UNIVERSITY OF CAMBRIDGE

Depth Uncertainty in Neural Networks

UDL Workshop, ICML 2020

Javier Antorán, James Urquhart Allingham, José Miguel Hernández-Lobato

About Us

Javier Antorán
ja666@cam.ac.uk

James Urquhart Allingham
jua23@cam.ac.uk

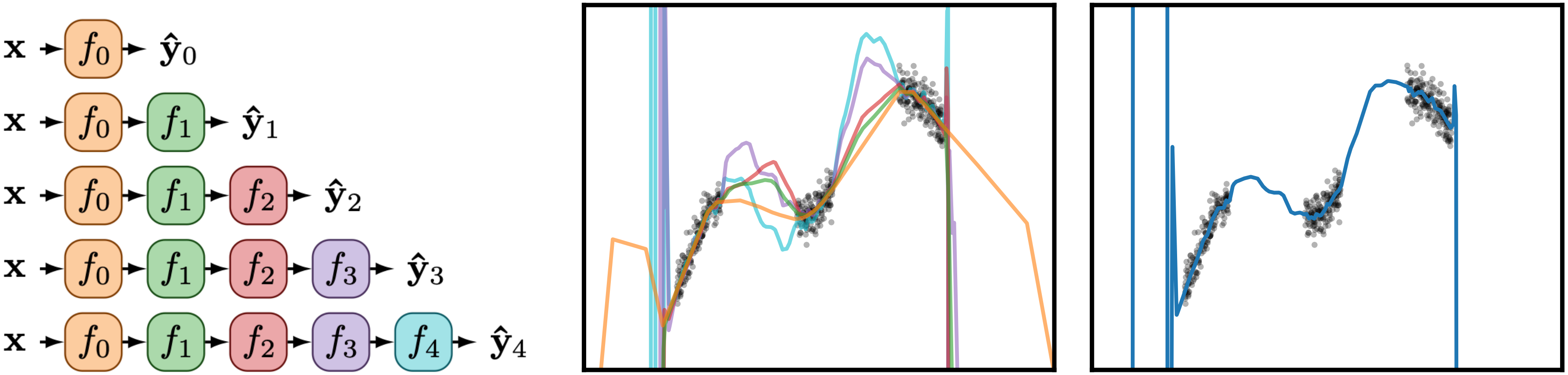
José Miguel Hernández-Lobato
jmh233@cam.ac.uk

Uncertainty in Deep Learning…

People saying AI will take over the world:

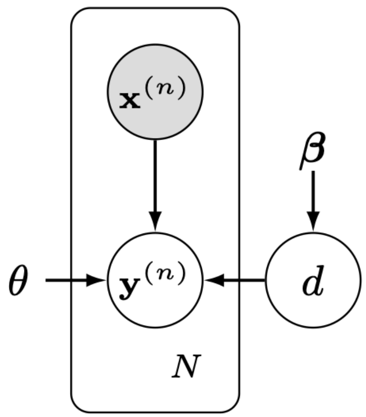
Meanwhile, my Deep Neural Network:

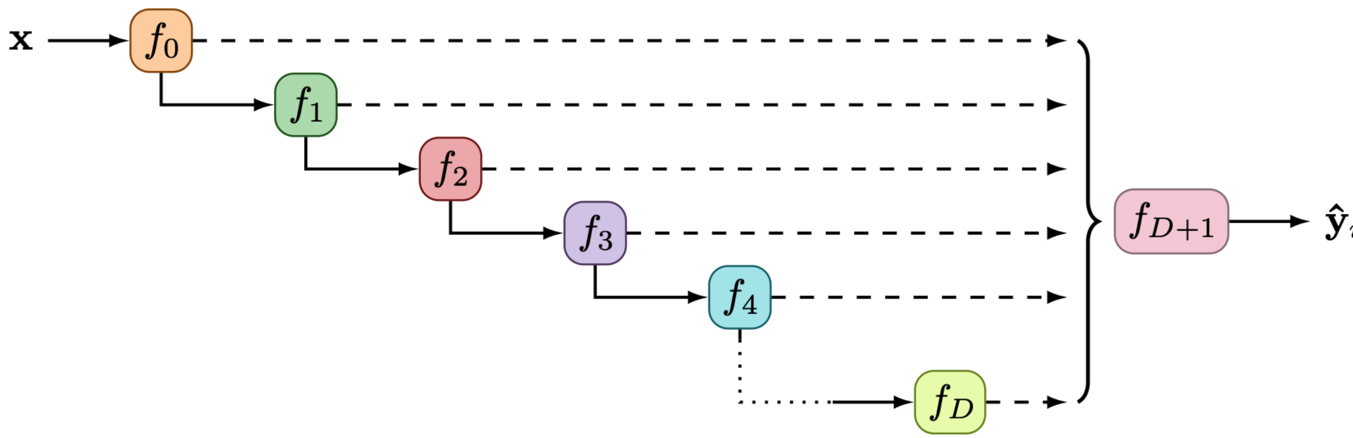
DUN: Intuition



- Depth Uncertainty Networks (DUNs), transform uncertainty over depth into predictive uncertainty.

DUN: Inference with a Single Forward Pass





$d \sim \text{Cat}(d; \beta)$

$$\log p(\mathfrak{D}; \boldsymbol{\theta}) = \log \sum_{i=0}^{D} \left(p_{\boldsymbol{\beta}}(d{=}i) \cdot \prod_{n=1}^{N} p(\mathbf{y}^{(n)} | \mathbf{x}^{(n)}, d{=}i; \boldsymbol{\theta}) \right)$$

$$\geq \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\theta}) = \sum_{n=1}^{N} \mathbb{E}_{q_{\boldsymbol{\alpha}}(d)} \left[\log p(\mathbf{y}^{(n)} | \mathbf{x}^{(n)}, d; \boldsymbol{\theta}) \right] - \mathrm{KL}(q_{\boldsymbol{\alpha}}(d) \,\|\, p_{\boldsymbol{\beta}}(d))$$

But Why VI?

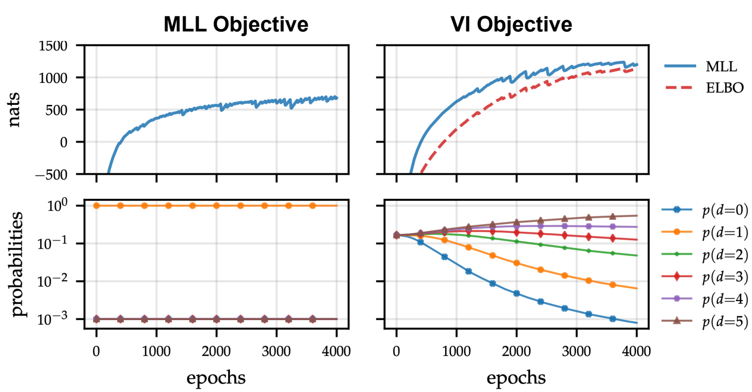
- With the ML objective, the rich get richer and the posterior is prone to collapse

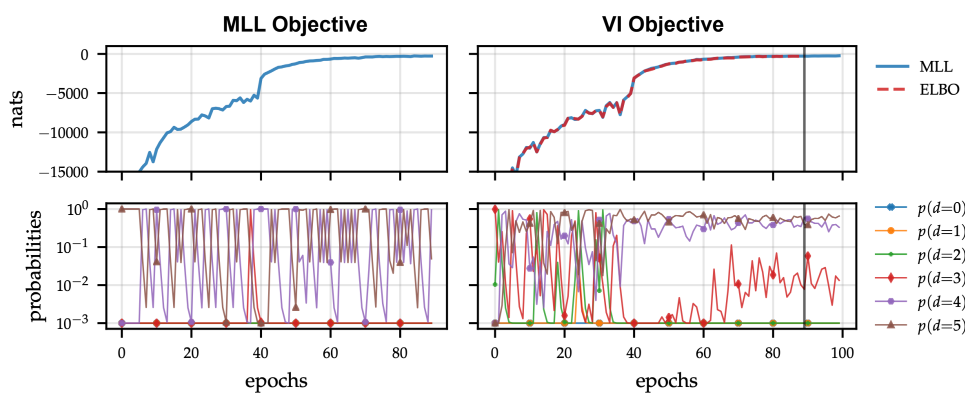
$$\frac{\partial}{\partial \boldsymbol{\theta}} \log p(\mathfrak{D}; \boldsymbol{\theta}) = \mathbb{E}_{p(d|\mathfrak{D};\boldsymbol{\theta})}\left[\frac{\partial}{\partial \boldsymbol{\theta}} \log p(\mathfrak{D}|d; \boldsymbol{\theta})\right]$$

- With VI, the optimization of variational parameters and model weights is decoupled

$$\frac{\partial}{\partial \boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{i=0}^{D} q_{\boldsymbol{\alpha}}(d{=}i) \frac{\partial}{\partial \boldsymbol{\theta}} \log p(\mathfrak{D}|d{=}i; \boldsymbol{\theta})$$

$$\frac{\partial}{\partial \alpha_i} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \log p(\mathfrak{D}|d{=}i; \boldsymbol{\theta}) \frac{\partial}{\partial \alpha_i} q_{\boldsymbol{\alpha}}(d{=}i) - (\log q_{\boldsymbol{\alpha}}(d{=}i) - \log p(d{=}i) + 1) \frac{\partial}{\partial \alpha_i} q_{\boldsymbol{\alpha}}(d{=}i)$$





Toy Examples!

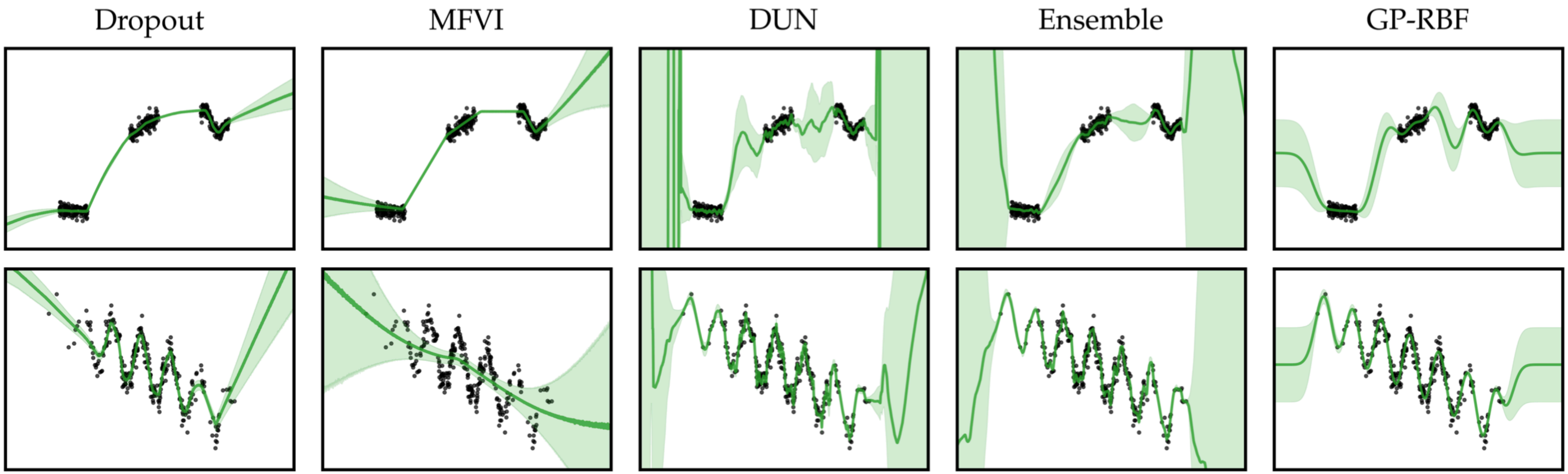
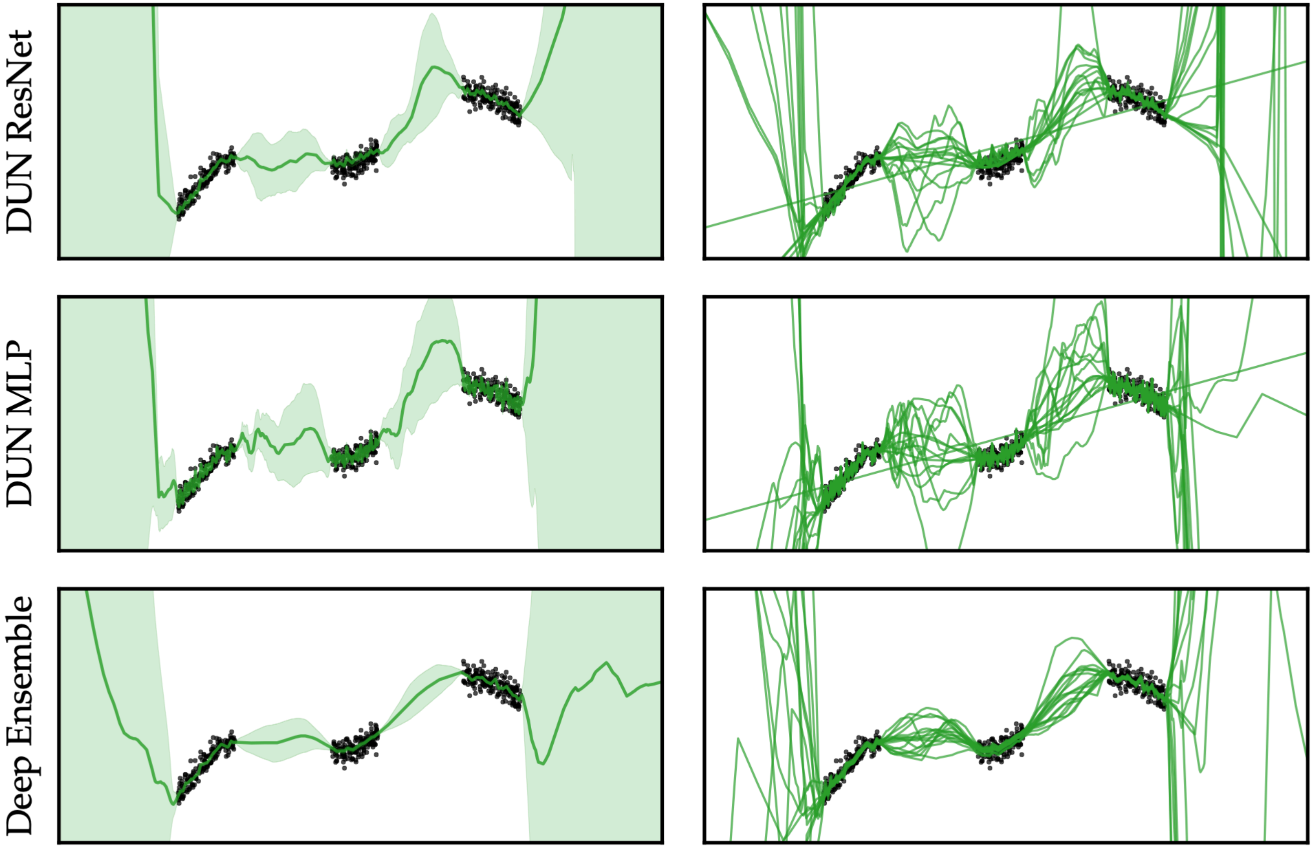


Figure 4: Top row: toy dataset from Izmailov et al. (2019). Bottom: Wiggle dataset. Black dots denote data points. Error bars represent standard deviation among mean predictions.

