information Sciences* **643**: 119177.

[58] Xiao , F., Wen , J., & Pedrycz , W. (2022) Generalized divergence-based decision making method with an application to pattern classification. *IEEE transactions on knowledge and data engineering* **35**(7):6941–6956.

[59] Xu , K., Yang , X., Yin , B., & Lau , R. W. (2020) Learning to restore low-light images via decomposition-and-enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* pages 2281–2290.

[60] Xu , P., Deng , Y., Su , X., & Mahadevan , S. (2013) A new method to determine basic probability assignment from training data. *Know.-Based Syst.* **46**:69–80.

[61] Xu , P., Davoine , F., Zha , H., & Denoeux , T. (2016) Evidential calibration of binary svm classifiers. *International Journal of Approximate Reasoning* **72**:55–70.

[62] Xu , X., Wang , R., Fu , C.-W., & Jia , J. (2022) Snr-aware low-light image enhancement. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* pages 17693–17703.

[63] Xu , Z., You , K., Guo , Q., Wang , X., & He , Z. (2024) Bkdsnn: Enhancing the performance of learning-based spiking neural networks training with blurred knowledge distillation. In *European Conference on Computer Vision* pages 106–123. Springer.

[64] Yang , W., Wang , S., Fang , Y., Wang , Y., & Liu , J. (2021.) Band representation-based semi-supervised low-light image enhancement: Bridging the gap between signal fidelity and perceptual quality. *IEEE Transactions on Image Processing* **30**:3461–3473.

[65] Yang , W., Wang , W., Huang , H., Wang , S., & Liu , J. (2021.) Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing* **30**:2072–2086.

[66] Yao , X., Li , F., Mo , Z., & Cheng , J. (2022) Glif: A unified gated leaky integrate-and-fire neuron for spiking neural networks. *Advances in Neural Information Processing Systems* **35**:32160–32171.

[67] Zaidi , S., Zela , A., Elsken , T., Holmes , C. C., Hutter , F., & Teh , Y. (2021) Neural ensemble search for uncertainty estimation and dataset shift. *Advances in Neural Information Processing Systems* **34**:7898–7911.

[68] Zamir , S. W., Arora , A., Khan , S., Hayat , M., Khan , F. S., Yang , M.-H., & Shao , L. (2022) Learning enriched features for fast image restoration and enhancement. *IEEE transactions on pattern analysis and machine intelligence* **45**(2):1934–1948.

[69] Zhang , L. & Xiao , F. (2022) A novel belief χ^2 divergence for multisource information fusion and its application in pattern classification. *International Journal of Intelligent Systems* **37**(10):7968–7991.

[70] Zhang , L., Qi , L., Yang , X., Qiao , H., Yang , M.-H., & Liu , Z. (2023) Automatically discovering novel visual categories with adaptive prototype learning. *IEEE transactions on pattern analysis and machine intelligence* **46**(4):2533–2544.

[71] Zhao , J., Xue , R., Dong , Z., Tang , D., & Wei , W. (2020) Evaluating the reliability of sources of evidence with a two-perspective approach in classification problems based on evidence theory. *Information Sciences* **507**: 313–338.

 https://doi.org/10.52202/085713-0409

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 contributions and scope of this paper are clearly presented in both the abstract and the introduction as well as verified in the experiment section.

   Guidelines:

   - The answer NA 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 NA 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: In the experimental part, we find that the proposed method has limitations on some datasets but still has room for improvement, which is discussed in the experimental analysis section.

   Guidelines:

   - The answer NA 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: For the theories mentioned in the paper, we provide examples. The theory-related properties provide complete (correct) proofs, as shown in the technical appendices.

Guidelines:

- The answer NA 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: Relevant information and details are described in the experimental results section and more details on the experimental settings section in the technical appendices, and the code is publicly available.

Guidelines:

- The answer NA 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.

 https://doi.org/10.52202/085713-0409

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: [Yes]

   Justification: The source code has been released, and the link is reflected in the abstract.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.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://nips.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, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: All relevant settings are described in the experimental results section, and more details on the experimental settings are provided in the technical appendices.

   Guidelines:

   - The answer NA 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: [No]

   Justification: The experimental results in this paper require no further error analysis to be reported.

