Abstract

Recently machine learning approaches based on deep learning have led to state-of-the-art results in many application domains: image recognition, semantic segmentation, super resolution, among many others. However, deep networks can be susceptible to overfitting. To alleviate overfitting, meta-algorithms, such as ensembling, are used to reduce the bias of the models. Mixture of experts is one such approach that consists of individual models that are trained to be “experts” on subsets of the data, and a gating network that provides weights to output a combination of the expert predictions. Mixture of experts models do not see wide use due to difficulty in training diverse experts and high computational requirements. This work presents modifications of the mixture of experts formulation that use domain knowledge to improve training, and incorporate parameter sharing among experts to reduce computational requirements.

First an application of mixture of experts to predict visual saliency is presented. Visual saliency is often represented as a spatial map that gives where humans are likely to look in a given image. A computational visual saliency model attempts to predict this map given an input image. Visual saliency models have recently begun to incorporate deep learning to achieve predictive capacity much greater than previous unsupervised methods. However, most existing models predict saliency without explicit knowledge of
This work proposes a model (MxSalNet) that incorporates global scene semantic information in addition to local information gathered by a convolutional neural network. The proposed model is formulated as a mixture of experts. Each expert network is trained to predict saliency for a set of closely related images. The final saliency map is computed as a weighted mixture of the expert networks’ output, with weights determined by a separate gating network. This gating network is guided by global scene information to predict weights. The expert networks and the gating network are trained simultaneously in an end-to-end manner. The proposed mixture formulation leads to improvement in performance over an otherwise identical non-mixture model that does not incorporate global scene information. Additionally, the proposed model achieves better performance than several other visual saliency models.

This work also presents an application of mixture of experts models for quality robust visual recognition. Experiments show that networks fine-tuned on distorted data greatly outperform the original networks when tested on distorted data. However, it is difficult for a single fine-tuned network to perform well across several distortion types. An ensemble method is proposed, MixQualNet, that is robust to different types of distortions. The “experts” in the proposed model are trained on a particular type of distortion. The output of the model is a weighted sum of the expert models, where the weights are determined by a separate gating network. The gating network is trained to predict weights for a particular distortion type and level. During testing, the network is blind to the distortion level and type, yet can still assign appropriate weights to the expert models. Finally, we propose MixQualNets with weight sharing. This work utilizes the TreeNet weight sharing architecture as well as introduce the Inverted TreeNet architecture. While both weight sharing architectures reduce memory requirements, the proposed Inverted TreeNet also achieves improved accuracy.