Evidential Uncertainty Quantification: A Variance-Based Perspective

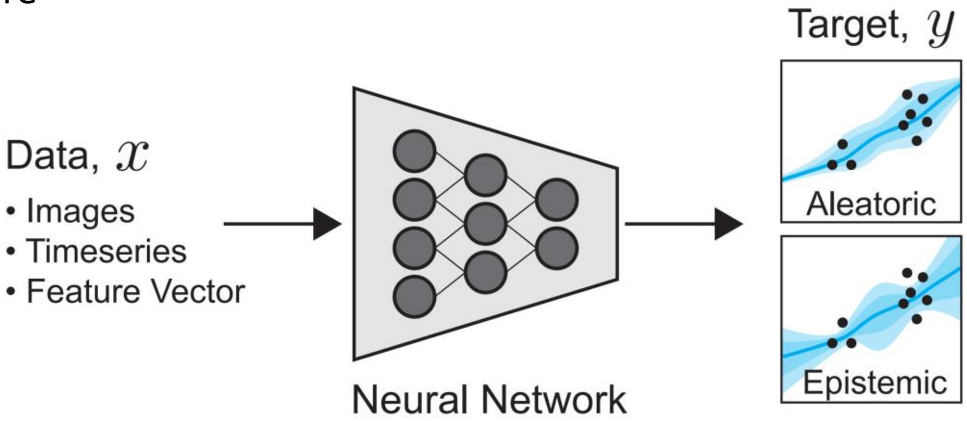
Authors: Ruxiao Duan[1], Brian Caffo[2,3], Harrison X. Bai[4], Haris I. Sair[3,4], and Craig Jones[3,4,5]

Presenter: Ruxiao Duan[1]

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

Evidential Deep Learning

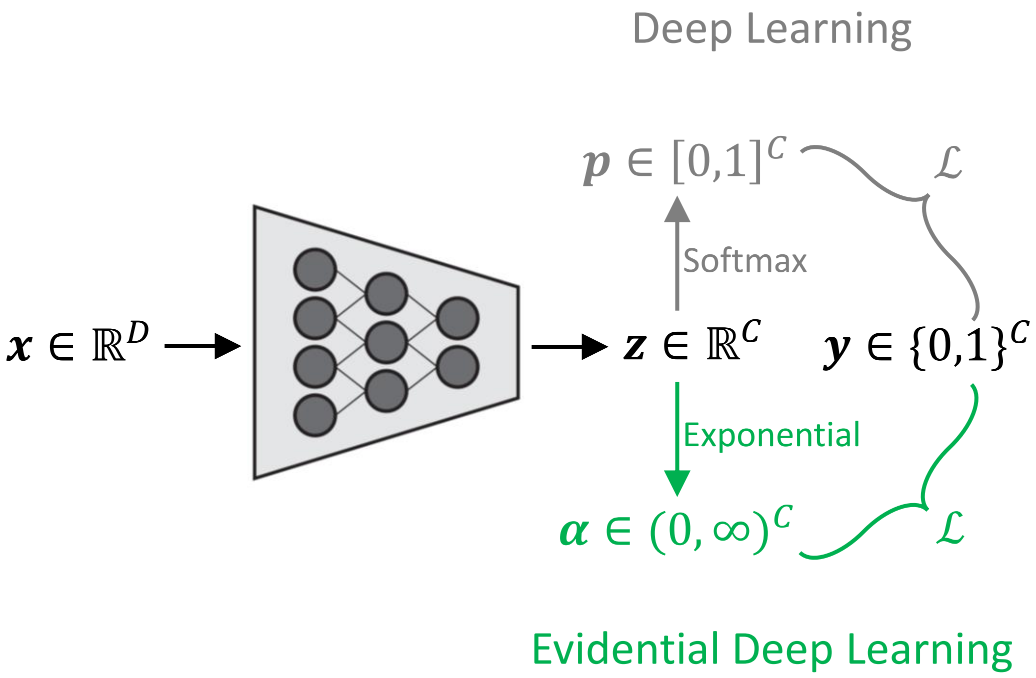
- **Evidential deep learning (EDL)**: a non-Bayesian approach to quantify predictive uncertainty in deep learning by modeling the evidential prior of the target distribution
  - Quantifies uncertainties directly with a single forward propagation
  - Disentangles aleatoric/data uncertainty (AU) and epistemic/model uncertainty (EU)
  - No change in model architecture
  - No sampling during inference



Amini, et al., 2020. *Advances in Neural Information Processing Systems*, *33*, 14927-14937.

