Andrew Watson

Fourth Research/Programming Assignment

Abstract

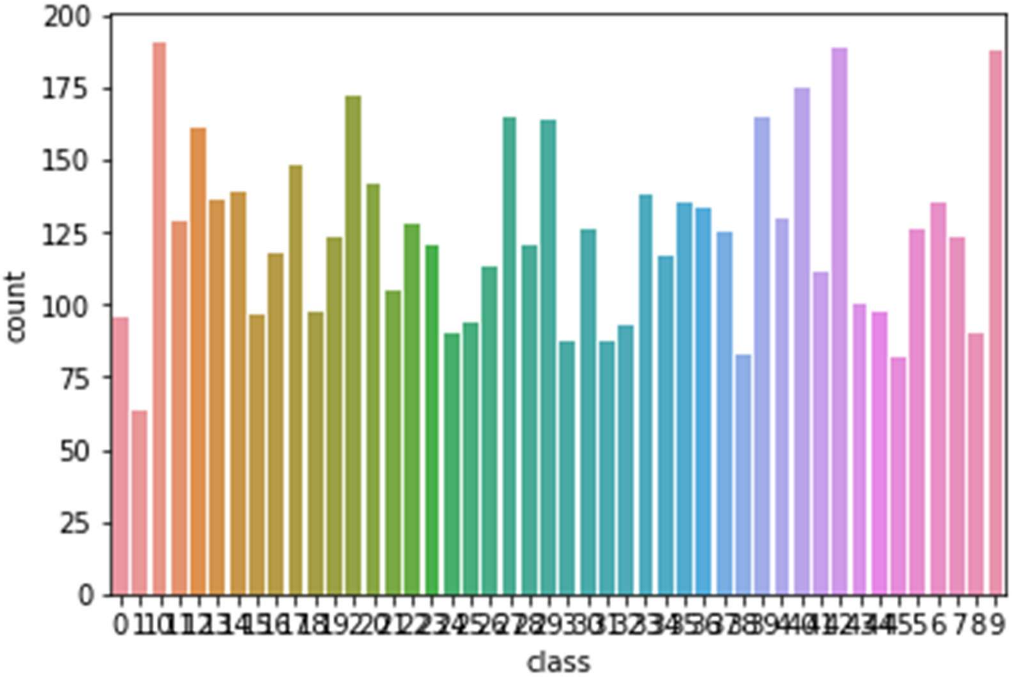
Transfer learning was applied to improve image classification for the Arboles in Chile dataset. Base convnets were trained on the larger Plant Village dataset. The base was then frozen at different layers to identify the ideal model architecture. Image classification did improve as a result of transfer learning, but improvement in accuracy was modest and subject to overfitting.

Introduction

Arboles en Chile Dataset

The Arboles en Chile dataset is a collection of tree images taken from the Metropolitan Region of Chile. With 6,120 train photos and 655 test photos, the labeled dataset includes samples from 46 different species. This low ratio of images per category presents a challenge for classification. Considering the model must distinguish between the nuances of leaves, it is expected to struggle. In fact, the author himself cannot easily distinguish between 46 tree species, so this study is attempting to build a model that exceeds typical human abilities. Additionally, not all the species in the dataset contain the same number of photos. Acaricia caven, for example, is only present in 68 images. With relatively little data for training, it is necessary to find methods to augment the model's learning capabilities.

Arboles en Chile Train Dataset Count by Species Class



The many species and limited sample size of some classes makes image classification challenging.

Transfer Learning and the Plant Village Dataset

Transfer learning can be applied as a possible work-around when dealing with with smaller datasets. The basic idea behind this technique is that a base model can be trained on a much larger set of data. Learning from a data-rich environment, the convolutional base can identify many image features that are insightful for classification. Once fully trained, the weights of the base model are frozen, and the top layers for classification are swapped out for untrained replacements. Now, the small dataset is fed through the frozen base model and fits the new classifier. Transfer learning grants the model access to all the learned features in the frozen base, but also allows the classifier to discriminate based on how they appear in the smaller dataset.

