ICML
International Conference
On Machine Learning

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Variational Inference for Infinitely Deep Neural Networks

Achille Nazaret, David Blei

Why infinitely deep? To study the depth!

- Deep models are excellent on many tasks

Variational Inference for Infinitely Deep Neural Network is | a cool paper.
at ICML 2022.

- Yet finding the optimal depth for a task is not easy: depends on many aspects of the data
  - Dimensionality, quantity, complexity

- **Can the depth be learned from the data during inference?**

Yes!
With the Unbounded Depth Neural Network (UDN)

Mathematically:

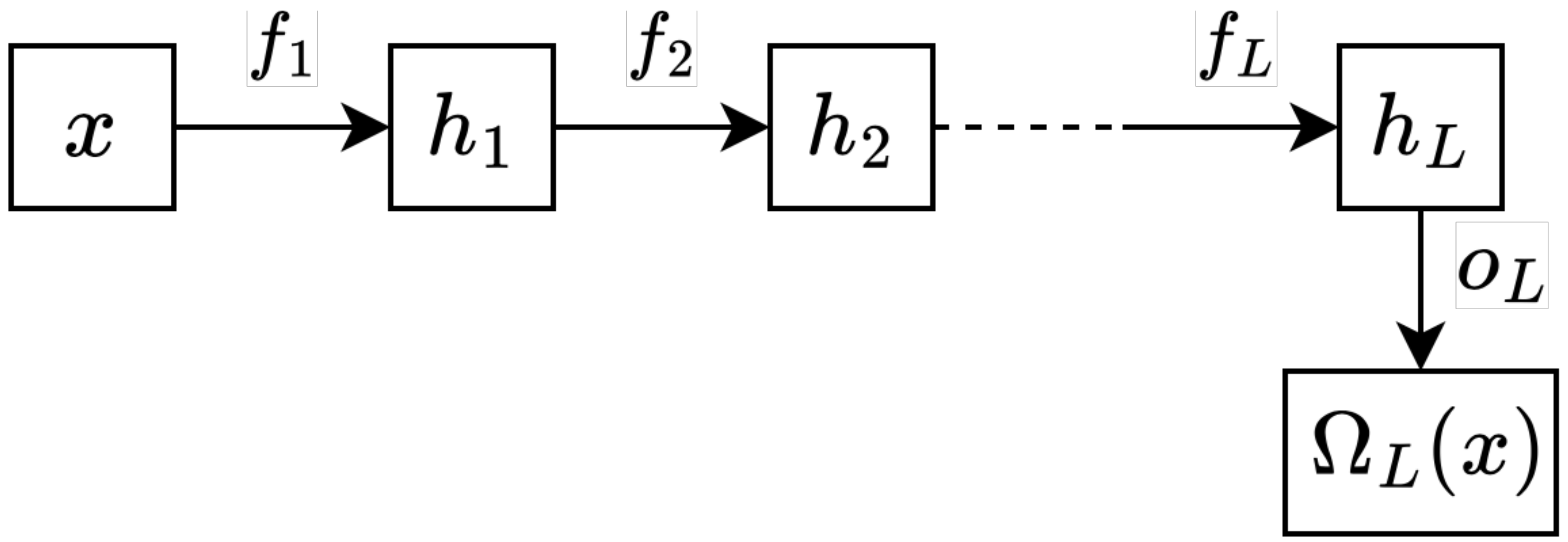
Infinite neural network trained with special variational inference

Concretely:

Neural network which adapts its depth to the data during training.

