Uncertainty in AI: A Variance-Based Approach

Ruxiao Duan[1], Brian Caffo[2], Harrison Bai[3], Haris Sair[3], and Craig Jones[3,4,5]

[1]Department of Applied Mathematics and Statistics, [2]Department of Biostatistics, [3]Department of Radiology and Radiological Science, [4]Department of Computer Science, [5]Malone Center for Engineering in Healthcare

Whiting School of Engineering, School of Public Health, School of Medicine | Johns Hopkins University | Baltimore MD

Introduction

- Uncertainty quantification (UQ) of deep neural networks is crucial in safety-critical applications. Evidential deep learning (EDL) [1] is a non-Bayesian approach to estimate aleatoric uncertainty (AU) and epistemic uncertainty (EU) by modeling the evidential prior of the target label distribution.
- Traditional classification EDL adopts an **entropy-based** approach [2] to quantify uncertainty and the class-level uncertainty is lost.
- We propose a **variance-based** approach to estimate AU and EU in deep learning classification, achieving class-wise uncertainty estimation and class correlation quantification.
- We test our approach in an active domain adaptation (ADA) algorithm to demonstrate the effectiveness of uncertainty quantification in a downstream application.

Method: Variance-Based Uncertainty Quantification

- In a C-class classification problem, let $\mathbf{y}$ be the one-hot target label vector and $\mathbf{p}$ be the class probability vector. Assume

$$\mathbf{y} \sim \mathrm{Categorical}(\mathbf{p}) \quad \text{and} \quad \mathbf{p} \sim \mathrm{Dirichlet}(\boldsymbol{\alpha})$$

- An evidential neural network, trained with negative log-likelihood function, can predict $\boldsymbol{\alpha}$ given the input $\mathbf{x}$.
- The covariance matrix of $\mathbf{y}$ can be decomposed by the law of total covariance

$$\mathrm{Cov}[\mathbf{y}] \coloneqq \mathrm{E}[(\mathbf{y}-\mathrm{E}[\mathbf{y}])(\mathbf{y}-\mathrm{E}[\mathbf{y}])^{\mathrm{T}}] = \mathrm{E}\big[\mathrm{Cov}[\mathbf{y}|\mathbf{p}]\big] + \mathrm{Cov}\big[\mathrm{E}[\mathbf{y}|\mathbf{p}]\big]$$

- Based on the distribution assumption, it can be derived that

$$\mathrm{Cov}[\mathbf{y}]^{\mathrm{alea}} \coloneqq \mathrm{E}\big[\mathrm{Cov}[\mathbf{y}|\mathbf{p}]\big] = \frac{\alpha_0}{\alpha_0+1}\left(\mathrm{Diag}(\bar{\mathbf{p}}) - \bar{\mathbf{p}}\bar{\mathbf{p}}^{\mathrm{T}}\right)$$

$$\mathrm{Cov}[\mathbf{y}]^{\mathrm{epis}} \coloneqq \mathrm{Cov}\big[\mathrm{E}[\mathbf{y}|\mathbf{p}]\big] = \frac{1}{\alpha_0+1}\left(\mathrm{Diag}(\bar{\mathbf{p}}) - \bar{\mathbf{p}}\bar{\mathbf{p}}^{\mathrm{T}}\right)$$

where $\alpha_0 \coloneqq \sum_{i=1}^{C}\alpha_i$ and $\bar{\mathbf{p}} \coloneqq \mathrm{E}[\mathbf{p}] = \frac{\boldsymbol{\alpha}}{\alpha_0}$.

- Class-wise evidential uncertainties of class i can be obtained by

$$U_i^{\mathrm{alea}} \coloneqq \mathrm{Cov}[\mathbf{y}]_{i,i}^{\mathrm{alea}} = \frac{\alpha_0}{\alpha_0+1}\bar{\mathbf{p}}_i(1-\bar{\mathbf{p}}_i)$$

$$U_i^{\mathrm{epis}} \coloneqq \mathrm{Cov}[\mathbf{y}]_{i,i}^{\mathrm{epis}} = \frac{1}{\alpha_0+1}\bar{\mathbf{p}}_i(1-\bar{\mathbf{p}}_i)$$

- Sample-wise uncertainties can be quantified as

$$U^{\mathrm{alea}} \coloneqq \sum_{i=1}^{C} U_i^{\mathrm{alea}} = \frac{\alpha_0}{\alpha_0+1}\left(1-\sum_{i=1}^{C}\bar{\mathbf{p}}_i^2\right)$$

$$U^{\mathrm{epis}} \coloneqq \sum_{i=1}^{C} U_i^{\mathrm{epis}} = \frac{1}{\alpha_0+1}\left(1-\sum_{i=1}^{C}\bar{\mathbf{p}}_i^2\right)$$

- Class correlation matrix can be calculated as

$$\mathrm{Corr}[\mathbf{y}] \coloneqq \frac{\mathrm{Cov}[\mathbf{y}]}{\sigma(\mathbf{y})\,\sigma(\mathbf{y})^{\mathrm{T}}}$$

where $\sigma(\mathbf{y}) \coloneqq \sqrt{\mathrm{diag}(\mathrm{Cov}[\mathbf{y}])}$.