   Guidelines:

   - The answer NA 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: [Yes]

   Justification: We describe the training settings for each experiment as described in the experimental results section and more details on the experimental settings section in the technical appendices.

   Guidelines:

   - The answer NA 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 read the NeurIPS Code of Ethics and have strictly adhered to it during the course of our research.

   Guidelines:

   - The answer NA 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: [Yes]

    Justification: We explore some of the possible social implications of our approach in the concluding section and technical appendices.

 https://doi.org/10.52202/085713-0409

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA 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., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: There is no security risk in this paper.

    Guidelines:

    - The answer NA 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 relevant information used in this paper has been explicitly mentioned and properly respected.

    Guidelines:

    - The answer NA 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: [Yes]

    Justification: The source code has been released, and the link is reflected in the abstract.

    Guidelines:

    - The answer NA 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: [NA]

    Justification: This paper does not include crowdsourcing experiments or research with human subjects.

    Guidelines:

    - The answer NA 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: [NA]

    Justification: This paper does not include crowdsourcing experiments or research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

https://doi.org/10.52202/085713-0409

- 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.

A Appendix/supplemental material

A.1 More Details of Preparatory Knowledge

A.1.1 Dempster–Shafer evidence theory

Dempster-shafer evidence theory is a well-established general framework for uncertainty reasoning that was first proposed by Arthur P. Dempster [13] on the basis of statistical inference and later formalized and significantly extended by Glenn Shafer into a framework for simulating epistemic uncertainty[46]. Owing to its effectiveness in dealing with uncertain information, DSE theory is widely used in various fields.

Definition 1. Frame of discernment

Define a set of classes called discriminant frames:

$$\Omega = \{\theta_1, \theta_2, \theta_3, \ldots, \theta_N\} \tag{17}$$

where $\theta_i (i = 1, 2, ..., N)$ are mutually exclusive. The power set is defined as:

$$2^{\Omega} = \{\emptyset, \{\theta_1\}, \{\theta_2\}, \ldots, \{\theta_N\}, \{\theta_1 \cup \theta_2\}, \ldots, \Omega\} \tag{18}$$

where Ω is the whole set, $\emptyset$ is the empty set and the one containing $\cup$ is a multielement set [13, 46].

Definition 2. Basic belief assignment

The basic belief assignment (BBA), also known as the mass function, denoted $m(\cdot)$ is defined as a mapping of 2^{Ω} to the interval $[0, 1]$ [13, 46].

$$m : 2^{\Omega} \to [0, 1] \tag{19}$$

and is satisfied:

$$\sum_{A \in 2^{\Omega} \setminus \{\emptyset\}} m(A) = 1 \tag{20}$$

$$m(\emptyset) = 0 \tag{21}$$

where $A \in 2^{\Omega} \setminus \{\emptyset\}$, $\emptyset$ is the empty set.

Definition 3. Focal set

If $m(A) > 0$, A is called a focal set, and the value of $m(\cdot)$ indicates the level of support of the model[13, 46].

Definition 4. Dempster's combination rule

Different fusion algorithms can be used to fuse information from different sources, one of which is Dempster's combination rule. As it can process the fusion of different sources of evidence represented by a BBA, it is assumed that m_1 and m_2 are two mutually independent BBAs defined on the same recognition frame 2^{Ω}. Dempster's combination rule aims to derive a combined BBA, usually denoted as $m = m_1 \oplus m_2$. The combination frame is defined as follows[13]:

$$m(A_k) = (m_1 \oplus m_2)(A_k) = \frac{\sum_{A_i \cap A_j = A_k} m_1(A_i) m_2(A_j)}{1 - \mathcal{R}} \tag{22}$$

$$\mathcal{R} = \sum_{A_i \cap A_j = \emptyset} m_1(A_i) m_2(A_j) < 1 \tag{23}$$

where $A_i, A_j, A_k \in \Omega$ and $\mathcal{R}$ are the conflict coefficients of m_1 and m_2.

