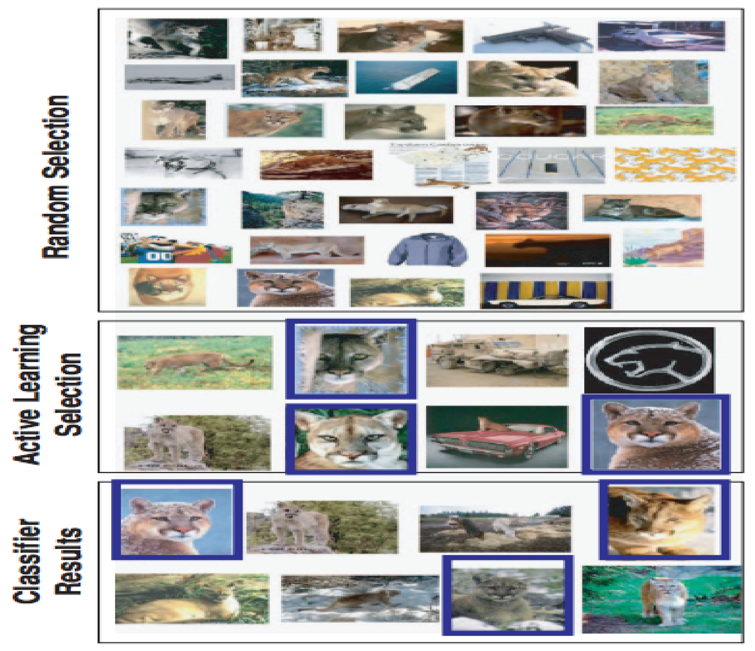
Active Learning with High Dimensional Inputs

Riashat Islam, Yarin Gal, Zoubin Ghahramani

University of Cambridge, UK

Model uncertainty in deep learning for information theoretic active learning

- **Bayesian approaches** to deep learning frameworks
- Bayesian Convolutional Neural Networks for **active learning** image data
- **Data-efficiency** in image recognition tasks



Using uncertainty for Bayesian Active Learning

1. **Pool-based Information Theoretic Active Learning**
   - Higher entropy implies classifier is uncertain about its class membership
2. **Representing model uncertainty in deep neural networks**
   - Dropout in DNNs as an approximation to the Gaussian Process (GP)
   - Dropout training in DNNs can be cast as approximate Bayesian inference
3. **Bayesian Convolutional Neural Networks (Bayesian CNNs)**
   - Dropout after every convolution layer at training
   - At test time - approximate the predictive posterior by averaging stochastic forward passes (Monte-Carlo Dropout)
   - Robust to over-fitting on small datasets, unlike traditional CNNs

Acquisition Functions

Dropout Bayesian Active Learning by Disagreement (Dropout BALD)

- Use expected information gain to quantify uncertainty in probability distribution

$$U(x) \approx H[\frac{1}{k}\sum_{i=1}^{k} P_i] - \frac{1}{k}\sum_{i=1}^{k} H[P_i] \qquad (1)$$

- Seek data point for which model has high uncertainty about the average output P_i obtained from each MC dropout probability estimates
- Seek data point for which model has low expected average uncertainty about its predictions

