Hierarchical Novelty Detection via Fine-Grained Evidence Allocation

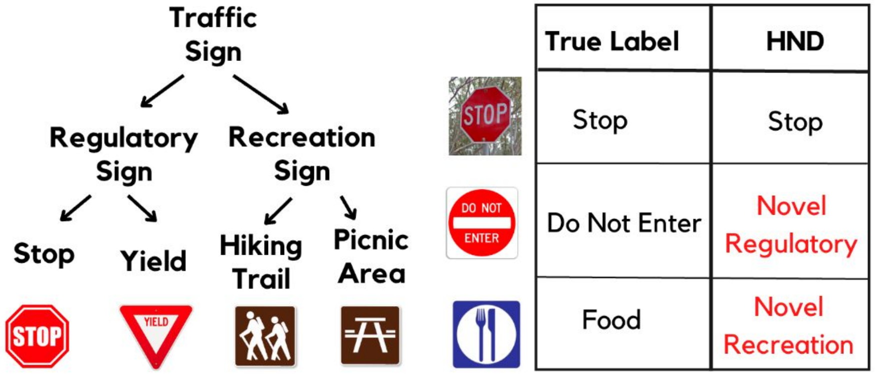
*Spandan Pyakurel and Qi Yu**

Rochester Institute of Technology

Hierarchical Novelty Detection

❑ Existing Novelty Detection methods only provide a binary detection result, indicating whether the sample is novel or not.

❑ With the help of a hierarchy of known classes, Hierarchical Novelty Detection (HND) can identify the class the novel sample is most similar to.

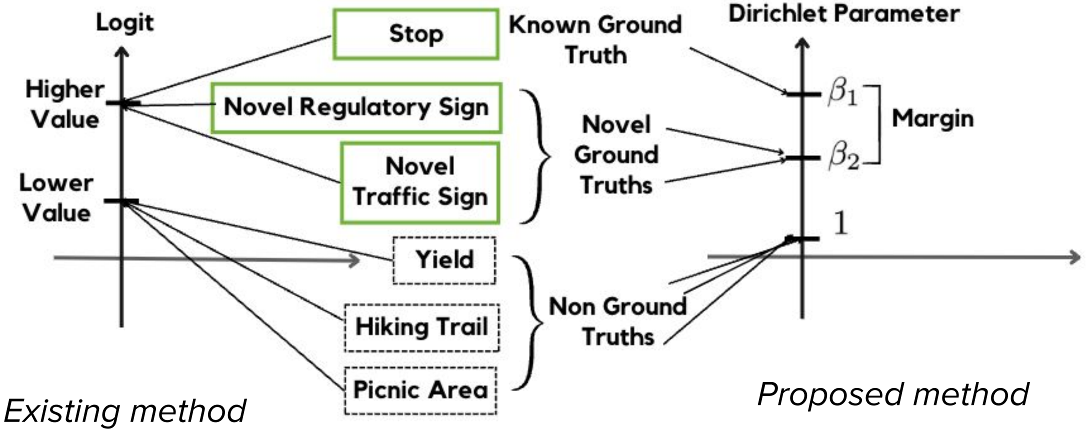


True Label	HND
Stop	Stop
Do Not Enter	Novel Regulatory
Food	Novel Recreation

An example of Hierarchical Novelty Detection

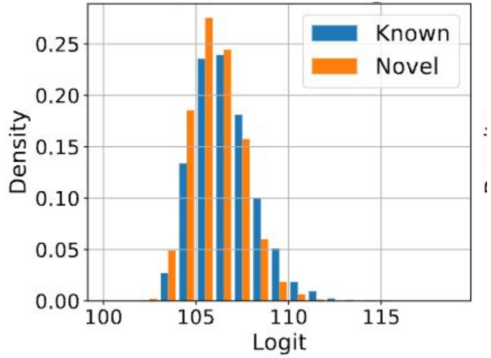
Issues with Existing Methods

- Existing method utilizes samples from known class as novel class in model training. Example: Model uses sample from **Stop sign** as **Novel Regulatory Sign** and **Novel Traffic Sign**.
- Existing method assigns higher logit values to both known and novel classes. As a result, they can not differentiate between novel and known class in testing.