The Plant Village dataset was selected to train the base model in this experiment due to its similarities with the Arboles in Chile dataset. This collection of images was originally concerned with identifying diseases on the leaves of plants typically found in a garden, like tomatoes, bell peppers, or apples. It contains 39 categories of plants/diseases in total and is composed of 11,699 leaf images. The features retained in the convolutional base are expected to be relevant to the classification of trees in the Arboles in Chile dataset.

Literature Review

This experiment was inspired by Cherti and Jitsev's study, "Effect of large-scale pre-training on full and few-shot transfer learning for natural and medical images," where they used transfer learning to detect COVID on chest X-Ray images. In particular, the researchers were concerned with the domain of images used to pre-train the base layer. They compared transfer learning results from when the model was originally trained on a natural set of images like ImageNet or a medical set of X-Ray images. While transfer learning from all domains allowed the researchers to accurately diagnose COVID with roughly 75% accuracy, they obtained their best results when the base model was pre-trained on medical X-Ray images. They also found that the improvement in accuracy for increasing the dataset size for pre-training is only observable when the domain matches the final model's objective.

Methods

This study aims to discover the ideal configuration of frozen and unfrozen layers for transfer learning on the Arboles in Chile dataset. All the convnets trained for the investigation contain the same architecture to ensure consistency. There are three convolutional/max pooling layers that increase from 32 to 64 to 128 nodes. From there, the data is flattened and passed to a dense layer with 384 nodes before a final softmax layer makes the classification. Regularization is employed with dropout after every convolutional/max pooling layer and an L2 regularizer in the dense layer. A summary of the model is listed below.