A.1.2 Probability transformation methods

Definition 5. Pignistic probability transformation

https://doi.org/10.52202/085713-0409

Let $\Omega = \{\theta_1, \theta_2, \theta_3, \ldots, \theta_N\}$ denote the discriminant framework. Given a corresponding BBA defined on Ω, denoted $m(\cdot)$, the pignistic probabilistic transformation of an element $\theta_i \in \Omega$, denoted P_{ppt}, is defined as follows:

$$P_{ppt}\left(\theta_i\right) = \sum_{\theta_i \in A | A \in 2^{\Omega}} \frac{m(A)}{|A|} \tag{24}$$

where A belongs to 2^{Ω}, $A \neq \emptyset$. $|A|$ is the cardinality of A.

Definition 6. Plausibility transformation method

Let $\Omega = \{\theta_1, \theta_2, \theta_3, \ldots, \theta_N\}$ denote the discriminant framework. Given a corresponding BBA defined on Ω, denoted $m(\cdot)$, the plausibility transformation of an element $\theta_i \in \Omega$, denoted P_{pt}, is defined as follows:

$$P_{pt}\left(\theta_i\right) = \frac{M\left(\theta_i\right)}{\sum_{i=1}^{N} M\left(\theta_i\right)} \tag{25}$$

where $M(\cdot)$ is the plausibility function:

$$M(A_i) = \sum_{A_i \cap A_j \not\subset \emptyset | A_i, A_j \subseteq 2^{\Omega}} m(A_j). \tag{26}$$

A.2 More Proof Details for Properties

When different quality functions are obtained, if they are simply combined according to Dempster's rule [48], the integration between different BBAs may be hindered because of conflicting information. For this reason, the idea of discounting techniques [46] is introduced. The high-order dynamic maximum mean difference (HODMMD) is defined as follows:

$$\text{HODMMD}^{\tau}\left(\widehat{m}_1, \widehat{m}_2\right) = \left\| \sum_{i=1}^{N} \left(\widehat{m}_{g_1}^{\tau}\left(A_i\right) - \widehat{m}_{g_2}^{\tau}\left(A_i\right) \right) \right\|_{\mathcal{H}} \tag{27}$$

where $\mathcal{H}$ is the regenerated kernel Hilbert space and the kernel function is computed via a Gaussian kernel function, $A_i \in 2^{\Omega} \backslash \{\emptyset\}$, and τ is the number of iterative splits. $\widehat{m}_g^{\tau}$ is τ order of $\widehat{m}_g$. The HODMMD has a variety of properties. The properties of the HODMMD and the corresponding proofs are as follows.

Property 1. When $\tau \to \infty$, the HODMMD is equivalent to measuring the difference between the pignistic probability transformations with $\widehat{m}_1$ and $\widehat{m}_2$ of the maximum mean difference.

$$\lim_{\tau \to \infty} \text{HODMMD}^{\tau}(\widehat{m}_1, \widehat{m}_2) = \left\| \sum_{i=1}^{N} \left(\text{Bet}P_1(\theta_i) - \text{Bet}P_2(\theta_i)\right) \right\|_{\mathcal{H}} \tag{28}$$

Proof. We denote by $\text{Bet}P_t$ the pignistic probability transformations of $\widehat{m}_1$ and $\widehat{m}_2$, $t = 1, 2$. That is,

$$\text{Bet}P_t(\theta_i) = \sum_{\theta_i \in A | A \in 2^{\Omega}} \frac{\widehat{m}_t(A)}{|A|} \tag{29}$$

If $A_i = \theta_i$, then the cardinality of A_i is 1. Thus, when $\tau \to \infty$ and $|A_i| = 1$,

$$\lim_{\tau \to \infty} \widehat{m}_{g_t}^{\tau}\left(A_i\right) = \lim_{\tau \to \infty} \frac{\sum_{A_i \subseteq A_j} \left(\frac{|A_i|}{|A_j|}\right)^{\frac{1}{\tau}} \frac{\tau^{|A_j|-|A_i|}}{(\tau+1)^{|A_j|}-\tau^{|A_j|}} \widehat{m}_t\left(A_j\right)}{\sum_{A_c \subseteq 2^{\Omega}} \sum_{A_c \subseteq A_j} \left(\frac{|A_c|}{|A_j|}\right)^{\frac{1}{\tau}} \frac{\tau^{|A_j|-|A_c|}}{(\tau+1)^{|A_j|}-\tau^{|A_j|}} \widehat{m}_t\left(A_j\right)} \tag{30}$$