Existing method

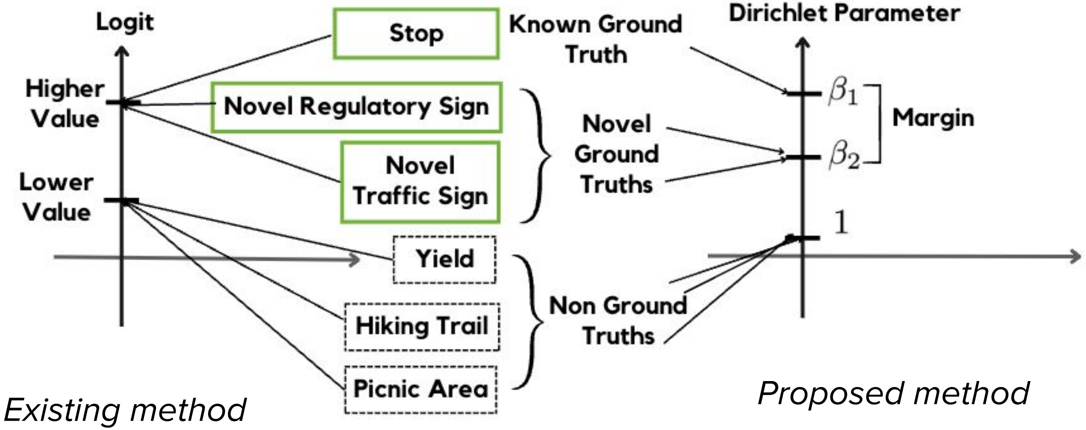
Proposed method



Existing method logit distribution

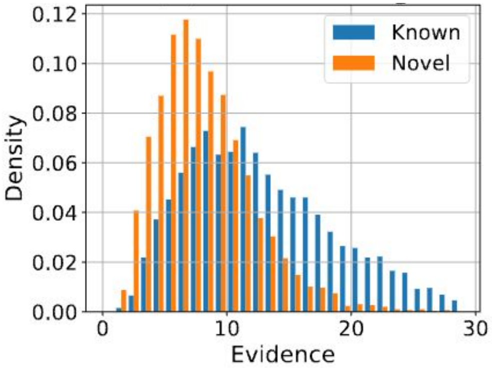
Proposed Method

- We propose a novel method, referred to as evidential hierarchical novelty detection (E-HND) that leverages fine-grained evidence to more precisely differentiate samples of known class from those of novel ones in the same hierarchy.
- We design a unique loss function that can create an evidence margin to ensure good separation of known and novel samples with sound theoretical guarantees.



Existing method

Proposed method



Proposed method evidence distribution

Learning Evidence Margin

The proposed loss function comprises two terms that work in a multitask fashion to allocate: (i) high evidence to the ground truth known leaf class and (ii) moderate evidence to the ground truth novel non-leaf classes.

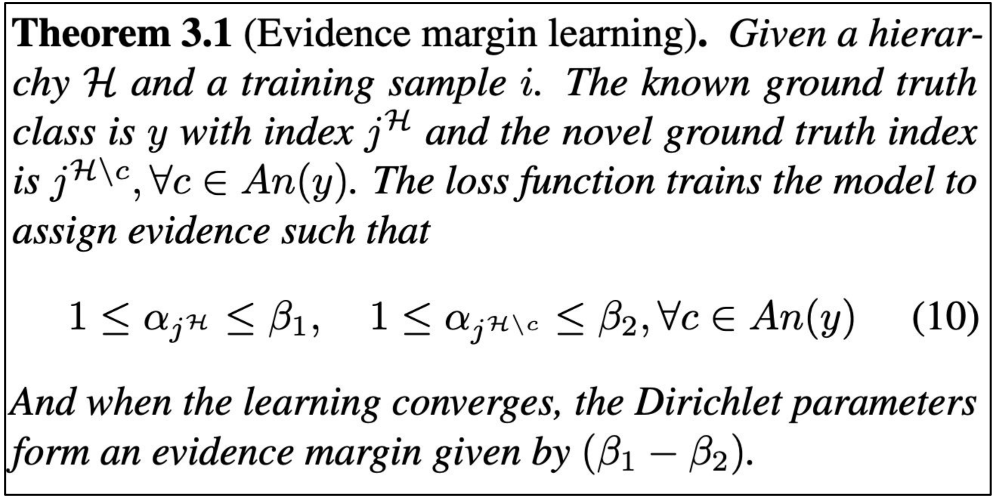
$$\mathcal{L}_i^{(1)}(\theta) = KL\left[D(\mathbf{p}_i|\boldsymbol{\alpha}_i;\theta_{Le(\mathcal{H})})||D(\mathbf{p}_i|\hat{\boldsymbol{\alpha}}_i;\theta_{Le(\mathcal{H})})\right]$$

