Uncertainty Aware Semi-Supervised Learning on Graph Data

Xujiang Zhao[1], Feng Chen[1], Shu Hu[2], Jin-Hee Cho[3]
[1]The University of Texas at Dallas, [2]University at Buffalo, SUNY, [3]Virginia Tech

Introduction

Consider a classifier only knows two classes: `car' and `road'

Incorrect prediction → High Uncertainty

Misclassify "car" as "road"

Out-of-distribution (OOD) → High Uncertainty

Misclassify "deer" (OOD object) as "car"

Quantifying predictive uncertainty is important for safely-critical applications

Multiple Uncertainties

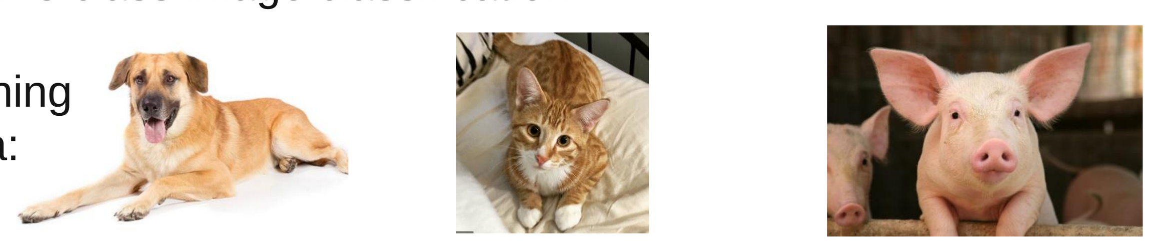
Probabilistic Uncertainty:
Epistemic (limited data)
Aleatoric (randomness)

Evidential Uncertainty:
Vacuity (lack of evidence)
Dissonance (conflicting evidence)

Motivation

Task: 3 class image classification

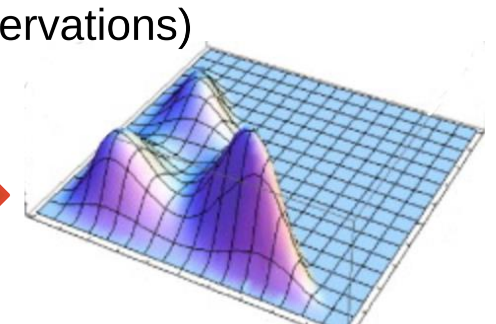
Training Data:



Dog (e_1 =10 images) Cat (e_2 = 10 images) Pig (e_3 = 10 images)

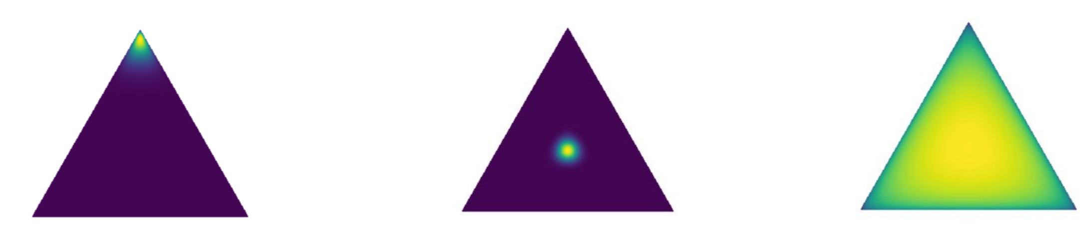
$e = [e_1, \dots, e_K]$ Evidence (Historical observations)

Subjective Opinion $\omega = (\boldsymbol{b}, u, \boldsymbol{a})$ ⟷ $\alpha = e + 1$ ⟷ **Dirichlet (α)**



Vacuity and Dissonance

Confidence Prediction | Conflict Prediction | Out-of-Distribution



Confidence Prediction	Conflict Prediction	Out-of-Distribution
α = [11, 1, 1]	α = [11, 11, 11]	α = [1, 1, 1]
p = [0.83, 0.08, 0.08]	p = [1/3, 1/3, 1/3]	p = [1/3, 1/3, 1/3]
Low Uncertainty	High Dissonance (conflicting evidence)	High Vacuity (lack of evidence)

