Evidential Deep Learning for Open Set Action Recognition
(ICCV-21 Oral)

Wentao Bao[1,2,3], Qi Yu[1,3], Yu Kong[1,2]

[1]Rochester Institute of Technology
[2]ActionLab, [3]MiningLab

Project: https://www.rit.edu/actionlab/dear

Content

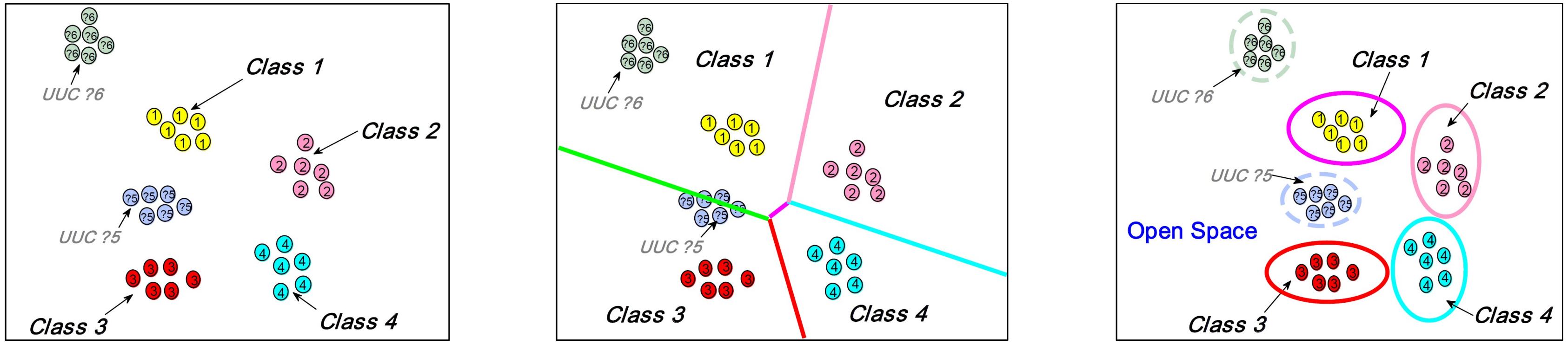
- Open Set Recognition
- Evidential Deep Learning
- The Proposed DEAR Model
- Experimental Results
- Conclusions and Discussions

Rochester Institute of Technology

Open Set Recognition

What is Open Set Recognition (OSR)?

- ❑ Classification model is only trained with known classes (closed set), but tested with any classes (open set).
- ❑ **Identify known** classes and **reject unknown** classes.



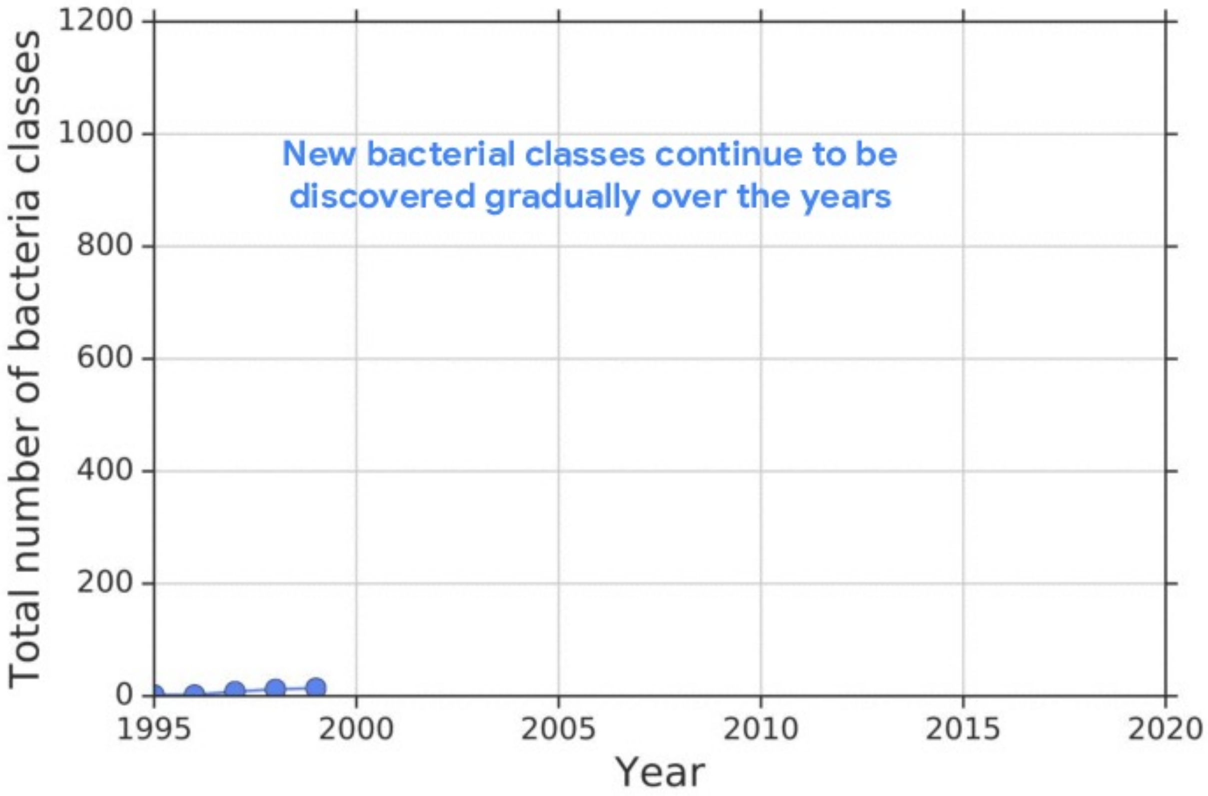
(a) Original Data Distribution (b) Closed Set Recognition (c) Open Set Recognition

Figures are from:

[1] Geng, Chuanxing, Sheng-jun Huang, and Songcan Chen. "Recent advances in open set recognition: A survey." *IEEE TPAMI 2020*.

Open Set Recognition

Why do we care about the UNKNOWN?



[1] Bacterial Classification (Ren *et al,* 2019)

[1] https://ai.googleblog.com/2019/12/improving-out-of-distribution-detection.html

Open Set Recognition

Why not representing the UNKNOWN as a separate class?

"There are Known Knowns…" ---- Donald Rumsheld[1]

- ❑ **Known Unknown:** labeled negative examples, not necessarily meaningful category.
- ❑ **Unknown Unknown:** classes unseen in the training, the most difficult situation.

The knowledge of the Unknown Class is always limited, increasing the difficulty of model learning.

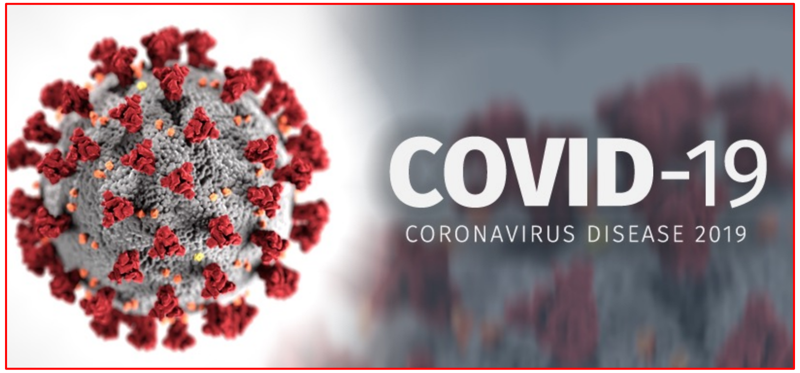
[1] https://en.wikipedia.org/wiki/There_are_known_knowns

Open Set Recognition

What are the benefits to reject the unknown?

Face/Identity Recognition[1]

Autonomous Driving[2]



Virus Diagnosis[3]

[1] https://www.securityinfowatch.com/access-identity/biometrics/facial-recognition-solutions/article/21152899/serious-advancements-in-facial-recognition-technologies.
[2] https://www.pri.org/stories/2016-08-23/stray-cattle-india-get-glow-dark-horns-prevent-crashes-vehicles.
[3] https://southkingstownri.com/998/COVID-19.

Related OSR Work

Early Works

- Prior to (TPAMI12)[1], OSR is only studied for evaluation, e.g., speaker recognition.