EDL Fundamental Assumptions

- **Classification notations**
  - C: number of classes
  - $\boldsymbol{x}$: input data
  - $\boldsymbol{y}$: target one-hot label vector
  - $\boldsymbol{p}$: class probability vector
  - $\boldsymbol{\alpha}$: prior distribution parameters
- **EDL probability assumptions**
  - $\boldsymbol{y} \sim \text{Multinomial}(1, \boldsymbol{p})$
  - $\boldsymbol{p} \sim \text{Dirichlet}(\boldsymbol{\alpha})$
- **Evidential neural network $\boldsymbol{f}$**
  - $\boldsymbol{f}(\boldsymbol{x}; \boldsymbol{\Theta}) = \boldsymbol{\alpha}$
  - Learning: $\mathcal{L}(\boldsymbol{\Theta}) = -\log \mathbb{P}(\boldsymbol{y}|\boldsymbol{\alpha})$



Sensoy, et al., 2018. *Advances in neural information processing systems, 31*.

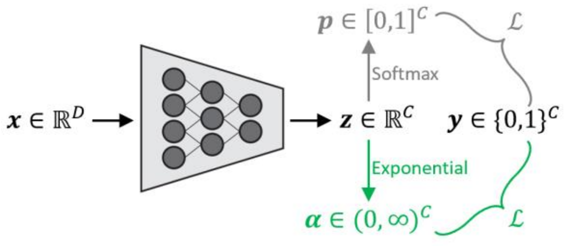
Traditional Entropy-Based EDL

- **Traditional entropy-based EDL for classification uncertainty quantification**
  - $U = -\sum_{c=1}^{C} \bar{p}_c \log \bar{p}_c$
  - $U^{alea} = \sum_{c=1}^{C} \bar{p}_c \big(\psi(\alpha_0 + 1) - \psi(\alpha_c + 1)\big)$
  - $U^{epis} = -\sum_{c=1}^{C} \bar{p}_c \big(\log \bar{p}_c + \psi(\alpha_0 + 1) - \psi(\alpha_c + 1)\big)$ $\quad \left(\alpha_0 \coloneqq \sum_{c=1}^{C} \alpha_c\,,\ \bar{\boldsymbol{p}} \coloneqq \frac{\boldsymbol{\alpha}}{\alpha_0}\right)$
- **Class-level uncertainties are lost**

• *Computer?*	*Probably*	*More contribution to total uncertainty*
• *Keyboard?*	*Probably*	
• *Monitor?*	*Probably*	
• *Mouse?*	*Probably*	
• *Calculator?*	*Not likely*	*Less contribution to total uncertainty*
• *Laptop?*	*Not likely*	
• *Refrigerator?*	*Not likely*	
• *TV?*	*Not likely*	

Malinin, et al., 2018. *Advances in neural information processing systems*, *31*.