$$\mathcal{L}_i^{(2)}(\theta) = \sum_{c\in An(y)} \mathcal{L}_{i,c}^{(2)}(\theta) \qquad \mathcal{L}_{i,c}^{(2)}(\theta) = KL\left[D(\mathbf{p}_i|\boldsymbol{\alpha}_i;\theta_{Le'(\mathcal{H}\backslash c)})||D(\mathbf{p}_i|\tilde{\boldsymbol{\alpha}}_i;\theta_{Le'(\mathcal{H}\backslash c)})\right]$$

$$\hat{\alpha}_{ik} = \begin{cases} \beta_1 \gg 1, & \text{if } k = j^{\mathcal{H}} \\ 1 & \text{otherwise} \end{cases}$$

$$\tilde{\alpha}_{ik} = \begin{cases} 1 < \beta_2 < \beta_1, & \text{if } k = j^{\mathcal{H}\backslash c} \\ 1 & \text{otherwise} \end{cases}$$

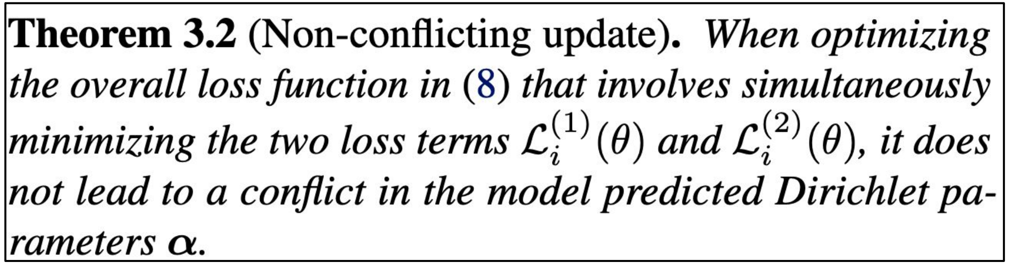
Theoretical Support



Theorem 3.1 (Evidence margin learning). *Given a hierarchy $\mathcal{H}$ and a training sample i. The known ground truth class is y with index $j^{\mathcal{H}}$ and the novel ground truth index is $j^{\mathcal{H}\setminus c}, \forall c \in An(y)$. The loss function trains the model to assign evidence such that*

$$1 \leq \alpha_{j^{\mathcal{H}}} \leq \beta_1, \quad 1 \leq \alpha_{j^{\mathcal{H}\setminus c}} \leq \beta_2, \forall c \in An(y) \qquad (10)$$

And when the learning converges, the Dirichlet parameters form an evidence margin given by $(\beta_1 - \beta_2)$.

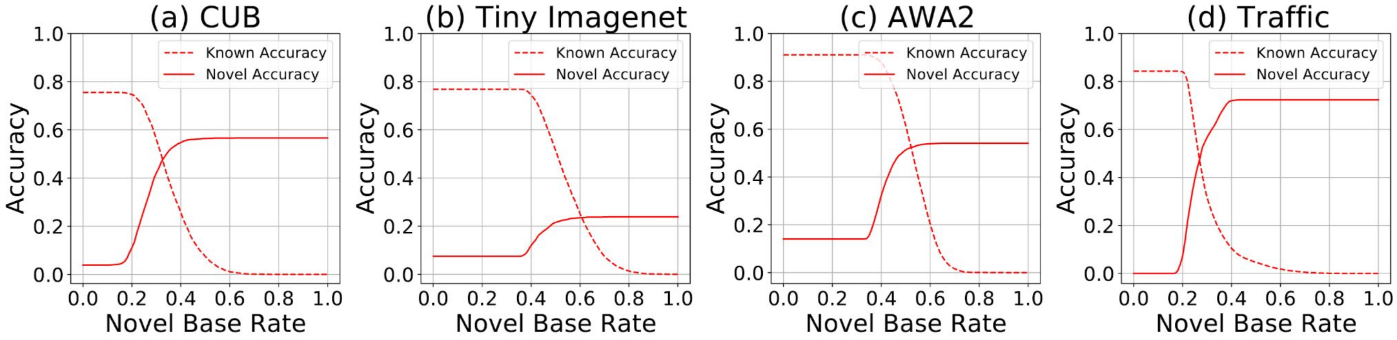


Theorem 3.2 (Non-conflicting update). *When optimizing the overall loss function in (8) that involves simultaneously minimizing the two loss terms $\mathcal{L}_i^{(1)}(\theta)$ and $\mathcal{L}_i^{(2)}(\theta)$, it does not lead to a conflict in the model predicted Dirichlet parameters $\boldsymbol{\alpha}$.*

Incorporating the Prior Belief

- The evidential theory allows us to encode a prior belief in the form of base rate distributions. Base rate for each class denotes the prior probability of a data sample belonging to that class when no evidence is observed.
- Higher base rate for the known classes denote the belief of completeness of the hierarchy, and a test sample will more likely be assigned to one of the known leaf classes.