OSR for Images

- **Discriminative**: Add new class boundary
  - W-SVM (tpami14), PI-SVM (eccv14), OpenMax (cvpr16), G-OpenMax (bmvc17), PROSER (cvpr21),…
- **Generative**: Learn a large reconstruction distance (or low density)
  - C2AE (cvpr19), CROSR (cvpr19), CGDL (cvpr20)…
- **Prototype Learning:** Learn the prototypical representation of each known class
  - GCPL (cvpr18), RPL (eccv20), ARPL (tpami21), …

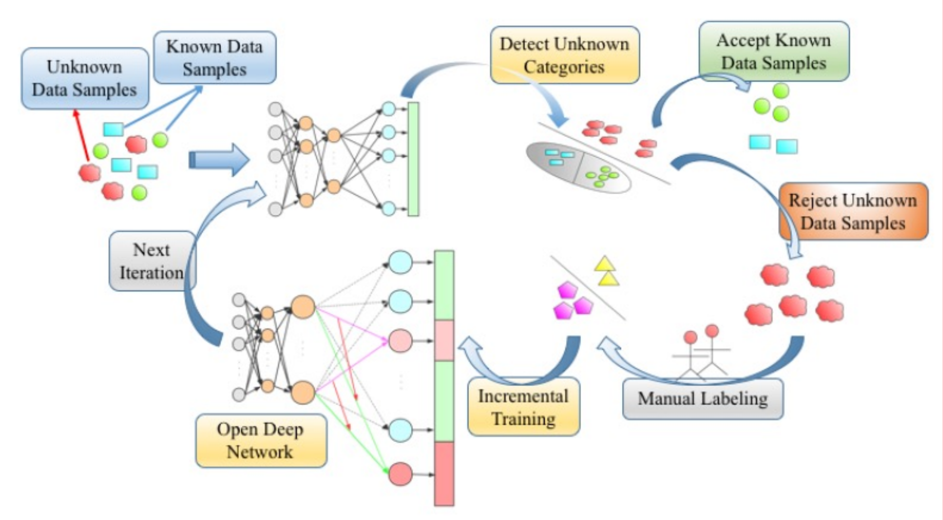
A recent survey by Geng et al (tpami'21)[2], and an awesome github repo:
https://github.com/iCGY96/awesome_OpenSetRecognition_list

[1] Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boult. Toward open set recognition. *IEEE TPAMI*, 35(7):1757–1772, 2012.
[2] Geng, Chuanxing, Sheng-jun Huang, and Songcan Chen. "Recent advances in open set recognition: A survey." *IEEE TPAMI (Early Access)*, 2020.

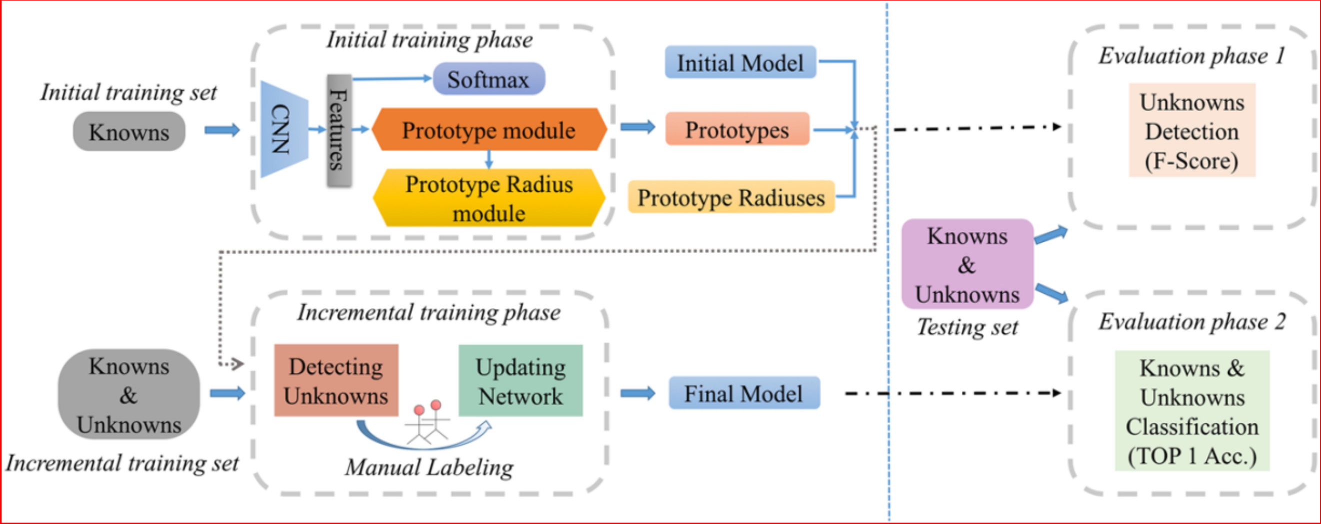
Related OSR Work

◻ OSR for Video

- ODN[1]: incrementally learn new class boundary.
- P-ODN[2]: learn class prototypes and use “fake unknown” labels in the training.



ODN[1]



P-ODN[2]

[1] Shu, Yu, Yemin Shi, Yaowei Wang, Yixiong Zou, Qingsheng Yuan, and Yonghong Tian. “ODN: Opening the deep network for open-set action recognition." In ICME, 2018.
[2] Shu, Yu, Yemin Shi, Yaowei Wang, Tiejun Huang, and Yonghong Tian. “P-ODN: prototype-based open Deep network for open Set Recognition." *Scientific Reports*, 2020.

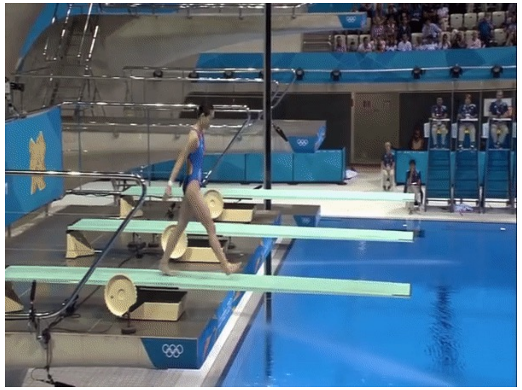
OSR in Video Action Recognition

What makes OSR unique for videos?

Temporal Dynamics

- Distinguishing between known and unknown actions relies on temporal dynamics.
- Temporal features are uncertain.

['Forward', '15som', 'NoTwis', 'PIKE']



['Reverse', '15som', '25Twis', 'FREE']

Static Bias

- DNNs could be over-fitted to static cues, from which the model incorrectly recognizes the unknown as the known.
- Scene bias[1,2], concurrent bias (objects, actor)

Singing in a **baseball field**

Playing the **piano**

Marching with **military uniforms**

[1] Videos are credited to Diving-48 dataset: http://www.svcl.ucsd.edu/projects/resound/dataset.html

[2] Choi, Jinwoo, Chen Gao, Joseph CE Messou, and Jia-Bin Huang. "Why Can't I Dance in the Mall? Learning to Mitigate Scene Bias in Action Recognition." in NeurIPS, 2019

Our Motivations

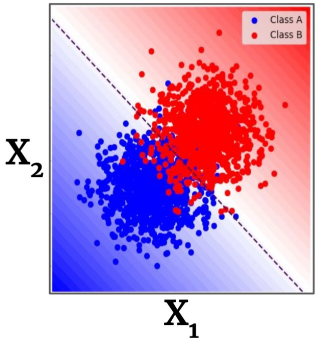
Modeling the unknown as Out-of-Distribution (OOD) data

- **Conventional I.I.D. Assumption**

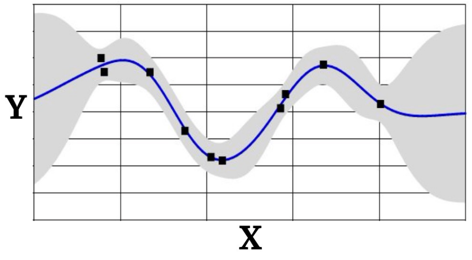
$$p_{\text{test}}(x, y) = p_{\text{train}}(x, y)$$

- **O.O.D. Assumption**

$$p_{\text{test}}(x, y) \neq p_{\text{train}}(x, y)$$



Classification



Regression

- **Examples of OOD Cases**[1]
  - Covariate Shift: $p(y|x)$ is fixed, but feature distribution $p(x)$ changes.
  - Label Shift: $p(x|y)$ is fixed, but label distribution $p(y)$ changes.
  - **Open Set Recognition**: new class appears.

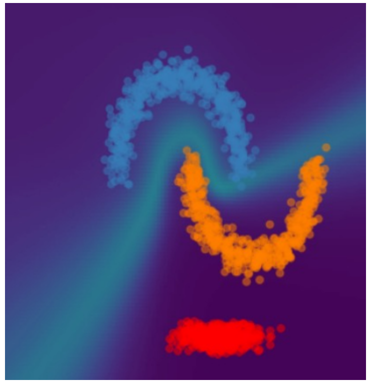
[1] Dustin Tran, et. al., “Practical Uncertainty Estimation & Out-of-Distribution Robustness in Deep Learning” in NeurIPS Workshop, 2020.

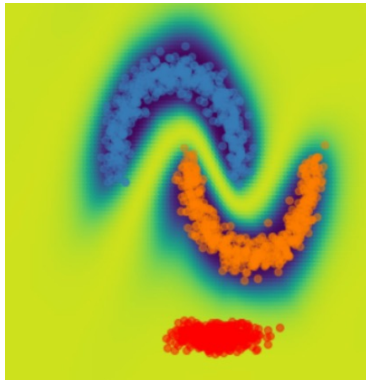
Our Motivations

Modeling the unknown as Out-of-Distribution (OOD) data

- **Uncertainty-based OOD detection for OSR**
  - When x^* is close to $p_{\text{train}}(x, y)$, trust the classification (Low Uncertainty expected).
  - When x^* is far from $p_{\text{train}}(x, y)$, reject it as the unknown (High Uncertainty expected).

