Mediated Experts for Deep Convolutional Networks

Sebastian Agethen & Winston H. Hsu

Communications and Multimedia Lab
National Taiwan University

Introduction

Experts systems are known to improve the classification accuracy of neural networks and have been studied extensively.

However, previous work is flawed in two aspects: First, computational complexity is a multiple of that of a traditional model, as all experts are fully executed in parallel. Second, any addition of data or classes mandates retraining of all experts, a scenario that is frequently encountered in a world of ever-growing datasets. The later problem is of importance in *Incremental Learning*, and was previously approached in [2].

Main Objectives

1. Improve image classification with experts
2. Enable *Incremental Learning* for experts
3. Limit computational overhead of experts

Method

The work in [2] approaches the problem of Incremental Learning with the help of a branching model, identifying a *superclass* first and then selecting an expert for fine-grained classification accordingly. However, the branching error is large and cannot be recovered: Even for $N = 2$ superclasses, we measure 6.6% on ImageNet.

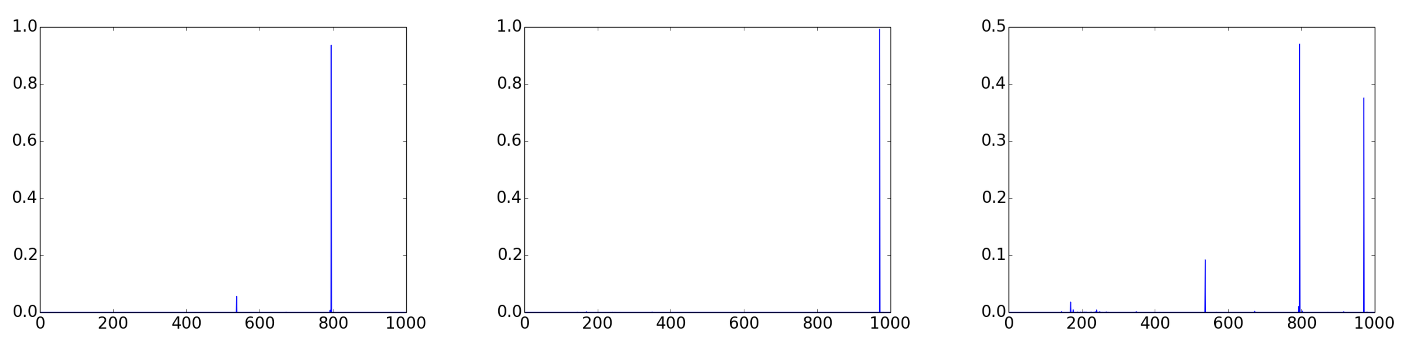

Figure 1: Softmax activations of experts A (left) and B (center) contradict each other. The mediator (right) is able to solve this conflict.

We present our proposed *Mediated Mixture of Experts* (MMoE) in Figure 2. The branching decision is moved to a confidence module, placed in higher convolutional layers, i.e., `Conv4`. All experts are executed in parallel up to this point – however, as lower layer features are generic (as shown in, e.g., [3]), we can share them between all experts. To handle cases where at least two experts remain confident, we propose an arbitrating mechanism, which we term *mediator*.

Mathematical Section

Confidence is computed in each confidence module IC_j: Each expert tracks his *score* $s^j = \mathbf{u}_j^{C_j}$, defined as the j-th component of his confidence activation vector. We then deem those experts i confident for which:

$$\max_{k \neq i}(s_k) - s_i \leq T, \qquad 1 \leq i, k \leq N, \quad T \in \mathbb{R}$$

Results

We evaluate on the ImageNet 1K classes dataset, and consider the case of $N = 2$ superclasses, see Table 1 and Figure 3.

Configuration	Accuracy	Early stop.
MMoE Slim 7L	**56.18%**	37.25%
MMoE Slim 8L	**52.96%**	33.5%
MMoE Default	**58.55%**	39.39%
Unmed. Shared	50.52%	39.23%
MMoE Shared	53.61%	39.23%
Baseline Slim 7L	53.44%	–
Baseline Slim 8L	49.48%	–
Baseline Default	55.84% (57.4%)	–

Table 1: Top-1 accuracy and early stopping probabilities of our MMoE system based on three different configurations of AlexNet[1] for $T = 4$.