New Approach: From a Variance-Based Perspective



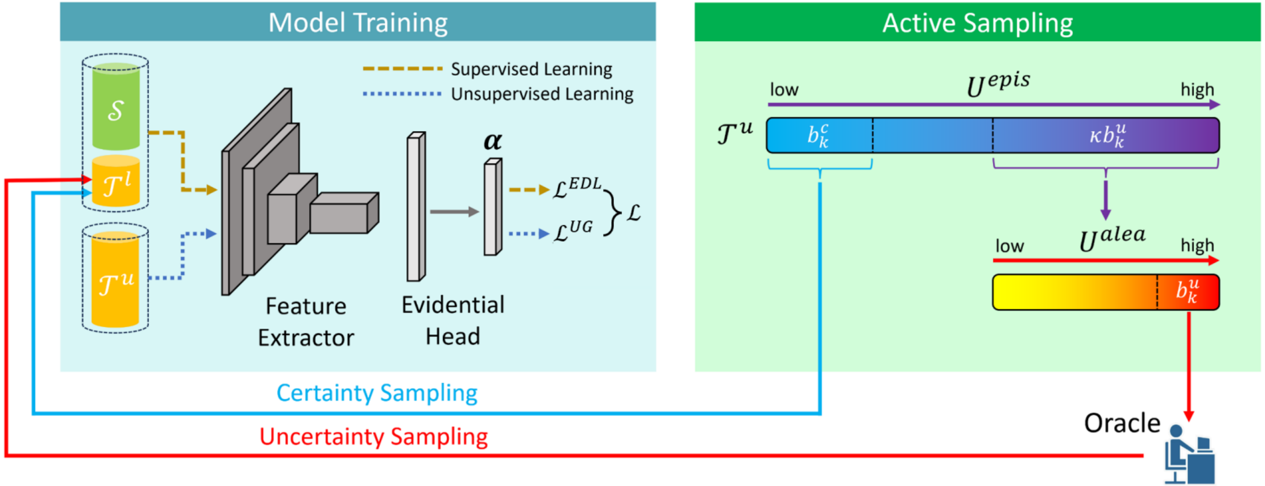
- **Covariance matrix decomposition**
  - $\mathrm{Cov}[\boldsymbol{y}] \coloneqq \mathbb{E}\left[(\boldsymbol{y}-\mathbb{E}[\boldsymbol{y}])(\boldsymbol{y}-\mathbb{E}[\boldsymbol{y}])^{\mathrm{T}}\right] = \mathrm{Diag}(\overline{\boldsymbol{p}}) - \overline{\boldsymbol{p}}\overline{\boldsymbol{p}}^{\mathrm{T}}$
  - $\mathrm{Cov}[\boldsymbol{y}]^{alea} \coloneqq \mathbb{E}\left[\mathrm{Cov}[\boldsymbol{y}|\boldsymbol{p}]\right] = \frac{\alpha_0}{\alpha_0+1}\left(\mathrm{Diag}(\overline{\boldsymbol{p}}) - \overline{\boldsymbol{p}}\overline{\boldsymbol{p}}^{\mathrm{T}}\right)$
  - $\mathrm{Cov}[\boldsymbol{y}]^{epis} \coloneqq \mathrm{Cov}\left[\mathbb{E}[\boldsymbol{y}|\boldsymbol{p}]\right] = \frac{1}{\alpha_0+1}\left(\mathrm{Diag}(\overline{\boldsymbol{p}}) - \overline{\boldsymbol{p}}\overline{\boldsymbol{p}}^{\mathrm{T}}\right)$

- **Class-level uncertainties**
  - $U_c \coloneqq \mathrm{Cov}[\boldsymbol{y}]_{c,c} = \overline{\boldsymbol{p}}_c(1-\overline{\boldsymbol{p}}_c)$
  - $U_c^{alea} \coloneqq \mathrm{Cov}[\boldsymbol{y}]_{c,c}^{alea} = \frac{\alpha_0}{\alpha_0+1}\overline{\boldsymbol{p}}_c(1-\overline{\boldsymbol{p}}_c)$
  - $U_c^{epis} \coloneqq \mathrm{Cov}[\boldsymbol{y}]_{c,c}^{epis} = \frac{1}{\alpha_0+1}\overline{\boldsymbol{p}}_c(1-\overline{\boldsymbol{p}}_c)$

- **Sample-level uncertainties**
  - $U \coloneqq \sum_{c=1}^{C} U_c = 1 - \sum_{c=1}^{C} \overline{\boldsymbol{p}}_c^2$
  - $U^{alea} \coloneqq \sum_{c=1}^{C} U_c^{alea} = \frac{\alpha_0}{\alpha_0+1}\left(1 - \sum_{c=1}^{C} \overline{\boldsymbol{p}}_c^2\right)$
  - $U^{epis} \coloneqq \sum_{c=1}^{C} U_c^{epis} = \frac{1}{\alpha_0+1}\left(1 - \sum_{c=1}^{C} \overline{\boldsymbol{p}}_c^2\right)$

- **Between-class correlation matrix**
  - $\mathrm{Corr}[\boldsymbol{y}] \coloneqq \frac{\mathrm{Cov}[\boldsymbol{y}]}{\sigma(\boldsymbol{y})\sigma(\boldsymbol{y})^{\mathrm{T}}}, \quad \sigma(\mathbf{y}) \coloneqq \sqrt{\mathrm{diag}(\mathrm{Cov}[\mathbf{y}])}$

Application: Active Domain Adaptation



EDL-based active domain adaptation (ADA) framework with concurrent certainty and uncertainty sampling

- **Domain gap reduction**: model trained by uncertainties of unlabeled target dataset
- **Uncertainty sampling**: two-round sampling strategy based on evidential uncertainties
- **Certainty sampling**: certain samples identified by EU for semi-supervised learning

Xie, et al., 2023. *arXiv preprint arXiv:2302.13824*.