- **Challenges**
  - Existing DNNs are **over-confident** in their predictions.
  - *Do not know when they don't know*.


existing DNNs


ideal density

Images are credited to Liu et. al., "Simple and principled uncertainty estimation with deterministic deep learning via distance awareness" in NeurIPS, 2020.

Content

- Open Set Recognition
- Evidential Deep Learning
- The Proposed DEAR Model
- Experimental Results
- Conclusions and Discussions

Rochester Institute of Technology

Deep Learning Uncertainty

Taxonomy of Deep Learning Uncertainty

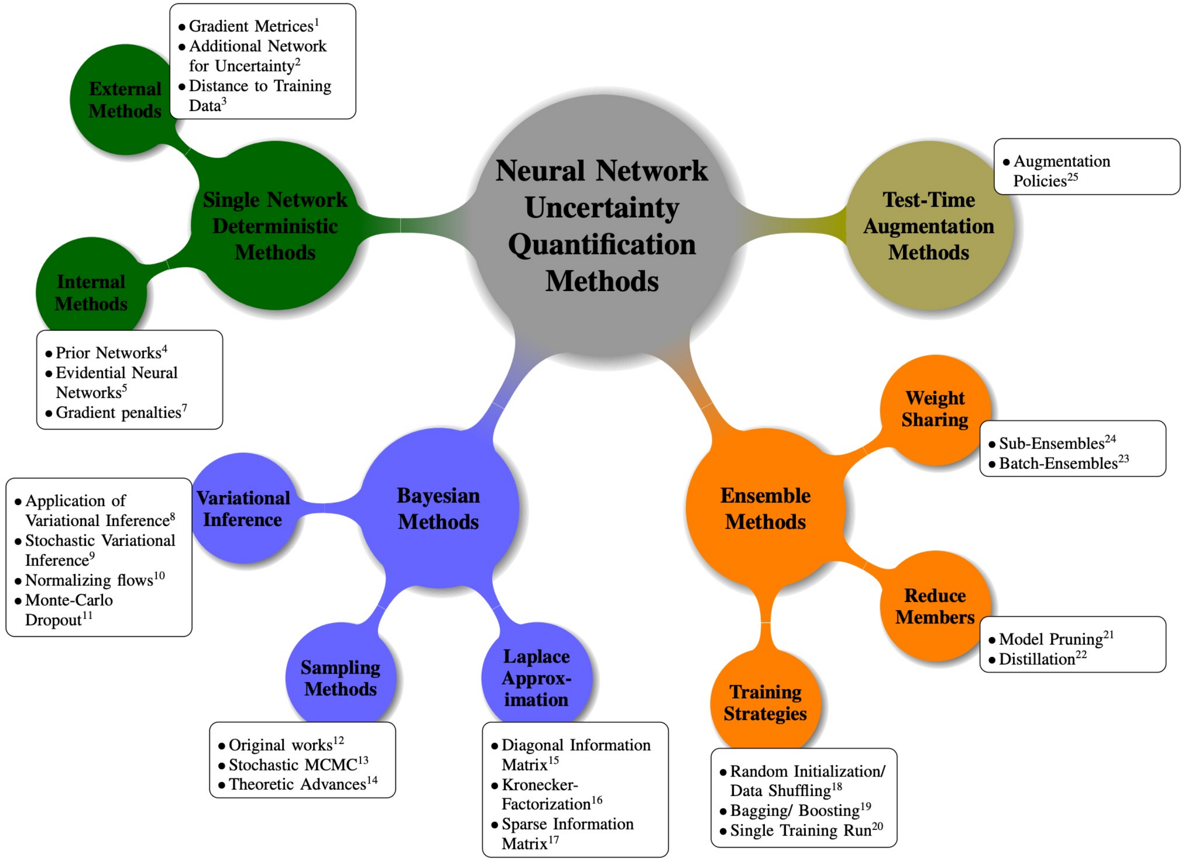
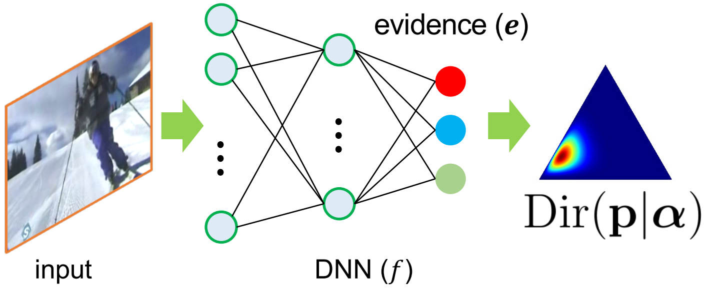


Figure is credited to Gawlikowski et. al., "A Survey of Uncertainty in Deep Neural Networks." arXiv preprint arXiv:2107.03342 (2021).

Evidential Deep Learning

- **Evidential Neural Networks (ENNs)[1]**
  - ENNs assume a Dirichlet Prior on the categorical probabilities.
  - A deterministic mapping is learned, i.e., $\boldsymbol{\alpha} = f(\boldsymbol{x})$, where $\boldsymbol{p} \sim D(\boldsymbol{p}|\boldsymbol{\alpha})$.

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



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

$$\boldsymbol{\alpha} = \boldsymbol{e} + \boldsymbol{1} \qquad \boldsymbol{b} = \frac{\boldsymbol{e}}{\sum_k \boldsymbol{\alpha}_k}$$

  - Expectation of Prediction:

$$\mathbb{E}[p] = \frac{\boldsymbol{\alpha}}{\sum_k \boldsymbol{\alpha}_k}, k = 1, \ldots, K$$

  - Uncertainty (vacuity):

$$u = \frac{K}{\sum_k \boldsymbol{\alpha}_k} \qquad \boxed{u + \sum_k \boldsymbol{b}_k = 1}$$

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

Evidential Deep Learning

- **Training**
  - **Cross-entropy loss** by minimizing the negative log-likelihood (NLL) objective $\mathrm{Mult}(y_i|p_i)$

$$\mathcal{L}_i(\Theta) = -\log\left(\int \prod_{j=1}^{K} p_{ij}^{y_{ij}} \frac{1}{B(\alpha_i)} \prod_{j=1}^{K} p_{ij}^{\alpha_{ij}-1} d\mathbf{p}_i\right) = \sum_{j=1}^{K} y_{ij}\Big(\log(S_i) - \log(\alpha_{ij})\Big) \quad (3)$$

  - Other two alternative forms can be found in [1].
  - Training ENNs is equivalent to gathering evidence to support for correct classification.

- **Inference**
  - Given an input data, an ENN model predicts evidence $\boldsymbol{e}$.
  - The categorical probabilities, and multi-dimensional uncertainties can be derived.

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

Evidential Deep Learning

How do ENNs Quantify Uncertainty?

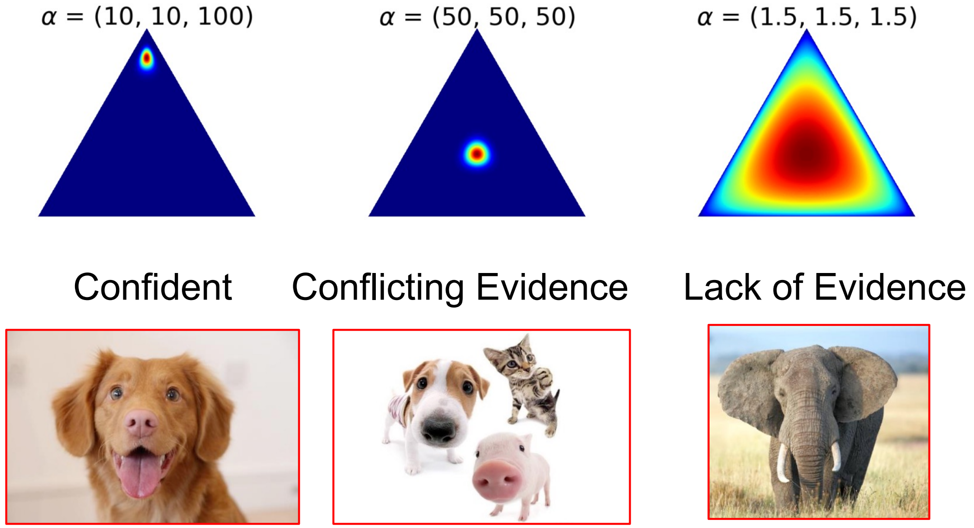
- **ENNs for Classification[1]**
  - **Vacuity**: due to lack of evidence

$$u_v = \frac{K}{\sum_k \boldsymbol{\alpha}_k}$$

  - **Dissonance**: due to conflicting evidence

$$u_d = \sum_k \frac{b_k \sum_{i \neq k} b_i \mathrm{Bal}(b_i, b_k)}{\sum_{i \neq k} b_i}$$

$$\mathrm{Bal}(b_i, b_k) = \begin{cases} 1 - \frac{|b_i - b_k|}{b_i + b_k}, & \text{if } b_i b_k \neq 0 \\ 0, & \text{else.} \end{cases}$$



- **ENNs for Regression[2]:** Aleatoric & Epistemic Uncertainties.

