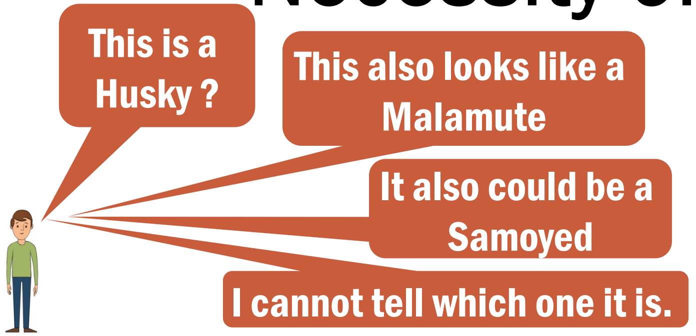
Hyper Evidential Deep Learning to Quantify Composite Classification Uncertainty

Changbin Li, Kangshuo Li, Yuzhe Ou, Lance M. Kaplan, Audun Jøsang, Jin-Hee Cho, Dong Hyun Jeong, Feng Chen

ICLR 2024

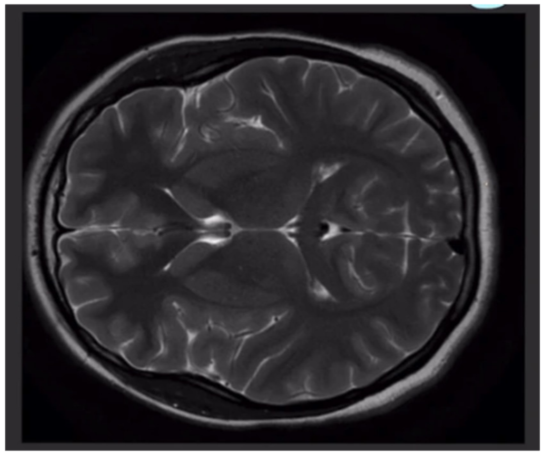
Necessity of composite label learning and prediction



Q: What the dog it is?

A dog image x_1 with
Y_1={Husky, <u>Malamute</u>, Samoyed}

X-ray of someone's brain



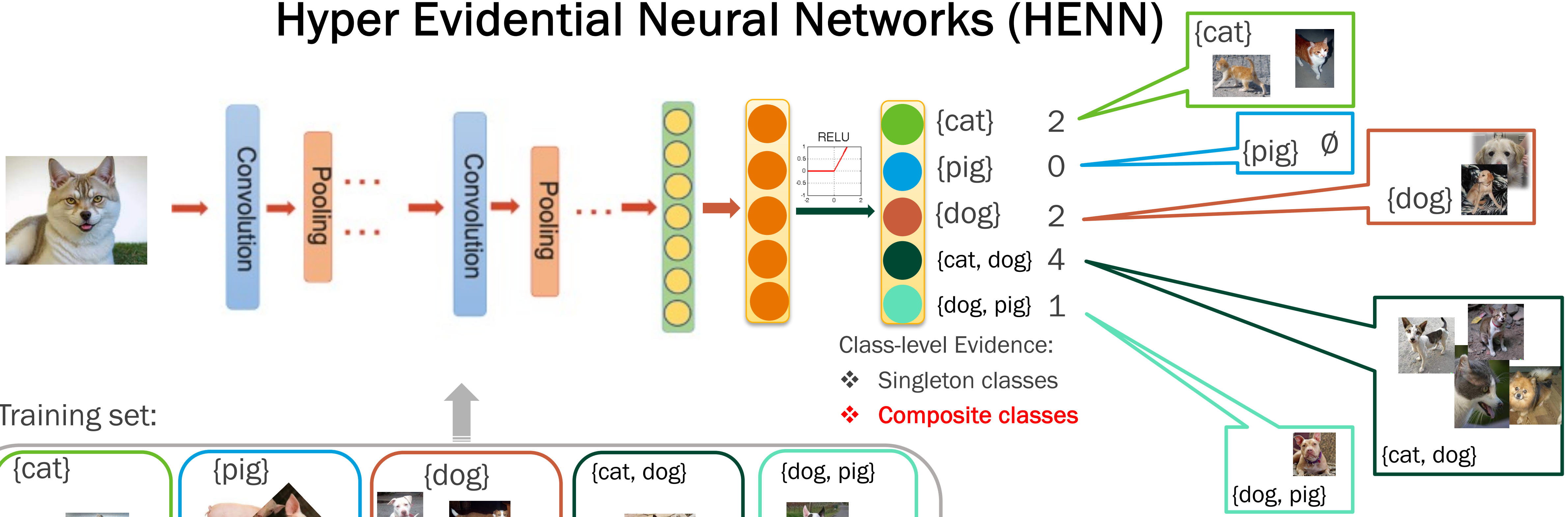
- A doctor making a high-stakes medical decision
- True diagnose ∈ {Normal, Concussion, Cancer}
- *Prob*[true diagnose ∈ {Normal, Concussion}] > *Prob*[true diagnose ∈ {Cancer}]

