Weakly-Supervised Residual Evidential Learning for Multi-Instance Uncertainty Estimation

Paper & Code

Pei Liu   Luping Ji ✉

School of Computer Science and Engineering, UESTC

ICML International Conference On Machine Learning

计算机科学与工程学院（网络空间安全学院）
School of Computer Science and Engineering (School of Cyber Security)

Motivation

- **Uncertainty Estimation** (UE), as an effective means to quantify predictive uncertainty, is crucial for safe and reliable decision-making.
- Existing UE methods often assume there are **completely-labeled** data with which neural networks can be trained to estimate $p(\boldsymbol{\omega}|D)$ and capture epistemic (model) uncertainty for accurate UE:

$$P(Y^*|X^*,\mathcal{D}) = \int \underbrace{P(Y^*|X^*,\boldsymbol{\omega})}_{\text{aleatoric}} \underbrace{p(\boldsymbol{\omega}|\mathcal{D})}_{\text{epistemic}} d\boldsymbol{\omega}.$$

- In fact, there are many practical tasks involving **weakly-annotated** data. A typical problem is Multi-Instance Learning (MIL), in which
  - ☐ Each sample is a bag of multiple instances: $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K\}$.
  - ☐ Instance label y is unknow and only an overall bag-level statement is given: $Y = \max\{y_1, y_2, \dots, y_K\}$.
- This motivates us to study a new problem of **Multi-Instance Uncertainty Estimation (MIUE).**

Challenges

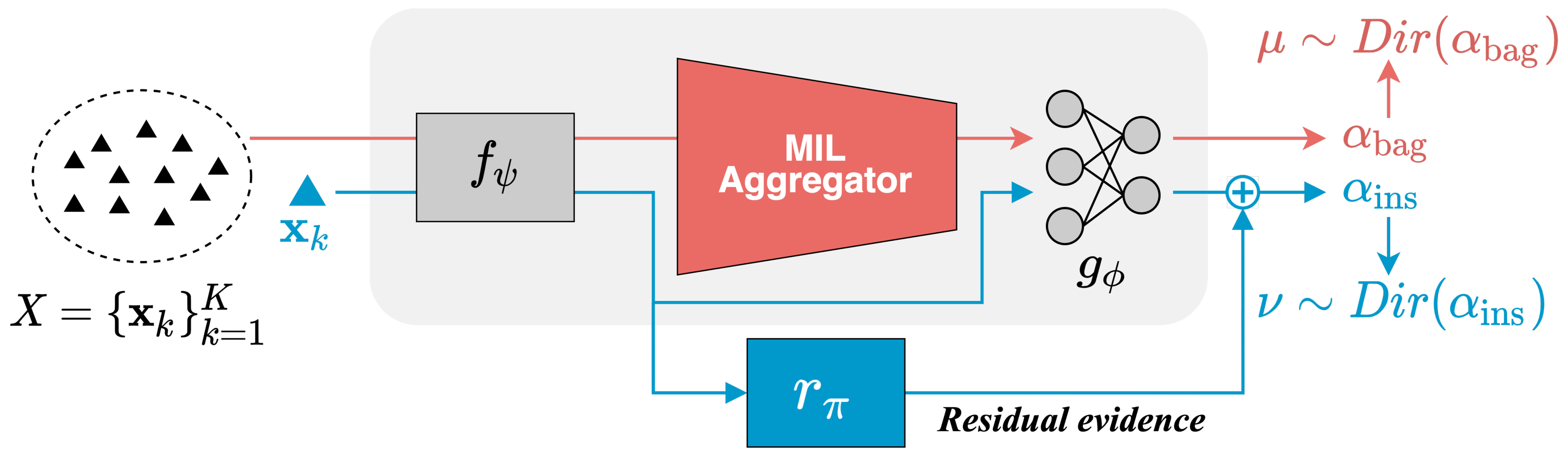
- Tackling MIUE requires a MIL model to
  - ☐ (**Bag-level UE**) learn $p(\boldsymbol{\omega}|D)$ from the multi-instance bags with variable sizes
  - ☐ (**Weakly-supervised instance-level UE**) and meanwhile jointly estimate a new weakly-supervised posterior $p(\boldsymbol{\theta}_{\text{w}}|D)$ with weakly-annotated instances.

$$P(y^*|\mathbf{x}^*,\mathcal{D}) = \int P(y^*|\mathbf{x}^*,\boldsymbol{\theta}_{\text{w}}) p(\boldsymbol{\theta}_{\text{w}}|\mathcal{D}) d\boldsymbol{\theta}_{\text{w}}$$

* $\boldsymbol{\theta}_{\text{w}}$ parameterizes the mapping from $\mathbf{x}$ to y (weak labels)

Method

- We propose to model bag-level and instance-level predictive probability with posterior **Dirichlet distributions**, inspired by evidential deep learning (EDL).