Dropout Maximum Entropy

- Seek data point which maximizes entropy

$$U(x) \approx H[\frac{1}{k}\sum_{i=1}^{k} P_i] \qquad (2)$$

Dropout Variation Ratio

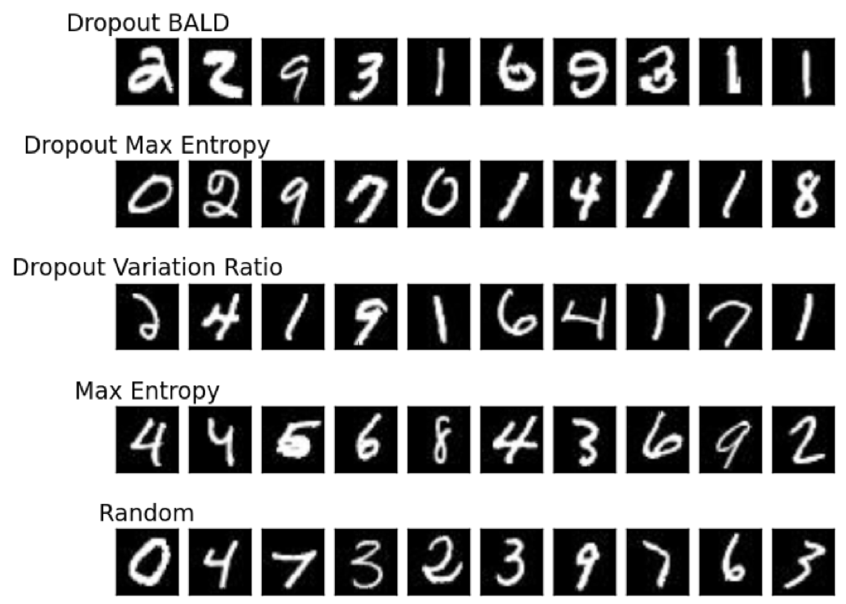
- Variation ratio as a measure of statistical dispersion

$$v = 1 - \frac{\text{Frequency of Mode}}{\text{Total Number of Dropout MC samples}} \qquad (3)$$

Dropout Bayes Segnet

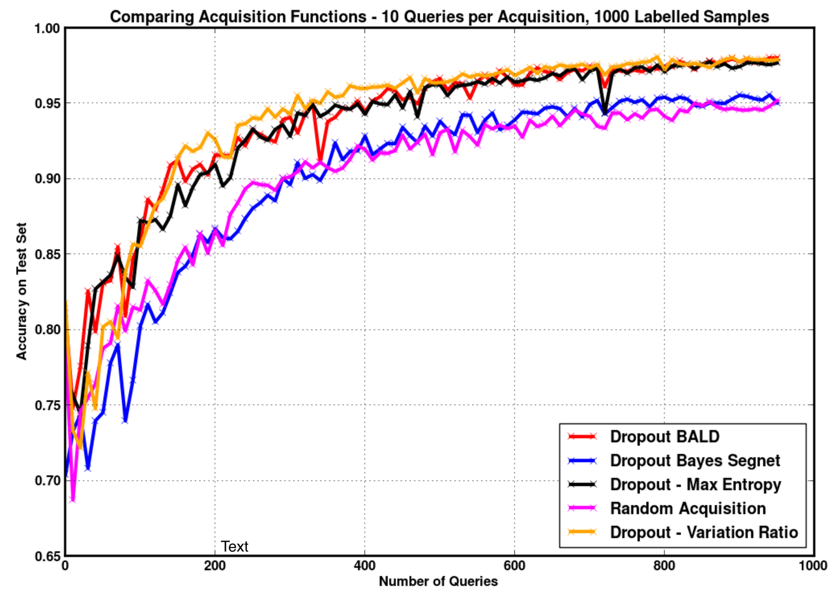
- Seek data point with maximum σ by computing σ of predicted probabilities
- But standard deviations of probabilities is NOT a good uncertainty measure!

