Bayesian Inference for Deep Learning

Inference and modern trends for Bayesian Neural Networks: Final Considerations

Simone Rossi and Maurizio Filippone

Data Science Department, EURECOM (France)

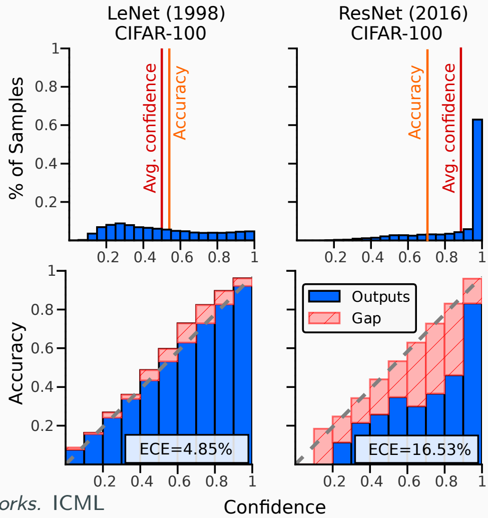
Quality of the uncertainty estimation

Deep neural networks are poorly calibrated

Expected calibration error:

$$\mathrm{ECE} = \sum_{b=1}^{\mathcal{B}} \frac{n_b}{N} \left|\mathrm{acc}(b) - \mathrm{conf}(b)\right|$$

- For regularized-loss training, deeper models are more accurate, more confident but less calibrated
- Uncertainty provided by the Bayesian models are well calibrated



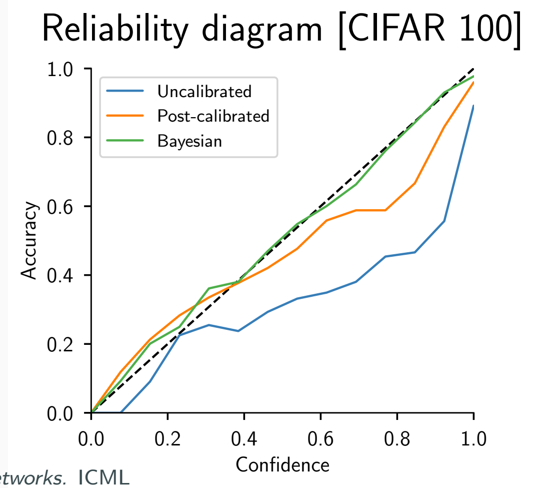
Guo et al. (2017). *On Calibration of Modern Neural Networks.* ICML

Deep neural networks are poorly calibrated

Expected calibration error:

$$\mathrm{ECE} = \sum_{b=1}^{\mathcal{B}} \frac{n_b}{N} \left|\mathrm{acc}(b) - \mathrm{conf}(b)\right|$$

- For regularized-loss training, deeper models are more accurate, more confident but less calibrated
- Uncertainty provided by the Bayesian models are well calibrated



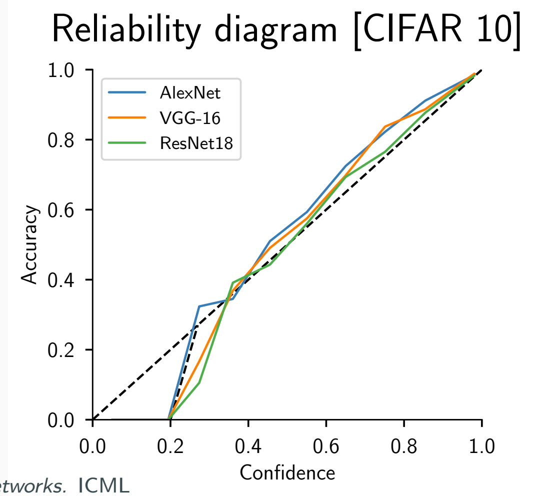
Guo et al. (2017). *On Calibration of Modern Neural Networks.* ICML

Deep neural networks are poorly calibrated

Expected calibration error:

$$\mathrm{ECE} = \sum_{b=1}^{\mathcal{B}} \frac{n_b}{N} \left|\mathrm{acc}(b) - \mathrm{conf}(b)\right|$$