[1] Josang, Audun, Jin-Hee Cho, and Feng Chen. "Uncertainty characteristics of subjective opinions." In FUSION, pp. 1998-2005. IEEE, 2018.
[2] Amini, Alexander, Wilko Schwarting, Ava Soleimany, and Daniela Rus. "Deep evidential regression." in NeurIPS, 2020.

Evidential Deep Learning

What evidential uncertainty do we need for OOD data?

- **Conclusion from Existing Literature**[1]

Special relations on the OOD and the CP.

(a) *For an OOD sample with a uniform prediction (i.e.,* $\alpha = [1, \ldots, 1]$*), we have*

$$\boxed{1 = u_v} = u_{en} > u_{alea} > u_{epis} > \boxed{u_{diss} = 0}$$

(b) *For an in-distribution sample with a conflicting prediction (i.e.,* $\alpha = [\alpha_1, \ldots, \alpha_K]$ *with* $\alpha_1 = \alpha_2 = \cdots = \alpha_K$*, if* $S \to \infty$*), we have*

$$u_{en} = 1, \boxed{\lim_{S \to \infty} u_{diss}} = \lim_{S \to \infty} u_{alea} = 1, \boxed{\lim_{S \to \infty} u_v =} \lim_{S \to \infty} u_{epis} = 0$$

with $u_{en} > u_{alea} > u_{diss} > u_v > u_{epis}$.

<u>Vacuity and Dissonance can clearly distinguish OOD from Conflicting Prediction (CP)</u>

[1] Zhao, Xujiang, Feng Chen, Shu Hu, and Jin-Hee Cho. "Uncertainty Aware Semi-Supervised Learning on Graph Data." in NeurIPS, 2020

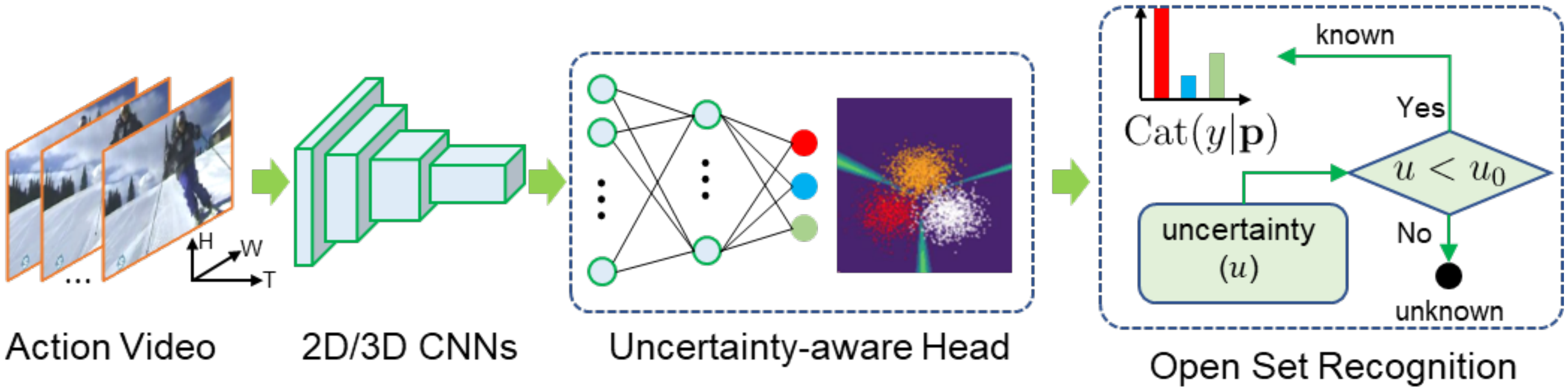
Content

- Open Set Recognition
- Evidential Deep Learning
- The Proposed DEAR Model
- Experimental Results
- Conclusions and Discussions

Rochester Institute of Technology

Proposed DEAR Model

- **Deep Evidential Action Recognition (DEAR)**
  - **Vanilla Framework**



  - **Training Loss**

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

Proposed DEAR Model

Limitations of the Vanilla DEAR

- Could be **over-fitting** due to minimizing only the NLL objective[1,2].
  - OSR model requires high generalization capability to reject the unknowns.
  - We propose to calibrate the uncertainty in training.

- Does not address the **uniqueness of video** data in OSAR setting.
  - Temporal dynamics
  - Static bias (scene bias, object bias, human bias)
  - We propose to debias the evidence by using spatial and temporal features.

[1] Chuan Guo, et. al., On calibration of modern neural networks. In ICML, 2017.
[2] Jishnu Mukhoti, et. al., Calibrating deep neural networks using focal loss. In NeurIPS, 2020.

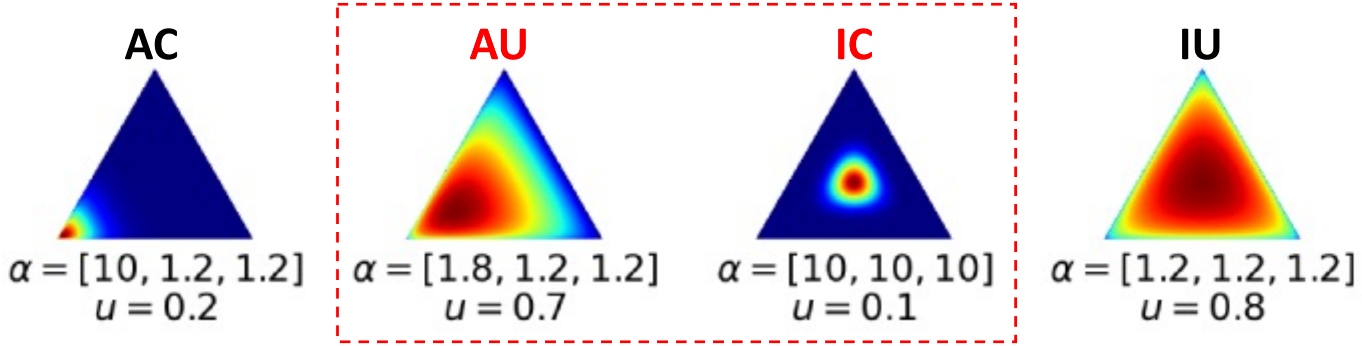
Proposed DEAR Model

Evidential Uncertainty Calibration (EUC)

- “A well-calibrated model should be <u>confident in its accurate prediction</u>, and be <u>uncertain about inaccurate</u> ones” [1,2].
- The goal is to maximizing the *Accuracy versus Uncertainty* (AvU) utility.

$$\mathrm{AvU} = \frac{n_{AC} + n_{IU}}{n_{AC} + n_{AU} + n_{IC} + n_{IU}}$$

- Toy examples of Dirichlet simplex for the four cases:



[1] Ranganath Krishnan and Omesh Tickoo. Improving model calibration with accuracy versus uncertainty optimization. In NeurIPS, 2020.
[2] Jishnu Mukhoti and Yarin Gal. Evaluating bayesian deep learning methods for semantic segmentation. arXiv preprint arXiv:1811.12709, 2018

Proposed DEAR Model

Evidential Uncertainty Calibration (EUC)

- Proposed EUC regularizer in a logarithm form:

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

where the confidence $p_i = \max\{p_i^{(1)}, p_i^{(2)}, \ldots, p_i^{(K)}\}$, and λ_t is an annealing factor controlled by training epoch (t):

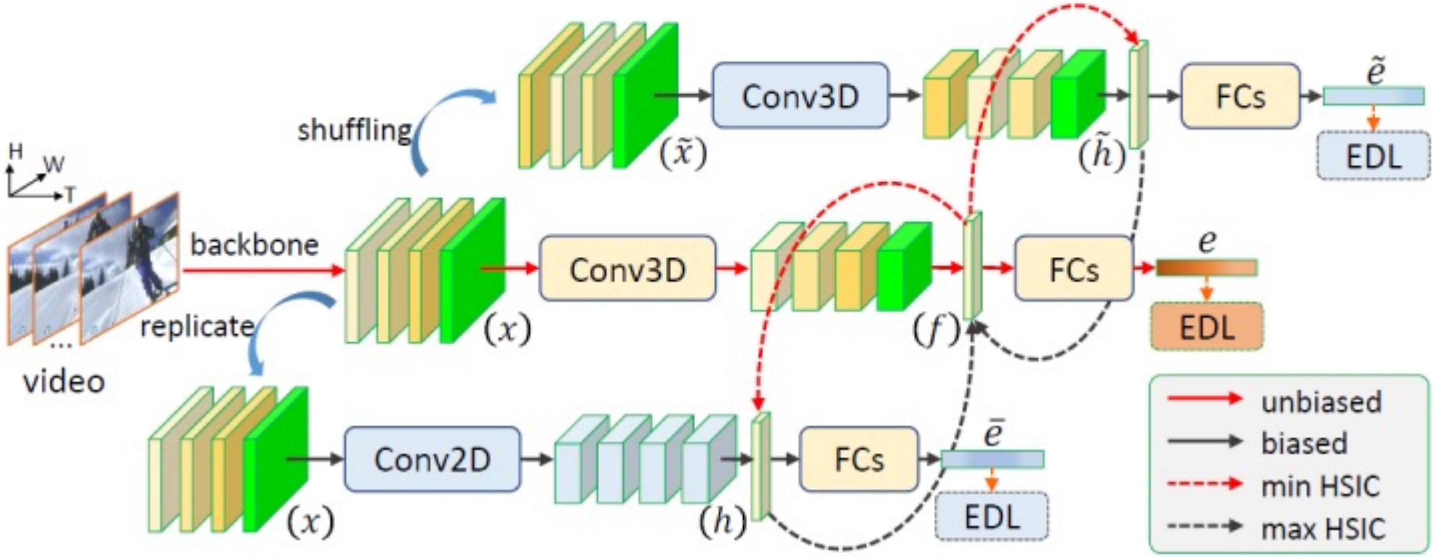
$$\lambda_t = \lambda_0 \cdot \exp\left\{-\frac{\ln \lambda_0}{T} t\right\}$$

- λ_t will be monotonically increasing from a small constant λ_0 to 1.0 within T epochs.

