ETSMI010-2017 An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification

Abstract

The availability of medical imaging data from clinical archives, research literature, and clinical manuals, coupled with recent advances in computer vision offer the opportunity for image-based diagnosis, teaching, and biomedical research. However, the content and semantics of an image can vary depending on its modality and as such the identification of image modality is an important preliminary step. The key challenge for automatically classifying the modality of a medical image is due to the visual characteristics of different modalities: some are visually distinct while others may have only subtle differences. This challenge is compounded by variations in the appearance of images based on the diseases depicted and a lack of sufficient training data for some modalities. In this paper, we introduce a new method for classifying medical images that uses an ensemble of different convolutional neural network (CNN) architectures. CNNs are a state-of-the-art image classification technique that learns the optimal image features for a given classification task. We hypothesise that different CNN architectures learn different levels of semantic image representation and thus an ensemble of CNNs will enable higher quality features to be extracted. Our method develops a new feature extractor by fine-tuning CNNs that have been initialised on a large dataset of natural images. The fine-tuning process leverages the generic image features from natural images that are fundamental for all images and optimises them for the variety of medical imaging modalities. These features are used to train numerous multi-class classifiers whose posterior probabilities are fused to predict the modalities of unseen images. Our experiments on the Image CLEF 2016 medical image public dataset (30 modalities; 6776 training images and 4166 test images) show that our ensemble of fine-tuned CNNs achieves a higher accuracy than established CNNs. Our ensemble also achieves a higher accuracy than methods in the literature evaluated on the same benchmark dataset and is only overtaken by those methods that source additional training data.