- Let $T = g \circ f$, our **residual instance estimator** is expressed as

$$R(\mathbf{x}) = T(\mathbf{x}) + r_\pi(\mathbf{h}) = g_\phi(f_\psi(\mathbf{x})) + \boxed{r_\pi(f_\psi(\mathbf{x}))}$$ ***Residual to learn***

Proposition 1. *Let $S(\cdot)$ be a bag classifier and satisfy $S(X) = g(\sum_k f(\mathbf{x}_k))$. For any bag X and its label $Y \in \{0,1\}$, further assume S can predict bags precisely: $S(X) = Y$. Then, there exists an estimator with* $\boldsymbol{T = g \circ f}$ *for any instance* $\mathbf{x}$*, such that $T(\mathbf{x}) = y$, where $y \in \{0,1\}$ is the label of* $\mathbf{x}$.

- Our **optimization strategy for $R(\mathbf{x})$**: (***i***) for $Y = 0$, all instances are negative, directly used for supervision; (***ii***) for $Y = 1$, we multiply the evidence of $\mathbf{x}_k$ (written as α_k) by different weights and aggregate them into a single one for supervision:

$$\mathcal{L}_{\text{ins}}^{+} = \mathbb{E}_{\boldsymbol{\nu} \sim Dir(\tilde{\boldsymbol{\alpha}})}\left[-\log p(Y=1|\boldsymbol{\nu}, \tilde{\boldsymbol{\alpha}}, \sigma^2)\right],$$

$$\tilde{\boldsymbol{\alpha}} = \sum_k \frac{w_k}{\sum_\tau w_\tau}\boldsymbol{\alpha}_k, \quad w_k = \mathbb{E}_{\boldsymbol{\nu} \sim Dir\left(\boldsymbol{\alpha}_k^{(T)}\right)} p(y=1|\boldsymbol{\nu}),$$

and $\alpha_k^{(T)}$ is the evidence of $\mathbf{x}_k$ derived from $T(\mathbf{x})$.

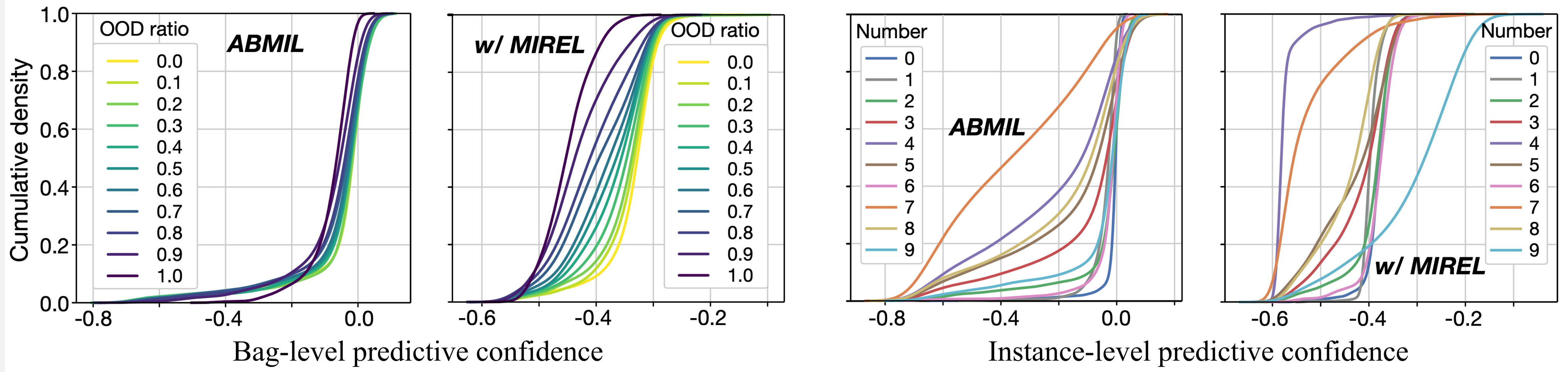
Justification for $\mathcal{L}_{\text{ins}}^{+}$: *it provides* ***a tighter upper bound*** *of the ideal (fully-supervised) instance loss function than common weakly-supervised optimization strategies under mild conditions, suggesting a more suitable* $\widehat{\boldsymbol{\theta}}_{\text{w}}$ *such that there is* $p(\boldsymbol{\theta}_{\text{w}}|D) \approx \delta(\boldsymbol{\theta}_{\text{w}} - \widehat{\boldsymbol{\theta}}_{\text{w}})$ *for accurate instance UE.*

Experiments & Results

- **Main results** on MNIST-bags. OOD-F and OOD-K mean that FMNIST and KMNIST are used for generated OOD bags, respectively. $\overline{UE}$ is the metric averaged on three UE tasks.