Proposed DEAR Model

Contrastive Evidence Debiasing (CED)

- Inspired by ReBias[1], HSIC is used to debias the features.
- Overview of CED.



- The CED module is only used in training.

[1] Hyojin Bahng, Sanghyuk Chun, Sangdoo Yun, Jaegul Choo, and Seong Joon Oh. Learning de-biased representations with biased representations. In ICML, 2020.

Proposed DEAR Model

Contrastive Evidence Debiasing (CED)

- Learn a discriminative and unbiased feature ($\boldsymbol{f}$):

$$\mathcal{L}(\theta_f, \phi_f) = \underbrace{\mathcal{L}_{EDL}(\mathbf{y}, \mathbf{e}; \theta_f, \phi_f)}_{\boldsymbol{f} \text{ is discriminative}} + \underbrace{\lambda \sum_{\mathbf{h} \in \Omega} \mathrm{HSIC}(\mathbf{f}, \mathbf{h}; \theta_f)}_{\boldsymbol{f} \text{ is pushed away from } \boldsymbol{h}},$$

- Learn biased features ($\boldsymbol{h}$) by *Conv2D* and *Temporal Shuffling*

$$\mathcal{L}(\theta_h, \phi_h) = \sum_{\mathbf{h} \in \Omega} \{\underbrace{\mathcal{L}_{EDL}(\mathbf{y}, \mathbf{e}_h; \theta_h, \phi_h)}_{\boldsymbol{h} \text{ is discriminative}} \underbrace{- \lambda \mathrm{HSIC}(\mathbf{f}, \mathbf{h}; \theta_h)}_{\boldsymbol{h} \text{ catches up to } \boldsymbol{f}}\}$$

- Alternative training vs. Joint Training

Hilbert-Schmidt Independence Criterion (HSIC)

- HSIC is commonly used to measure the dependency of two high-dimensional variables.
- Unbiased HSIC Estimation[1]:

$$\mathrm{HSIC}^{k,l}(U,V) = \frac{1}{m(m-3)}\left[\mathrm{tr}(\tilde{U}\tilde{V}^T) + \frac{\mathbf{1}^T\tilde{U}\mathbf{1}\mathbf{1}^T\tilde{V}\mathbf{1}}{(m-1)(m-2)} - \frac{2}{m-2}\mathbf{1}^T\tilde{U}\tilde{V}^T\mathbf{1}\right],$$

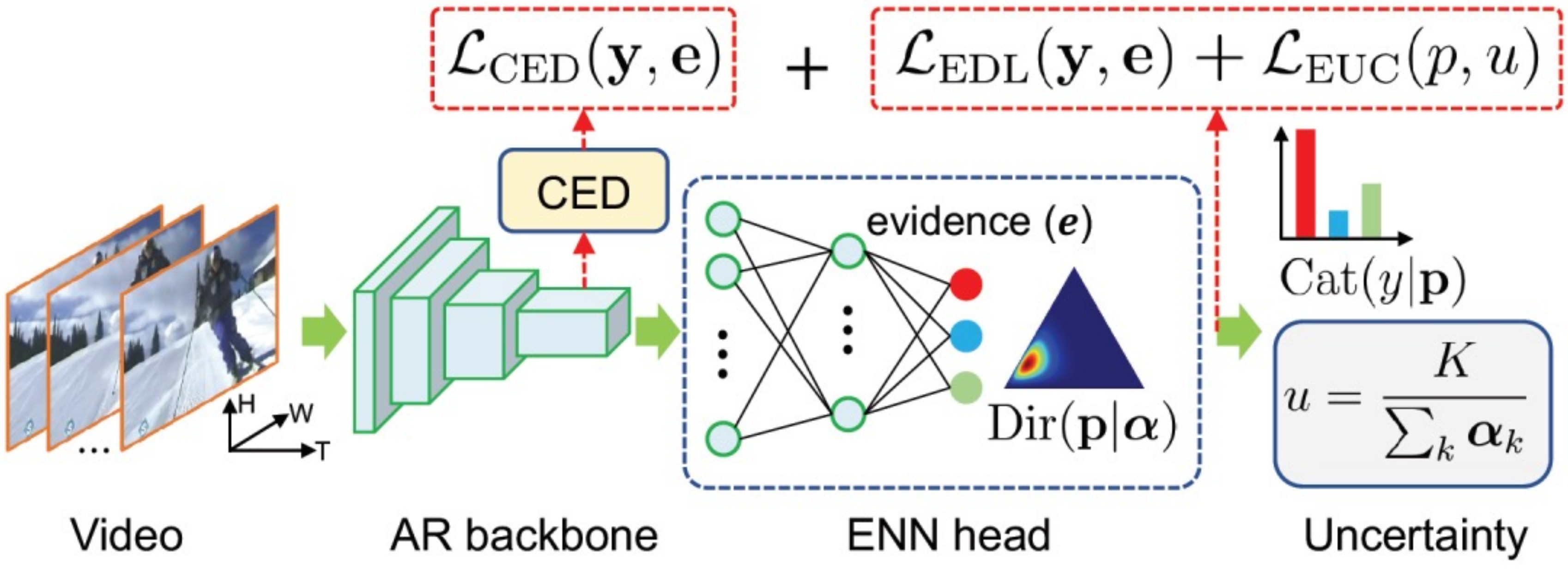
- where $\widetilde{U}$ is the kernelized matrix of U with **RBF kernel** k by $\widetilde{U}_{ij} = \left(1-\delta_{ij}\right)k(u_i, u_j)$.

- HSIC is fully differentiable in training.
- Smaller HSIC indicates U is more independent of V.

[1] Le Song, Alex Smola, Arthur Gretton, Justin Bedo, and Karsten Borgwardt. Feature selection via dependence maximization. *JMLR*, 13(5), 2012.

Proposed DEAR Model

□ Summary of the Complete DEAR Model



Content

- Open Set Recognition
- Evidential Deep Learning
- The Proposed DEAR Model
- Experimental Results
- Conclusions and Discussions

Rochester Institute of Technology

Experimental Results

Video Action Datasets

- **UCF-101**
  - ✓ Contains 101 classes.
  - ✓ For model training, closed-set testing.
- **HMDB-51**
  - ✓ Contains 51 classes
  - ✓ <u>For small-scale unknown testing</u>
- **MiT-v2**
  - ✓ Contains 305 classes, ~20x larger than HMDB-51.
  - ✓ <u>For large-scale unknown testing</u>
- **Kinetics & Mimetics**
  - ✓ Mimetics are out-of-context version of Kinetics, sharing the same class.
  - ✓ 10 same classes of each dataset are selected following [1].
  - ✓ For validating the CED performance.

[1] Hyojin Bahng, Sanghyuk Chun, Sangdoo Yun, Jaegul Choo, and Seong Joon Oh. Learning de-biased representations with biased representations. In ICML, 2020.

Experimental Results

Evaluation Protocols

- Open macro-F1 Score

$$\text{Open maF1} = \frac{\sum_i \omega_O^{(i)} \cdot F_1^{(i)}}{\sum_i \omega_O^{(i)}}$$

where $\omega_O^{(i)}$ is the **openness[1]** when there are i novel classes are used as the unknown.

