Hyperparameter Optimization in Deep Learning-based Side-Channel Analysis: Three Directions

by Doreen Mulder | CSE3000 | Supervised by Stjepan Picek and Marina Krcek

The performance of a deep learning-based side-channel attack is greatly influenced by the chosen hyperparameters. Yet there is still a lack of information and documentation on which configurations are most appropriate for the side-channel analysis problem, and the reasoning behind this.
This literature study explored methods to systematize the deep learning techniques used in profiled side-channel analysis.

Learning Curve Extrapolation

- Hyperparameter configuration is tested before the training process is completed.
- After this a decision is made on whether to terminate training or not.
- This makes it possible to **quickly test and gauge the quality of many configurations.**

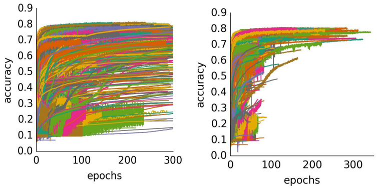

Figure 1: Learning curves of fully connected networks trained on CIFAR-10. The plots contain all learning curves from 10 runs with the SMAC hyperparameter optimizer, as well as 10 runs with the TPE hyperparameter optimizer, On the right, an early termination algorithm was used.[1]

- Future work in the side-channel analysis field should look into these techniques, as knowledge is still limited about which architecture configurations are most appropriate for the side-channel analysis problem.[2]

Bayesian Optimization

- Iterative algorithm using a probabilistic model and acquisition function to determine which configurations to evaluate.
- Maximizing the optimization performance is challenging, making it difficult for non-experts.

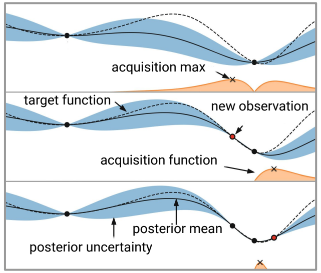

Figure 2: Bayesian optimization on a one-dimensional function.[3]

- Results show that it **performs well in tuning for image classification, speech recognition, and neural language modeling.**[3] These fields all deal with different datatypes. **It is worthwhile to find out if it performs well in side-channel analysis too.**

Transfer Learning

- Trained models are used as starting points for models with similar functions.
- Using these techniques in a side-channel analysis context is interesting, as there exists a number of trained machine learning models in this field already.

Ensemble Learning

- Ensemble learning is a type of transfer learning that combines multiple configurations into one.
- Ensemble models are expected to perform better classification compared to single models. This stems from the idea that in statistics, combined measurements can lead to more reliable estimations, because the influence of outliers and random fluctuations in the single measurements is reduced.
- Combining models can be done in multiple ways. **The bagging method has been tried in SCA[4], and these ensembles had a success rate that is at least as good as the best single model.**
- The use of ensembles relaxes the need to carefully select hyperparameters, but it cannot replace hyperparameter search methods.
- Future work may look into the boosting and stacking methods to create ensembles.

[1] Domhan, T., Springenberg, J. T., and Hutter, F. Speeding up automatic hyperparameter optimization of deep neural networks by extrapolation of learning curves.
[2] Tubbing, R. An analysis of deep learning based profiled side-channel attacks.
[3] Hutter, F., Kotthoff, L., and Vanschoren, J. *Automated Machine Learning.*
[4] Perin, G. Deep learning model generalization in side-channel analysis analysing class probabilities , metrics and ensembles.