- ConvNet: 3 configurations of AlexNet
- Performance depends on T
- Diminishing effects for $T \geq 6$
- Early stopping better than branching
- Clear impact of mediator: 3.11%

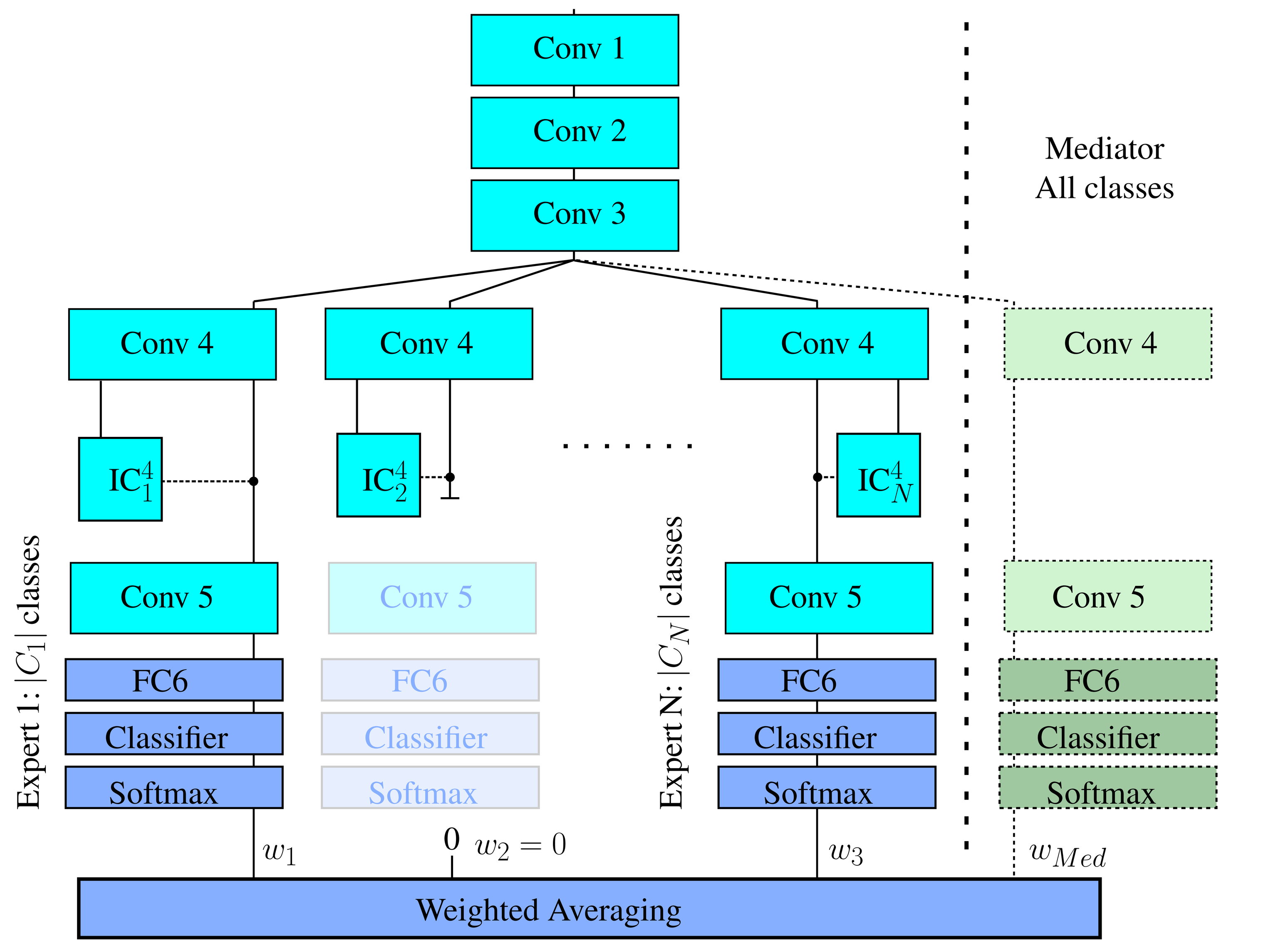

Figure 2: MMoE with N superclasses. Low-level features are computed once and then fed into all N experts. Each expert decides whether he is confident to give a correct prediction. In this example, the shading of expert two indicates that he was not confident and stopped early. Given that more than one expert remains, an arbitration process is necessary, and implemented in the form of a mediator.