Method	Bag-level Acc.	Bag-level Conf.	Bag-level OOD-F	Bag-level OOD-K	Bag-level $\overline{UE}$	Instance-level Acc.	Instance-level Conf.	Instance-level OOD-F	Instance-level OOD-K	Instance-level $\overline{UE}$
- Combined with deep MIL networks										
Mean	93.38 ± 0.90	**87.02** ± 1.04	**77.57** ± 2.46	54.66 ± 2.62	73.08	86.52 ± 0.97	66.49 ± 1.37	**79.36** ± 1.95	**57.43** ± 1.50	67.76
Mean + MIREL	93.50 ± 0.53	87.01 ± 1.04	75.26 ± 1.52	**57.69** ± 6.28	**73.32**	92.45 ± 1.22	**91.49** ± 1.76	69.98 ± 4.41	56.70 ± 4.97	**72.72**
Max	94.56 ± 0.46	87.82 ± 1.49	75.23 ± 1.32	62.44 ± 3.00	75.17	92.53 ± 0.54	81.86 ± 1.54	76.97 ± 1.71	**62.53** ± 1.61	73.79
Max + MIREL	95.96 ± 0.29	**87.85** ± 2.23	**84.17** ± 3.32	**66.75** ± 5.70	**79.59**	96.82 ± 0.27	**84.22** ± 0.43	**80.81** ± 4.88	61.15 ± 3.28	**75.40**
DSMIL	96.22 ± 0.17	**87.56** ± 0.95	71.13 ± 5.20	60.71 ± 7.91	73.13	70.16 ± 3.56	64.64 ± 0.49	59.75 ± 2.35	57.50 ± 2.55	60.63
DSMIL + MIREL	96.50 ± 0.37	87.26 ± 2.66	**87.27** ± 4.27	**62.03** ± 7.78	**78.85**	97.19 ± 0.29	**73.79** ± 15.68	**73.29** ± 10.85	**57.58** ± 3.44	**68.22**
ABMIL	95.74 ± 0.38	**86.91** ± 0.98	82.93 ± 4.81	74.37 ± 4.84	81.41	75.03 ± 0.28	61.28 ± 0.86	63.68 ± 1.00	52.63 ± 1.07	59.20
ABMIL + MIREL	96.48 ± 0.22	86.63 ± 1.32	**92.84** ± 0.60	**79.95** ± 4.12	**86.47**	87.71 ± 0.67	**90.73** ± 1.31	**78.13** ± 2.19	**67.02** ± 1.94	**78.63**
- Compared with related UE methods using ABMIL as the base MIL network										
Deep Ensemble	96.06 ± 0.35	87.36 ± 0.59	80.07 ± 2.57	74.33 ± 3.97	80.59	75.56 ± 0.32	71.89 ± 0.91	70.48 ± 0.53	55.22 ± 1.16	65.87
MC Dropout	96.28 ± 0.41	**88.46** ± 1.82	89.57 ± 3.84	78.24 ± 4.89	85.42	75.61 ± 0.66	68.40 ± 1.54	68.34 ± 1.06	58.61 ± 1.38	65.12
$\mathcal{I}$-EDL	96.08 ± 0.20	86.78 ± 0.87	85.51 ± 7.56	73.15 ± 3.87	81.82	75.45 ± 0.13	60.72 ± 1.46	63.91 ± 1.31	54.14 ± 2.19	59.59
Bayes-MIL	96.44 ± 0.33	85.63 ± 1.53	81.02 ± 11.71	57.04 ± 12.61	74.57	91.64 ± 1.25	82.24 ± 1.85	60.77 ± 6.59	42.06 ± 2.84	61.69
MIREL	96.48 ± 0.22	86.63 ± 1.32	**92.84** ± 0.60	**79.95** ± 4.12	**86.47**	87.71 ± 0.67	**90.73** ± 1.31	**78.13** ± 2.19	**67.02** ± 1.94	**78.63**

- Bag-level and instance-level **uncertainty analysis** on MNIST-bags.



- Understand **$R(\mathbf{x})$'s UE behavior** using a synthetic 2D MIL dataset: