Enabling Uncertainty Estimation in Iterative Neural Networks

Nikita Durasov · Doruk Oner · Jonathan Donier · Hieu Le · Pascal Fua

EPFL NEURAL CONCEPT

ICML
International Conference On Machine Learning

Uncertainty Estimation

- Uncertainty estimation is well studied for classical ML, but not so well studied for deep learning.
- Assessing the reliability of neural networks remains a challenge, whereas several important applications rely on it (such as robotics, autonomous driving, and medical fields).

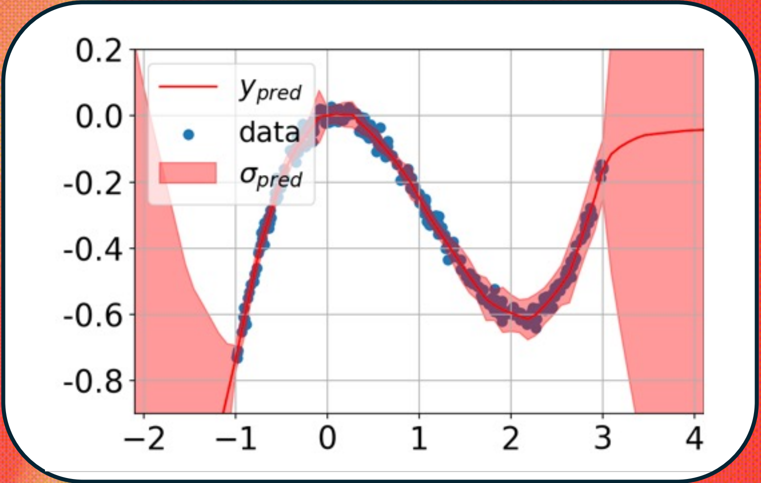


Uncertainty Estimation in DL

- From a technical perspective, uncertainty estimation for a model can be formalized as providing additional "confidence" in the predictions.

$$f_\theta(x) = (y, \sigma), \text{where } \sigma \text{ is prediction's confidence}$$

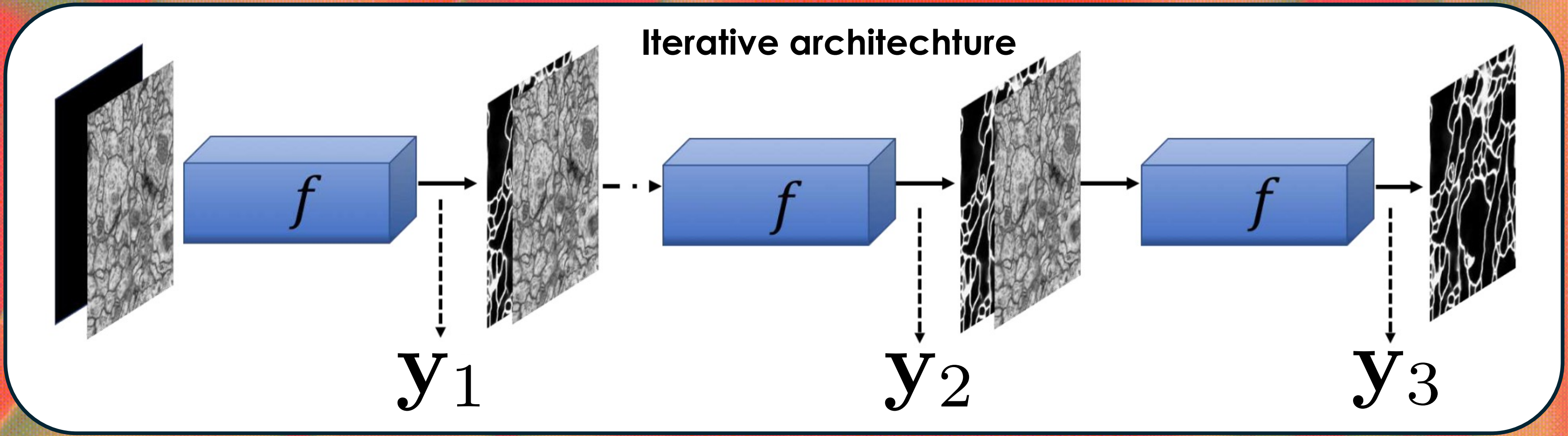
- Ensembling approaches are popular for deep learning uncertainty. Here, confidence is measured by the variance in the model's predictions for a given input.



Ensemble

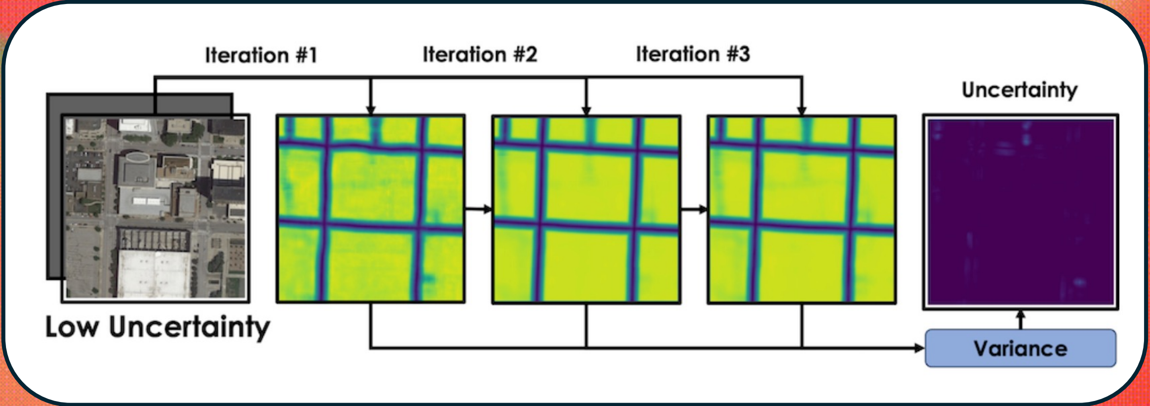
Iterative Models

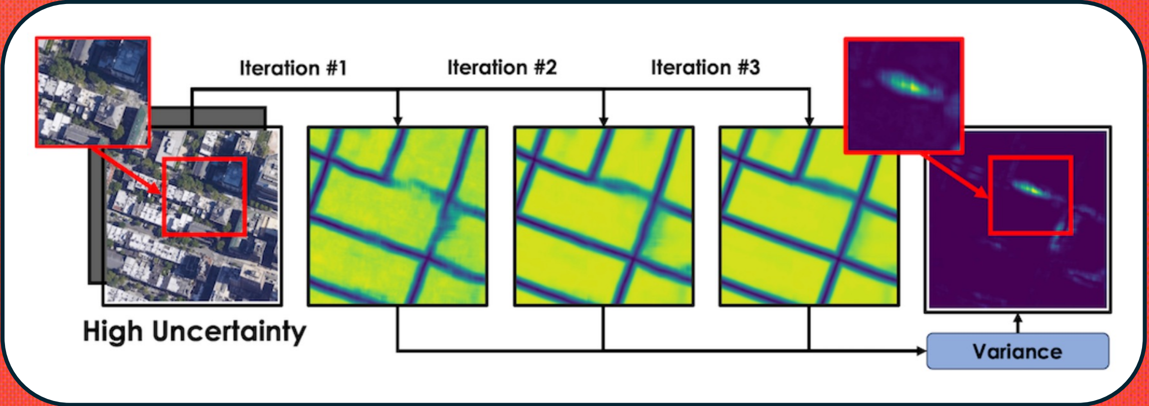
- An iterative model uses its own output as input in subsequent iterations to refine predictions, boosting performance and reducing the need for labeled data.
- This approach is used in semantic segmentation, pose estimation, depth estimation, multi-task learning, and natural language processing.



Iterative Uncertainty

- The iterative model refines its predictions over successive iterations. The speed of convergence indicates the difficulty of making a prediction: faster convergence means higher confidence, while slower convergence means greater uncertainty. **To measure uncertainty, we can compute the variance of successive iterations.**





Theoretical Justification

- Let us consider an iterative model that takes two variables as inputs, where the second input represents an output from the previous step.

$$\mathbf{y}_{i+1} = f_{\Theta}(\mathbf{x}, \mathbf{y}_i)$$

- By looking at this model as a denoising autoencoder, we can describe the iteration dynamics through the following formula from [1].

$$\mathbf{y}_{i+1} - \mathbf{y}_i = f_{\Theta}(\mathbf{x}, \mathbf{y}_i) - \mathbf{y}_i \propto \frac{\partial \log p(\mathbf{y}_i|\mathbf{x})}{\partial \mathbf{y}_i}$$

[1] Alain, Guillaume, and Yoshua Bengio. "What regularized auto-encoders learn from the data-generating distribution." JMLR 2014

Theoretical Justification (Example)

- Let us consider two different inputs, for which outputs have different distributions: with lower and higher variance

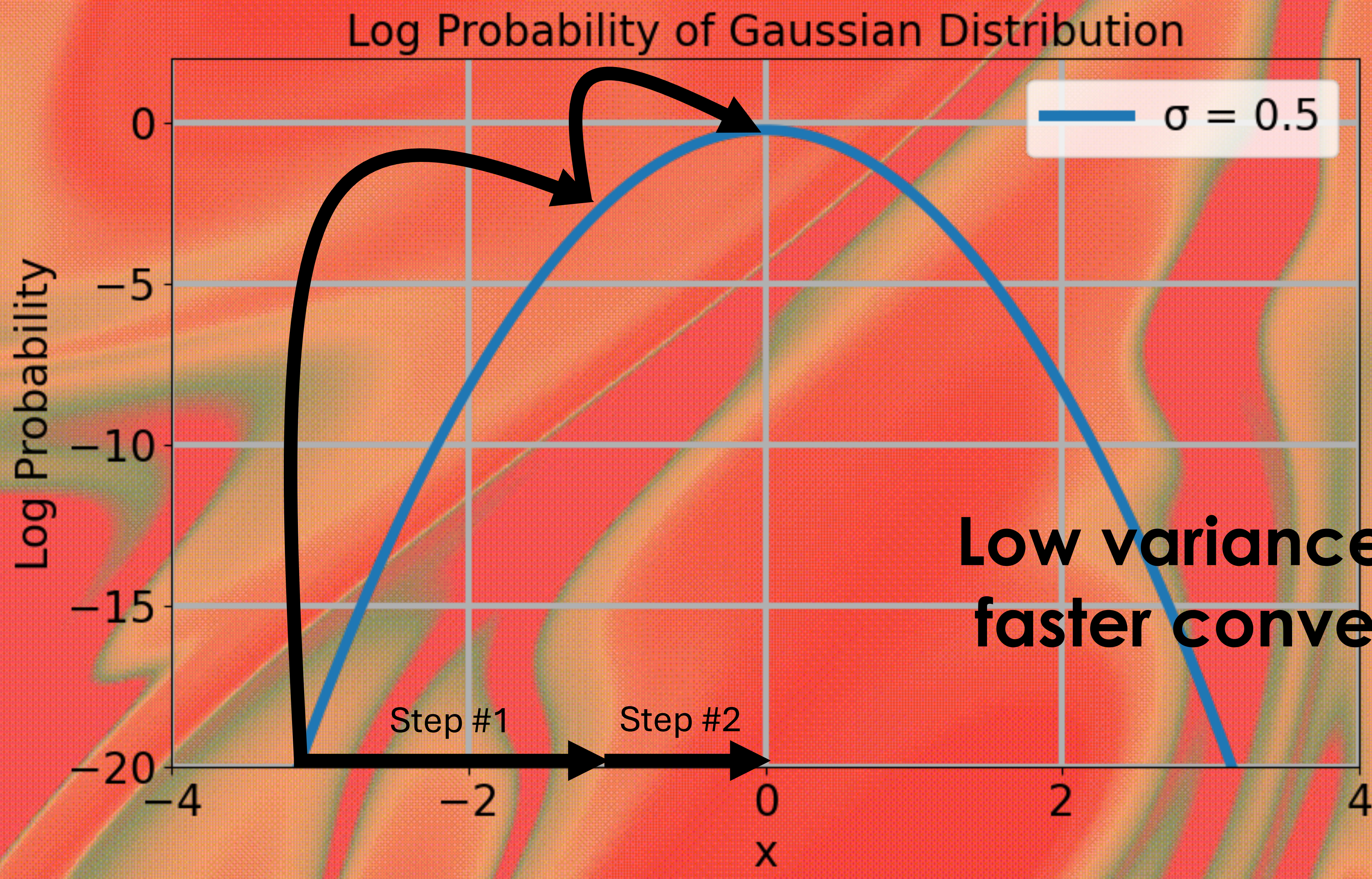
$$p(\mathbf{y}|\mathbf{x}_0) \sim \mathcal{N}(0, \sigma_0) \qquad p(\mathbf{y}|\mathbf{x}_1) \sim \mathcal{N}(0, \sigma_1) \qquad \sigma_0 < \sigma_1$$

- For these two cases, the first represents lower uncertainty and the second higher uncertainty. The iteration formulas are written as:

$$\mathbf{y}_{i+1} - \mathbf{y}_i = \begin{cases} f_\Theta(\mathbf{x_0}, \mathbf{y}_i) - \mathbf{y}_i \propto -\mathbf{y}_i/\sigma_0^2 \\ f_\Theta(\mathbf{x_1}, \mathbf{y}_i) - \mathbf{y}_i \propto -\mathbf{y}_i/\sigma_1^2 \end{cases}$$

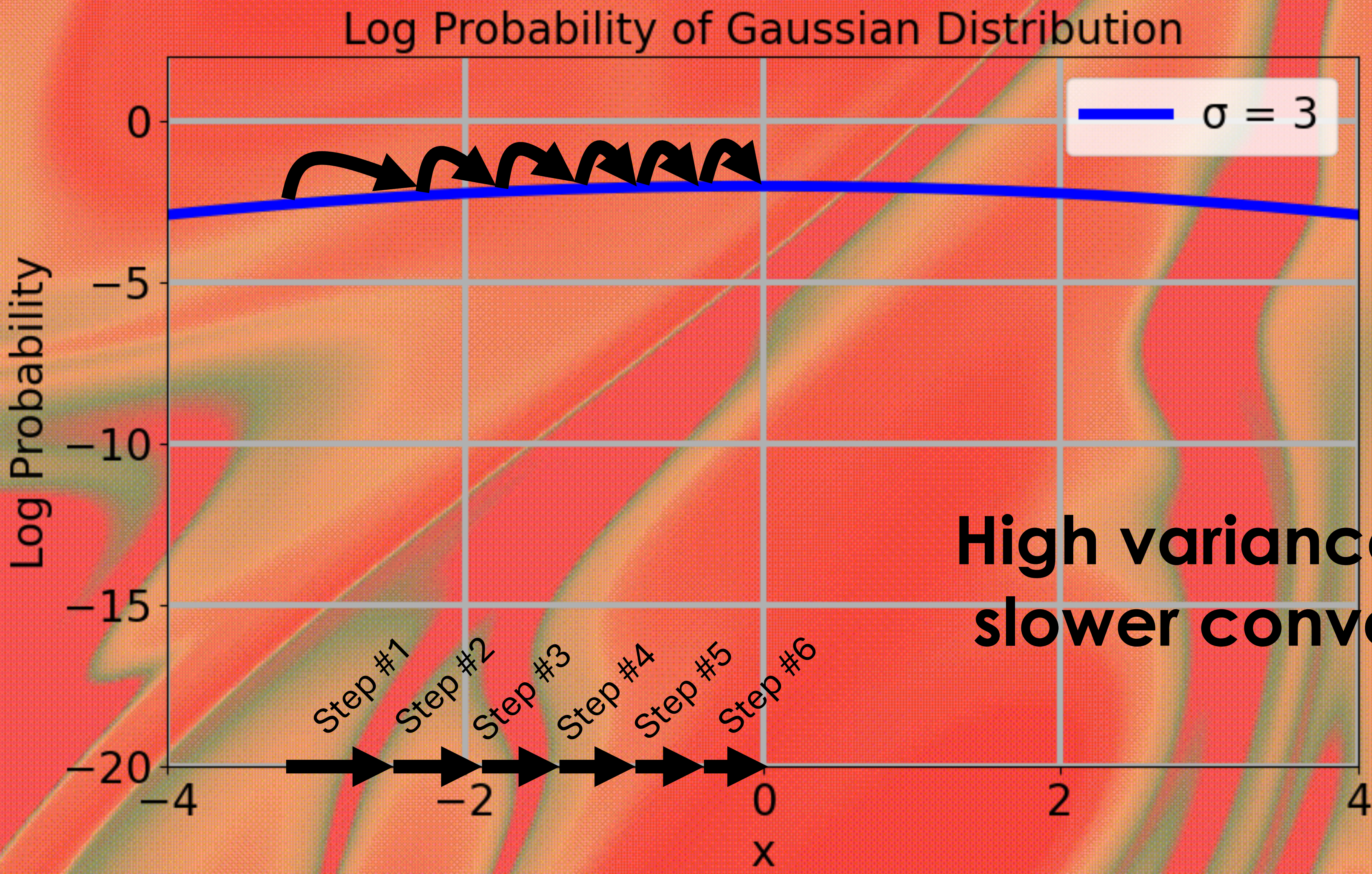
Higher variance, slower convergence

Low Uncertainty



Low variance means faster convergence

High Uncertainty



High variance means slower convergence

Results

- We demonstrate the effectiveness of our approach on two very different real-world tasks: road detection in aerial images and the estimation of aerodynamic properties of 2D and 3D shapes.
- Out method brings:

Better aleatoric unc.

	rAULC	Corr	Train	Inf	
MC-Dropout	30.18	59.72	**1x**	5x	RT
Deep Ensembles	**72.19**	**79.42**	5x	5x	RT
Ours	69.23	74.73	2.8x	**2.7x**	RT
MC-Dropout	19.56	32.50	**1x**	5x	MS
Deep Ensembles	78.65	76.39	5x	5x	MS
Ours	**79.27**	**87.46**	2.8x	**2.7x**	MS