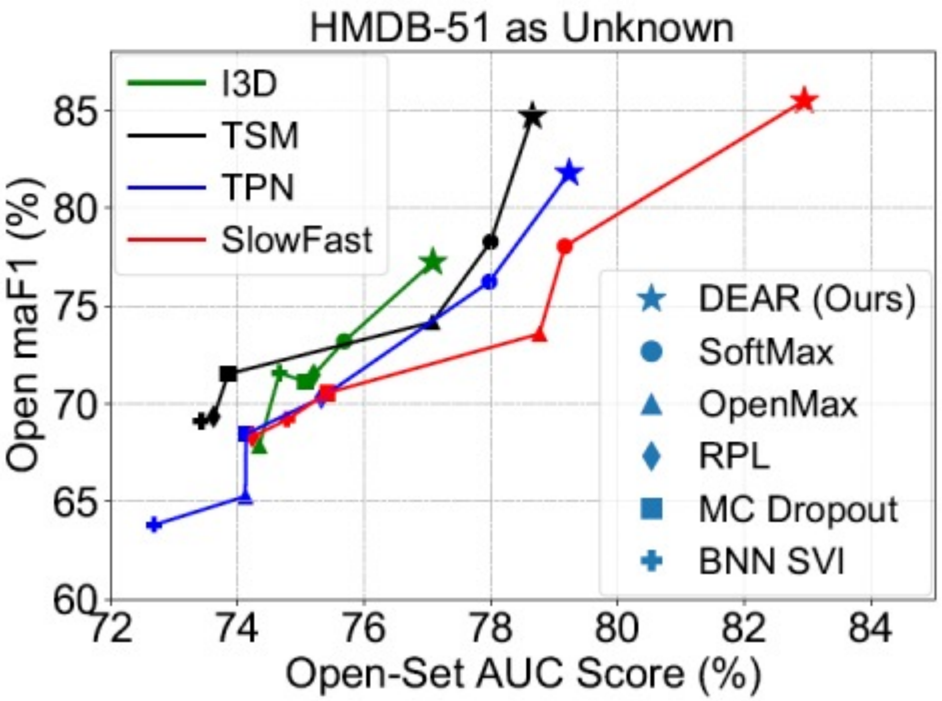
$$\omega_O^{(i)} = 1 - \sqrt{2K/(2K+i)}$$

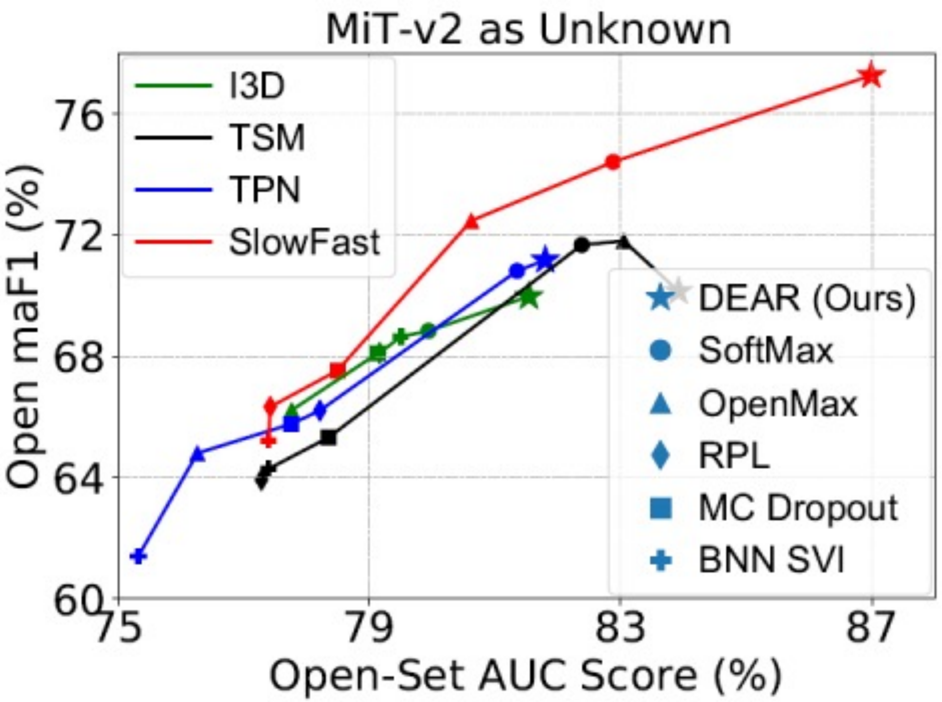
- Open Set AUC
  - Area Under the Curve (AUC) of ROC for distinguishing the known and unknown.
- Closed Set Accuracy
  - Mean accuracy of all K known classes.

[1] Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boult. Toward open set recognition. *TPAMI*, 35(7):1757–1772, 2012.