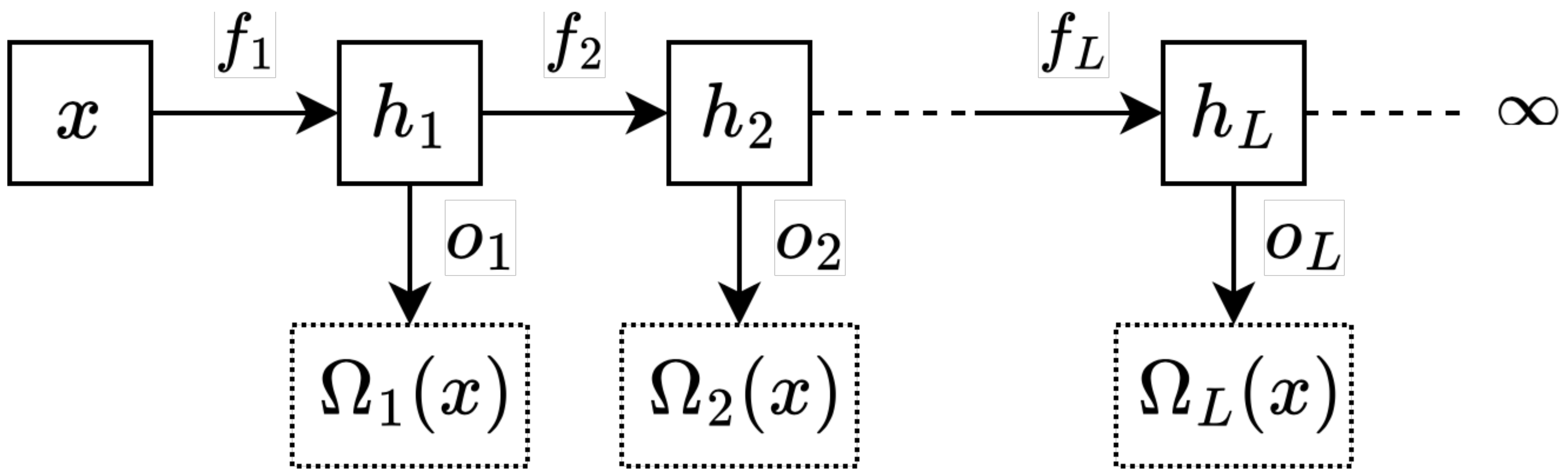
The Unbounded Depth Neural Network model

- Classic:



$$p(y \mid x) = p(y; \Omega_L(x))$$

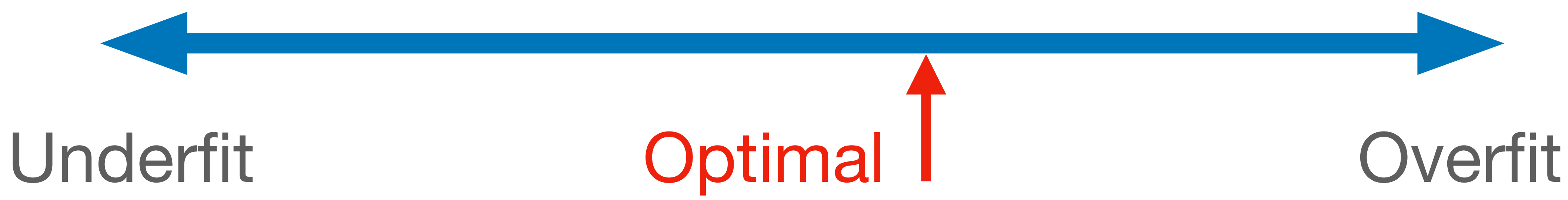
- Unbounded Depth:



Truncation ℓ

$$\Omega = \Omega_\ell$$

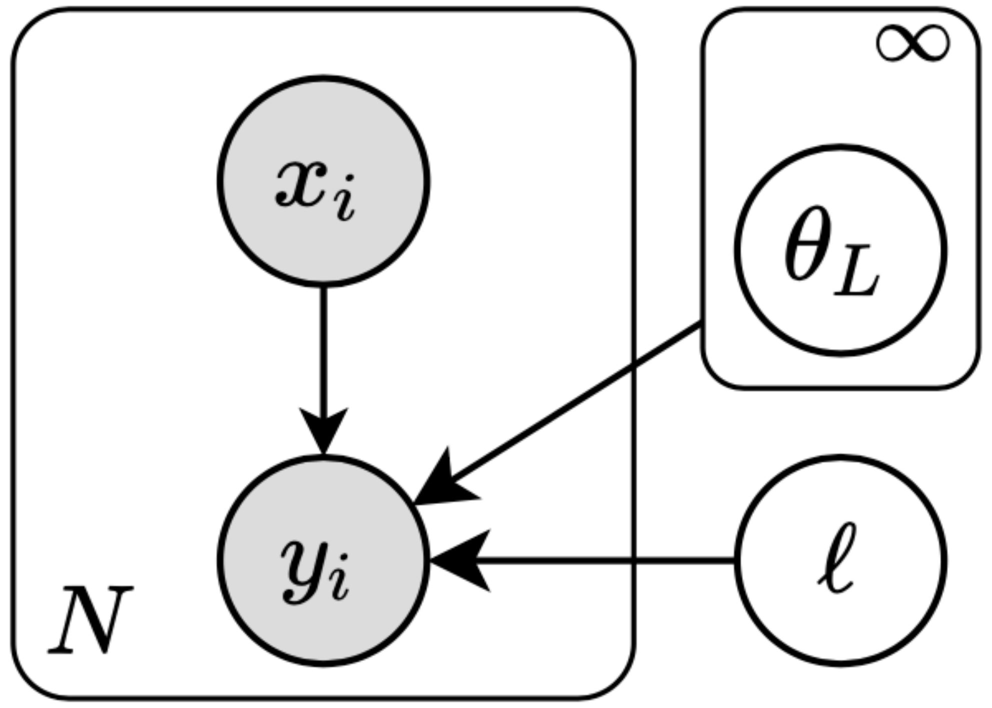
$$p(y \mid x) = p(y; \Omega(x))$$



The Unbounded Depth Neural Network model

- The generative model for the UDN:

$$\theta \sim \rho(\theta) \qquad \triangleright \text{ network weights}$$
$$\ell \sim \mu(\ell) \qquad \triangleright \text{ truncation}$$
$$y_i | x_i, \theta, \ell \sim p(y_i; \Omega_\ell(x_i; \theta)) \qquad \triangleright \text{ response}$$



- Can we fit such a model?

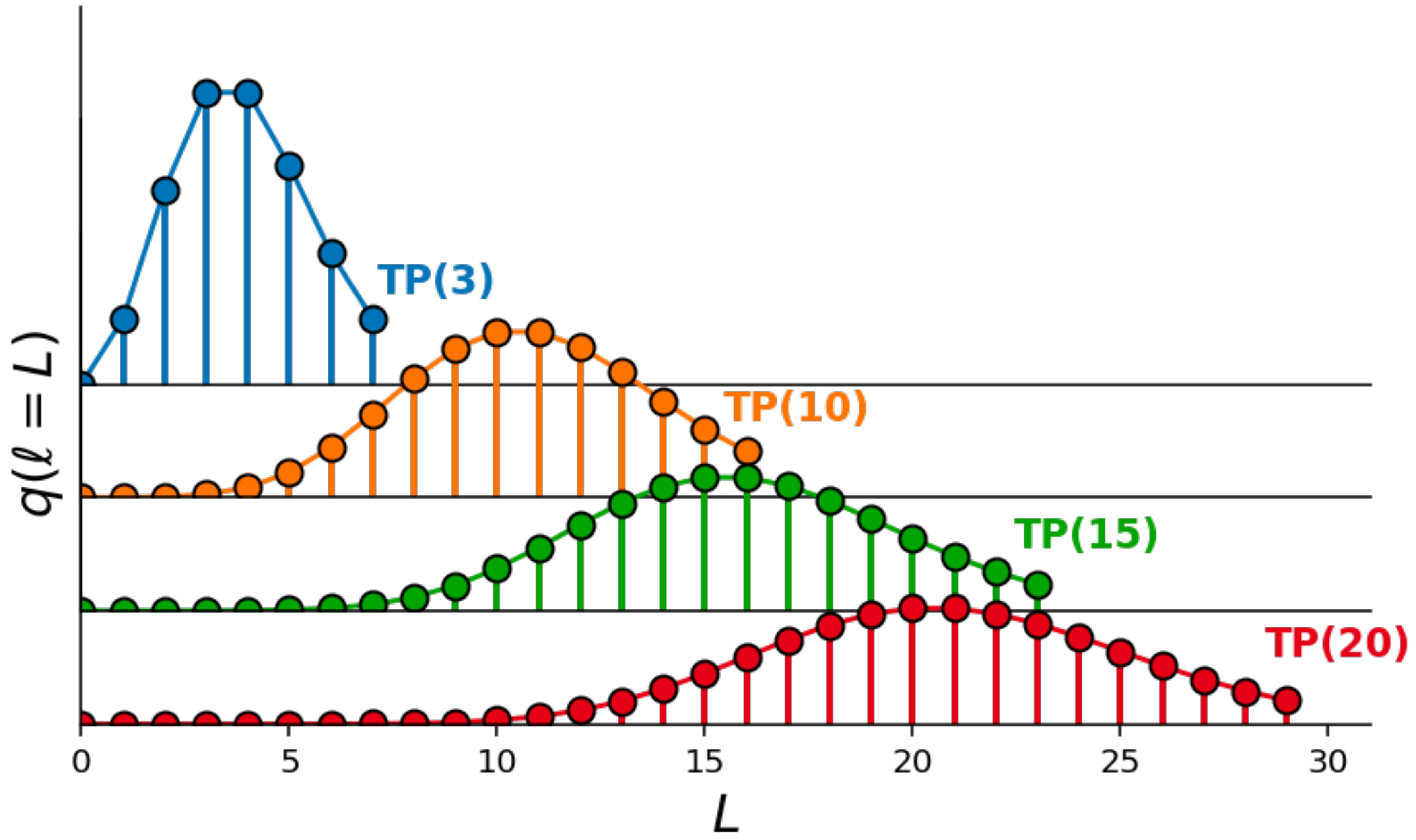
- Given observed data (x_i, y_i) —> **posterior** $p(\theta, \ell \,|\, (x_i, y_i))$
  Indicates which depth ℓ is more likely to have generated the data