Experiment: Active Domain Adaptation

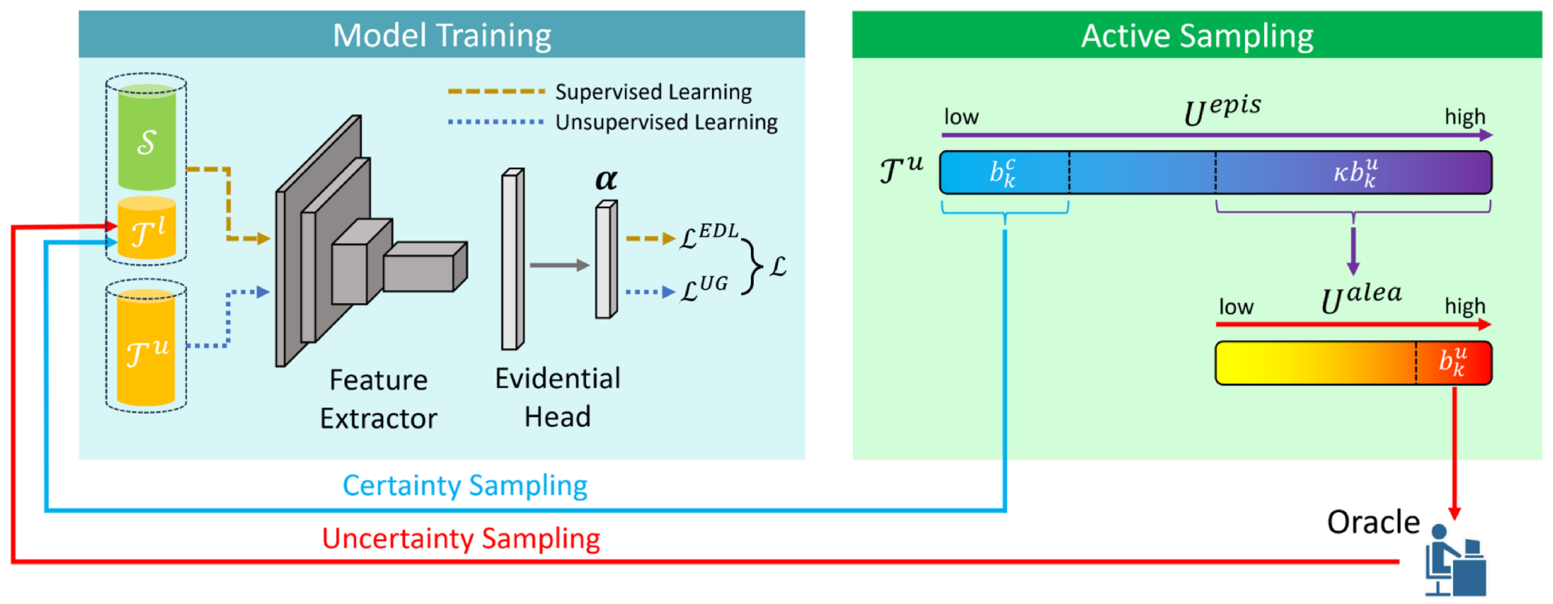


Figure 1: ADA with DUC algorithm [3] and certainty sampling.

- The uncertainty quantification method is applied in the state-of-the-art active domain adaptation algorithm, DUC [3]. An improved version of DUC with a certainty sampling technique is implemented.
- In each active selection round, certain samples and uncertain samples in the unlabeled target domain dataset are first identified by EU. Certain samples are directly moved to the labeled dataset with their predictions as pseudo labels. Uncertain samples are then sorted by their AU and the oracle annotates the most uncertain ones based on the budget and adds them to the labeled pool.
- Two multi-domain image classification datasets, Office-Home and Visda-2017, are employed for testing.

Results: Sample-Level Uncertainty Quantification

UQ	Office-Home		Visda-2017	
	w/o CS	w/ CS	w/o CS	w/ CS
Entropy	78.1	79.3	89.3	89.3
Variance	78.2	79.5	89.4	89.5

Table 1: ADA accuracy (%). (CS: certainty sampling.)

UQ	Office-Home		Visda-2017	
	AU	EU	AU	EU
Entropy	83.0	81.7	75.6	73.8
Variance	83.3	82.0	75.3	73.8

Table 2: Misclassification detection AUC (%).

