CS5260 Neural Network and Deep Learning II

Share Tasks among Experts: MTL-MOE Model for Fine-grained Image Classification

Wang Kai, Lao Guoyin, Hu Weilin, Ma Haozhe, School of Computing, National University of Singapore

INTRODUCTION

Fine-grained image classification is a challenging computer vision problem due to subtle differences in the overall appearance between various classes and large pose and appearance variations in the same class.

Figure 1. Example images from CUB 200-2011 dataset which exhibits large intra-class variations and low inter-class variations. Each column represents a unique class.

Mixture of experts (MOE) has been proposed to solve above problems. By grouping similar images and assigning them to corresponding expert neural networks, input space is partitioned so that an expert network can better learn the subtle differences between similar samples.

In this work, for each input image, we aim to assign it to experts **automatically** by applying **sparsely-gated mixture-of-experts** layer, so that each expert will focus on certain type of pose and appearance across all sub-categories, which will decrease intra-class variation and increase inter-class variation for each expert. Our method avoids both manual partitioning of input images and overfitting for each expert. Empirical evaluations show that this approach outperforms single expert model in terms of classification accuracy. We also proved that **transfer learning** in MOE does help increase learning speed.

Neural-based **multi-task learning (MTL)** has been successfully used in many real-world large-scale applications. In this work, based on the assumptions that region detection can facilitate fine-grained feature learning, we introduced **discriminative region localization** task to help us achieve better fine-grained classification accuracy. We conducted comprehensive experiments and show that our **MTL-MOE** model achieves better classification accuracy and converges faster during training than both single-expert model and single-task model. Over the dataset CUB 200-2011, our best result (80.44%) is better than most of the models published before 2018.

METHOD

Our MTL-MOE model has two tasks at hands: **fine-grained classification task** and **object bounding box coordinate prediction task**. In the bottom layers, we established an MOE model with two different gated network. In the high layers, we built two towers for two task separately to further aggregate the extracted features and map the features to their specific output. During training process, we updated the parameters using the integrated loss between complete-IoU loss in the regression task and the Cross-entropy loss in the classification task. In this way, we can achieve discriminative region localization and image feature representation in a mutually reinforced way.

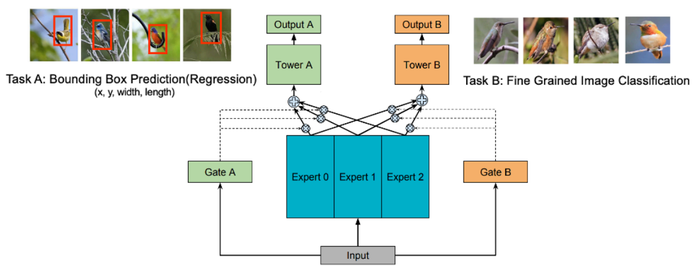

Figure 2. Overall structure of our MTL-MOE model

EXPERIMENTS

Experiment 1: MOE for Fine-Grained Classification

First, we check whether MOE model works better than a single model in the fine-grained classification task. The model structure is shown in Figure 3.

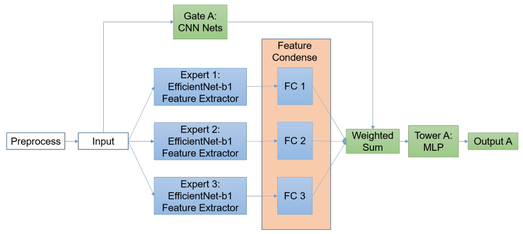

Figure 3. MOE model for experiment 1

In this experiment, we trained a single-expert model and a MOE model with 6 experts on 10 classes of the dataset. The results are shown in Figure 3.

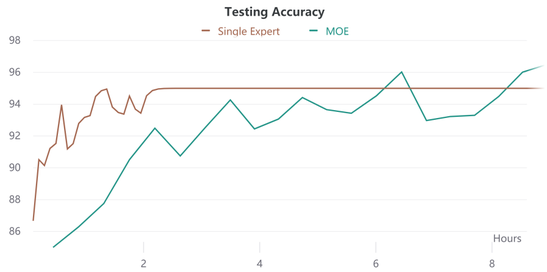

Figure 4. Testing accuracy of MOE and single-expert model

From the experiment result, we can conclude that MOE model indeed improves the classification accuracy, however, sacrifices the training speed.

Then, we adopted transfer learning strategy during training process, aiming to increase the learning speed for our MOE model. The experiment settings are shown in the table below and the experiment results are shown in Figure 5.