- Learned with variational inference:

$$p(\theta, \ell \,|\, (x_i, y_i)) \approx q(\ell, \theta)$$

Variational inference $q(\ell, \theta) = q_\nu(\theta \mid \ell) q_\lambda(\ell)$

- How to deal with an infinite number of parameters?
- With a special variational family $Q = \{q_\lambda(\ell, \theta)\}$
- The q_λ have different number of variational parameters
- For the depth: Truncated Poisson (TP) family $q_\lambda = TP^{0.95}(\lambda)$

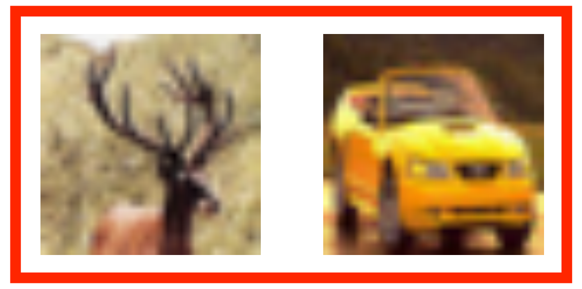


- support$(q_\lambda(\ell))$ is bounded
- support$(q_\lambda(\ell)) \xrightarrow[\lambda \to +\infty]{} \mathbb{N}^*$

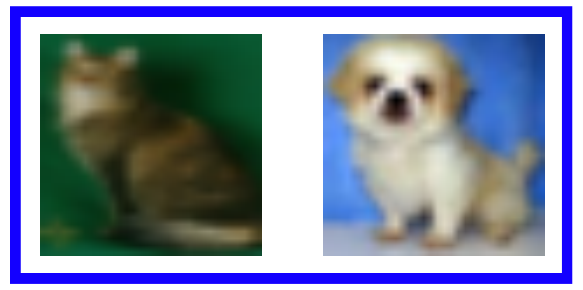
Experiments

On image classification (CIFAR10)

