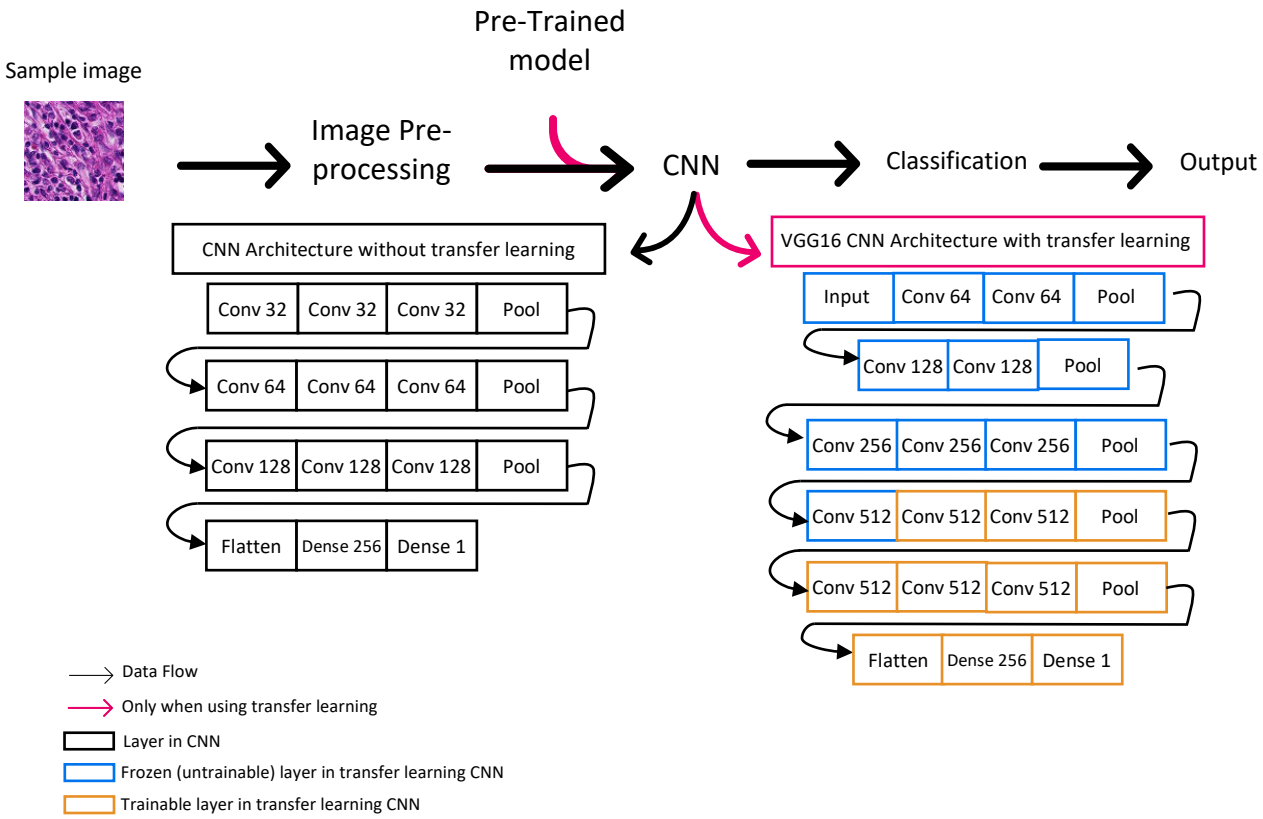
Ali Farahmand – CS department of Earlham College

Cancer Metastasis using CNN and Transfer Learning

Artificial Intelligence (AI) has been used extensively in the field of medicine. More recently, advanced machine learning algorithms have become a big part of oncology as they assist with detection and diagnosis of cancer. Convolutional Neural Networks (CNN) are common in image analysis and they offer great power for detection, diagnosis and staging of cancerous regions in radiology images. Convolutional Neural Networks get more accurate results, and more importantly, need less training data with transfer learning, which is the practice of using pre-trained models and **fine-tuning** them for specific problems. This project utilizes transfer learning along with CNNs for staging cancer diagnoses. Randomly initialized CNNs are compared with CNNs that used transfer learning to determine the extent of improvement that transfer learning can offer with cancer staging and metastasis detection. Additionally, the model utilizing transfer learning was trained with a **smaller** subset of the **dataset** to determine if using transfer learning reduced the need for a large dataset to get improved results.



Results

The models were saved after training and used on a test dataset. This test dataset is 8000 images of each class that were not used for training and validation steps. The saved models were used for evaluating the models on their performance outside of the training process. Table 5 represents the results of this evaluation. Based on the results, using transfer learning not only improved the final results by roughly three percent it also had improved results when only half of the dataset was used for training.

Model	Evaluation results
Basic CNN	89.68%
VGG CNN using transfer learning	92.12%
VGG CNN using transfer learning with half of the dataset	91.84%

The process of acquiring images and pre-processing the data is no different than other cancer detection CNNs, I will be using the Breast Histopathologic Cancer Detection dataset from. This dataset includes over 200,000 images of breast pathology scans and it includes both images of metastatic cancer which was spread to other tissue and benign tumors that have not spread to other places. Determining if cancer has spread to other regions is a crucial part of **staging** the cancer in the National Cancer Institution's TNM staging system.

The testing dataset included 16000 images which were 8000 images of each class. This **testing** dataset was not used in the **training** and **validation** steps, meaning the dataset was never seen before by the models. The benefit of using new images for evaluation is two-fold: first, it shows that the training **didn't overfit** the model to only the training and validation datasets and second, it makes sure that the model is extracting features and classifying the images based on the **correct features** and not a coincidental set of features in the training and validation datasets.

Special Thanks to: Charles Peck and David Barbella

For citations and more details visit: https://portfolios.cs.earlham.edu/index.php/author/afarah18/