Different Evidential Uncertainty

Same probability

Test images

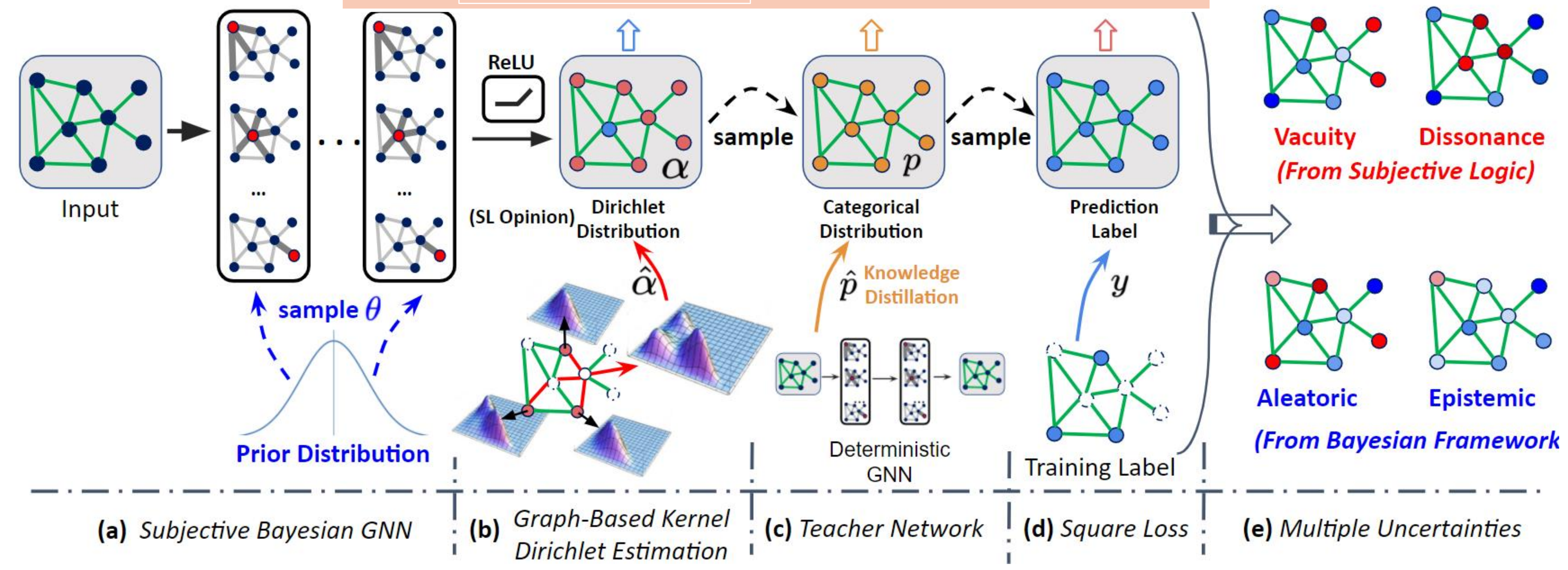
Multidimensional Uncertainty Framework

Given: graph $\mathcal{G} = (\mathbb{V}, \mathbb{E}, \mathbf{r}, \mathbf{y}_{\mathbb{L}})$

$\mathbb{V}$: Node $\mathbb{E}$: Edge $\mathbf{r}$: node-level feature $\mathbf{y}_{\mathbb{L}}$: training labels

Goal: Predict
- Class probabilities $\boldsymbol{p}$
- Multidimensional uncertainty $\boldsymbol{u}$