Challenge:
How to quantify predictive uncertainty in the presence of composite set labels in the training set?

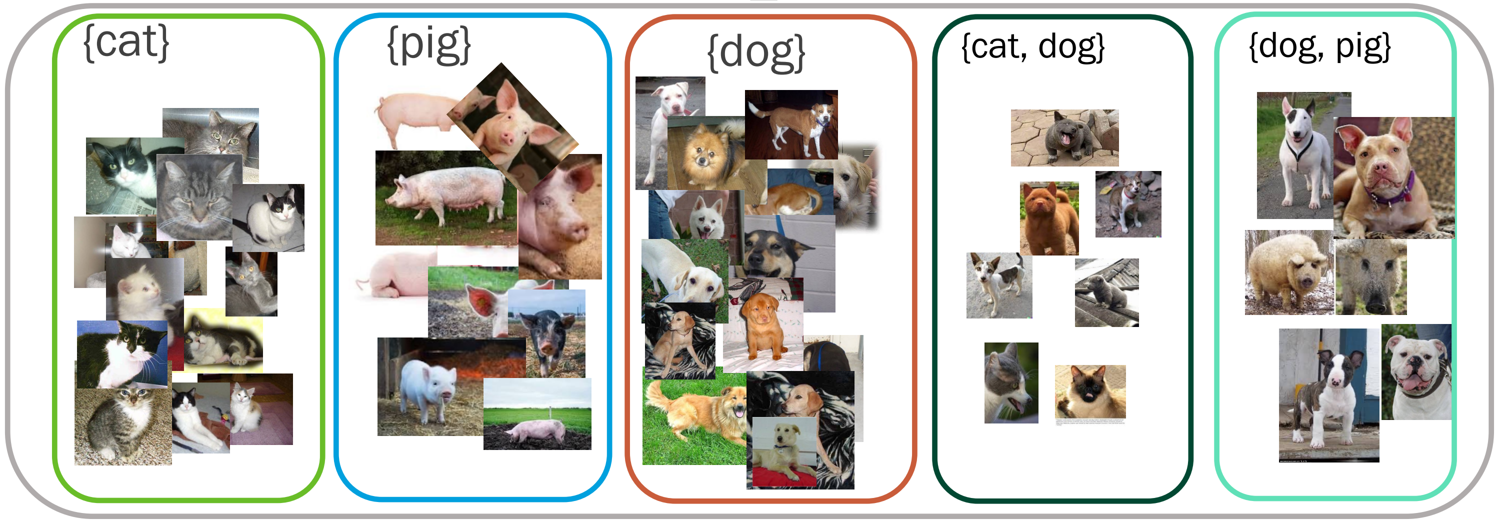
Source: https://iclr.cc/media/iclr-2022/Slides/6038.pdf

Our Approach

Hyper Evidential Neural Networks (HENN)