Experimental Results

☐ Diagram Overview





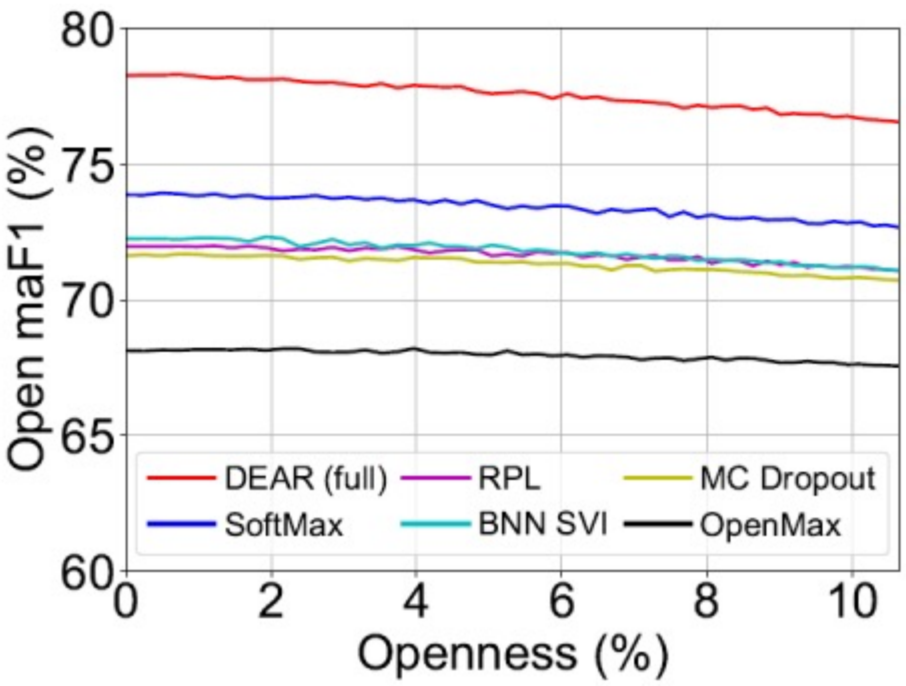
Experimental Results

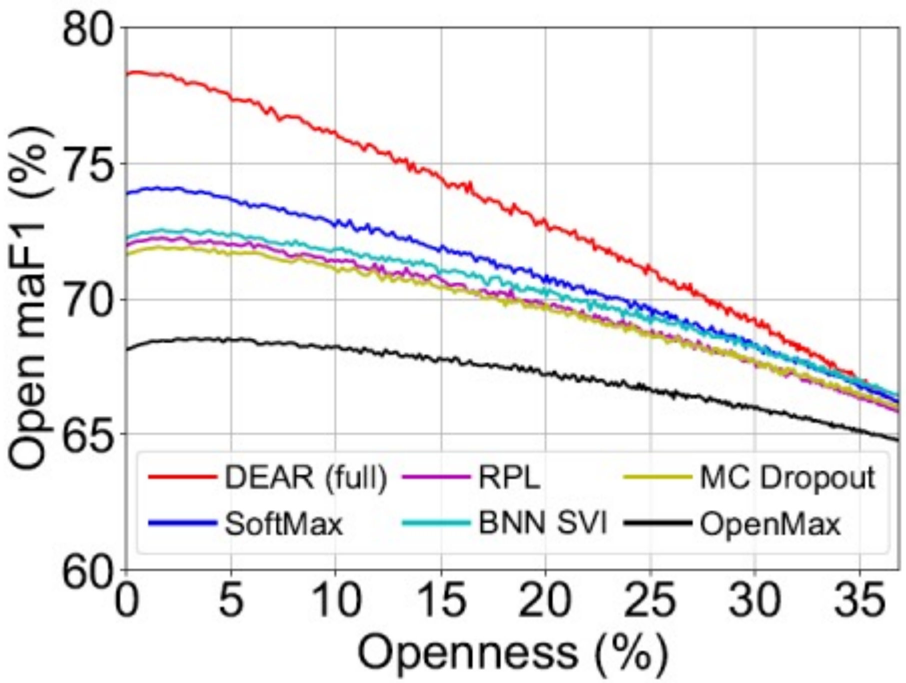
Detailed Results

Models	OSAR Methods	UCF-101 [54] + HMDB-51 [31]		UCF-101 [54] + MiT-v2 [39]		Closed Set Accuracy (%) **(For reference only)**
		Open maF1 (%)	Open Set AUC (%)	Open maF1 (%)	Open Set AUC (%)	
I3D [8]	OpenMax [5]	67.85 ± 0.12	74.34	66.22 ± 0.16	77.76	56.60
	MC Dropout	71.13 ± 0.15	75.07	68.11 ± 0.20	79.14	94.11
	BNN SVI [27]	71.57 ± 0.17	74.66	68.65 ± 0.21	79.50	93.89
	SoftMax	73.19 ± 0.17	75.68	68.84 ± 0.23	79.94	94.11
	RPL [10]	71.48 ± 0.15	75.20	68.11 ± 0.20	79.16	94.26
	DEAR (ours)	**77.24** ± 0.18	**77.08**	**69.98** ± 0.23	**81.54**	93.89
TSM [35]	OpenMax [5]	74.17 ± 0.17	77.07	**71.81** ± 0.20	83.05	65.48
	MC Dropout	71.52 ± 0.18	73.85	65.32 ± 0.25	78.35	95.06
	BNN SVI [27]	69.11 ± 0.16	73.42	64.28 ± 0.23	77.39	94.71
	SoftMax	78.27 ± 0.20	77.99	71.68 ± 0.27	82.38	95.03
	RPL [10]	69.34 ± 0.17	73.62	63.92 ± 0.25	77.28	95.59
	DEAR (ours)	**84.69** ± 0.20	**78.65**	70.15 ± 0.30	**83.92**	94.48
SlowFast [14]	OpenMax [5]	73.57 ± 0.10	78.76	72.48 ± 0.12	80.62	62.09
	MC Dropout	70.55 ± 0.14	75.41	67.53 ± 0.17	78.49	96.75
	BNN SVI [27]	69.19 ± 0.13	74.78	65.22 ± 0.21	77.39	96.43
	SoftMax	78.04 ± 0.16	79.16	74.42 ± 0.22	82.88	96.70
	RPL [10]	68.32 ± 0.13	74.23	66.33 ± 0.17	77.42	96.93
	DEAR (ours)	**85.48** ± 0.19	**82.94**	**77.28** ± 0.26	**86.99**	96.48
TPN [61]	OpenMax [5]	65.27 ± 0.09	74.12	64.80 ± 0.10	76.26	53.24
	MC Dropout	68.45 ± 0.12	74.13	65.77 ± 0.17	77.76	95.43
	BNN SVI [27]	63.81 ± 0.11	72.68	61.40 ± 0.15	75.32	94.61
	SoftMax	76.23 ± 0.14	77.97	70.82 ± 0.21	81.35	95.51
	RPL [10]	70.31 ± 0.13	75.32	66.21 ± 0.21	78.21	95.48
	DEAR (ours)	**81.79** ± 0.15	**79.23**	**71.18** ± 0.23	**81.80**	96.30

Experimental Results

- **Gradually Increasing the Openness**
  - I3D model is used as the backbone.
  - For each openness point, 10 random trials are performed to select the unknown.

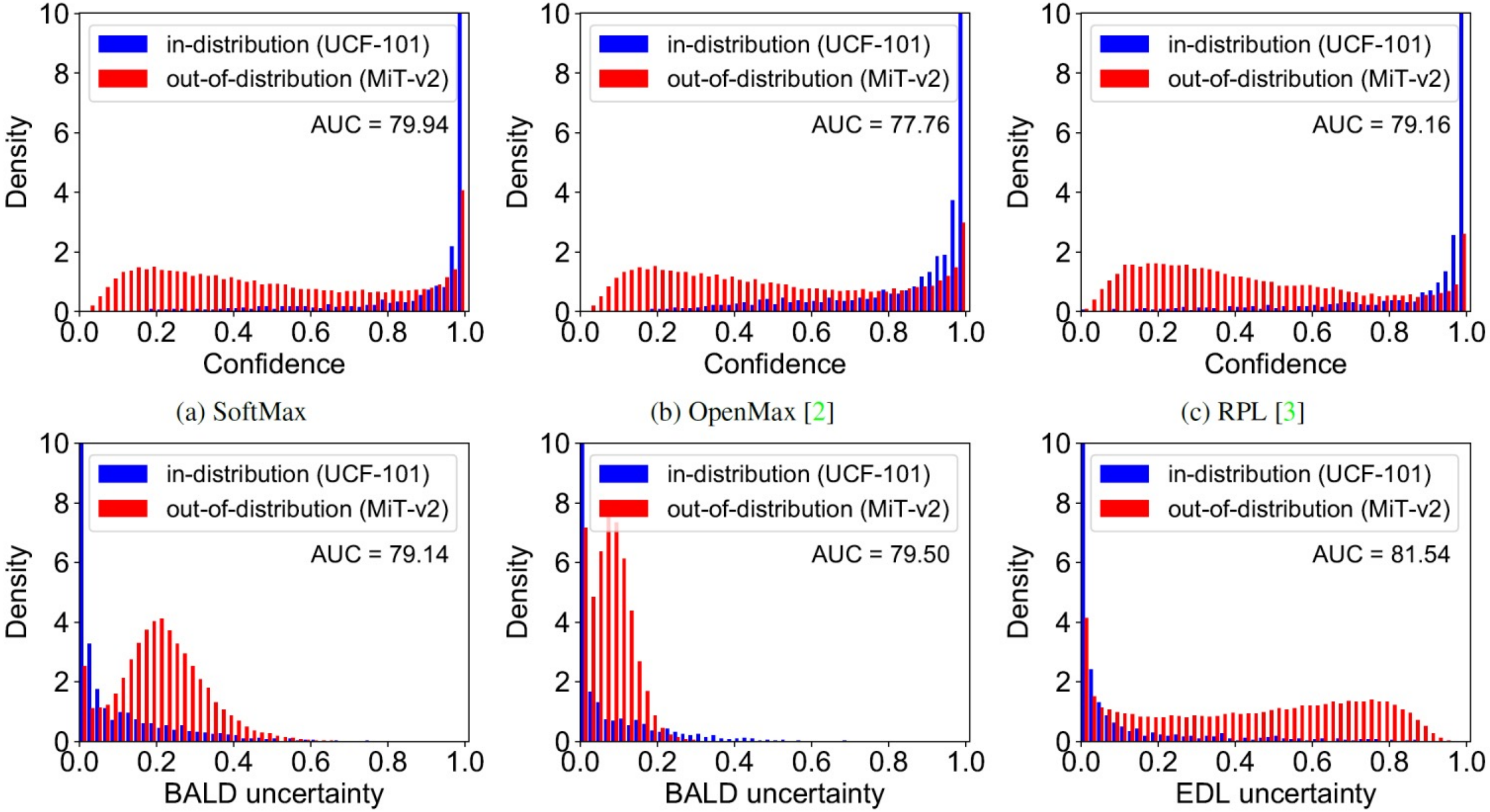




Experimental Results

Out-of-Distribution (OOD) Detection

- I3D model is used as the backbone, and MiT-v2 is used as the OOD data.



(a) SoftMax (b) OpenMax [2] (c) RPL [3]

(d) MC Dropout (e) BNN SVI [6] (f) DEAR (full)

Experimental Results

Ablation Study

- TPN model is used as the backbone, and HMDB-51 is used as the unknown.

$\mathcal{L}_{EUC}$	CED	Joint Train	Open maF1 (%)	OS-AUC (%)
✗	✗	✓	74.95 ± 0.18	77.12
✓	✗	✓	75.88 ± 0.16	77.49
✓	✓	✗	81.18 ± 0.15	79.02
✓	✓	✓	$\mathbf{81.79 \pm 0.15}$	**79.23**

Experimental Results

Are the performance gains of EUC benefited from better Uncertainty Calibration?

- **Validate the Uncertainty Calibration**
  - Expected Calibration Error (ECE) is adopted to evaluate calibration performance.
  - Smaller ECE indicates better calibration.

Model variants	Open Set (K+1)	Open Set (2)	Closed Set (K)
DEAR (w/o $\mathcal{L}_{EUC}$)	0.284	0.256	0.030
DEAR (full)	**0.268**	**0.239**	**0.029**

  - Calibration effect is more significant in OSR setting than Closed Set setting.

Experimental Results

Are the performance gains of CED rooted in better Representation Debiasing?

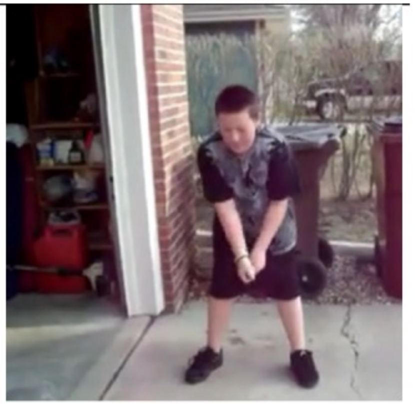
- **Validate the Representation Debiasing**
  - Models are only trained on biased dataset, i.e., Kinetics

Methods	Biased (Kinetics)		Unbiased (Mimetics)	
	top-1	top-5	top-1	top-5
DEAR (w/o CED)	91.18	99.30	26.56	69.53
DEAR (full)	**91.18**	**99.54**	**34.38**	**75.00**

  - Models trained on biased data (Kinetics) are vulnerable when testing with unbiased data (Mimetics).
  - Our CED module can significantly improve performance on unbiased data, while even slightly improve the performance on biased data.

