Evidential Deep Learning for Open Set Action Recognition

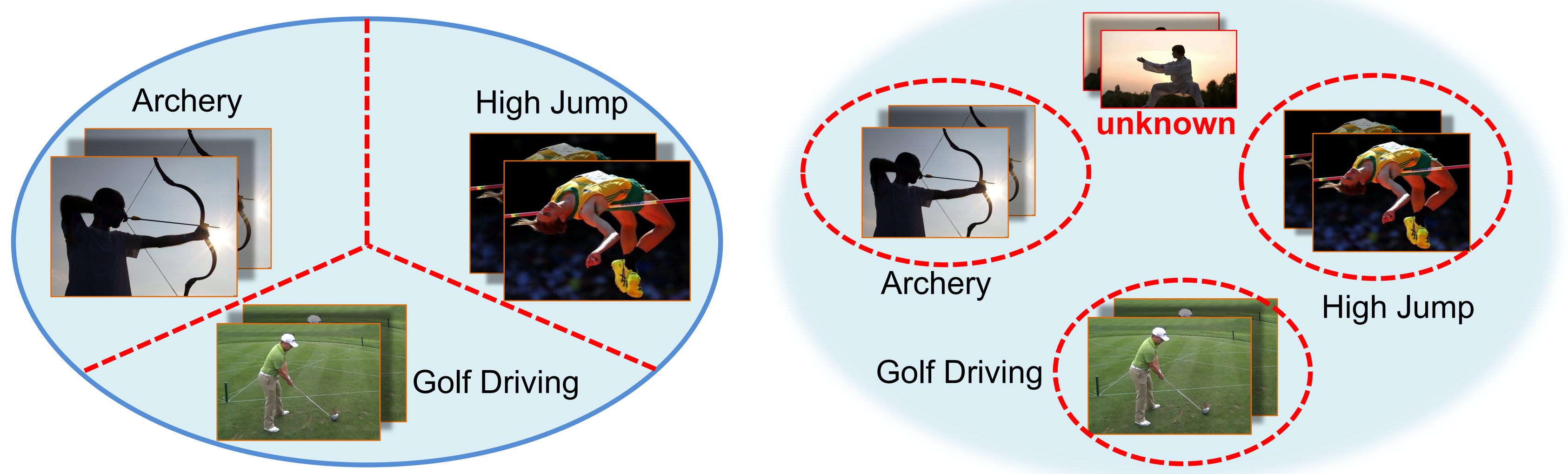
Wentao Bao, Qi Yu, Yu Kong

Golisano College of Computing and Information Sciences, Rochester Institute of Technology (RIT)

ICCV 2021 OCTOBER 11-17 VIRTUAL

Introduction

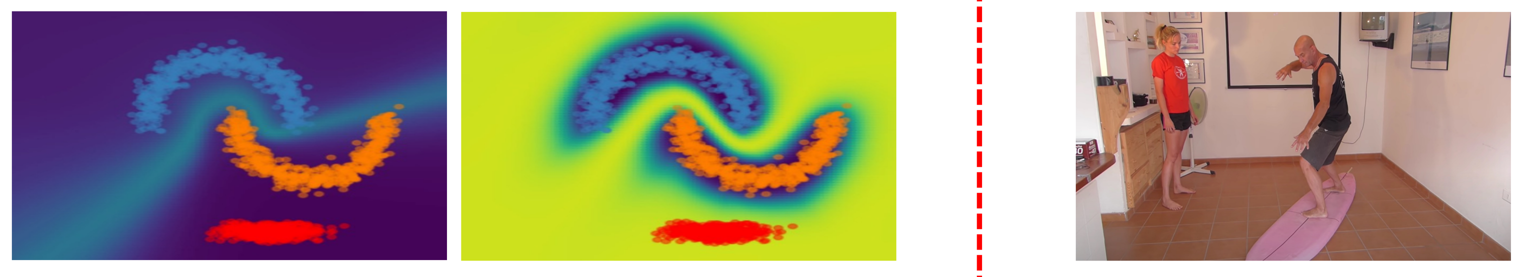
Open Set Action Recognition (OSAR) requests for: (1) **action classification** (2) **reject the unknown.**



Closed Set Action Recognition Open Set Action Recognition

Motivations:

- DNNs are **over-confident** in their predictions, and do not know that they don't know.
- DNNs tend to learn spurious correlation from **static bias** of video data without learning from the human actions.