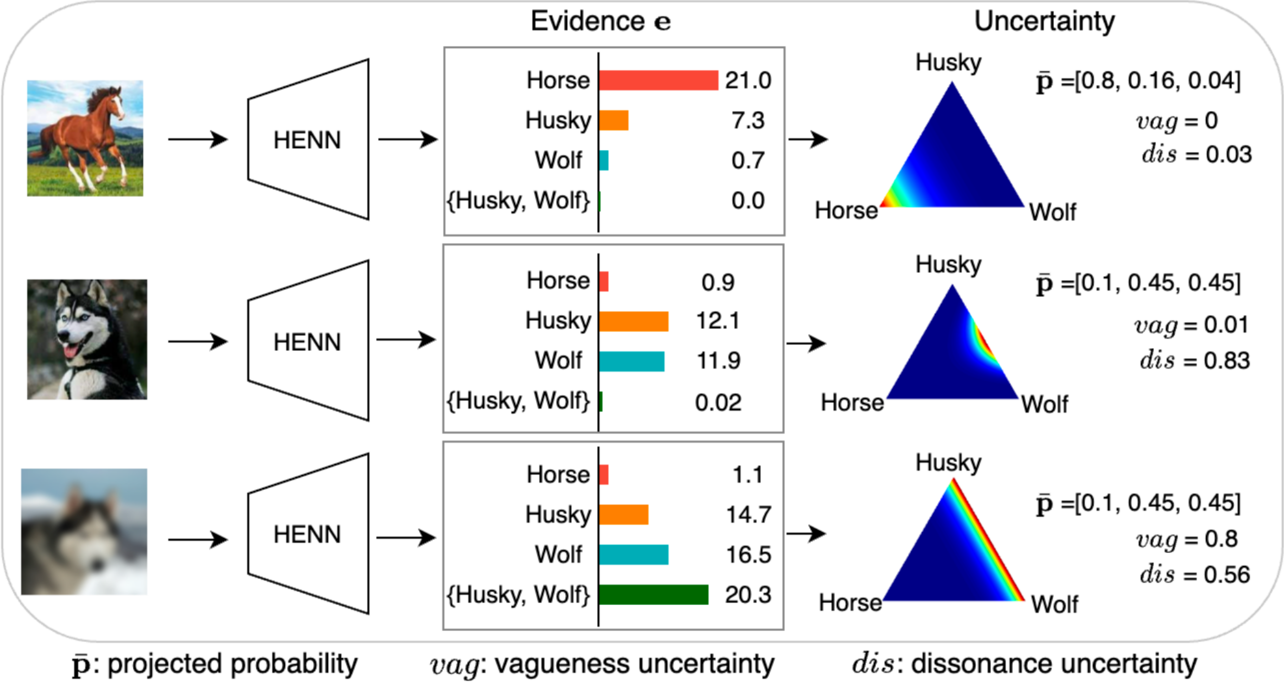


Training set:



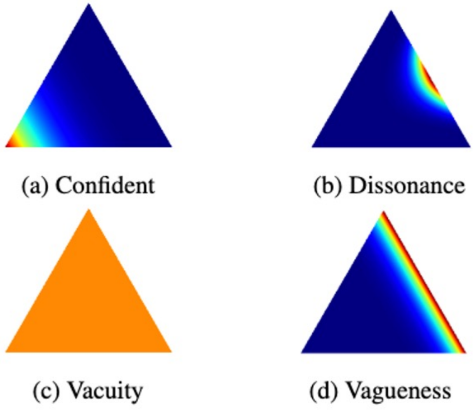
We propose a novel HENN designed to **predict hyper-opinions** and quantify predictive classification uncertainty caused by composite class labels.

Inference process of HENN



Different predictive uncertainties from HENN.

This framework can identify
either a singleton class or a composite set
with the highest belief, and it can
predict the singleton class with the greatest projected class probability.



Examples of different uncertainties.

