Appendix of "Reliable Conflictive Multi-view Learning"

Anonymous submission

We show the proof and derivation, implementation details, hyper-parameters setting and performance comparison with uncertainty-aware methods in this appendix.

Proof and Derivation

This section provides the simplified calculation and explanation of fusion methods, as well as the derivation for KL divergence between two Dirichlet distributions.

Proposition 1 The averaging belief fusion is equal to compute the average of the evidence parameters by Dirichlet distribution.

Proof 1 Let $\boldsymbol{w}^i = (\boldsymbol{b}^i, u^i, \boldsymbol{a}^i)$ and $\boldsymbol{w}^j = (\boldsymbol{b}^j, u^j, \boldsymbol{a}^j)$ be the multinomial opinion corresponding to the i-th and j-th view of the same sample respectively. After averaging belief fusion, $\boldsymbol{w} = (\boldsymbol{b}, u, \boldsymbol{a})$ is the integrated multinomial opinion from the multinomial opinion $\boldsymbol{w}^i$ and $\boldsymbol{w}^j$. Correspondingly, e_k^i, e_k^j and e_k are the k-th category of evidence for the i-th, j-th and integrated view of the same sample respectively.

$$b_k = \frac{e_k}{S}, u = \frac{K}{S}, S = \textstyle\sum_{k=1}^{K}\left(e_k + 1\right), \tag{1}$$

$$b_k = \frac{b_k^i u^j + b_k^j u^i}{u^i u^j}, \tag{2}$$

$$u = \frac{2u^i u^j}{u^i + u^j}, a_k = \frac{a_k^i + a_k^j}{2}. \tag{3}$$

According to equation (1,2,3), e_k is updated as:

$$\begin{aligned}
e_k = b_k S &= \frac{b_k K}{u} \\
&= \frac{b_k^i u^j + b_k^j u^i}{u^i + u^j} \cdot \frac{K(u^i + u^j)}{2u^i u^j} \\
&= \frac{K}{2} \cdot \frac{\frac{K e_k^i}{S^i S^j} + \frac{K e_k^j}{S^i S^j}}{\frac{K}{S^i} \cdot \frac{K}{S^j}} \\
&= \frac{e_k^i + e_k^j}{2}.
\end{aligned}$$

Proposition 2 For the conflictive opinion aggregation, after aggregating a new opinion $\boldsymbol{w}^a$ into the original opinion $\boldsymbol{w}^o$, if the uncertain mass of the new opinion u^a is smaller than the original uncertain mass u^o, the uncertain mass of the aggregated opinion u would be smaller than the original one; conversely, it would be larger.

Proof 2

$$u = \frac{2u^o u^a}{u^o + u^a} = \frac{1}{\frac{1}{2}(1 + \frac{u^o}{u^a})} \cdot u^o.$$

The uncertain mass of the aggregated opinion and the original opinion is:

$$\begin{cases} u < u^o, & if\ u^a < u^o, \\ u = u^o, & if\ u^a = u^o, \\ u > u^o, & if\ u^a > u^o. \end{cases}$$

Derivation 1 The KL divergence between two Dirichlet distributions $D(\mathbf{p}|\boldsymbol{\alpha})$ and $D(\mathbf{p}|\boldsymbol{\beta})$ can be obtained in closed form as follows:

$$\begin{aligned}
& KL[D(\mathbf{p}|\boldsymbol{\alpha}) \parallel D(\mathbf{p}|\boldsymbol{\beta})] \\
& = \mathbb{E}_{D(\mathbf{p}|\boldsymbol{\alpha})}[\ln D(\mathbf{p}|\boldsymbol{\alpha}) - \ln D(\mathbf{p}|\boldsymbol{\beta})] \\
& = \ln\Gamma(\alpha_0) - \ln\Gamma(\beta_0) + \sum_{k=1}^{K} \ln\Gamma(\beta_k) - \ln\Gamma(\alpha_k) \\
& + \sum_{k=1}^{K} (\alpha_k - \beta_k)\mathbb{E}_{D(\mathbf{p}|\boldsymbol{\alpha})}\left[\ln p_k\right] \\
& = \ln\Gamma(\alpha_0) - \ln\Gamma(\beta_0) + \sum_{k=1}^{K} \ln\Gamma(\beta_k) - \ln\Gamma(\alpha_k) \\
& + \sum_{k=1}^{K} (\alpha_k - \beta_k)(\psi(\alpha_k) - \psi(\alpha_0))
\end{aligned} \tag{4}$$

where $\alpha_0 = \sum_{k=1}^{K} \alpha_k$ and $\beta_0 = \sum_{k=1}^{K} \beta_k$.

Supplementary Experimental Content

Implementation Details

The Handwritten Dataset, CUB Dataset, HMDB Dataset, Scene15 Dataset, Caltech101 Dataset and PIE Dataset are composed of pre-extracted vectorized features. Then, we utilize fully connected networks with a ReLU layer to extract

Data	MCDO	DE	UA+	EDL	Ours	△%
HandWritten	97.37±0.38	<u>98.55±0.36</u>	97.10±0.82	97.25±0.45	$\mathbf{99.40 \pm 0.00}$	0.86
CUB	89.76±0.56	<u>90.32±0.31</u>	86.73±1.77	89.43±1.20	$\mathbf{98.50 \pm 2.75}$	9.06
HMDB	53.06±0.97	57.39±1.23	53.37±1.33	<u>58.90±1.18</u>	$\mathbf{90.84 \pm 1.86}$	54.23
Scene15	<u>52.93±1.34</u>	39.29±1.08	41.23±1.11	46.57±0.52	$\mathbf{76.19 \pm 0.12}$	43.94
Caltech101	91.56±0.51	91.35±1.12	<u>92.43±1.14</u>	90.83±0.54	$\mathbf{95.36 \pm 0.38}$	3.17
PIE	83.96±1.04	70.23±2.43	80.76±2.17	<u>84.33±0.86</u>	$\mathbf{94.71 \pm 0.02}$	12.31

Table 1: Accuracy (%) on normal test sets. The best and the second best results are highlighted by **boldface** and <u>underlined</u> respectively. △% denotes the performance improvement of ECML over the best baseline.