Results: Class-Level Uncertainty Quantification

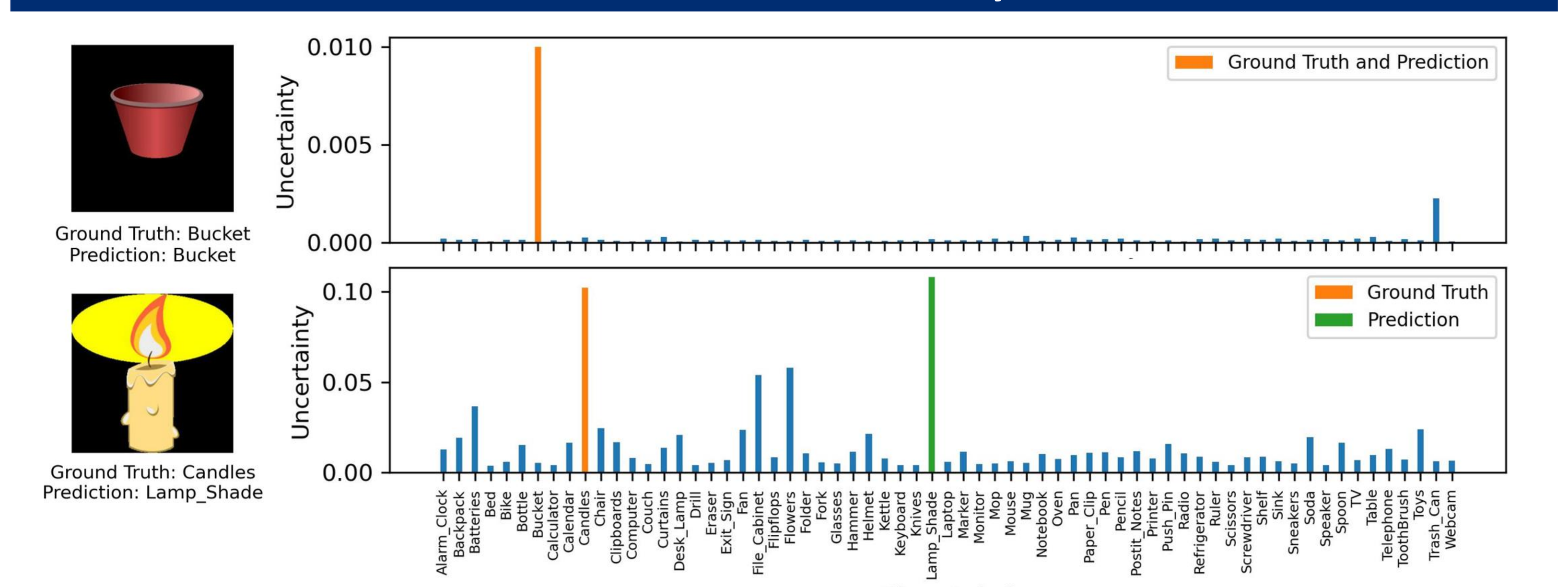


Figure 2: Class-level uncertainty of two Office-Home images.