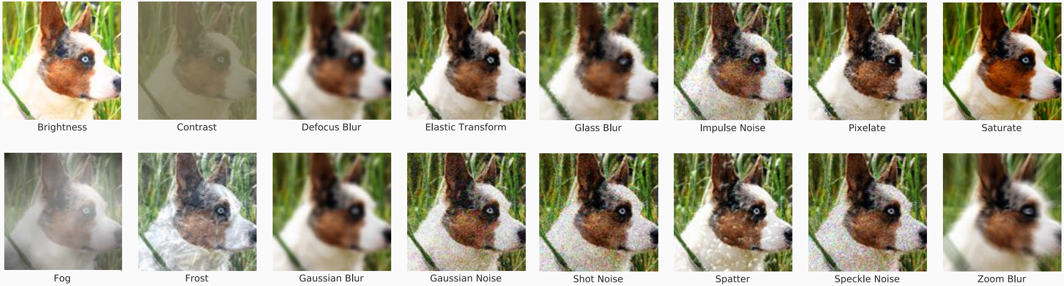
- For regularized-loss training, deeper models are more accurate, more confident but less calibrated
- Uncertainty provided by the Bayesian models are well calibrated



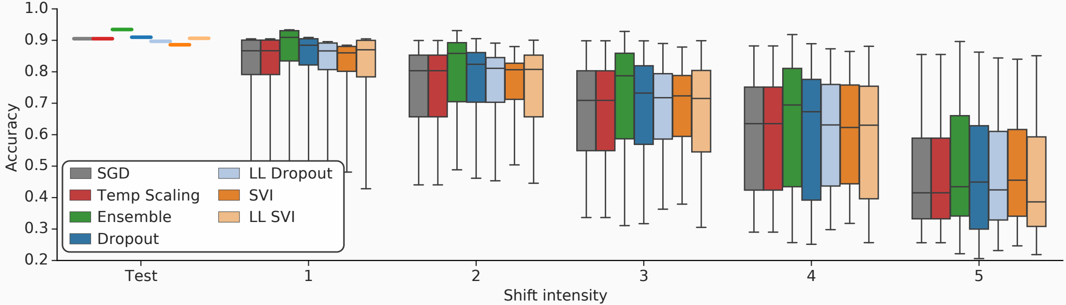
Guo et al. (2017). *On Calibration of Modern Neural Networks.* ICML

Things become more difficult to evaluate under dataset shift

Consider CIFAR10 and ImageNet 16 different random perturbation at 5 different intensity level.



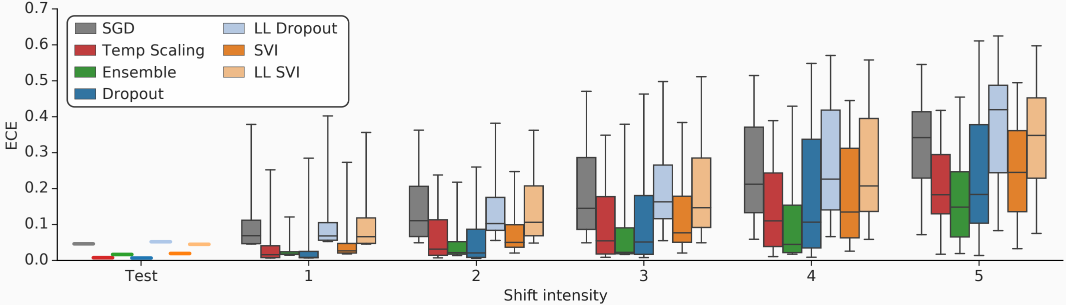
Out-of-Distribution analysis of probabilistic models on CIFAR10



- Accuracy decreases as similar rate . . .

Ovadia et al. (2019). *Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty Under Dataset Shift*. NeurIPS

Out-of-Distribution analysis of probabilistic models on CIFAR10

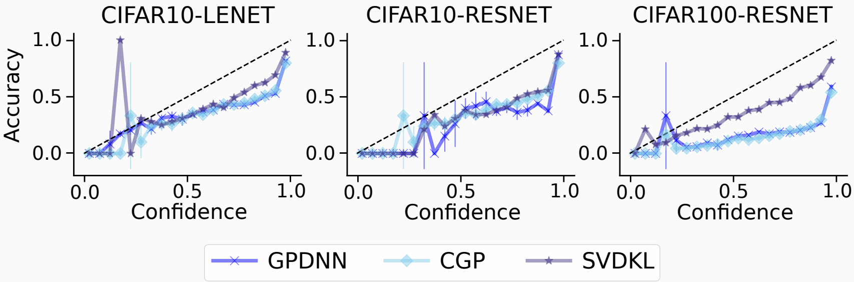


- Accuracy decreases as similar rate ...
- ... but even probabilistic models become over-confident (still better than point-estimates)

Ovadia et al. (2019). *Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty Under Dataset Shift*. NeurIPS

