SUPPLEMENTARY MATERIAL

Author(s) Name(s)

Author Affiliation(s)

A Details of the Evidence-Based Loss Function L_{eb}

Evidential Deep Learning (EDL) [1] was developed to mitigate limitations in traditional softmax-based models, particularly regarding inaccurate uncertainty estimation and overconfident incorrect predictions. Missing modalities can further exacerbate these issues. To enhance model robustness, we integrate the EDL framework [1] into our approach. Under the framework of Subjective Logic [2], the predictive probability distribution across K possible labels can be modeled as a Dirichlet distribution [2]. Specifically, Subjective Logic [2] represents each of the K mutually exclusive labels as a belief mass b_k for class $k = 1, \cdots, K$, with an additional uncertainty mass u:

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

where $u \geq 0$ and $b_k \geq 0$. The belief mass b_k is derived from the evidence associated with class k. Let $e_k \geq 0$ be the evidence for class k, then the belief b_k and uncertainty u are computed as follows:

$$b_k = \frac{e_k}{S} \text{ and } u = \frac{K}{S}, \tag{2}$$

where $S = \sum_{i=1}^{K}(e_i+1)$. The belief assignment, also known as a subjective opinion, corresponds to a Dirichlet distribution parameterized by $\alpha_k = e_k+1$, characterized by K parameters $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_K]$, and is expressed as:

$$D(\boldsymbol{p}|\boldsymbol{\alpha}) = \begin{cases} \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^{K} p_i^{\alpha_i - 1} & \text{for } \boldsymbol{p} \in \boldsymbol{S}_K, \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where $\boldsymbol{S}_K$ is the K-dimensional unit simplex:

$$\boldsymbol{S}_K = \left\{ \boldsymbol{p} \mid \sum_{i=1}^{K} p_i = 1 \text{ and } 0 \leq p_1, \cdots, p_K \leq 1 \right\}. \tag{4}$$

Following the EDL theory [1], the expected probability for the k-th class is:

$$p_k = \frac{\alpha_k}{S}, \tag{5}$$

where $\boldsymbol{p} = \frac{\boldsymbol{\alpha}}{S} \in [0,1]^K$. In our model, after the backbone network (e.g., ViLT [3]) outputs logits before softmax, we follow the EDL approach [1] by treating the evidence as $ReLU(\text{logits})$, i.e., $\boldsymbol{e} = ReLU(\text{logits})$. Treating $D(\boldsymbol{P}|\boldsymbol{\alpha})$ as prior information, the Bayes risk formulation for cross-entropy, which constitutes the L_{eb} loss, is expressed as:

$$\begin{aligned} L_{eb} &= \int \left[\sum_{j=1}^{K} -y_j \log p_j \right] \frac{\prod_{j=1}^{K} p_j^{\alpha_j - 1}}{B(\boldsymbol{\alpha})} \, d\boldsymbol{p} \\ &= \sum_{j=1}^{K} y_j (\psi(S) - \psi(\alpha_j)), \end{aligned} \tag{6}$$

where y_j, p_j, and α_j represent the j-th element of the label vector $\boldsymbol{y}$, predictive probability vector $\boldsymbol{p}$, and subjective opinion vector $\boldsymbol{\alpha}$, respectively. Minimizing L_{eb} with respect to the parameters α_j effectively optimizes the predictive probability vector $\boldsymbol{p} = [p_1, \cdots, p_K]$, resulting in the final prediction.

B Detailed Computation of L_{KL}

In Section 2.2, we introduced an additional Kullback-Leibler (KL) divergence term to prevent incorrect labels from generating higher evidence [1]. The term is formulated as:

$$L_{KL} = KL\left[D(\boldsymbol{p}|\tilde{\boldsymbol{\alpha}}) \parallel D(\boldsymbol{p}|\mathbf{1})\right], \tag{7}$$

where $\tilde{\boldsymbol{\alpha}} = \mathbf{y} + (\mathbf{1} - \mathbf{y}) \odot \boldsymbol{\alpha}$. Based on the definition of KL divergence and Equation (3), the expression becomes:

$$\begin{aligned} &KL\left[D(\boldsymbol{p}|\tilde{\boldsymbol{\alpha}}) \parallel D(\boldsymbol{p}|\mathbf{1})\right] \\ &= \log \left(\frac{\Gamma(\sum_{k=1}^{K} \tilde{\alpha}_k)}{\Gamma(K) \prod_{k=1}^{K} \Gamma(\tilde{\alpha}_k)} \right) + \sum_{k=1}^{K} (\tilde{\alpha}_k - 1) \left[\psi(\tilde{\alpha}_k) - \psi\left(\sum_{j=1}^{K} \tilde{\alpha}_j \right) \right], \end{aligned} \tag{8}$$

where $\Gamma(\cdot)$ represents the gamma function and $\psi(\cdot)$ denotes the digamma function. This regularization term helps alleviate the adverse effects of misleading evidence.

C References

[1] Murat Sensoy, Lance Kaplan, and Melih Kandemir, "Evidential deep learning to quantify classification uncertainty," in *NeurIPS*, 2018.

[2] Audun Jøsang, *Subjective Logic: A Formalism for Reasoning Under Uncertainty*, Springer Publishing Company, Incorporated, 1st edition, 2016.

[3] Wonjae Kim, Bokyung Son, and Ildoo Kim, "Vilt: Vision-and-language transformer without convolution or region supervision," in *ICLR*, 2021.