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_3 (Conv2D)            (None, 62, 62, 32)        896
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 31, 31, 32)        0
_________________________________________________________________
dropout_3 (Dropout)          (None, 31, 31, 32)        0
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 29, 29, 64)        18496
_________________________________________________________________
max_pooling2d_4 (MaxPooling2 (None, 14, 14, 64)        0
_________________________________________________________________
dropout_4 (Dropout)          (None, 14, 14, 64)        0
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 12, 12, 128)       73856
_________________________________________________________________
max_pooling2d_5 (MaxPooling2 (None, 6, 6, 128)         0
_________________________________________________________________
dropout_5 (Dropout)          (None, 6, 6, 128)         0
_________________________________________________________________
flatten_1 (Flatten)          (None, 4608)              0
_________________________________________________________________
dense_2 (Dense)              (None, 384)               1769856
_________________________________________________________________
dense_3 (Dense)              (None, 46)                17710
=================================================================
Total params: 1,880,814
Trainable params: 1,880,814
Non-trainable params: 0
```

Experiment 1

The Arboles en Chile dataset is trained on a convnet to establish baseline accuracy for future experiments.

Experiment 2

The Plant Village dataset is trained on a convnet to ensure that it is successfully able to classify images.

Experiment 3

A convolution base is pre-trained on the Plant Village dataset. All three convolution/max pooling layers are frozen to preserve their weights. The Arboles in Chile dataset is then fed through the model, and the dense layer is fit to the new collection of images.

Experiment 4

A convolution base is pre-trained on the Plant Village dataset. The first two convolution/max pooling layers are frozen to preserve their weights. The Arboles in Chile dataset is then fed through the model, and final convolution/max pooling layer along with the dense layer are fit to the new collection of images.

Experiment 5

A convolution base is pre-trained on the Plant Village dataset. The first convolution/max pooling layer is frozen to preserve its weights. The Arboles in Chile dataset is then fed through the model, and the final two convolution/max pooling layers along with the dense layer are fit to the new collection of images.

Experiment 6

This is the only trial in which the convnet model's architecture varies slightly. A convolution base is pre-trained on the Plant Village dataset. All three convolution/max pooling layers are frozen to preserve their weights. An additional convolution/max pooling layer with 256 nodes is attached to the base. The Arboles in Chile dataset is then fed through the model, and the new convolution/max pooling layer along with the dense layer are fit to the new collection of images.

Results

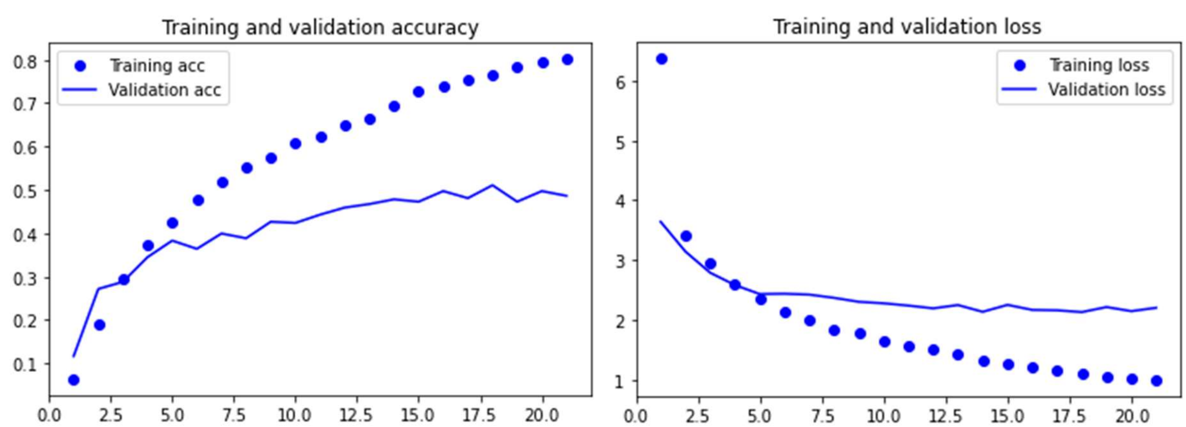
Experiment Number	**Network Description**	**Train Accuracy (%)**	**Validation Accuracy (%)**	**Test Accuracy (%)**
1	Arboles en Chile- no transfer learning	45.66	34.51	35.42
2	Plant Village- no transfer learning	90.55	80.60	84.57
3	3 frozen convolution layers	71.36	48.64	42.14
4	2 frozen convolution layers	81.36	48.64	50.08
5	1 frozen convolution layers	82.00	45.65	49.77
6	3 frozen convolution layers + 1 unfrozen convolution layer	57.80	42.39	47.48

