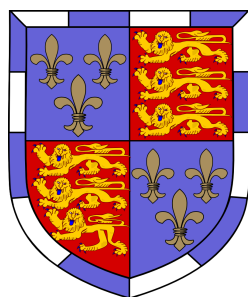




# Architecture Search for Bayesian Neural Networks



**Michael Hutchinson**

Supervisor: Dr. Richard E. Turner

Department of Engineering  
University of Cambridge

This report is submitted for the degree of  
*Master of Engineering*

# Architecture Search for Bayesian Neural Networks

Michael Hutchinson

Standard neural networks have revolutionised the machine learning field, improving performance by orders of magnitude over previous methods in a wide variety of applications, particularly in tasks with significant quantities of data. A key part of the success of these methods has come from designing bespoke architectures, which also work well with deep learning optimisation methods. As the size of neural networks has increased significantly, the task of designing these networks by hand has become increasingly more difficult and time consuming.

Automated methods of performing architecture search have therefore been developed, and this topic has seen a significant body of work in the last 3 years. None of this work however has been to be applied to Bayesian Neural Networks. These probabilistic counterparts hold several key advantages over regular neural networks. They provide well calibrated uncertainty estimates in their predictions, they are significantly more robust to tampering via adversarial samples, and a range of useful additions, such as continual learning, distributed learning and active learning can be applied readily without any modification. Additionally, there has been little systematic study into how the various hyper-parameters associated with Bayesian Neural Networks affect model performance.

This report has two main groups of contributions.

First, a detailed study of the effects of various hyper-parameters on the performance of MLP structure BNNs, and identification of systematic trends. We find that layer width and depth are both important factors. Networks that are too large can be detrimental to performance. The effect of the amount of data in the training set has on model optimisation is also investigated. It is shown that the balance between model size and data size is important in controlling the under-fitting or over-fitting of models. Finally the effects of pruning in VI BNNs investigated, and the manner in which the optimiser makes the trade off between reconstruction loss and prior fit terms. We find that in larger networks, many of the weights are pruned completely out the model, reducing the model to a minimum descriptive length, or in some cases further.

Second, is a new method for searching for optimal MLP structure networks in Bayesian Neural Networks, a task currently untackled. This method robustly outperforms the random search on the majority of tasks, excepting cases where the noise from randomness in training BNNs is too great. In most tasks it is as good as, or marginally better than simply performing Bayesian optimisation over the final performance of the BNNs.