Let

$$\widehat{m}_g^{\tau}\left(A_i\right) = \sum_{A_i \subseteq A_j} \left(\frac{|A_i|}{|A_j|}\right)^{\frac{1}{\tau}} \frac{\tau^{|A_j|-|A_i|}}{(\tau+1)^{|A_j|}-\tau^{|A_j|}} \widehat{m}_g^{\tau-1}\left(A_j\right) \tag{31}$$

$$\lim_{\tau\to\infty}\widehat{m}_{g_t}^{\tau}\left(A_i\right)=\lim_{\tau\to\infty}\sum_{A_i\subseteq A_j}\left(\frac{|A_i|}{|A_j|}\right)^{\frac{1}{\tau}}\frac{\tau^{|A_j|-|A_i|}}{(\tau+1)^{|A_j|}-\tau^{|A_j|}}\widehat{m}_t\left(A_j\right)$$
$$=\lim_{\tau\to\infty}\sum_{A_i\subseteq A_j}\left(\frac{1}{|A_j|}\right)^{\frac{1}{\tau}}\frac{\tau^{|A_j|-1}}{(\tau+1)^{|A_j|}-\tau^{|A_j|}}\widehat{m}_t\left(A_j\right) \tag{32}$$

where A_i is a subset of A_j. When $|A_i|=1$, there must be $|A_j|\geq 1$; hence, $|A_j|-1\geq 0$ and $\tau^{|A_j|-1}>0$. The numerator and denominator are equally divisible by $\tau^{|A_j|-1}$

$$\lim_{\tau\to\infty}\widehat{m}_{g_t}^{\tau}\left(A_i\right)=\lim_{\tau\to\infty}\sum_{A_i\subseteq A_j}\left(\frac{1}{|A_j|}\right)^{\frac{1}{\tau}}\frac{1}{\tau(\frac{1}{\tau}+1)^{|A_j|}-\tau}\widehat{m}_t\left(A_j\right) \tag{33}$$

Let $\rho=\frac{1}{\tau}$, when $\tau\to\infty$, have $\frac{1}{\tau}\to 0$, which is $\rho\to 0$. A Taylor expansion of $(\rho+1)^{|A_j|}$ has

$$(\rho+1)^{|A_j|}=1+|A_j|\rho+o(\rho) \tag{34}$$

where $o(\rho)$ is the infinitesimal of ρ, have

$$\lim_{\rho\to 0}\widehat{m}_{g_t}^{\rho}\left(A_i\right)=\lim_{\rho\to 0}\sum_{A_i\subseteq A_j}\left(\frac{1}{|A_j|}\right)^{\rho}\frac{\tau^{|A_j|-1}}{(\tau+1)^{|A_j|}-\tau^{|A_j|}}\widehat{m}_t\left(A_j\right)$$
$$=\lim_{\rho\to 0}\sum_{A_i\subseteq A_j}\left(\frac{1}{|A_j|}\right)^{\rho}\frac{1}{\frac{1}{\rho}(1+|A_j|^{\rho}+o(\rho))-\frac{1}{\rho}}\widehat{m}_t\left(A_j\right)$$
$$=\sum_{A_i\subseteq A_j}\frac{\widehat{m}_t(A_j)}{|A_j|} \tag{35}$$

Therefore, when $\tau\to\infty$

$$\lim_{\tau\to\infty}\widehat{m}_{g_t}^{\tau}\left(A_i\right)=\frac{\sum_{A_i\subseteq A_j}\frac{\widehat{m}_t(A_j)}{|A_j|}}{\sum_{A_k\in 2^{\Omega}}\sum_{A_i\subseteq A_j}\frac{\widehat{m}_t(A_j)}{|A_j|}}$$
$$=\sum_{A_i\subseteq A_j}\frac{\widehat{m}_t(A_j)}{|A_j|}$$
$$=\sum_{\theta_i\in A|A\in 2^{\Omega}}\frac{\widehat{m}_t(A)}{|A|}$$
$$=\text{Bet}P_t(A_i)$$
$$=\text{Bet}P_t(\theta_i) \tag{36}$$