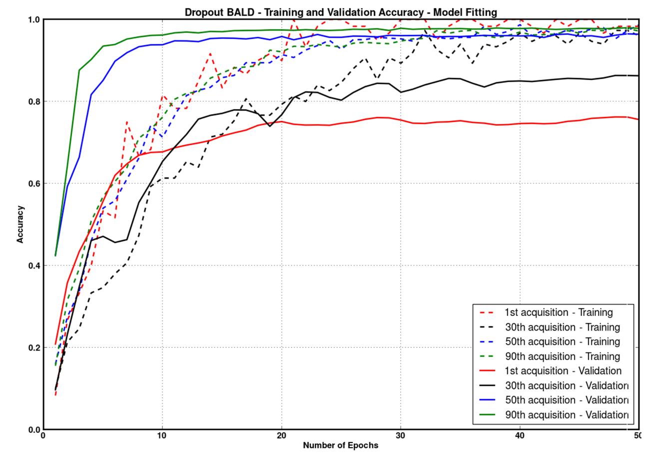
Pooled Images



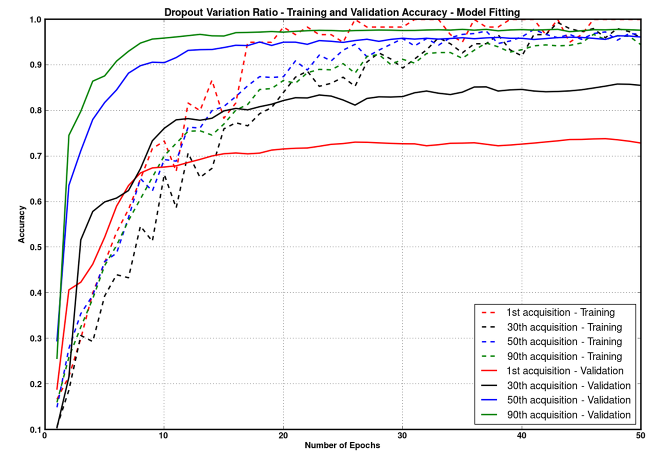
Performance of Active Learners

Test set accuracy with Number of Queries



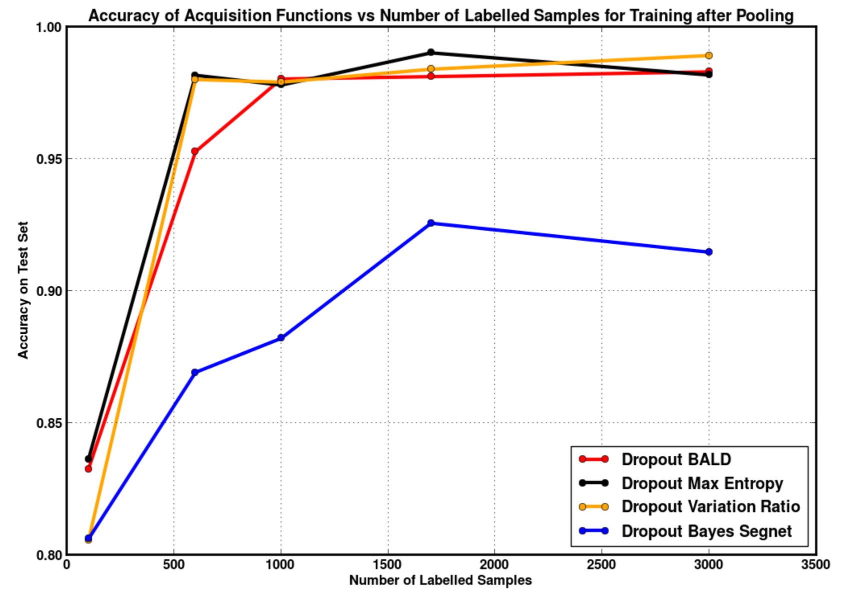


Dropout BALD Model Fitting



Dropout Variation Ratio Model Fitting

Test accuracy of **98.40%** with 1600 labelled samples and **98.92%** with 3000 labelled samples after pooling, on **10,000** test samples



Test Error Results on MNIST for 100 and 1000 labelled training samples

Test error % with number of used labels	100	1000
Semi-sup. Embedding (Weston et al., 2012)	16.86	5.73
MTC (Rifai et al., 2011)	12.03	3.64
Pseudo-label (Lee, 2013)	10.49	3.46
AtlasRBF (Pitelis et al.,2014)	8.10	3.68
DGN (Kingma et al., 2014)	3.33	2.40
Virtual Adversarial (Miyato et al., 2015)	2.12	1.32
SSL with Ladder Networks (Rasmus et al., 2015)	**1.04**	**0.84**
Active Learning with Bayesian CNN (Q=1)	in progress	in progress
Active Learning with Bayesian CNN (Q=5)	14.26*	in progress
Active Learning with Bayesian CNN (Q=10)	16.74*	**1.97**
Active Learning with Bayesian CNN (Q=100)	19.97*	**2.29**

* Starting with 20 training samples only, Bayesian ConvNet model prone to overfitting with 100 training labelled samples (very small dataset)

References

[1] Gal, Y, and Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning, 2015.

[2] Gal, Y, and Ghahramani, Z. Bayesian Convolutional Neural Networks with Bernoulli Approximate Variational Inference, 2015

[3] Houlsby, N, and Ghahramani, Z. Bayesian Active Learning for Classification and Preference Learning, 2011