Function Diversity in DUNs



Regression (ranks across 9 UCI datasets)

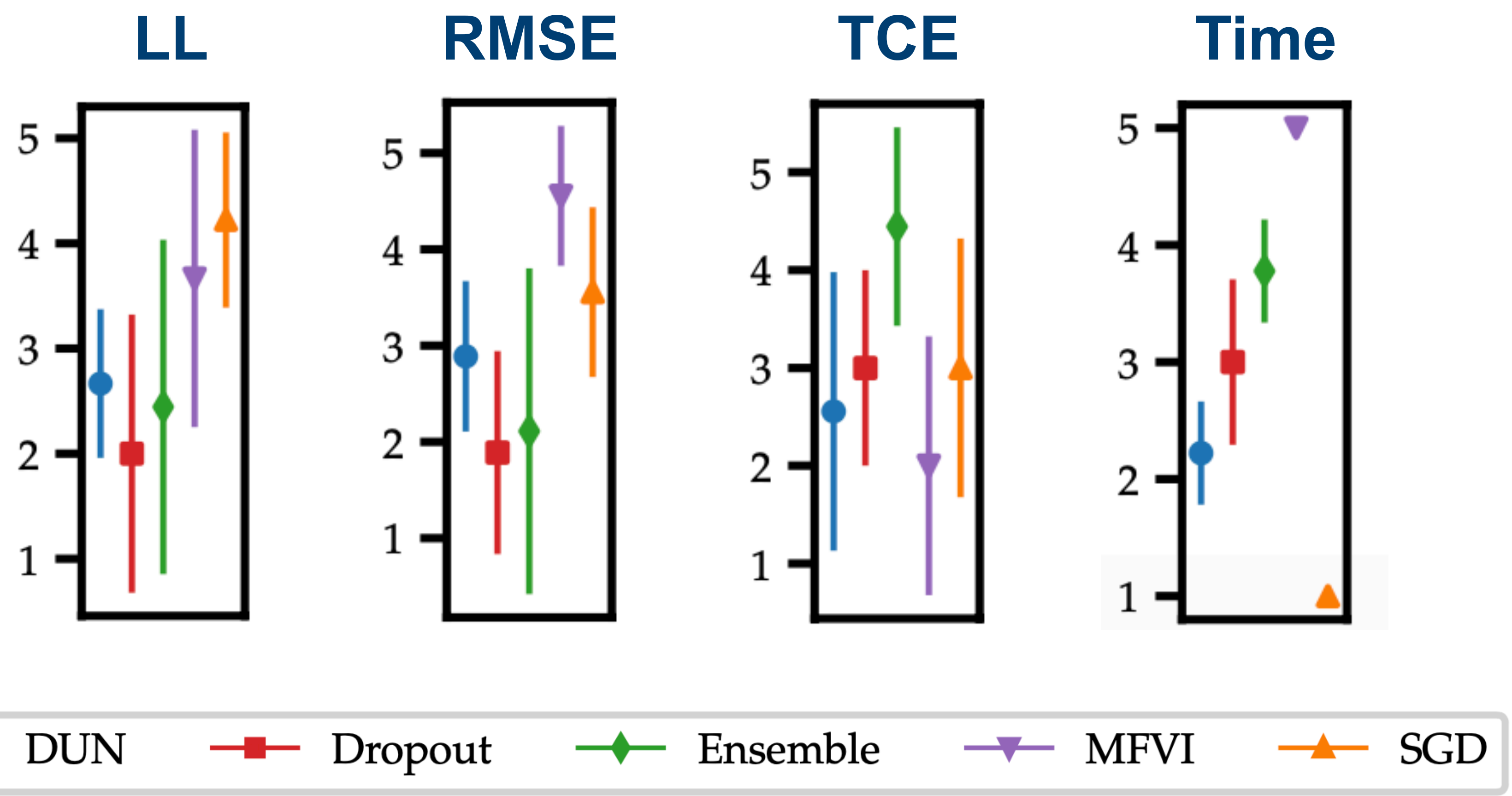