Thus,

$$\lim_{\tau\to\infty}\text{HODMMD}^{\tau}(\widehat{m}_1,\widehat{m}_2)=\left\|\sum_{i=1}^{N}\left(\text{Bet}P_1(\theta_i)-\text{Bet}P_2(\theta_i)\right)\right\|_{\mathcal{H}} \tag{37}$$

□

Property 2. When $\widehat{m}_1$ and $\widehat{m}_2$ degenerate into probability distributions, that is, $U=(u_1,u_2,\ldots,u_N)$ and $V=(v_1,v_2,\ldots,v_N)$, the proposed HODMMD degenerates into a maximum mean difference.

$$\text{HODMMD}^{\tau}\left(\widehat{m}_1,\widehat{m}_2\right)=\text{MMD}(\widehat{m}_1,\widehat{m}_2) \tag{38}$$

Proof. When $\widehat{m}_1$ and $\widehat{m}_2$ are probability distribution,

$$\forall A_i\in 2^{\Omega}\quad \widehat{m}_1(A_i)=u_i,\ \ \widehat{m}_2(A_i)=v_i \tag{39}$$

https://doi.org/10.52202/085713-0409

and it satisfies $\sum_{i=1}^{N} \widehat{m}_1(A_i) = 1$, $\sum_{i=1}^{N} \widehat{m}_2(A_i) = 1$.

$$\begin{aligned}\widehat{m}_{g_t}^{\tau}\left(A_i\right) &= \lim_{\tau\to\infty} \frac{\sum_{A_i\subseteq A_j}\left(\frac{|A_i|}{|A_j|}\right)^{\frac{1}{\tau}} \frac{\tau^{|A_j|-|A_i|}}{(\tau+1)^{|A_j|}-\tau^{|A_j|}}\widehat{m}_t\left(A_j\right)}{\sum_{A_c\subseteq 2^{\Omega}}\sum_{A_c\subseteq A_j}\left(\frac{|A_c|}{|A_j|}\right)^{\frac{1}{\tau}} \frac{\tau^{|A_j|-|A_c|}}{(\tau+1)^{|A_j|}-\tau^{|A_j|}}\widehat{m}_t\left(A_j\right)} \\ &= \widehat{m}_t(A_i)\end{aligned} \tag{40}$$

therefore,

$$\begin{aligned}\text{HODMMD}^{\tau}\left(\widehat{m}_1,\widehat{m}_2\right) &= \left\|\sum_{i=1}^{N}\left(\widehat{m}_{g_1}^{\tau}\left(A_i\right)-\widehat{m}_{g_2}^{\tau}\left(A_i\right)\right)\right\|_{\mathcal{H}} \\ &= \left\|\sum_{i=1}^{N}\left(\widehat{m}_1\left(A_i\right)-\widehat{m}_2\left(A_i\right)\right)\right\|_{\mathcal{H}} \\ &= \left\|\sum_{i=1}^{N}\left(u_i-v_i\right)\right\|_{\mathcal{H}} \\ &= \text{MMD}\left(U,V\right)\end{aligned} \tag{41}$$

□

Property 3. HODMMD$^{\tau}\left(\widehat{m}_1,\widehat{m}_2\right)$ and HODMMD$^{\tau}\left(\widehat{m}_2,\widehat{m}_1\right)$ are equivalent.

$$\text{HODMMD}^{\tau}\left(\widehat{m}_1,\widehat{m}_2\right) = \text{HODMMD}^{\tau}\left(\widehat{m}_2,\widehat{m}_1\right) \tag{42}$$

Proof.

$$\begin{aligned}\text{HODMMD}^{\tau}\left(\widehat{m}_1,\widehat{m}_2\right) &= \left\|\sum_{i=1}^{N}\left(\widehat{m}_{g_1}^{\tau}\left(A_i\right)-\widehat{m}_{g_2}^{\tau}\left(A_i\right)\right)\right\|_{\mathcal{H}} \\ &= \left\|\sum_{i=1}^{N}\left(\widehat{m}_1\left(A_i\right)-\widehat{m}_2\left(A_i\right)\right)\right\|_{\mathcal{H}}\end{aligned} \tag{43}$$

$$\begin{aligned}\text{HODMMD}^{\tau}\left(\widehat{m}_2,\widehat{m}_1\right) &= \left\|\sum_{i=1}^{N}\left(\widehat{m}_{g_2}^{\tau}\left(A_i\right)-\widehat{m}_{g_1}^{\tau}\left(A_i\right)\right)\right\|_{\mathcal{H}} \\ &= \left\|\sum_{i=1}^{N}\left(\widehat{m}_2\left(A_i\right)-\widehat{m}_1\left(A_i\right)\right)\right\|_{\mathcal{H}}\end{aligned} \tag{44}$$