Being Bayesian on the last-layer only might not sufficient

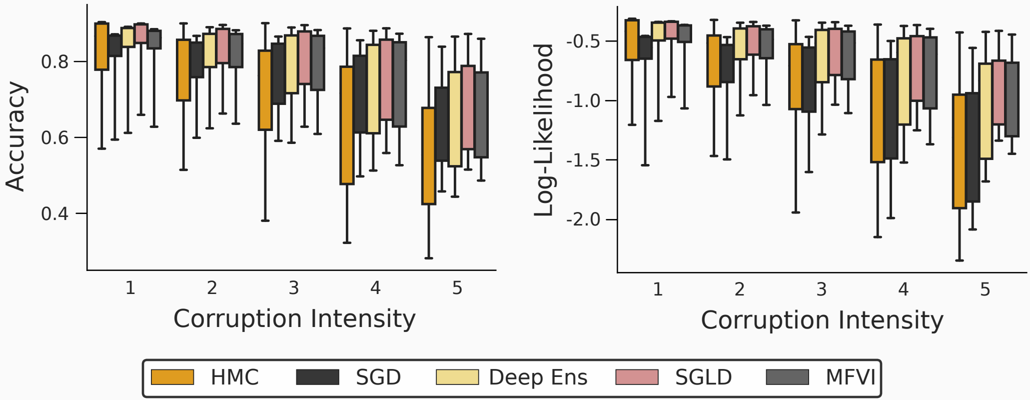
Combinations of Conv.Nets and Bayesian layers (like Gaussian processes) can still be over-confident.



Note on Laplace approximation: Kristiadi et al. (2020). *Being Bayesian, Even Just a Bit, Fixes Overconfidence in ReLU Networks*. ICML

Tran et al. (2019). *Calibrating Deep Convolutional Gaussian Processes*. AISTATS

Full posteriors are worse than anything else (including non-Bayesian)



Izmailov et al. (2021). *What Are Bayesian Neural Network Posteriors Really Like?* ICML

Conclusions

Bayesian Inference for Deep Learning

1. How can we work with intractable posterior?
2. How can we handle millions to billions of parameters? Scalability to big datasets?
3. What kind of priors should we use for these models? How can we do model selection?
4. Can we trust the uncertainty quantification of Bayesian inference?

Bayesian Inference for Deep Learning

1. How can we work with intractable posterior?
   - Approximation with variational inference, Laplace or ensembles
   - Sampling the true posterior with MCMC methods
2. How can we handle millions to billions of parameters? Scalability to big datasets?
3. What kind of priors should we use for these models? How can we do model selection?
4. Can we trust the uncertainty quantification of Bayesian inference?

Bayesian Inference for Deep Learning

1. How can we work with intractable posterior?
    - Approximation with variational inference, Laplace or ensembles
    - Sampling the true posterior with MCMC methods
2. How can we handle millions to billions of parameters? Scalability to big datasets?
    - Stochastic gradient methods and Monte Carlo integration
3. What kind of priors should we use for these models? How can we do model selection?
4. Can we trust the uncertainty quantification of Bayesian inference?

Bayesian Inference for Deep Learning

1. How can we work with intractable posterior?
   - Approximation with variational inference, Laplace or ensembles
   - Sampling the true posterior with MCMC methods
2. How can we handle millions to billions of parameters? Scalability to big datasets?
   - Stochastic gradient methods and Monte Carlo integration
3. What kind of priors should we use for these models? How can we do model selection?
   - Functional priors or empirical Bayes methods are possible solutions
4. Can we trust the uncertainty quantification of Bayesian inference?

Bayesian Inference for Deep Learning

1. How can we work with intractable posterior?
   - Approximation with variational inference, Laplace or ensembles
   - Sampling the true posterior with MCMC methods
2. How can we handle millions to billions of parameters? Scalability to big datasets?
   - Stochastic gradient methods and Monte Carlo integration
3. What kind of priors should we use for these models? How can we do model selection?
   - Functional priors or empirical Bayes methods are possible solutions
4. Can we trust the uncertainty quantification of Bayesian inference?
   - Yes (mostly), but the prior belief plays an important role

Bayesian Inference for Deep Learning

1. How can we work with intractable posterior?
   - Approximation with variational inference, Laplace or ensembles
   - Sampling the true posterior with MCMC methods
2. How can we handle millions to billions of parameters? Scalability to big datasets?
   - Stochastic gradient methods and Monte Carlo integration
3. What kind of priors should we use for these models? How can we do model selection?
   - Functional priors or empirical Bayes methods are possible solutions
4. Can we trust the uncertainty quantification of Bayesian inference?
   - Yes (mostly), but the prior belief plays an important role

---

"Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise."

J. Tukey (1962). *The Future of Data Analysis*. Ann. Math. Stat

Acknowledgments

A thanks to all the collaborators and colleagues for helping with the material of this tutorial.

Maurizio Filippone

Pietro Michiardi

Simone Rossi

Dimitrios Milios

Ba-Hien Tran

Giulio Franzese

Jonas Wacker