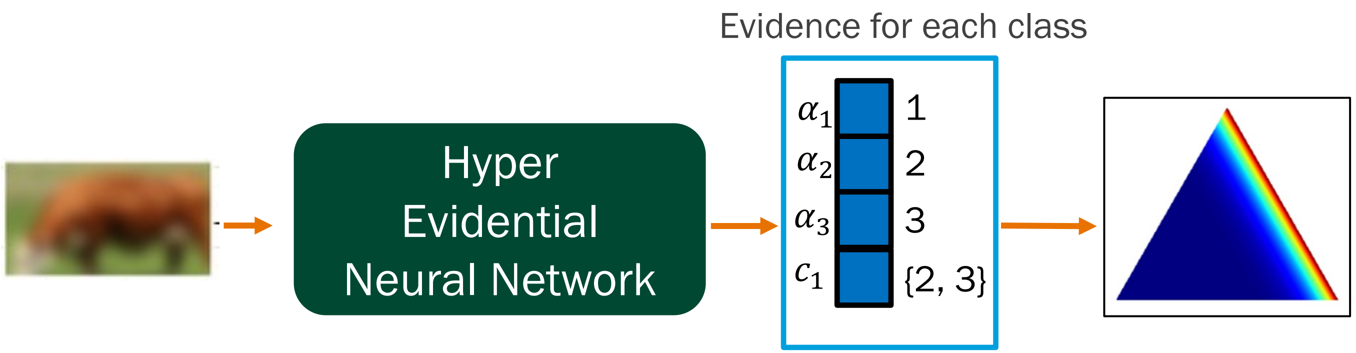
Introduce Vagueness

$$TotalVagueness = \sum_{y_j \in C(\mathbb{Y})} b_j$$

$$b_j = \frac{\alpha_j}{K + \sum_{l=1}^{K} \alpha_l}$$

K : the number of classes

Model Training



$x^{(i)}$

$\mathrm{GDD}(\mathbf{p}^{(i)}; \boldsymbol{\alpha}, \boldsymbol{c})$: Grouped Dirichlet Distribution

$$\mathrm{GDD}(\mathbf{p}|\boldsymbol{\alpha}, \mathbf{c}) = Z^{-1} \prod_{k=1}^{K} p_k^{\alpha_k - 1} \prod_{j=1}^{\eta} \Big(\sum_{l \in \mathcal{S}_j} p_l \Big)^{c_j}, \text{ for } \mathbf{p} \in \Delta_K$$

$\tilde{y}^1 = [1, 0, 0]$

$\tilde{y}^2 = [0, 1, 1]$

$\tilde{y}^3 = [1, 0, 0]$

$\tilde{y}^4 = [0, 1, 0]$

$\tilde{y}^5 = [0, 1, 1]$

…

$$\mathcal{L}(\theta) = \frac{1}{N} \sum_{i}^{N} \mathbb{E}_{\mathbf{p}^{(i)} \sim \mathrm{GDD}(\mathbf{p}^{(i)}; \boldsymbol{\alpha}, \boldsymbol{c})} \Big(-\log \sum_{l:\, \tilde{y}_l^i = 1} \mathbf{p}_l\Big) + \lambda \cdot Reg\big(\mathrm{GDD}(\mathbf{p}^{(i)}; \boldsymbol{\alpha}, \boldsymbol{c})\big)$$

Expected cost

KL regularizer

Jaccard Similarity (JS)

$$\text{JS}(y, \hat{y}) = \frac{|y \cap \hat{y}|}{|y \cup \hat{y}|}$$

Backbone EfficientNet-b3

Table 2: Results (%) based on Gaussian kernel size: 3×3 on CIFAR100 and tinyImageNet. (The average of three runs is provided, and the confidence interval is included in the App. due [Set prediction] ions.)

