Improve Predictive Model Calibration for Safety-Critical Domains

Le Van Tuan Anh[*]
anh.lvt2416657@sis.hust.edu.vn

Nguyen Tran Minh Tri[*]
tri.ntm2400116@sis.hust.edu.vn

Le Xuan Nhat Khoi[*]
khoi.lxn2416711@sis.hust.edu.vn

Nguyen Duy Tung[*]
tung.nd2416762@sis.hust.edu.vn

1   Introduction

Deep neural networks have achieved strong performance on classification tasks. However, their predictive probabilities are often poorly calibrated, meaning the confidence scores do not reliably reflect true correctness likelihood. This limitation is particularly problematic in high-stakes domains, such as medical image analysis, where overconfident but incorrect predictions may lead to severe consequences.

Traditional deterministic neural networks typically provide point estimates without meaningful uncertainty information. As a result, they are unable to distinguish between confident and uncertain predictions, especially under data ambiguity, limited training samples, or domain shift. This limitation motivates the use of Bayesian and approximate Bayesian methods to capture predictive uncertainty and improve model reliability.

In this project, we focus on uncertainty-aware deep learning methods, specifically Monte Carlo Dropout (MC Dropout) and Bayesian Variational Inference (Bayesian VI), as scalable approximations to Bayesian neural networks. These methods allow uncertainty estimation without substantially changing the underlying model architecture or incurring prohibitive computational cost.

The goals of this project are threefold:

- Evaluate the ability of MC Dropout and Bayesian Variational Inference to capture predictive uncertainty and improve model calibration on standard image classification benchmarks, namely CIFAR-10 and CIFAR-100.
- Compare uncertainty-aware models against deterministic baselines using calibration-focused metrics (e.g., Expected Calibration Error) and reliability visualizations, in addition to standard accuracy-based performance.
- Apply the studied uncertainty estimation methods to the task of brain tumor classification from medical images, assessing whether improved calibration and uncertainty awareness translate to more reliable predictions in a safety-critical setting.

Overall, this project aims to demonstrate that incorporating uncertainty estimation into deep learning models can enhance not only predictive reliability but also practical applicability, particularly in domains where decision confidence is as important as raw accuracy.

2   Conclusion

---

[*]Group 9