Thus,

$$\text{HODMMD}^{\tau}\left(\widehat{m}_1,\widehat{m}_2\right) = \text{HODMMD}^{\tau}\left(\widehat{m}_2,\widehat{m}_1\right) \tag{45}$$

□

Property 4. When $\widehat{m}_1 = \widehat{m}_2$, the value of HODMMD is always equal to 0.

$$\text{HODMMD}^{\tau}\left(\widehat{m}_1,\widehat{m}_2\right) = 0 \tag{46}$$

Proof. When $\widehat{m}_1 = \widehat{m}_2$

$$\begin{aligned}
\text{HODMMD}^{\tau}\left(\widehat{m}_1, \widehat{m}_2\right) &= \left\| \sum_{i=1}^{N} \left(\widehat{m}_{g_2}^{\tau}\left(A_i\right) - \widehat{m}_{g_1}^{\tau}\left(A_i\right) \right) \right\|_{\mathcal{H}} \\
&= \left\| \sum_{i=1}^{N} \left(\widehat{m}_2\left(A_i\right) - \widehat{m}_1\left(A_i\right) \right) \right\|_{\mathcal{H}} \\
&= \left\| \sum_{i=1}^{N} \left(\widehat{m}_1\left(A_i\right) - \widehat{m}_1\left(A_i\right) \right) \right\|_{\mathcal{H}} \\
&= 0
\end{aligned} \tag{47}$$

□

A.3 Broader Impacts

This paper proposes a dynamic learning strategy based on nonuniform splitting mechanism and Hilbert space mapping to promote real-world applications. We apply the proposed method in pattern classification, image classification, and low-light image enhancement encountered in real life, which encourages research on their synergistic combination in real life. In addition, we are the first to introduce DSE theory to low-light image enhancement and achieve effective performance enhancement for low-light image enhancement tasks. This provides a new idea for the low-light image enhancement task. As far as this paper is concerned, we believe that the proposed method does not have any significant negative impact.

A.4 Additional details regarding the experiment

A.4.1 Dataset and experimental settings

To rigorously evaluate the performance of the proposed method, we conducted extensive experiments and applied the proposed algorithm in three experiments. For pattern classification, experiments were conducted on the Iris [18], Heart [27], Hepatitis [1], Parkinsons [36], Australian [42], Segment [2], and Connectionist Bench (CBench)[16] datasets, and the details of these datasets are shown in Table 4. In this evaluation phase, we compare the proposed method with a class of classical classifiers: a Bayes theorem-based classifier (NaB) [23], a k-nearest neighbor classification method (kNN) [12], a decision tree algorithm (REPTree) [19], a support vector machine classifier (SVM) [5], an SVM method with a radial basis function (SVM-RBF) [5], a multilayer perceptron method (MLP) [4], and a radial basis function network (RBFN) [8]. Another class is evidence theory-based classifiers: a DS theory-based kNN method (kNN-DST) [15], a data probability distribution-based method (NDC) [60], an evidence calibration method (EvC) [61], and a generalized divergence-based decision-making method (DMA) [58].

Table 4: Dataset information.

Dataset	Instances	Class	Features	Missing Values
Iris[18]	150	3	4	No
Heart[27]	270	2	13	No
Hepatitis[1]	155	2	19	Yes
Parkinsons[36]	197	2	22	No
Australian[42]	690	2	14	Yes
Segment[2]	2310	7	19	No
CBench[16]	990	11	10	No

For image classification, experiments were conducted on the publicly available datasets CIFAR-10[31] and CIFAR-100[31], and the details of these datasets are shown in Table 5. Furthermore, performance comparisons were made with current state-of-the-art methods, including DIR-Net [40],

 https://doi.org/10.52202/085713-0409

MST[51], ReSTE[54], ADMM[9], SML[14], UDSP[20], BiPer [50], TAB[28], APL[70], ESNN[47], Dspike[34], GLIF[66], Diet-SNN[43], PASNN[17], MPBN [22], MS-ResNet[24], and BKDSNN[63].