Loss = $KL[Dir(\alpha)||Dir(\hat{\alpha})]$ + $KL[p||\hat{p}]$ + $||y - p||_2^2$

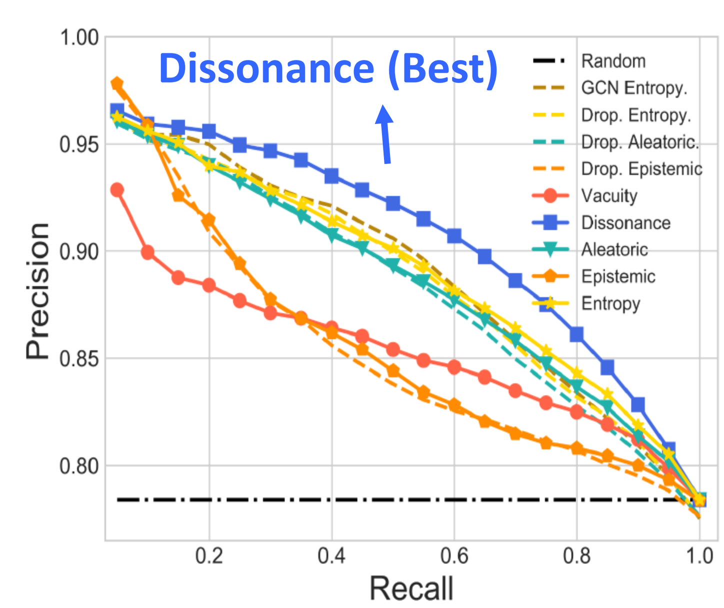


Key properties of this model:

(a) S-BGNN: providing the multiple uncertainties
(b) GKDE: predicting Dirichlet more accurately
(c) Teacher Network: refining the class probability
(d) Square loss: minimizing prediction error and variance