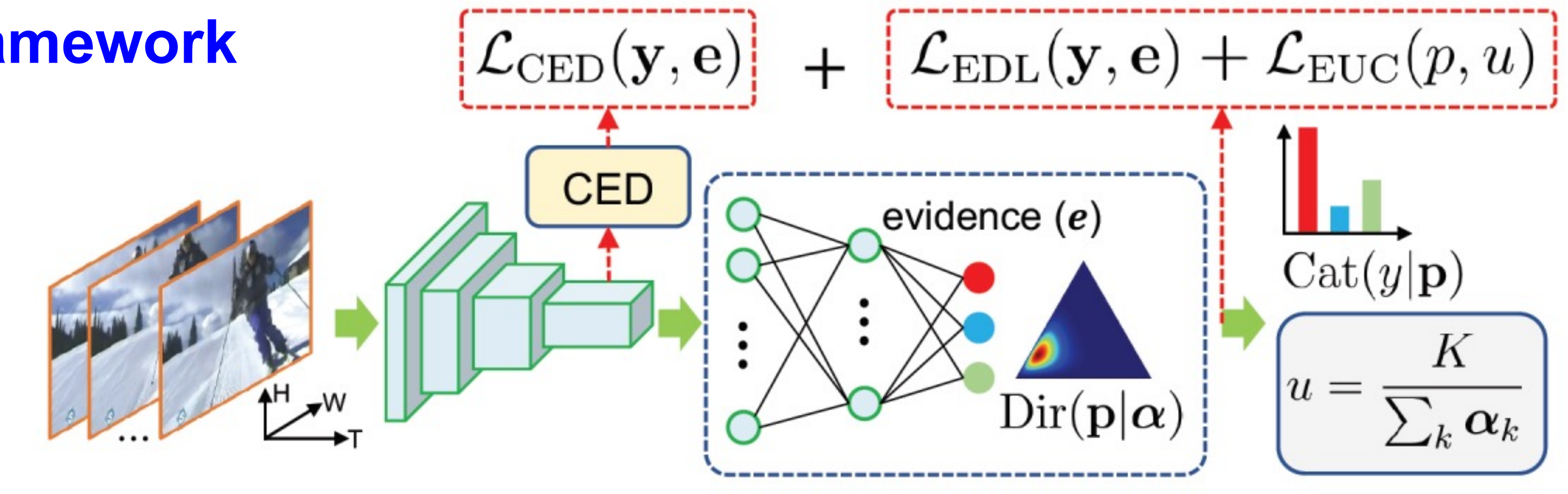
over-confident well-calibrated surfing water

Contribution

- The first Evidential Deep Learning (**EDL**) method for video action recognition in an open world.
- **DEAR:** An open set action recognition model with principled and efficient uncertainty estimation.
- We further proposed **EUC** loss to mitigate over-confident predictions, and **CED** module to eliminate static bias in videos.

DEAR Model

Framework



Evidential Deep Learning (EDL)

- EDL assumes a **Dirichlet Prior** on categorical probabilities, and the strength $\boldsymbol{\alpha}$ of the Dirichlet posterior is learned by DNNs.

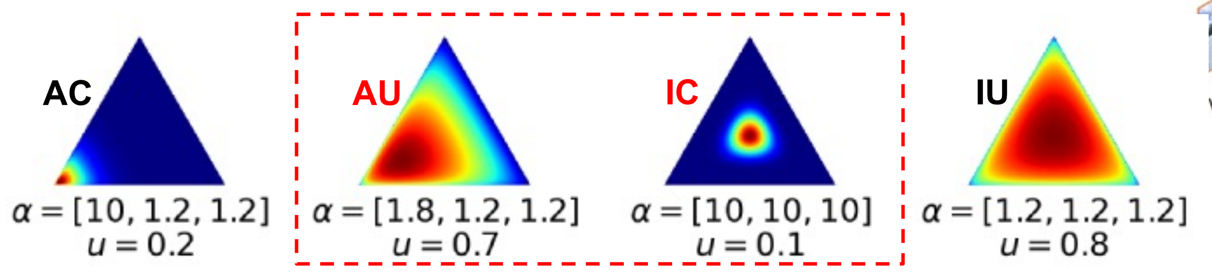
$$\mathcal{L}_{EDL}^{(i)}(\mathbf{y}^{(i)}, \mathbf{e}^{(i)}; \theta) = \sum_{k=1}^{K} \mathbf{y}_k^{(i)} \left(\log S^{(i)} - \log(\mathbf{e}_k^{(i)} + 1) \right)$$