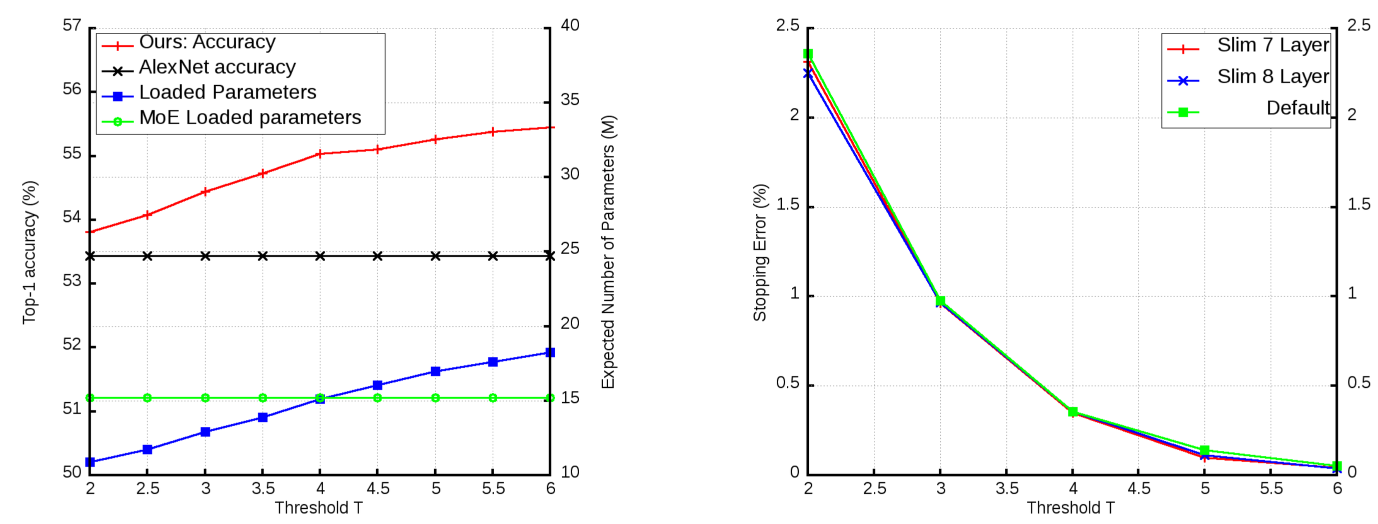

Figure 3: Left: Impact of threshold T on framework accuracy vs. a traditional singular model and number of loaded parameters vs. a traditional mixture of experts system. Right: Probability that the true expert is falsely stopped (branching error).

Conclusions

- Simplified learning from partitioned dataset
- Allows addition of new data or classes
- Conflicts between experts solved by *mediator*
- Reduced computational complexity due early stopping and feature sharing

Future Research

The concepts of the confidence module and the mediator can be further developed. Improving the mediator allows to reduce computation costs even further, while a better structure for the confidence module avoids further retraining.

References

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[2] Tianjun Xiao, Jiaxing Zhang, Kuiyuan Yang, Yuxin Peng, and Zheng Zhang. Error-driven incremental learning in deep convolutional neural network for large-scale image classification. In *Proceedings of the ACM International Conference on Multimedia*, MM '14, pages 177–186, New York, NY, USA, 2014. ACM.

[3] Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. *CoRR*, abs/1311.2901, 2013.

Acknowledgements

This is work is supported in part by MOST grant 104-2622-8-002-002 and MediaTek.

CMLab
Communications & Multimedia Lab