Experimental Results

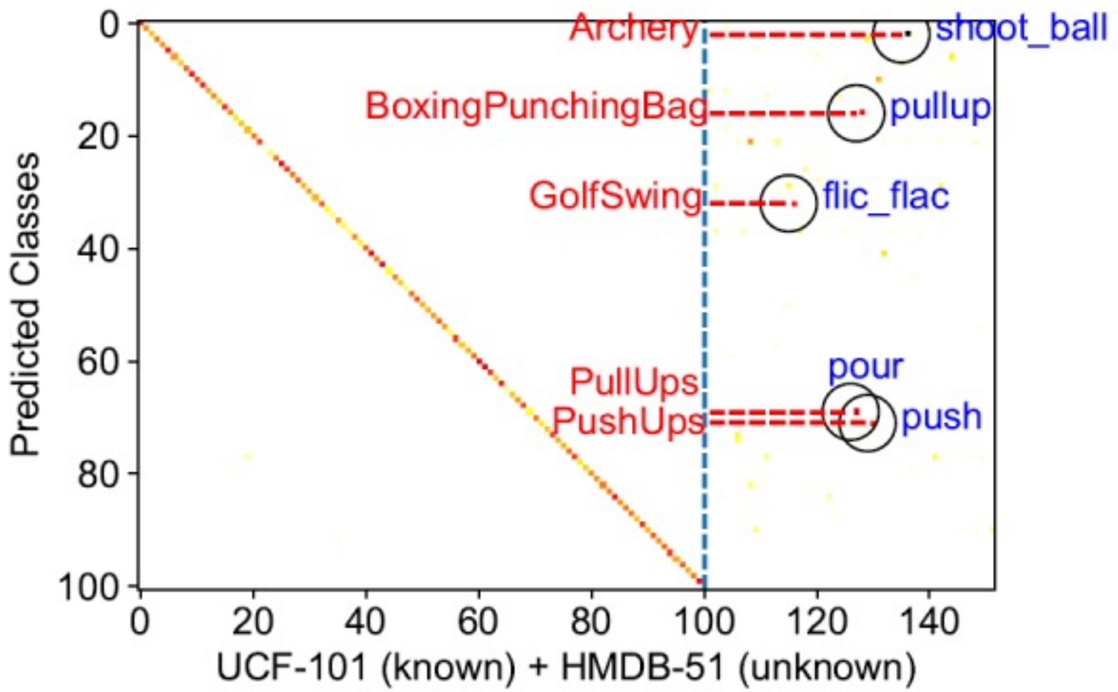
Some Visualizations of the Debiasing

Kinetics (Biased)			
DEAR (w/o CED) DEAR (full)	Playing Volleyball (✗) Playing Piano (✓)	Opening Bottle (✗) Writing (✓)	Shooting Soccer Goal (✗) Golf Driving (✓)
Mimetics (Unbiased)			
DEAR (w/o CED) DEAR (full)	Golf Driving (✗) Playing Piano (✓)	Golf Driving (✗) Writing (✓)	Opening Bottle (✗) Golf Driving (✓)

Experimental Results

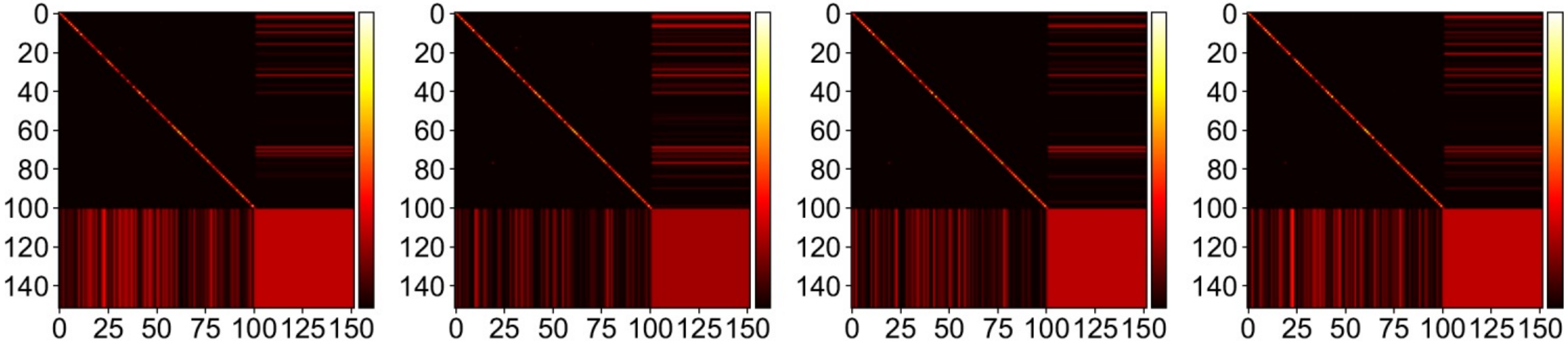
What types of unknown are more easily mis-classified?

- **Confusion Matrix**
  - The row represents the predicted action class.
  - The column indicates the ground-truth labels for both known and unknown actions.



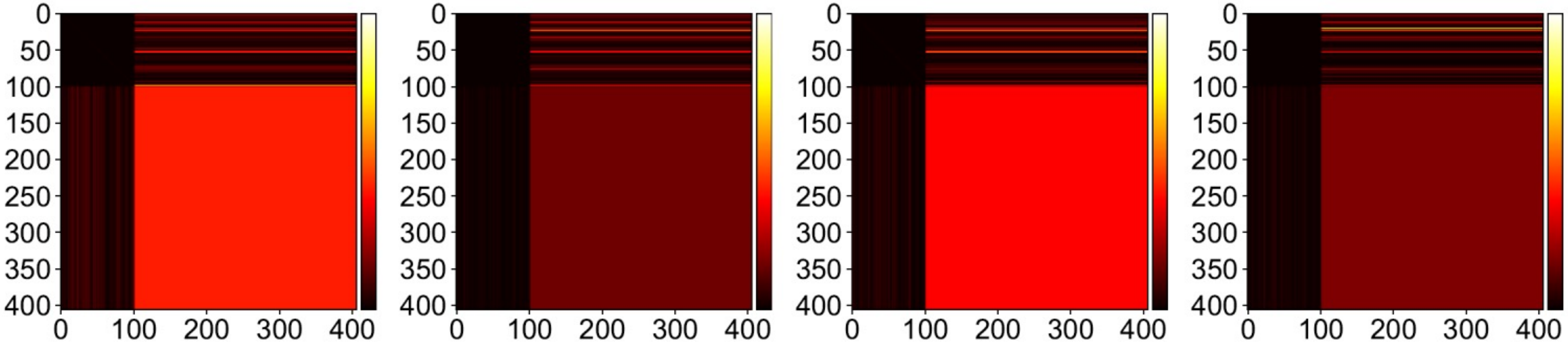
Experimental Results

- **More Complete Confusion Matrices**
  - HMDB-51 as Unknown



(a) I3D (b) TSM (c) SlowFast (d) TPN

  - MiT-v2 as Unknown



(a) I3D (b) TSM (c) SlowFast (d) TPN

Content

- Open Set Recognition
- Evidential Deep Learning
- The Proposed DEAR Model
- Experimental Results
- Conclusions and Discussions

RIT

Rochester Institute of Technology

Conclusions

- **The first to introduce Evidential Deep Learning to video understanding applications**
  - More efficient in training and inference than BNNs.
  - Distributional (2^{nd}-order) uncertainty is deterministically learned.

- **Uncertainty Calibration and Video Bias are explored in the context of EDL.**
  - Fundamental aspects to improve the generalization capability of video models.
  - Easy-to-use, plug-and-play.

- **Open Set Action Recognition task is comprehensively benchmarked**
  - Multiple mainstream action recognition models are benchmarked.
  - Thanks to PyTorch and MMAction2.

Discussions

- **Limitations of DEAR**
  - Similar to DNNs, ENNs also suffer from over-fitting problem.
  - Distinguish between the known and unknown is sensitive to thresholding.

- **Unexplored Research Questions**
  - What if our training data is limited in OSAR task? (Few-shot/Zero-shot Learning)
  - Can we use the learned evidence to discover new classes? (Generalized OSAR)
  - How do the types of unknown affect an OSAR model?
    - Long-tail, class hierarchy, noisy labeling, data in-the-wild, etc.
  - Generalize EDL to other vision tasks?
    - Video instance segmentation, action/event detection, 3D object detection, etc.

Our Labs

- **ActionLab**
  - Lead by Dr. Yu Kong (https://www.rit.edu/actionlab)
  - Our recent ICCV21 works:
    - Wentao Bao, Qi Yu, and Yu Kong: Evidential Deep Learning for Open Set Action Recognition. in **ICCV (Oral)**, 2021.
    - Wentao Bao, Qi Yu, and Yu Kong: DRIVE: Deep Reinforced Accident Anticipation with Visual Explanation. In **ICCV**, 2021.
    - Junwen Chen and Yu Kong: Explainable Video Entailment with Grounded Visual Evidence. In **ICCV**, 2021.

- **MiningLab**
  - Lead by Dr. Qi Yu (https://www.rit.edu/mining)
  - Related works:
    - Weishi Shi, Xujiang Zhao, Feng Chen, Qi Yu: Multifaceted Uncertainty Estimation for Label-Efficient Deep Learning. in **NeurIPS,** 2020.

THANKS

Q & A

Feel free to contact me via wb6219@rit.edu.