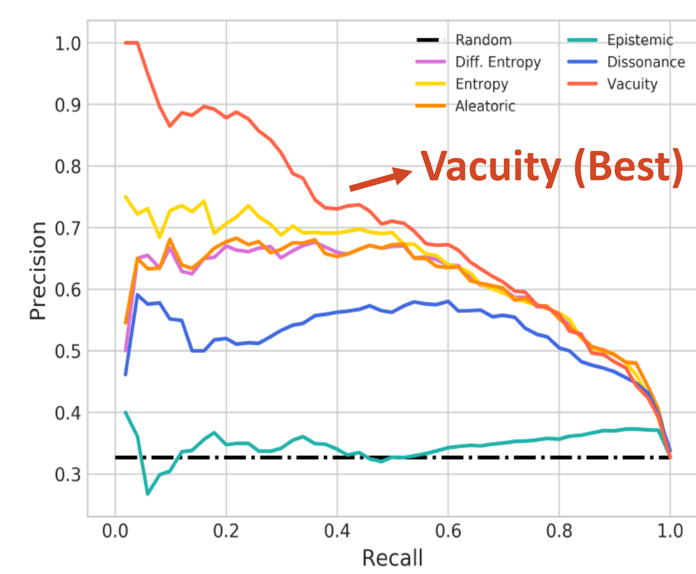
Experimental Results

Misclassification Detection



PR curves on Pubmed

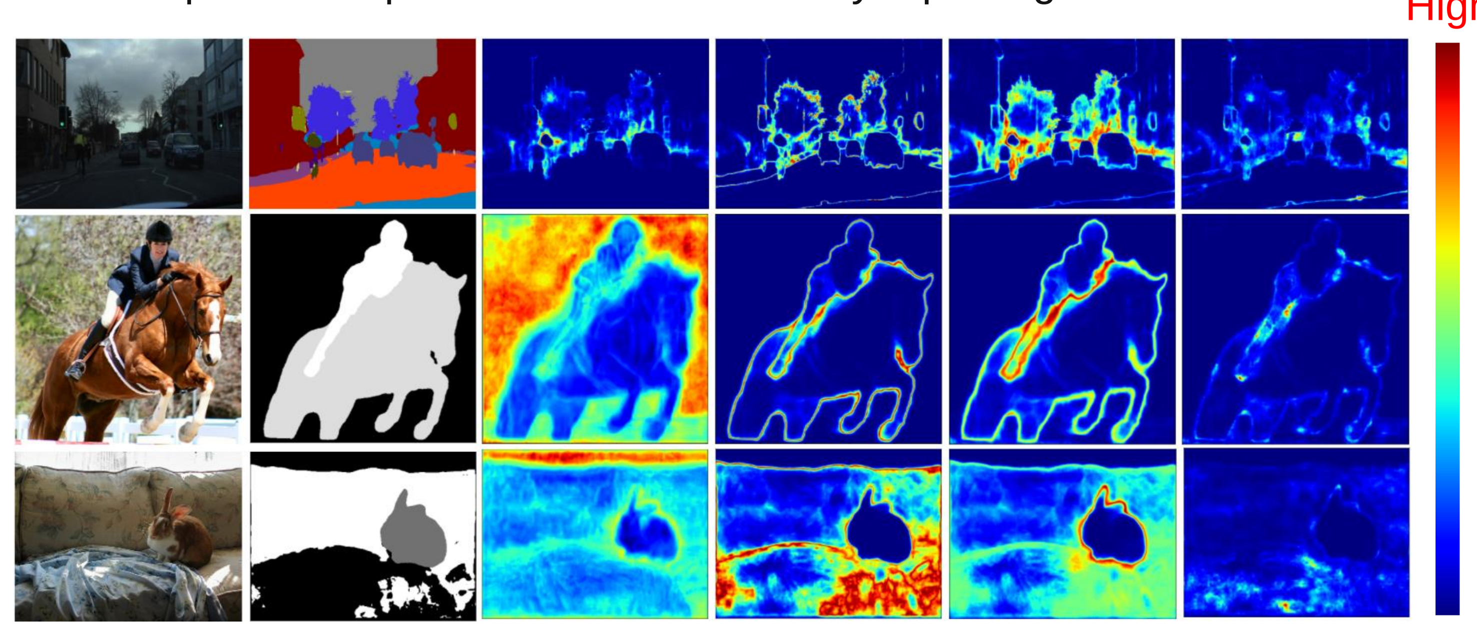
OOD Detection



PR curves on Amazon Computers

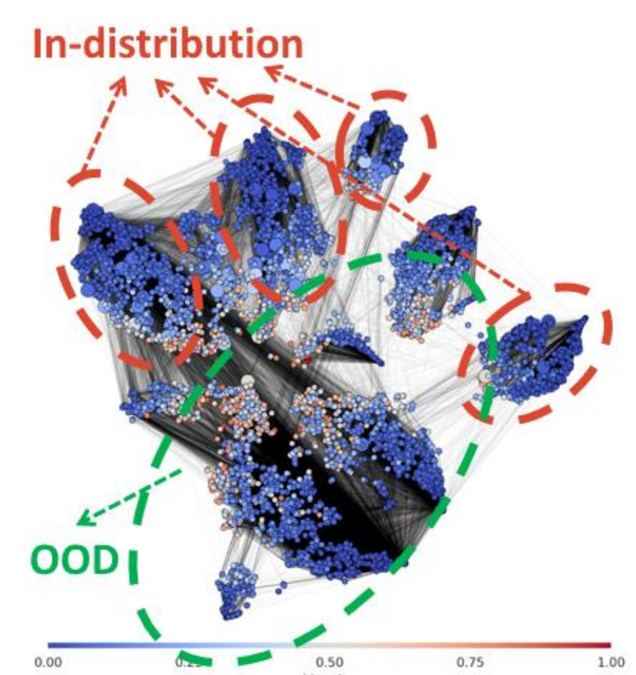
Extension to Other Deep Learning Model (CNN)

Provide pixel-level predictive uncertainties by replacing GNN with CNN



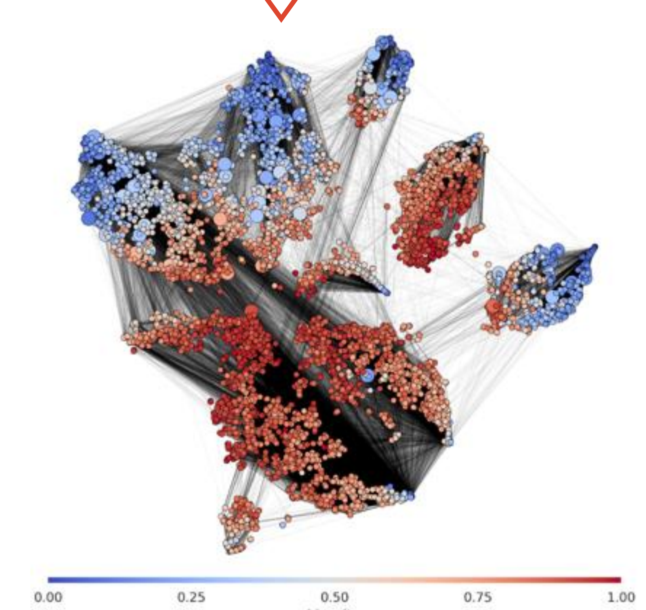
Input | Prediction | Vacuity | Dissonance | Aleatoric | Epistemic