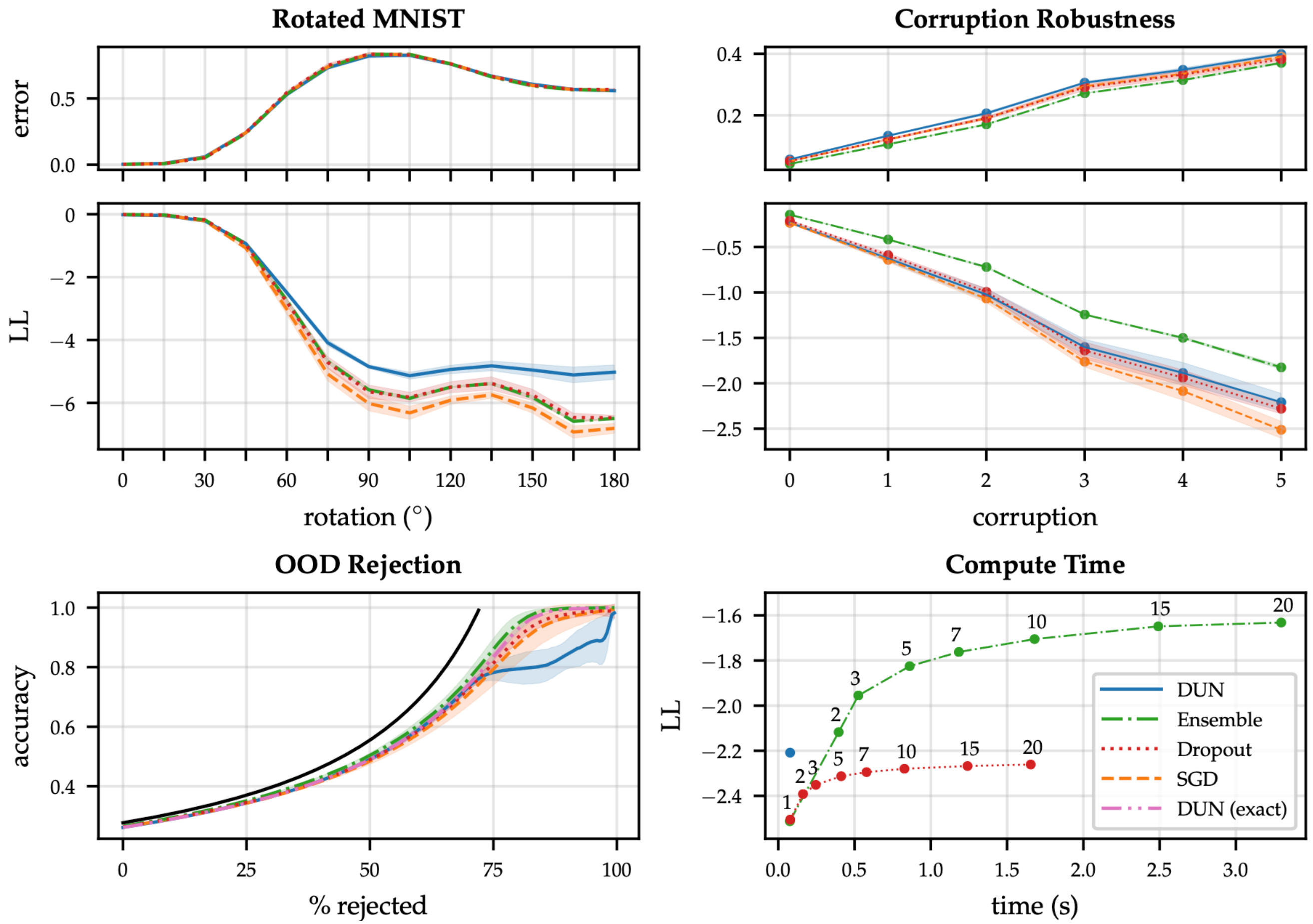
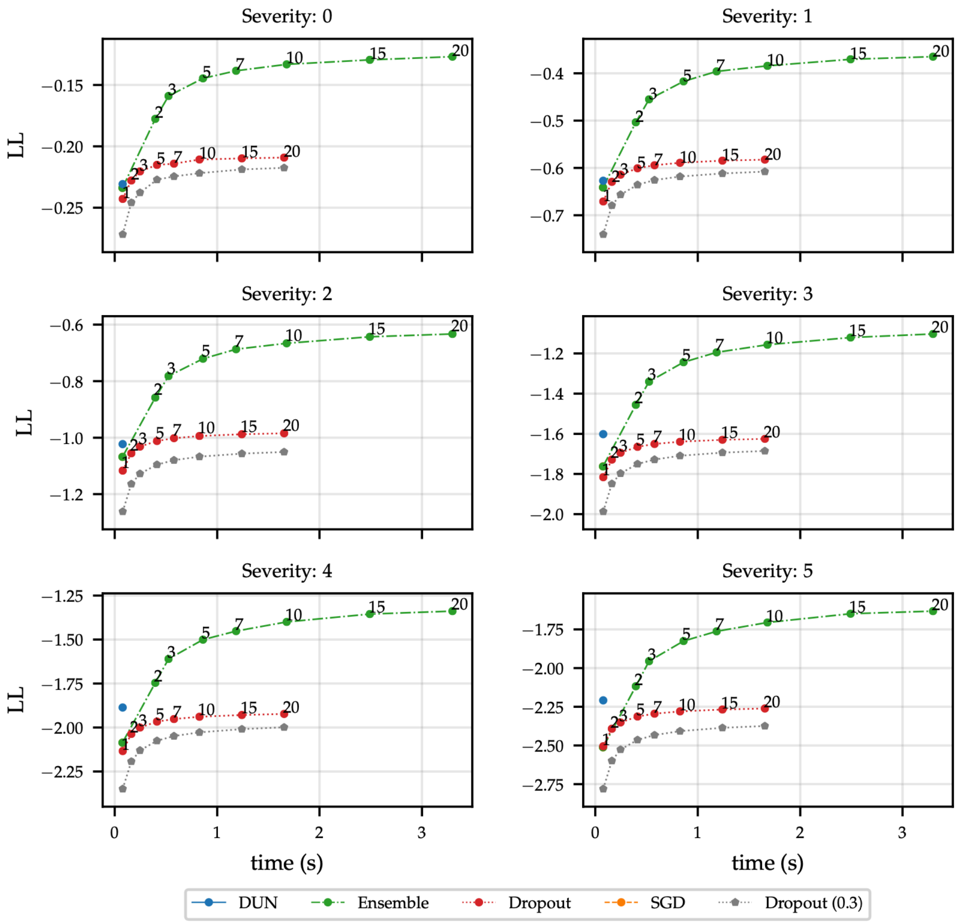


Image Classification (ResNet50)

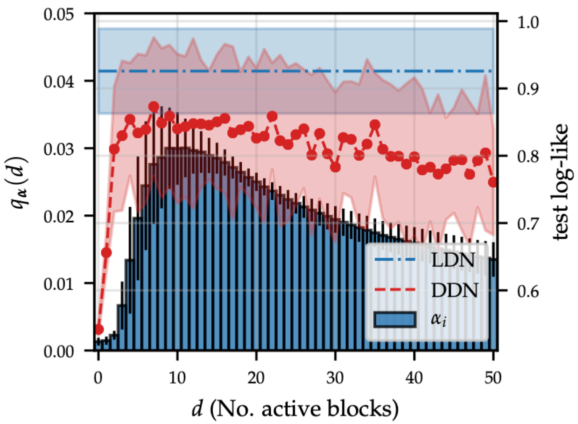


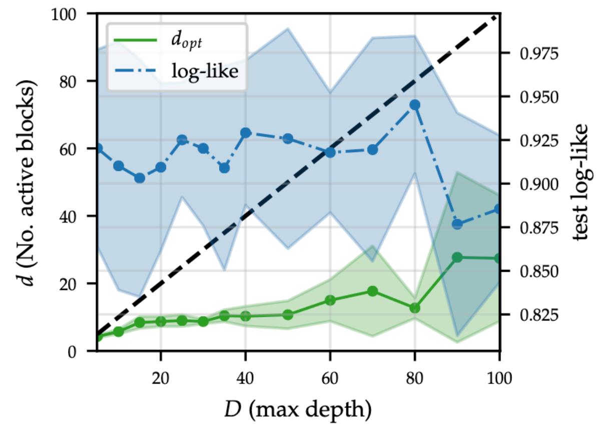
- DUNs provide best robustness vs compute time trade-off.

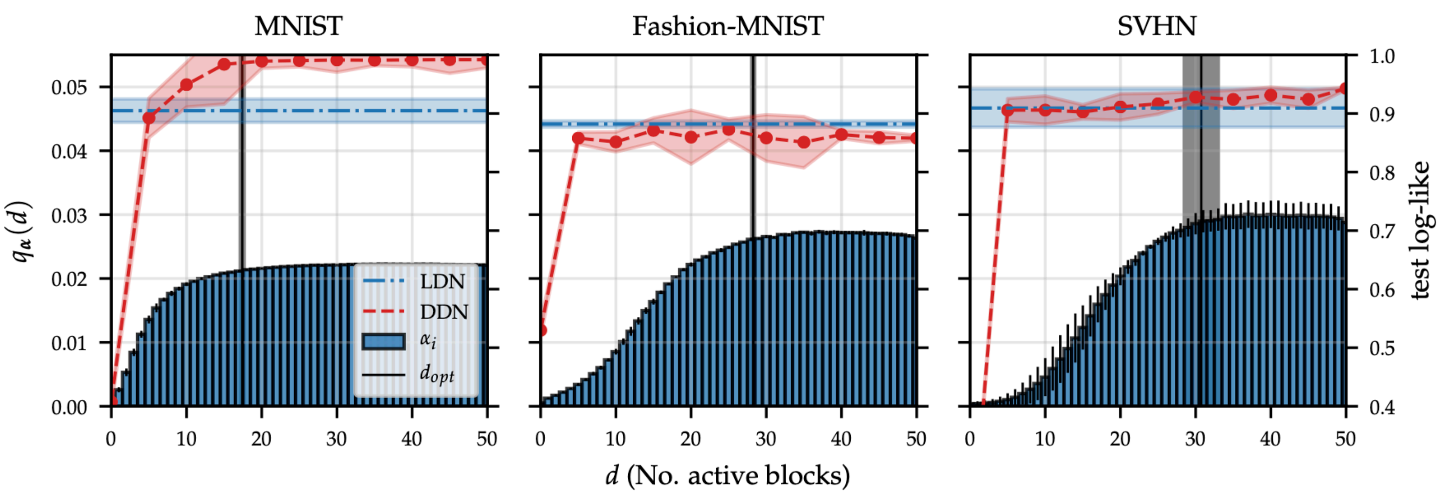
DUNs are Compute Efficient!



Architecture Search with DUNs

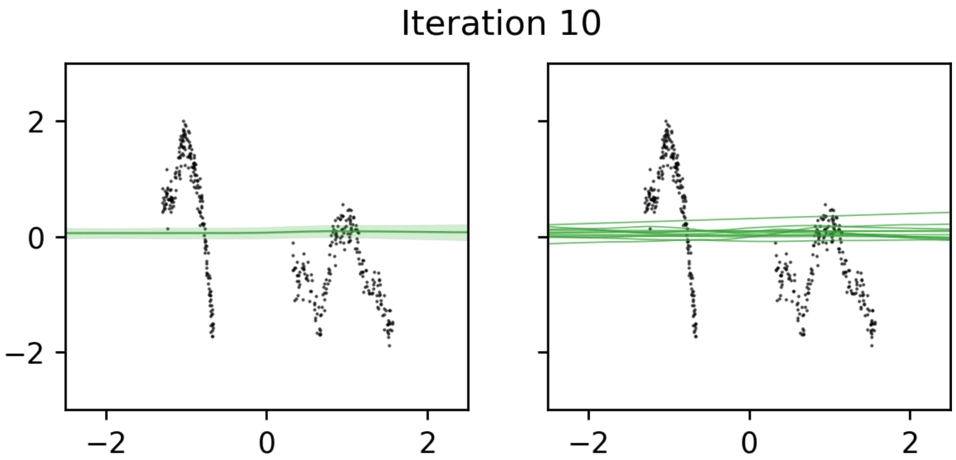






Summary

- Existing methods for estimating uncertainty in deep learning are computationally expensive.
- Depth Uncertainty Networks (DUNs), transform uncertainty over depth into predictive uncertainty in a single forward pass.
- DUNs provide the best robustness vs compute time trade-off in both classification and regression with modern architectures.



Code

github.com/cambridge-mlg/DUN