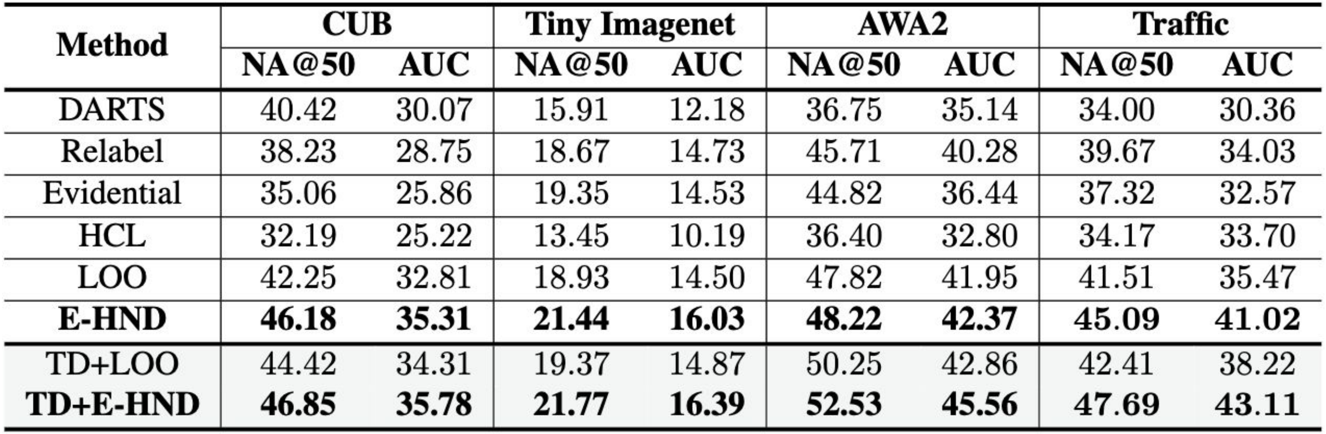
$$\underset{\underset{\text{Known Base Rate}}{\downarrow}}{\sum_{k=1}^{|Le(\mathcal{H})|} a_k^{(kn)}} + \underset{\underset{\text{Novel Base Rate}}{\downarrow}}{\sum_{k=1}^{|NLe(\mathcal{H})|} a_k^{(no)}} = 1$$



Impact of base rates

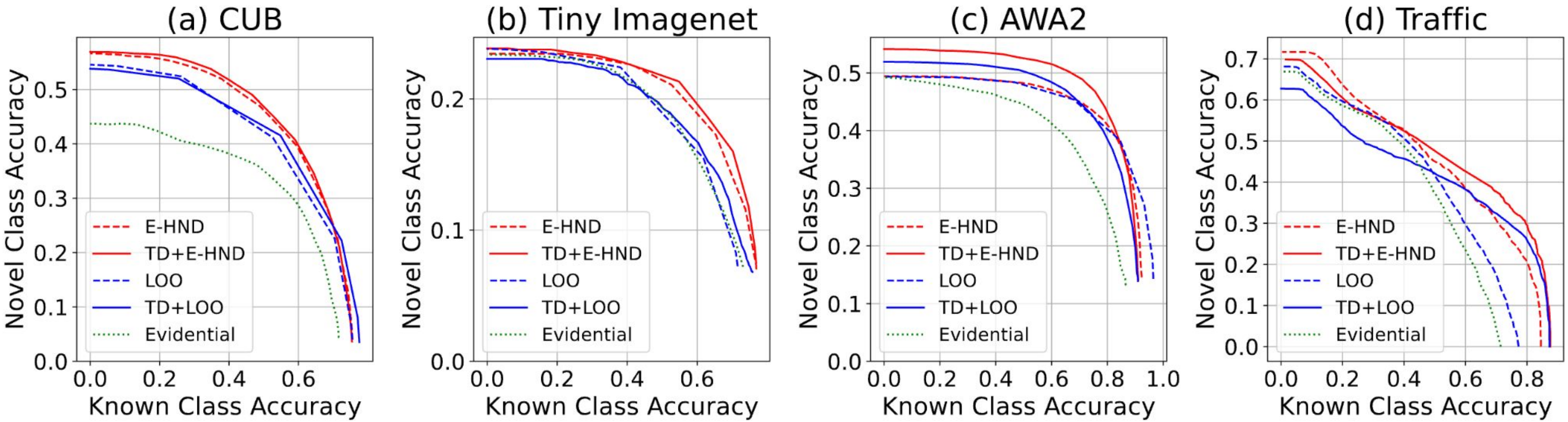
Experimental Results [1/3]