Better accuracy

	Corr	Comp	Qual	F1	APLS	
U-Net	85.2	59.5	54.3	21.1	65.04	RT
MC-Dropout	87.1	58.2	54.1	20.4	58.78	RT
Deep Ensembles	**87.4**	66.7	60.8	22.1	68.81	RT
Ours	85.2	**77.8**	**68.6**	**24.5**	**77.21**	RT
U-Net	81.5	91.4	77.8	13.8	65.42	MS
MC-Dropout	81.6	**92.3**	78.2	13.6	59.65	MS
Deep Ensembles	83.6	90.4	78.7	14.1	67.53	MS
Ours	**92.3**	86.7	**81.1**	**15.4**	**78.04**	MS

Better epistemic unc.

	ROC-AUC	PR-AUC	
MC-Dropout	61.25	62.64	RT
Deep Ensembles	67.03	67.85	RT
Ours	**67.09**	**72.11**	RT

Key Takeaways

- Utilizes the convergence rate / variance of successive outputs to estimate uncertainty.
- Provides state-of-the-art uncertainty estimates with lower computational cost compared to methods like Deep Ensembles.
- Does not require modifications to the original iterative model.
- Achieves high accuracy and reliable uncertainty estimates without the need for too many forward passes or training multiple networks.

Check the Code!

Project page!

EPFL

NEURAL CONCEPT

ICML
International Conference
On Machine Learning