Algorithm 1 Reliable Conflictive Multi-view Learning

Input: Multi-view dataset:$\{\{\mathbf{x}_n^v\}_{v=1}^V, \mathbf{y}_n\}$
Parameter: Hyper-parameter: $\gamma = 1, \beta$ = 1
/—Train—/
Output: networks parameters.

while not converged **do**
  for $v = 1 : V$ **do**
    $\mathbf{e}^v \leftarrow$ evidential network output.
    Calculate opinion $\mathbf{w}^v$ by Eq. 3.
  end for
  Calculate joint opinion $\mathbf{w}$ by Eq. 6.
  Calculate $\mathbf{e}$ by Eq. 3.
  Calculate $\boldsymbol{\alpha} = \mathbf{e} + \mathbf{1}$ by Eq. 3.
  Calculate the overall loss L by Eq. 12.
  Update the networks by gradient descent by L.
end while
return networks parameters.

/—Test—/
Output: the final prediction and corresponding uncertainty.

for $v = 1 : V$ **do**
  $\mathbf{e}^v \leftarrow$ evidential network output.
  Calculate opinion $\mathbf{w}^v$ by Eq. 3.
end for
Calculate joint opinion $\mathbf{w}$ by Eq. 6.
Calculate $\mathbf{e}, u, S$ by Eq. 3.
Calculate $\boldsymbol{\alpha} = \mathbf{e} + \mathbf{1}$ by Eq. 3.
Calculate $\mathbf{p} = \boldsymbol{\alpha}/S$.
return the decision $\mathbf{p}$ and corresponding uncertainty u.

---

view-specific evidence. In averaging belief fusion, the opinions are fused in order of uncertainty from high to low. The hyper-parameter γ, β for all datasets are set as 1, respectively. The parameter β represents the loss ratio of all single views to the fused view. Since the data has been normalized in advance, $\beta = 1$. The model is implemented by PyTorch 3.10 on one NVIDIA GeForce GTX 3070.

Hyper-parameters Analysis

Fig.1 shows the results using different hyper-parametersγ on the original Scene15 dataset, where γ = 0 means we just use the fusion strategy and do not consider the consistency loss among multiple views. We can find the experiment results best when $\gamma = 1$, but has slight negative effect with low values of γ.

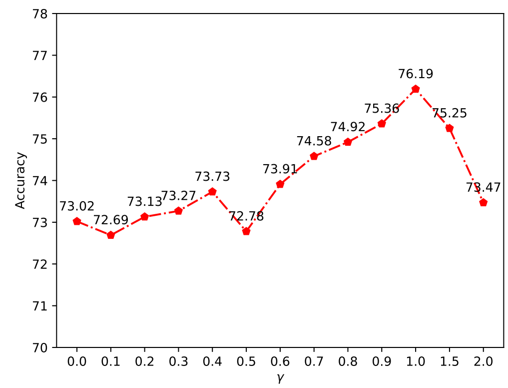

Figure 1: Hyper-parameters Analysis.

Uncertainty-aware deep learning methods.

To further evaluate the performance of our approach, we compared our proposed approach to several existing uncertainty-aware deep learning methods include: (1) MCDO (Monte Carlo Dropout) (Gal and Ghahramani 2016) (Gal and Ghahramani 2016) converts dropout network training to approximate inference in Bayesian neural networks; (2) DE (deep Ensemble) (Lakshminarayanan, Pritzel, and Blundell 2017) (Lakshminarayanan, Pritzel, and Blundell 2017) utilizes multiple independent sub-models and combines the predictions of the sub-models; (3) UA(Uncertainty Perceived Attention) (Heo et al. 2018) (Heo et al. 2018) generates attention weights that follow a Gaussian distribution based on the learned mean and variance, which allows heteroscedastic uncertainty to be captured and produces a more accurate calibration of prediction uncertainty; (4) EDL(Evidential Deep Learning)(Sensoy, Kaplan, and Kandemir 2018) (Sensoy, Kaplan, and Kandemir 2018) calculates evidences of all categories as the parameters of Dirichlet distribution to design the predictive distribution of the classification. Since some uncertainty-aware methods use single-view data, we concatenate the original features of multiple views for all the comparison methods. The results shown in Table 1 indicate that our method outperforms other methods on all datasets. Taking the results on HMDB and Scene15 as examples, our method improves the accuracy by about 31.94% and 23.26% compared to the second-best models (EDL/MCDO) in terms of accuracy respectively, which verifies the superiority of the proposed method.

References

Gal, Y.; and Ghahramani, Z. 2016. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. In *International Conference on Machine Learning*, volume 48, 1050–1059.

Heo, J.; Lee, H. B.; Kim, S.; Lee, J.; Kim, K. J.; Yang, E.; and Hwang, S. J. 2018. Uncertainty-Aware Attention for Reliable Interpretation and Prediction. In *Advances in Neural Information Processing Systems*, volume 31.

Lakshminarayanan, B.; Pritzel, A.; and Blundell, C. 2017. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. In *Advances in Neural Information Processing Systems*, volume 30.

Sensoy, M.; Kaplan, L.; and Kandemir, M. 2018. Evidential Deep Learning to Quantify Classification Uncertainty. In *Advances in Neural Information Processing Systems*, volume 31.