As expected, the convnet in Experiment 1 struggled to classify the Arboles en Chile dataset without any transfer learning. It barely obtained 35.42% test accuracy. On the other hand, the convnet in Experiment 2 proved to be aptly suited to the Plant Village dataset, reaching

84.57% accuracy. These results provided confidence that a convolutional base may be able to extract enough features that could be applied to transfer learning.

Experiments 3-6 show the modestly successful results of transfer learning. Each model demonstrated higher test accuracy than Experiment 1, with the values ranging from roughly 40%-50%. Experiment 4 was the high performer at 50.08%, which is still far too low to be considered for commercial use. Overfitting, despite the regularization employed, proved to be an issue for every model, and a substantial drop is observed between the train and validation accuracies.

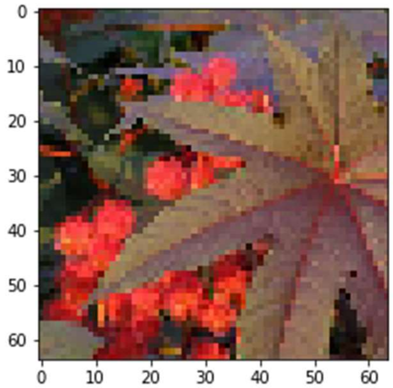
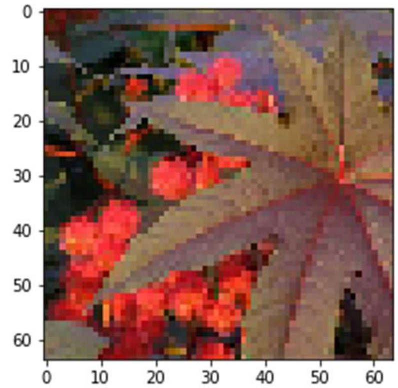
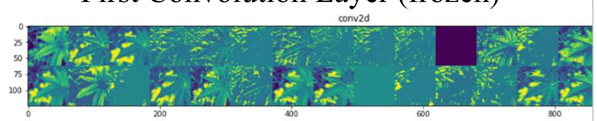
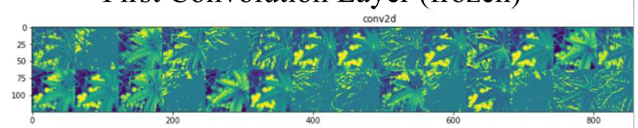
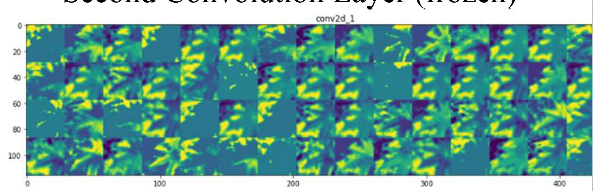
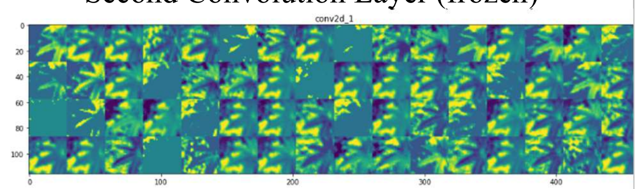
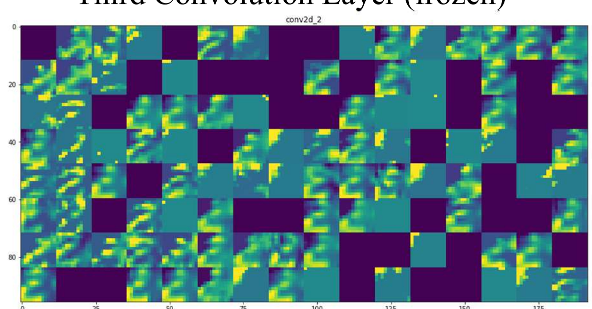
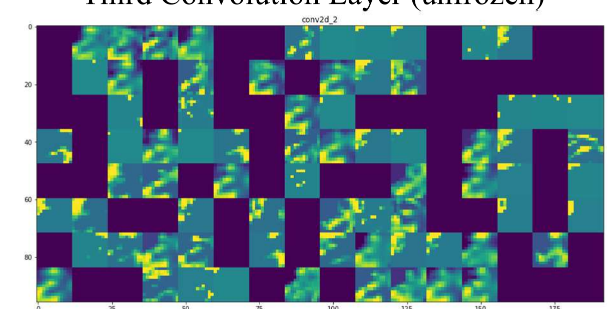
Training and Validation Discrepancies for Experiment 3



The divergence between training and validation for accuracy and loss can be observed after 5 epochs.

The difference in test accuracies between Experiments 3-6 can be challenging to explain when the convent does not explicitly report how and why it assigns weights. One possible theory behind Experiment 4's relative superiority over the other models relates to how identifies features in the convolution layers. The side-by-side set of images below illustrate how convolutional layers visualized the same image during Experiments 3 and 4.

Convolution Visualizations

Experiment 3	Experiment 4
Original Image 	Original Image 
First Convolution Layer (frozen) 	First Convolution Layer (frozen) 
Second Convolution Layer (frozen) 	Second Convolution Layer (frozen) 
Third Convolution Layer (frozen) 	Third Convolution Layer (unfrozen) 

One can observe that the visualizations for the first two convolution layers are similar for both models. Because these layers are frozen, both models are identifying roughly the same features. However, the models differ at the third layer where it is frozen in Experiment 3 and unfrozen in Experiment 4. The Experiment 3 model would still be searching for the specific features that were pertinent to classifying the Plant Village dataset. The unfrozen Experiment for model, on the other hand, would be looking for new features that are more relevant to the Arboles in Chile dataset. Transfer learning relies on a balance between retaining base knowledge and identifying new features that give insight into the new classification task on hand. Perhaps Experiment 3 fell right in sweet spot between both conflicting interests.

Conclusion

This study demonstrated that the Plant Village dataset can be used to train a convnet base for transfer learning to improve species classification in the Arboles in Chile dataset. Nonetheless, the best model was still only able to reach 50% accuracy and demonstrated troubling patterns of overfitting. Before any form of commercial implementation, the classifier would need to improve significantly. Avenues for further exploration include the use of a pretrained convnet like ImageNet, which contains millions of pretrained images. Additionally, data augmentation techniques like image rotation or zoom could be employed to avoid overfitting while training the base convnet or classifying the Arboles en Chile dataset.