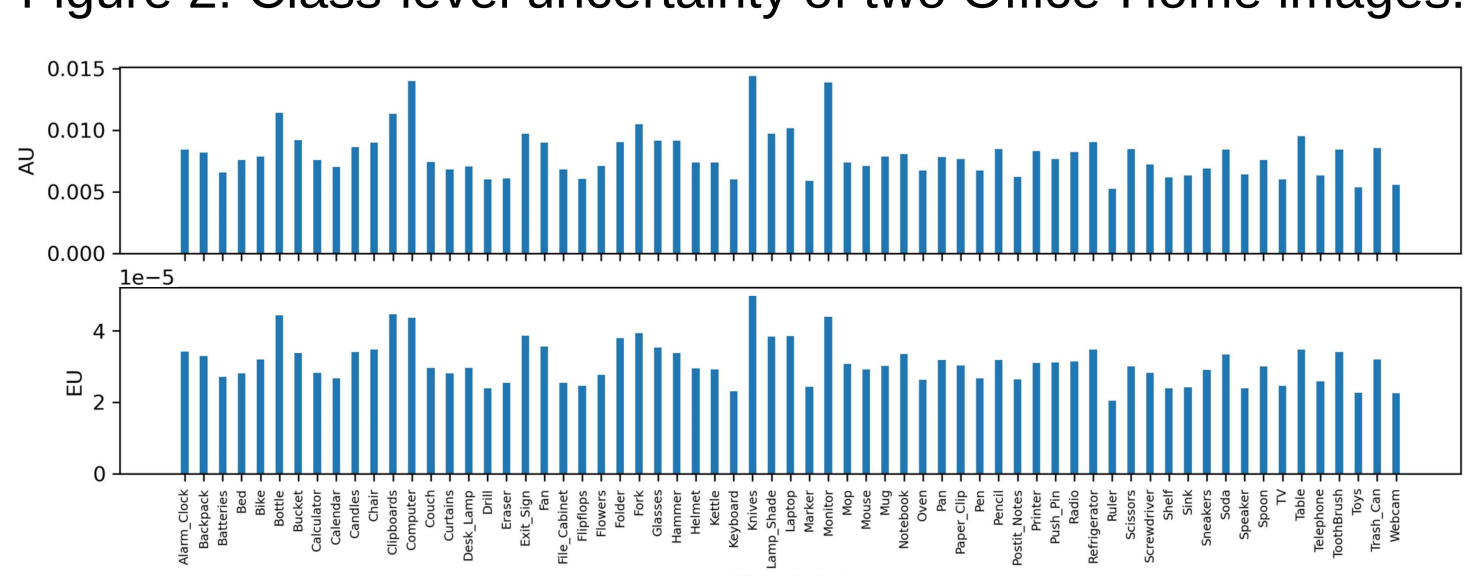


Figure 3: Average class-level uncertainty of Office-Home Clipart.

Results: Between-Class Correlation

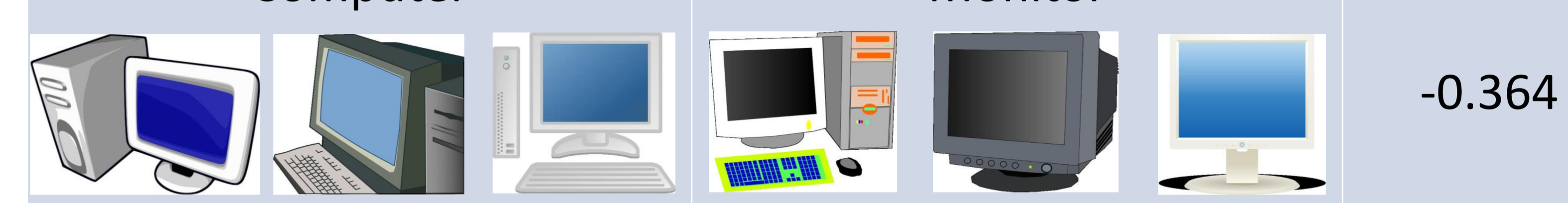
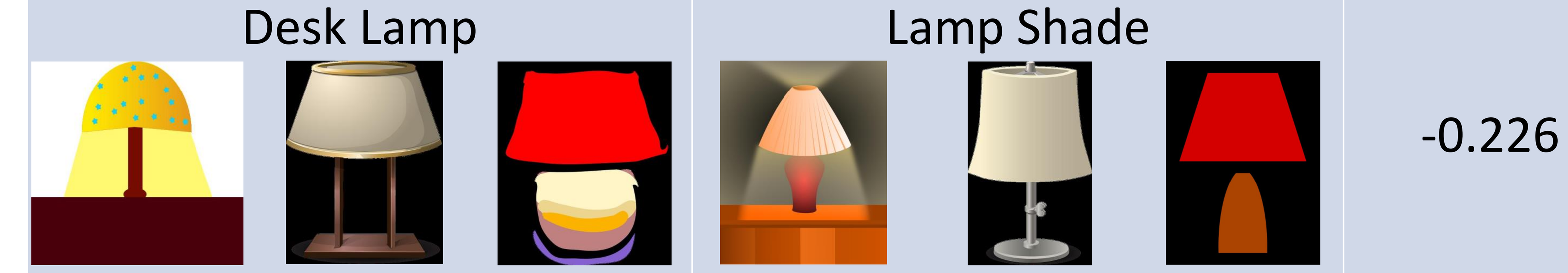
Class Pair		Correlation
Computer 	Monitor	-0.364
Bucket 	Trash Can	-0.239
Desk Lamp 	Lamp Shade	-0.226

Table 3: Three most correlated class pairs and their example images in Office-Home Clipart domain.

Conclusion

- The proposed variance-based method demonstrates similar performance as the entropy-based method on sample-level uncertainty estimation in ADA and misclassification detection.
- The variance-based approach further provides uncertainty information at the class level, which cannot be achieved by the entropy-based approach. This can be beneficial in applications that require accurate estimation of class-wise uncertainties.
- Between-class correlations can be quantified using our method, and similar class pairs confusing to the model can be identified.

References

- [1] Sensoy, M., Kaplan, L., & Kandemir, M. (2018). Evidential deep learning to quantify classification uncertainty. Advances in neural information processing systems, 31.
- [2] Malinin, A., & Gales, M. (2018). Predictive uncertainty estimation via prior networks. Advances in neural information processing systems, 31.
- [3] Xie, M., Li, S., Zhang, R., & Liu, C. H. (2023). Dirichlet-based Uncertainty Calibration for Active Domain Adaptation. arXiv preprint arXiv:2302.13824.