Set prediction

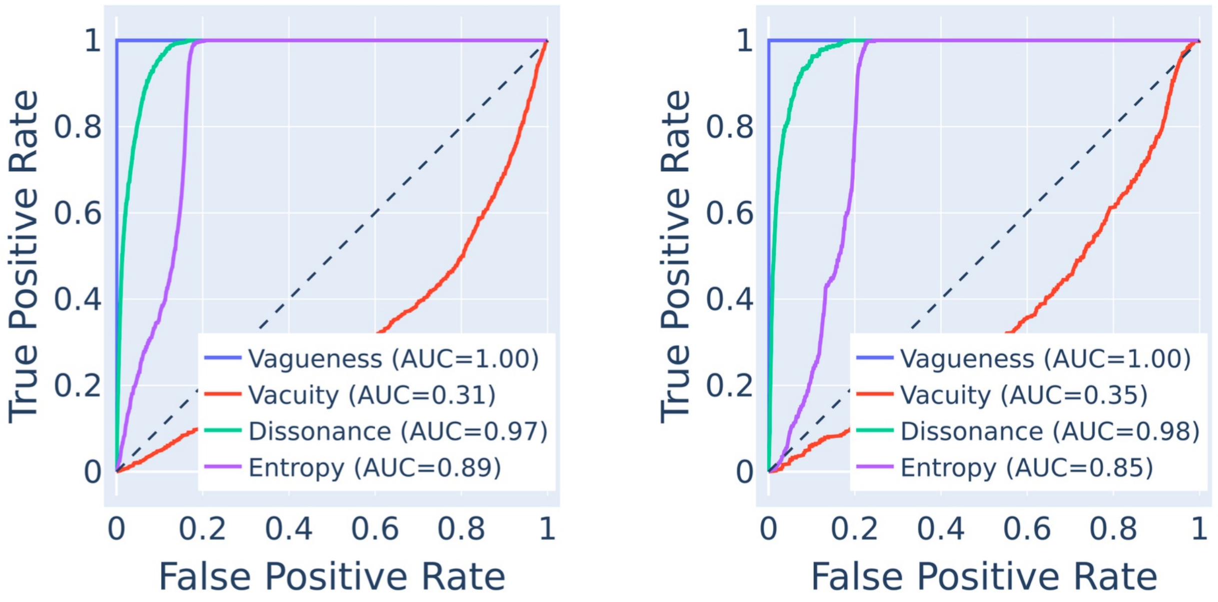
Singleton prediction

M	Methods	tinyImageNet OverJS	tinyImageNet CompJS	tinyImageNet Acc	living17 OverJS	living17 CompJS	living17 Acc	nonliving26 OverJS	nonliving26 CompJS	nonliving26 Acc
10	DNN (Tan & Le, 2019)	83.4	66.9	79.8	88.1	81.0	83.3	85.6	62.0	82.9
	ENN (Sensoy et al., 2018)	75.9	63.5	80.7	88.0	72.3	84.5	85.0	52.9	84.5
	E-CNN (Tong et al., 2021)	33.4	31.1	68.2	30.5	36.8	65.7	28.3	35.8	60.6
	RAPS (Angelopoulos et al., 2021)	73.1	43.6	79.8	86.4	61.3	83.3	82.7	46.3	82.9
	PiCO (Wang et al., 2022b)	57.2	35.6	64.3	62.5	43.7	65.2	61.8	42.6	64.8
	HENN (ours)	**84.4**	**93.4**	**82.5**	**88.8**	**96.5**	**85.6**	**86.9**	**96.8**	**85.4**
15	DNN (Tan & Le, 2019)	84.3	67.3	79.5	88.1	84.8	80.2	85.6	68.9	81.5
	ENN (Sensoy et al., 2018)	83.5	60.7	81.2	88.0	78.3	82.4	85.4	62.6	82.9
	E-CNN (Tong et al., 2021)	32.5	33.3	68.4	31.6	37.3	65.5	29.8	35.1	60.1
	RAPS (Angelopoulos et al., 2021)	68.1	45.6	79.5	85.5	66.5	80.2	83.8	56.1	81.5
	PiCO (Wang et al., 2022b)	56.8	35.3	64.6	61.4	43.1	64.8	61.5	42.5	64.6
	HENN (ours)	**84.6**	**90.6**	**81.6**	**88.8**	**96.6**	**85.7**	**86.9**	**96.2**	**84.1**

Key observations: (similar results observed in other backbones, such as ResNet and VGG-16)

- HENN outperforms other baselines in terms of accuracy for singleton prediction.
- HENN outperforms other baselines in terms of OverJS and CompJS for set prediction.

ROC curves on CIFAR100 and tinyImageNet



(a) CIFAR100 M=15 (b) tinyImageNet M=15

- Vagueness (HENN)
- Vacuity (ENN)
- Dissonance (ENN)
- Entropy (DNN)

Key observations: (similar results observed in other datasets, such as living17 and nonliving26)

- Vagueness from HENN is the best indicator to distinguish singleton examples and composite examples compared to other common uncertainties.

Thank you

Poster Session 7
Hall B
Fri 10 May, 10:45 a.m. - 12:45 p.m. CEST

SCAN ME