Table 5: Dataset information.

Dataset	Train set	Test set	Class
CIFAR-10 [31]	50000	10000	10
CIFAR-100 [31]	50000	10000	100

Table 6: Dataset information.

Dataset	Train set	Test set
LOL-v1 [53]	485	15
LOL-v2-real [65]	689	100
LOL-v2-syn [65]	900	100
SID [7]	2099	598
SMID [6]	20809	5046

In addition, we validated the effectiveness of the proposed method on a low-light image enhancement task and tested it on the publicly available datasets LOL-v1 [53], LOL-v2-real [65] and LOL-v2-syn [65]. The allocation of the datasets and related information is shown in Table 6. In this evaluation phase, our method was compared with several state-of-the-art low-light image enhancement methods, including MIRNet[68], FIDE[59], ZeroDCE[21], Sparse[53], DRBN[64], RUAS[37], ZeroDCE++[33], SCI[39] and Restormer[29]. In addition, to realistically demonstrate the superiority of our method, methods from recent years were selected as the baseline network, and only targeted training strategies were added to the original methods, including SNR [62], LLFlow-L [52], LLFlow-S [52] and Retinexformer [3]. For fairness, the parameter settings were kept the same as those in the original method. All the experiments were run on NVIDIA RTX 3090 GPUs.

A.4.2 Extended ablation studies

To verify the indispensability of the ADPT and HODMMD modules, we perform ablation studies on the image classification task of the CIFAR-10 dataset with the architecture of ResNet-18, and the results are shown in Table 7.

Table 7: Quantitative comparison of ablation study.

Dataset	Methods	Accuracy
	w/o ADPT	94.56%
	w/o HODMMD	93.55%
	Replace HODMMD with Euclidean distance	94.31%
CIFAR-10 [31]	Building BBAs using evidential neural networks [48]	94.57%
	Fusion via Dempster's combination rule	93.55%
	Replacing HODMMD with Euclidean distance	94.31%
	Ours	**95.61%**

The above experiments indicate that the model performance decreases to 94.56% when the ADPT module is removed. This shows that ADPT can use the prior information of the data to better address uncertainty information through nonuniform splitting, thereby significantly improving the classification accuracy. The accuracy of the model decreased by 2.06% after the HODMMD module

was removed. The accuracy of the model decreased by 1.3% when the Euclidean distance was used to replace the HODMMD. This finding indicates that our method can more accurately capture the nonlinear difference between complex data distributions, thereby more effectively assessing the conflict between evidence and more reliable fusion decisions.

A.4.3 Hyperparameter sensitivity experiment

To better compare model performance, we have supplemented the experiments with analyses of key hyperparameter sensitivity and model complexity. The numerical experiments are as follows.

$$\widehat{m_1} : \widehat{m_1}(\{\theta_1\}) = \frac{3}{20}, \widehat{m_1}(\{\theta_2\}) = \frac{1}{4}, \widehat{m_1}(\{\theta_1, \theta_2\}) = \frac{3}{10}, \widehat{m_1}(\{\theta_3\}) = \frac{1}{10}, \widehat{m_1}(\{\theta_1, \theta_3\}) = \frac{1}{5};$$

$$\widehat{m_2} : \widehat{m_2}(\{\theta_1\}) = \frac{1}{4}, \widehat{m_2}(\{\theta_2\}) = \frac{3}{20}, \widehat{m_2}(\{\theta_1, \theta_2\}) = \frac{1}{5}, \widehat{m_2}(\{\theta_3\}) = \frac{1}{5}, \widehat{m_2}(\{\theta_1, \theta_3\}) = \frac{1}{5}.$$

By fixing the kernel bandwidth $\sigma = 0.5$, we tested the variation in HODMMD in the range of [1,6], and the variation values of the mass function and HODMMD were plotted as a line graph. As τ increases, the value of the mass function of the uncertainty information gradually decrease is assigned to the mass function represents a single category, and the HODMMD value also decreases. In addition, $\tau = 1$ was fixed, and the variation of HODMMD in the range of [0.05, 10] for σ was tested. When σ is between [0.05, 0.1], the HODMMD rapidly decreases, and when σ is between [0.1, 2], it tends to increase and eventually stabilizes.