Transfer Learning EXP Settings	# of classes in Base model Training	# of classes in Fine-Tune Training	Training Strategy
MOE Direct Learning	N.A.	10	Directly train MOE model on last 10 classes
MOE Transfer Learning with freezing	190	10	Freeze parameters in experts and feature condense layer, only fine-tune other parts
MOE Transfer Learning w/o freezing	190	10	Update all parameters in MOE architecture

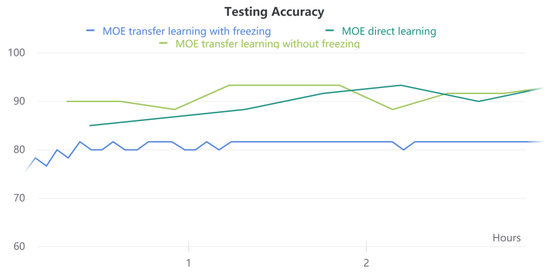

Figure 5. Testing Accuracy of Transfer Learning and Direct Learning

From the comparison between transfer learning without freezing and direct learning, we concluded that transfer learning achieved the same accuracy much faster than direct learning, which proves that transfer learning in MOE does help increase learning speed.

We can also conclude that during transfer learning, freezing the parameters in experts and feature condense layer will harm the model accuracy a lot, which probably because that the last 10 classes of images may have some features that cannot be extracted by the experts in the original MOE model. Only if the parameters in feature extractors get updated, the model can reach a higher accuracy.

EXPERIMENTS

Experiment 2: MTL-MOE Model for Fine-Grained Classification

Experiment 2 aims to research how multi-task learning strategy and the number of experts influence the model performance. The MTL-MOE model used in this experiment is shown in Figure 6.

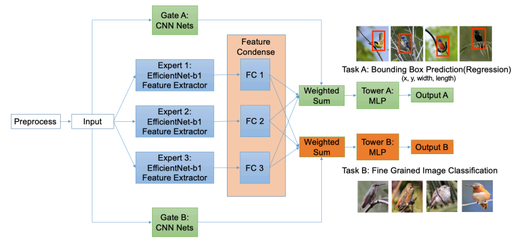

Figure 6. MTL-MOE model structure in experiment 2

We firstly compare the performance of single-task MOE model and multi-task MOE model, From the comparison of classification task accuracy in Figure 7, We know that the MTL strategy improves the overall classification accuracy by 1.4%, which proves our MTL-MOE effective.

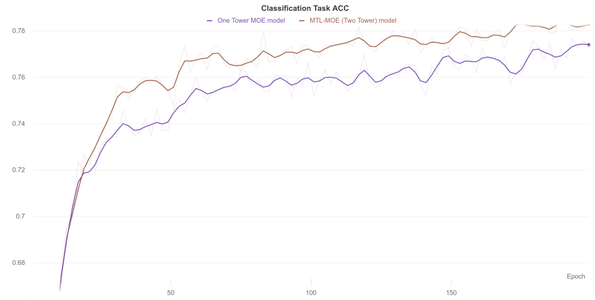

Figure 7. Classification Task Accuracy of MTL-MOE model and single-task MOE model

Then we try to figure out the influence of number of experts on the model performance, we tested MTL-MOE model with 1, 3 and 6 experts, the accuracy for the two tasks (regression and classification) are shown in the Table below.

	Batch Size	Learning Rate	# of Classes	Epoch	Best Regression Task CIoU Accuracy	Best Classification Task Accuracy
1-Expert MTL-MOE	16	1e-4	200	100	0.9337	0.7918
3-Experts MTL-MOE	16	1e-4	200	100	**0.9424**	**0.8044**
6-Experts MTL-MOE	8	1e-4	200	100	0.9391	0.8010

From the above table, 3 experts reaches the highest accuracy both in the regression task (94.43%) and the classification task (80.44%). Both the 3-experts model and 6-experts model have better performance than 1-expert model, which means that more experts are able to bring higher accuracy. However, there should be a trade-off between more representation ability and training difficulty.

CONCLUSION

In this project, we make efforts on using the Mixture of Experts model to solve fine-grained image classification tasks. The training strategy of transfer learning and the model design of multi-task learning are involved to improve our model.

Our MTL-MOE model, which combines the advantages of Multi-task-learning (MTL) and Mixture-of-Experts (MOE) and can decrease intra-class variation while enjoying the benefit of additional discriminative region localization tasks. The experiment results reflect that these approaches bring higher accuracy as well as making contributions to faster training in the fine-grained image classification task.