- Experiments of 4 hierarchical datasets: CUB, Tiny Imagenet, AWA2, Traffic
- To capture trade-off between known (K-ACC) and novel accuracies (N-ACC), add a bias term to the logit of novel classes, and obtain sets of K-ACC and N-ACC.
- Area Under the Curve (AUC): obtained by plotting K-ACC and N-ACC.
- NA@50: N-ACC, where the model has exactly 50% K-ACC.



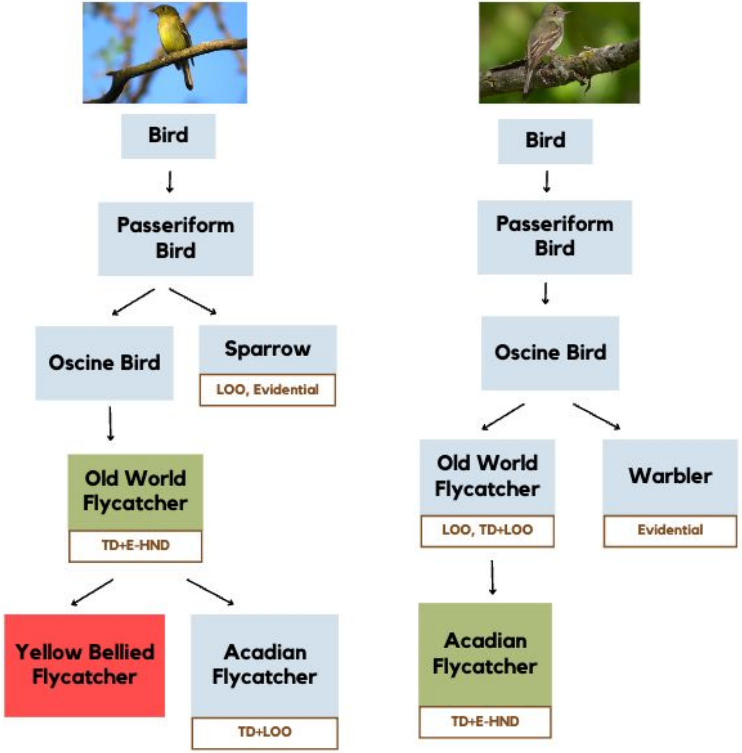
Method	**CUB**		**Tiny Imagenet**		**AWA2**		**Traffic**	
	NA@50	**AUC**	**NA@50**	**AUC**	**NA@50**	**AUC**	**NA@50**	**AUC**
DARTS	40.42	30.07	15.91	12.18	36.75	35.14	34.00	30.36
Relabel	38.23	28.75	18.67	14.73	45.71	40.28	39.67	34.03
Evidential	35.06	25.86	19.35	14.53	44.82	36.44	37.32	32.57
HCL	32.19	25.22	13.45	10.19	36.40	32.80	34.17	33.70
LOO	42.25	32.81	18.93	14.50	47.82	41.95	41.51	35.47
E-HND	**46.18**	**35.31**	**21.44**	**16.03**	**48.22**	**42.37**	**45.09**	**41.02**
TD+LOO	44.42	34.31	19.37	14.87	50.25	42.86	42.41	38.22
TD+E-HND	**46.85**	**35.78**	**21.77**	**16.39**	**52.53**	**45.56**	**47.69**	**43.11**

Experimental Results [2/3]



AUC curve

Experimental Results [3/3]



Bird
Passeriform Bird
Oscine Bird
Sparrow
LOO, Evidential
Old World Flycatcher
TD+E-HND
Yellow Bellied Flycatcher
Acadian Flycatcher
TD+LOO

Bird
Passeriform Bird
Oscine Bird
Old World Flycatcher
LOO, TD+LOO
Warbler
Evidential
Acadian Flycatcher
TD+E-HND

Qualitative study