Second, in the classification and LLIE tasks, τ controls the degree of nonuniform splitting. We tested the effect on task accuracy when τ was 1,2,3,4,5,6. The results show that when $\tau = 5$, the model performance remains stable and optimal. Similarly, for σ, we took the interval as 0.1 to test the effect on task accuracy between [0.1, 2]. The results show that when $\sigma = 0.5$, the model performance remains stable and optimal. In summary, we fix $\tau = 5$ and $\sigma = 0.5$.

A.4.4 Analysis of model complexity

We have added parameter (M), FLOPs (G) and FPS comparisons with the baseline method on the low-light image enhancement task. The results are shown in Table 8.

Table 8: Efficiency Comparison of different methods.

Methods	Param(M)	FLOPs(G)	FPS
SNR	4.01	26.35	1.175
SNR-TTS	5.07	37.56	1.172
LLFlow-L	37.68	287	0.813
LLFlow-L-TTS	38.74	298.21	0.812
LLFlow-S	4.97	37.86	0.943
LLFlow-S-TTS	6.03	49.07	0.942
Retinexformer	1.61	15.57	1.724
Retinexformer-TTS	2.67	26.78	1.718

The results indicate that the complexity and time of the model both increase after the introduction of the TTS. However, considering the improvement in model performance, this increase is within the acceptable range, which proves the actual application efficiency of our method.

A.4.5 Large dataset experiment

The experiments on the SID and SMID datasets to verify the effectiveness of the TTS by using the Retinexformer as the baseline networks.The results are shown in Table 9.

A.5 Discussion of Our Method with Machine Learning Methods and Deep Learning Methods

Although different methods are currently available to address decision-making problems, our method has the following advantages over machine learning methods and deep learning methods.

https://doi.org/10.52202/085713-0409

Table 9: Quantitative comparison of SID and SMID datasets.

Methods	SMID[6]		SID[7]	
	PSNR	SSIM	PSNR	SSIM
Retinexformer[3] ICCV'2023	29.15	0.815	24.44	0.680
Retinexformer-TTS	**29.23**(+0.08)	**0.816**(+0.001)	**24.62**(+0.18)	**0.682**(+0.002)

First, our approach has a better ability to handle uncertainty. Machine learning and deep learning usually produce only a single probability estimate and make predictions under the assumption of mapping relationships. This approach does not model uncertainty well in the face of conflicting or uncertain information. However, our method directly models uncertainty through BBA, which allows assigning confidence to composite propositions. Moreover, it can be dynamically adjusted according to the level of importance between different pieces of information, capturing the nonspecificity of the information and quantifying the discrepancies between pieces of evidence in a more flexible way.

Second, our method can fuse information from multiple sources in a more rational way. Machine learning and deep learning approaches tend to use network layers for feature learning and prediction and lack interpretability for information fusion decisions. However, our method can understand the reasons behind the decisions well. For example, the proposed method maps data into Hilbert space for computation, which is more responsive to the differences in the true distributions of complex data, reduces conflicts between different types of information, and is highly interpretable.

Third, our approach remains applicable in the presence of limited data. The training process of deep learning suffers from overfitting on small or unrepresentative data and relies on training complex models. However, our method's splitting mechanism, which is based on a priori information, can be more flexible in dealing with uncertainty and conflicting information and can work effectively with limited data.

We propose a dynamic learning strategy based on nonuniform splitting mechanism and Hilbert space mapping enhances the interpretability of the decision. This will promote the application of deep learning technology in a wider range of fields. In addition to classification and low-light image enhancement, our ideas can be applied in all uncertainty tasks, for example, image segmentation (uncertainty of the object boundary) and automatic driving (uncertainty in the fusion of multisensor information). Owing to time constraints, this work requires a large amount of computing resources and time. We will continue to explore the performance of this method in other fields.