Key Merits of GKDE



OOD Detection AUROC: 64%

⇩ add GKDE



OOD Detection **AUROC: 93%**

Key Theoretical Results

Relationships Between Multiple Types of Uncertainties

Consider a simple scenario:
y: a multinomial random variable
y ~ Cal(p): y follows a K-class categorical distribution
p ~ Dir(α): class probabilities p follows a Dirichlet distribution

General relations on all prediction scenarios

vacuity + dissonance ≤ 1 *vacuity > epistemic*

Special relations on the out-of-distribution (OOD)

1 = vacuity = entropy > aleatoric > epistemic > dissonance = 0

Special relations on the Conflicting Prediction (CP)

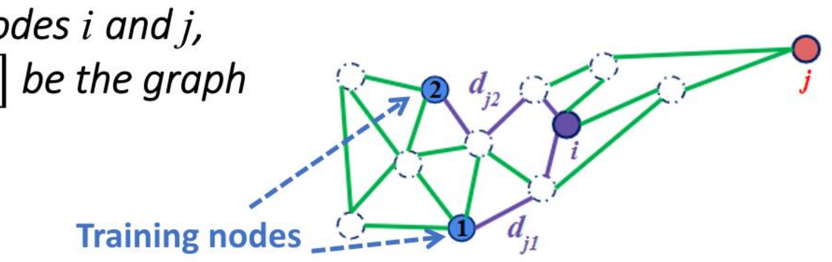
entropy = 1, dissonance → 1, aleatoric → 1, vacuity → 0, epistemic → 0

entropy > aleatoric > dissonance > vacuity > epistemic

- Higher vacuity leads to a lower dissonance, and vice versa.
- Vacuity indicates an upper bound of epistemic uncertainty.
- Entropy **cannot distinguish** different types of uncertainties caused by different root causes.
- **High aleatoric uncertainty** and **low epistemic uncertainty** are observed under both cases.
- Vacuity and dissonance can **clearly distinguish** OOD from a CP.

Impact of Graph-based Kernel Dirichlet Distribution Estimation

Given L training nodes and two testing nodes i and j, Let $\boldsymbol{d}_i = [d_{i1}, \dots, d_{iL}]$, $\boldsymbol{d}_i = [d_{j1}, \dots, d_{jL}]$ be the graph distances from training nodes.

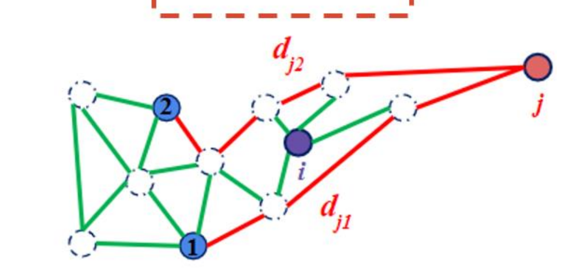


If for all $l \in \{1, \dots, L\}$, $d_{il} \le d_{jl}$, we have

$$\widehat{vacuity}_i \le \widehat{vacuity}_j \Leftarrow \begin{matrix} d_{i1} < d_{j1} \\ d_{i2} < d_{j2} \end{matrix}$$

estimated based on GKDE.

High vacuity occurs when testing node is far away from training nodes.



Summary

- Proposed a multi-source uncertainty framework of GNNs.
- Provided a theoretical analysis about the relationships between different types of uncertainties.
- Demonstrate the use of vacuity for OOD detection and dissonance for misclassification detection.

[paper] [code] [slides]