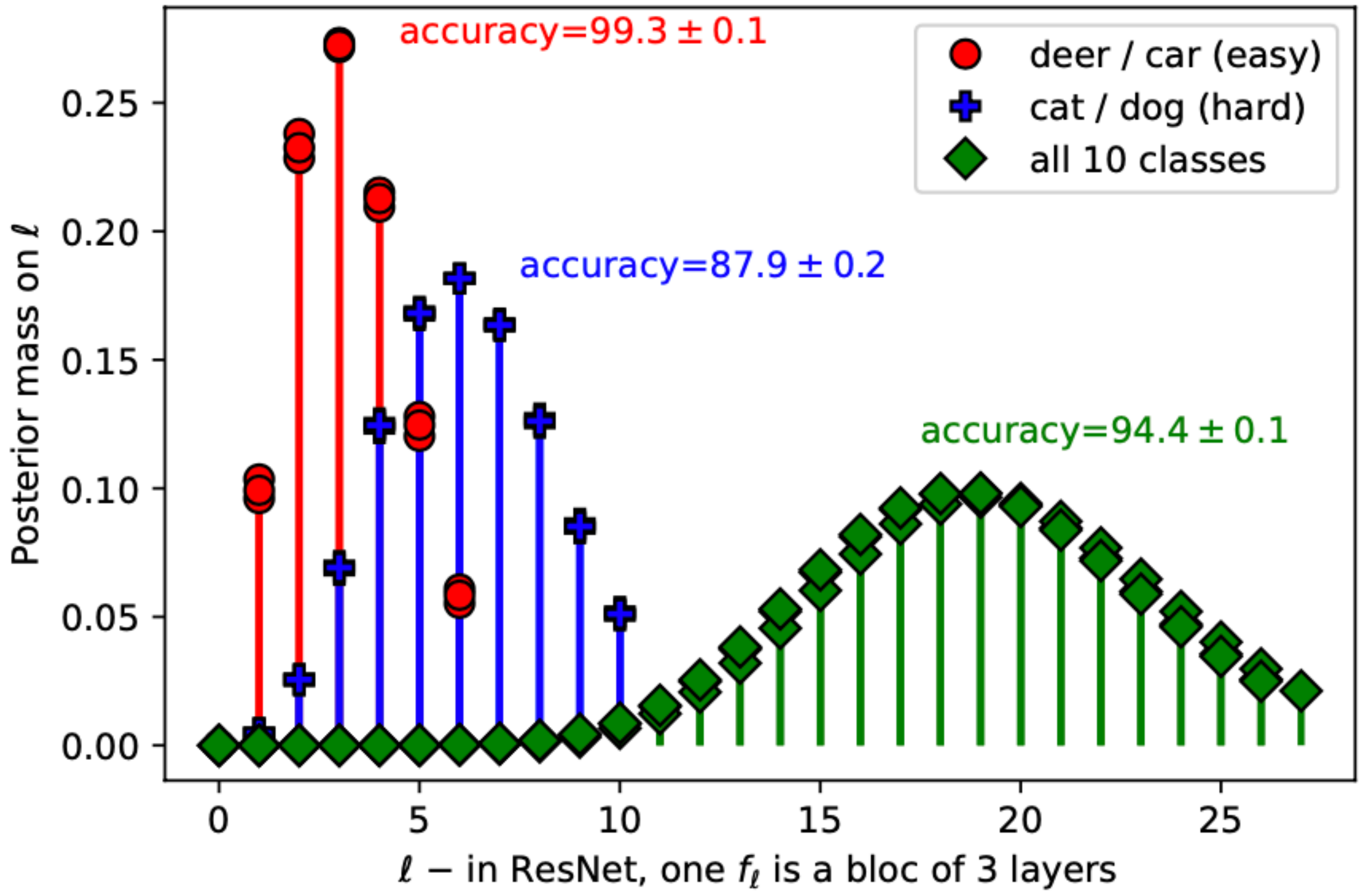
Easy task



Hard task



Model	Accuracy
ResNet-15, Ω_5	91.7$\pm$0.2
ResNet-18 (Bai et al., 2020)	92.9$\pm$0.2
ResNet-24, Ω_8	93.6$\pm$0.4
ResNet-30, Ω_{10}	94.0$\pm$0.2
ResNet-45, Ω_{15}	94.0$\pm$0.2
ResNet-60, Ω_{20}	93.9$\pm$0.1
ResNet-90, Ω_{30}	93.9$\pm$0.1
NODE (Dupont et al., 2019)	53.7$\pm$0.2
ANODE (Dupont et al., 2019)	60.6$\pm$0.4
MDEQ (Bai et al., 2020)	93.8$\pm$0.3
UDN with ResNet	**94.4$\pm$0.2**



- UDN outperforms corresponding finite models and other infinite neural networks formulations
- UDN adapts its depth for different dataset complexities, on the same task

The Unbounded Depth Neural Network

Conclusion

- A novel formulation of infinite neural networks
- A new variational inference method that can explore unbounded spaces

- In practice: Find the right depth in a single training, without extra computational complexity

Variational Inference for Infinitely Deep Neural Networks

Achille Nazaret, David Blei