- Dirichlet strength ($\boldsymbol{\alpha}$), Evidence ($\boldsymbol{e}$), Belief ($\boldsymbol{b}$), and uncertainty (u):

$$\boldsymbol{\alpha} = \boldsymbol{e} + \mathbf{1} \qquad \boldsymbol{b} = \frac{\boldsymbol{e}}{\sum_k \alpha_k} \qquad u = \frac{K}{\sum_k \alpha_k}$$

Evidential Uncertainty Calibration

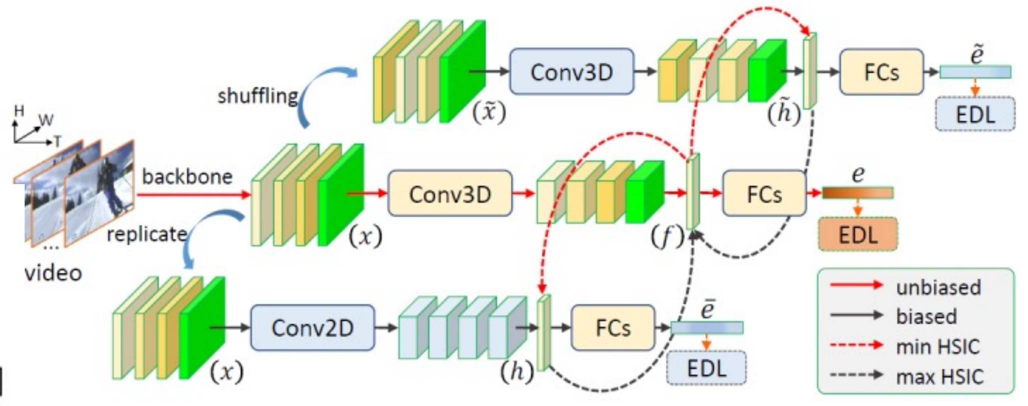
- Be confident in accurate predictions, and uncertain about inaccurate ones.



$$\mathcal{L}_{EUC} = -\lambda_t \sum_{i \in \{\hat{y}_i = y_i\}} p_i \log(1 - u_i) - (1 - \lambda_t) \sum_{i \in \{\hat{y}_i \neq y_i\}} (1 - p_i) \log(u_i)$$

- Here λ_t is exponentially increasing w.r.t. training epoch t from 0.01 to 1.0.

Contrastive Evidence Debiasing



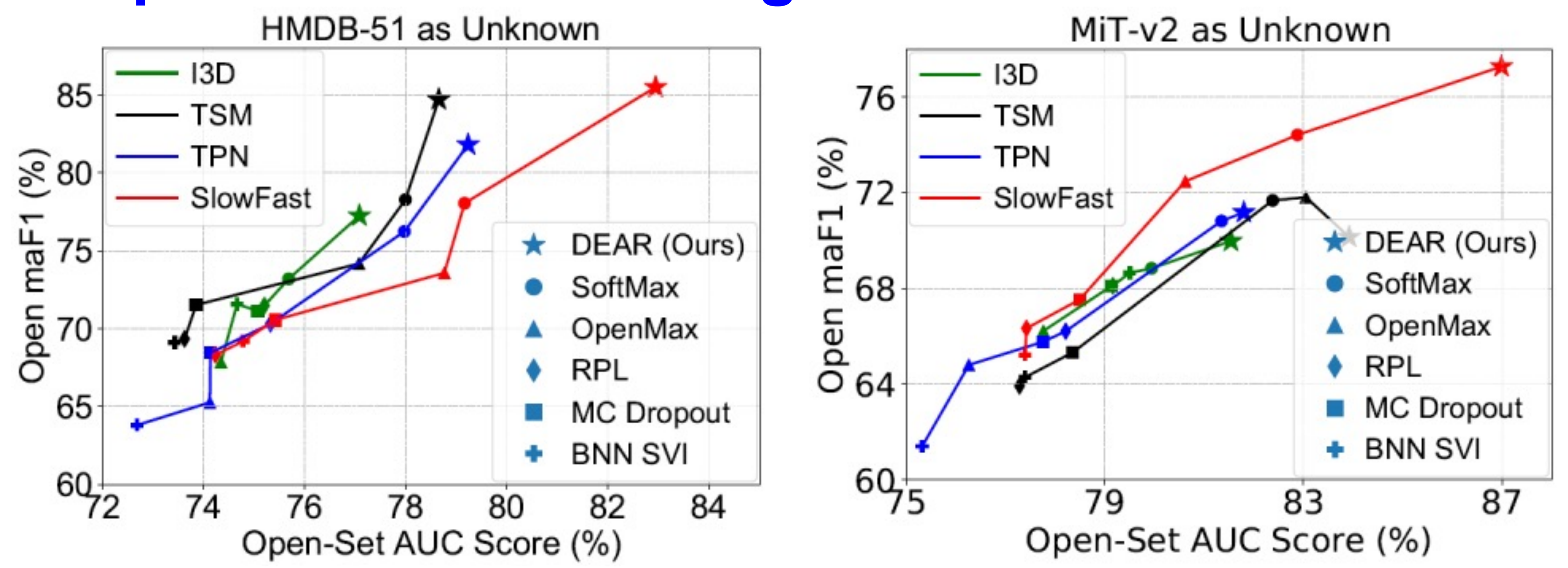
- Debias the evidential feature by two biased branches.
- Alternative optimization:

$$\mathcal{L}(\theta_f, \phi_f) = \mathcal{L}_{EDL}(\mathbf{y}, \mathbf{e}; \theta_f, \phi_f) + \lambda \sum_{\mathbf{h} \in \Omega} \text{HSIC}(\mathbf{f}, \mathbf{h}; \theta_f),$$

$$\mathcal{L}(\theta_h, \phi_h) = \sum_{\mathbf{h} \in \Omega} \{ \mathcal{L}_{EDL}(\mathbf{y}, \mathbf{e}_h; \theta_h, \phi_h) - \lambda \text{HSIC}(\mathbf{f}, \mathbf{h}; \theta_h) \}$$

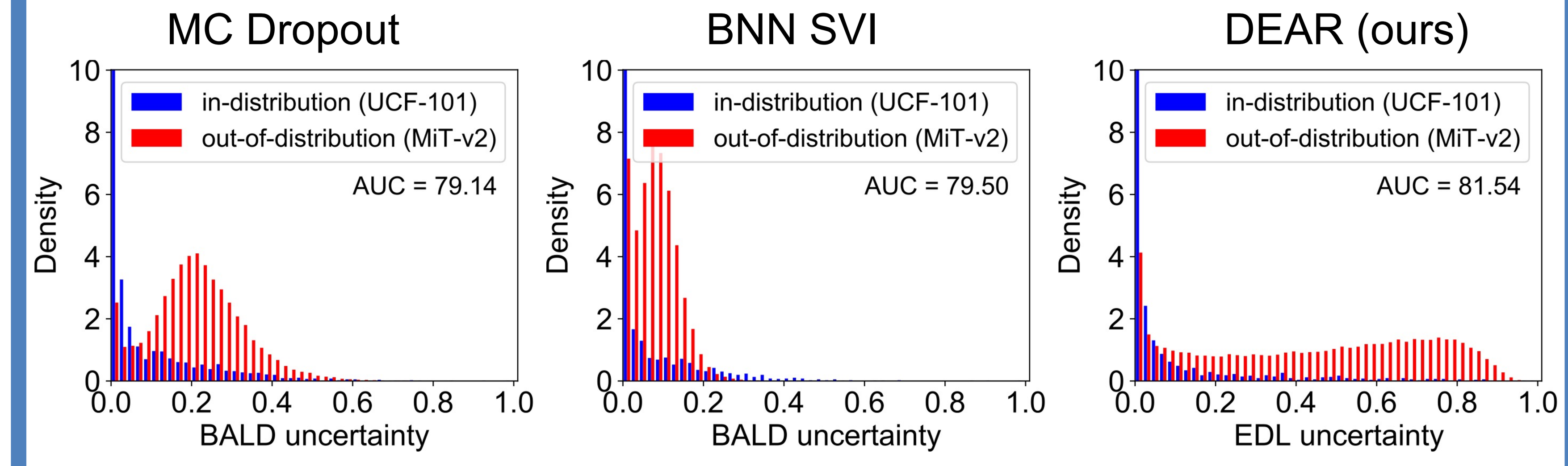
Experimental Results

Open Set Action Recognition



- DEAR method outperforms baselines by large margins.

Out-of-Distribution Detection



- DEAR could learn better uncertainty separation between I.D. and OOD human actions.

Open Resources

Paper Code Project

Feel free to contact **Wentao Bao** via wb6219@rit.edu