Spring 3-27-2019

Multi-path Sequential Fine-tuning of Pre-trained Models for Medical Image Classification Part I

David Zeng  
Dakota State University

James Boit  
Dakota State University

Rajesh Godasu  
Dakota State University

Seema Bhandari  
Dakota State University

Follow this and additional works at: https://scholar.dsu.edu/research-symposium

Recommended Citation
Zeng, David; Boit, James; Godasu, Rajesh; and Bhandari, Seema, "Multi-path Sequential Fine-tuning of Pre-trained Models for Medical Image Classification Part I" (2019). Annual Research Symposium. 22.  
https://scholar.dsu.edu/research-symposium/22

This Book is brought to you for free and open access by the University Publications at Beadle Scholar. It has been accepted for inclusion in Annual Research Symposium by an authorized administrator of Beadle Scholar. For more information, please contact repository@dsu.edu.
Abstract
While transfer learning techniques such as using a pre-trained model as a feature extractor and fine-tuning a few top layers of convolutional blocks to fit better with the classification tasks of interest have been successfully applied into solving medical image classification problems, less attention has been paid to effective utilizing the feature maps re-learned via fine-tuning at the end of conv layers. We propose a novel technical change in the architecture of deep neural network by sequentially feeding the feature maps via multiple paths of different layers before the transformed feature maps are flattened and feed into the dense layers for final classification. While a modest improvement in performance is achieved, a novel theoretic explanation could significantly contribute to this fast-advancing field.

Background
DenseNets: For each layer, the feature-maps of all preceding layers are used as inputs, and its own feature-maps are used as inputs into all subsequent layers. They require fewer parameters than an equivalent traditional CNN, as there is no need to learn redundant feature maps. Besides better parameter efficiency, one big advantage of DenseNets is their improved flow of information and gradients throughout the network, which makes them easy to train.

Inception: It is heavily engineered and used a lot of tricks to push performance. Salient parts in the image can have extremely large variation in size. Because of this huge variation in the location of the information, choosing the right kernel size for the convolution operation becomes tough. A larger kernel is preferred for information that is distributed more globally, and a smaller kernel is preferred for information that is distributed more locally. The solution: Having filters with multiple sizes operate on the same level. The network essentially would get a bit “wider” rather than “deeper”. Very deep networks are prone to overfitting. It also hard to pass gradient updates through the entire network.

Auxiliary Classifiers: To prevent the middle part of the network from dying out. They essentially applied softmax to the outputs of two of the inception modules, and computed an auxiliary loss over the same labels.

Literature Review
In a recent work on facial recognition, inception modules are integrated into dense blocks and in the transitional layer of DenseNets. And other experiments are done to learn the inter-dependencies of different feature maps, enhancing more useful feature maps while suppressing less important ones to enhance efficiency.

Methodology
A novel architecture is created with Dense Blocks embedded in the inception architecture to effectively learn multiscale features, characterizing medical images at various levels, to boost features and gradients flow, making deeper networks to characterize complex data distributions, and to learn the inter-dependencies of different feature maps. Three Versions of the DIM1 Network, depending on where the DIM is positioned. Multiple DIMs will be used in DIMx Networks.

Results
The validation accuracies (100 epochs, mini-batch: 32, sgd with momentum and adaptive learning rate: drop 4% every 8 epochs; model parallelism with 2 TITAN V GPUs, softmax operations are executed by CPU; models are implemented with Keras and TensorFlow-gpu as the backend.)

Conclusions
With DIM replacing auxiliary classifiers, DIM1 Networks appear to achieve higher performance with fewer parameters than the standard Inception Network, indicating higher efficiency than auxiliary classifiers to prevent early/middle layers of the network from dying out. Future work: strategically position the DIMs and an auxiliary classifier to shed more light on the behavior of the DIMx Networks and identify the optimal one. Theoretical explanations: encouraging interactions among learned feature maps is more efficient than changing loss functions. Then use it as the pre-trained model in medical image classification tasks such as diagnosis with cardiac and chest images.