Sample-Level Uncertainty Quantification

- **Variance-based approach has similar performance as entropy-based appraoch on sample-level uncertainty quantification**

* CS: certainty sampling

Uncertainty	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

Active domain adaptation accuracy (%)

Uncertainty	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

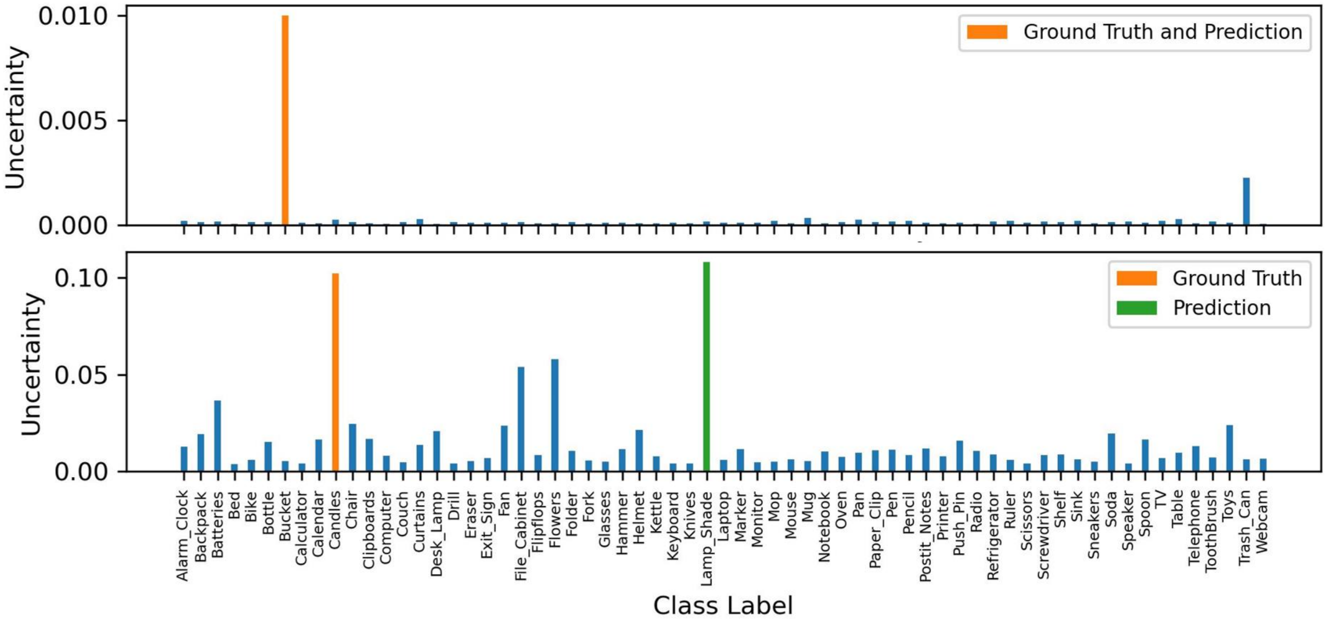
Misclassification detection AUC (%)

Class-Level Uncertainty Quantification

- **Variance-based approach further provides class-level uncertainties**

Ground Truth: Bucket
Prediction: Bucket

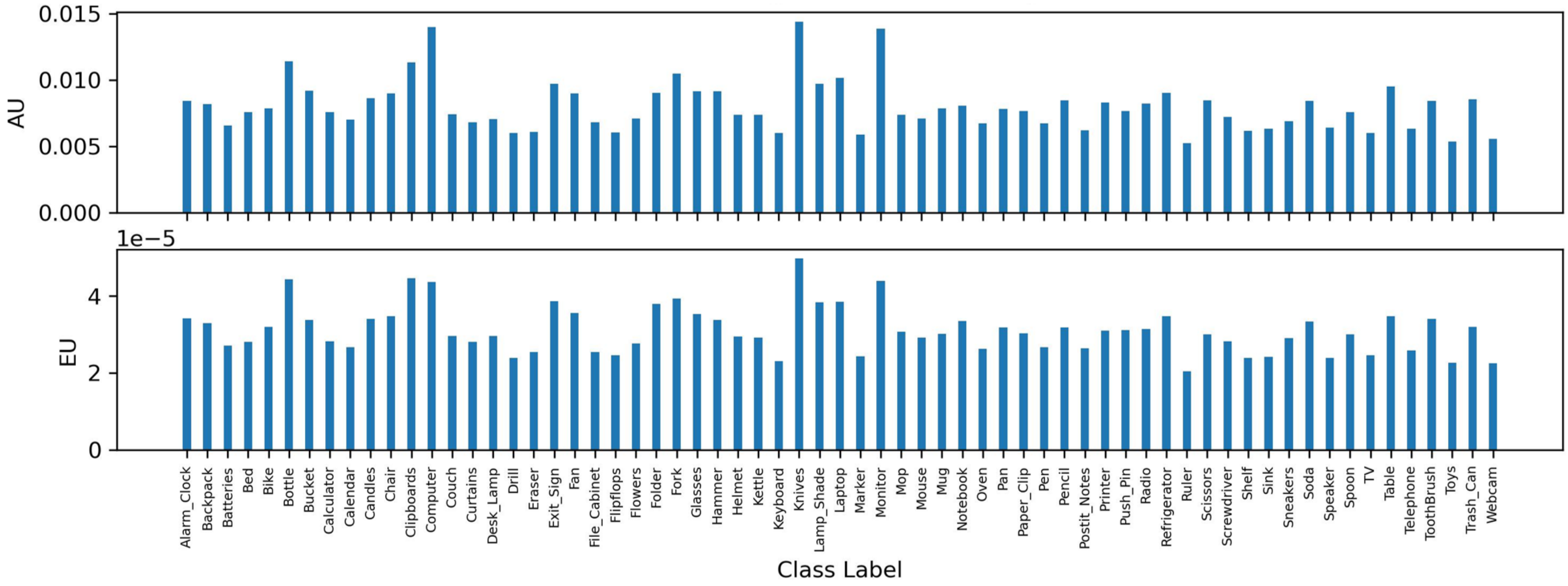
Ground Truth: Candles
Prediction: Lamp_Shade



Class-level uncertainties of two images in Office-Home

Class-Level Uncertainty Quantification

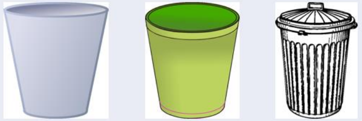
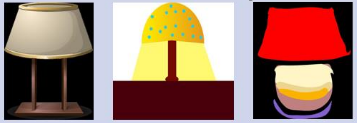
- **Variance-based approach further provides class-level uncertainties**



Class-level uncertainties of Office-Home Clipart

Between-Class Correlation Quantification

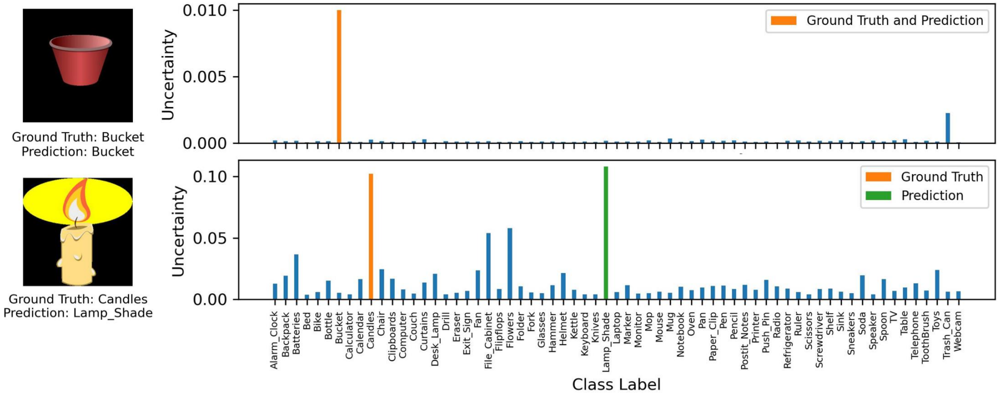
- **Variance-based approach further provides between-class correlations**

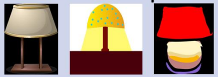
Class Pair		Correlation
Computer 	Monitor 	-0.364
Bucket 	Trash Can 	-0.239
Desk Lamp 	Lamp Shade 	-0.226

Top three correlated class pairs in Office-Home Clipart

Conclusion

- We introduced a variance-based uncertainty quantification approach for classification EDL, which can provide
  - sample-level evidential uncertainties
  - class-level evidential uncertainties
  - between-class correlations



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

References

- Amini, A., Schwarting, W., Soleimany, A., & Rus, D. (2020). Deep evidential regression. *Advances in Neural Information Processing Systems*, *33*, 14927-14937.
- Malinin, A., & Gales, M. (2018). Predictive uncertainty estimation via prior networks. *Advances in neural information processing systems*, *31*.
- Sensoy, M., Kaplan, L., & Kandemir, M. (2018). Evidential deep learning to quantify classification uncertainty. *Advances in neural information processing systems*, *31*.
- Xie, M., Li, S., Zhang, R., & Liu, C. H. (2023). Dirichlet-based Uncertainty Calibration for Active Domain Adaptation. *arXiv preprint arXiv:2302.13824*.

JOHNS HOPKINS

WHITING SCHOOL *of* ENGINEERING