Project Transfer Learning: Monkey Species Classifier

Introduction

Research into computer vision has made great leaps in the past few years with many innovations to improve learning and classification ability. Through a series of progressive network architecture innovations, the ability for a neural network to learn and classify a wide range of images has been improved. Two specific innovative networks were the ResNet and the VGG network. Both of these networks included breakthroughs in architecture that greatly improved performance on the ImageNet dataset.

Through transfer learning, a pretrained ResNet or VGG network can be used to classify any number of different image classes. By attaching a classifier that predicts the necessary number of classes, the pretrained networks weights and innovative architecture can be used to achieve high accuracy for any image recognition task.

Within computer vision, there are two general categories of classification tasks: general classification and fine-grain classification. A general classification task involves perceiving images and classifying them as one among a set of classes such as fruits or cars. A fine grain classification task involves classifying images as one amongst different types of apples. Theoretically, this classification task would require the neural network learning a more complex and detailed understanding of the image data. In a neural parallel, humans are able to perform this type of classification with ease after seeing a few examples.

A detailed fine-grain classification approach holds great promise for autonomous systems and image recognition. A medical imaging device that could perceive the distinction between subtypes of a type of cancer could greatly improve the accuracy and efficiency of a cancer diagnosis.

Data

The data for this project was obtained from Kaggle.com. The data includes a training and a validation dataset of which the training set was split further into a training and test dataset. Each dataset includes images from the ten different classes. The images are of varying dimensions and required standardizing through PyTorch transforms. More information on the dataset can be found at the Kaggle dataset page.

Kaggle Link: https://www.kaggle.com/slothkong/10-monkey-species

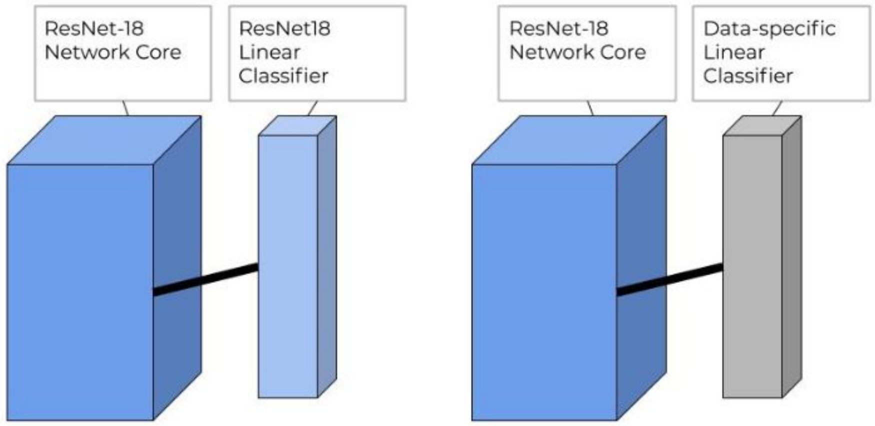
The total number of images in the training dataset was 1097. Interestingly, 1097 is a prime number, resulting in uneven batch sizes. Since the number of images per class was roughly the same, a random image was removed to create a total training set size of 1096. This allowed for a batch size of 137 per training dataset. For the test dataset, a random subset totalling 25% of the 1096 training images was isolated. In total, the training dataset included 822 images, the test data included 274 images, and the validation data included 274 images.

Transfer Learning

Transfer learning refers to the using a pretrained model to classify and analyze a novel dataset. In general, it is implemented by using a large neural network trained on a general dataset that is utilized as a fixed feature extractor for a novel dataset. This type of transfer learning involves defining and training a new linear classifier that adjoins the pretrained network.

The transfer learning models used in this project extended different ResNet and VGG networks. Theoretically, transfer learning can be visualized as below. For visual simplicity, the architecture of the ResNet network and the linear classifiers is represented as a single layer. The data-specific linear classifier is what is defined dependent on the number of predicted classes. The benefit of this approach is that the network weights are pre-trained, reducing the complexity of the training process. In addition, transfer learning allows one to rapidly adapt an innovative research image classification network such as ResNet to any number of datasets without requiring the complicated and resource intensive training of the entire ResNet network. In this project, the data specific linear classifier involved two fully connected layers that output a prediction value for the ten classes of monkey species.

Figure 1: Transfer Learning Architecture



There were multiple distinct transfer learning models for each ResNet and VGG based on the different number of layers in the extended ResNet and VGG models. The abbreviation BN refers to Batch Normalization. These models are listed below.

ResNet
Model R1: extends ResNet-18
Model R2: extends ResNet-3
Model R3: extends ResNet-50
Model R4: extends ResNet-101
Model R5: extends ResNet-152

VGG
Model V1: extends VGG-13
Model V2: extends VGG-16
Model V3: extends VGG-19
Model V4: extends VGG-13 with BN
Model V5: extends VGG-16 with BN
Model V6: extends VGG-19 with BN

ResNet or Residual Network was an advanced neural network architecture proposed in 2015 by researchers working at Microsoft Research. The key breakthrough of this architecture was the use of shortcut connections to connect a layer to a later layer in the network. These connections enhanced the network's learning potential and overcame the vanishing gradient problem encountered with multi-layer deep networks. The original paper describing the ResNet architecture included performance data for ResNet with 18, 34, 50, 101, and 152 layers [1].

VGG is an advanced neural network architecture proposed in 2014 by Oxford University's Visual Geometry Group. The network was innovative because it demonstrated that a deeper network with more layers performed better than one with fewer layers. The result that deeper models performed better became a core aspect of all later neural network architectural innovations. The original paper describing the VGG network architecture discussed and included the model details for VGG 16 and 19 [2].

Data Preprocessing

The image data was loaded with PyTorch's ImageFolder class. The training, testing, validation data were resized to 256x256 size and normalized tensor values to a range of [-1,1]. Data augmentation was not included because the larger input image size along with the network's pretrained status resulted in high test accuracy and low training, test, and validation losses. It is possible that Data augmentation would improve network reliability on unknown or novel data. In addition, if the image size were reduced further then it is highly likely that data augmentation would be required for good network performance.

Training the Model

The models were trained on data preprocessed as described in the earlier section. Before iterating through the training data, the data specific classifier was defined in a Classifier class and attached to an imported model, like ResNet 18 for example. The parameters of the imported model were frozen by setting the requires_grad value to False.

This is a critical step because the imported model is already pretrained and any further training would be resource-intensive and alter the pretrained weights. With the imported model's features frozen, the training tunes the linear classifier to predict among the image classes for that specific dataset.

The ResNet models were trained for 40 epochs and the VGG models were trained for 20 epochs. The VGG models V1-3 were trained with a learning rate of 0.001. The other models were trained with a learning rate of 0.01. After visualizing the training and validation losses, the test accuracy was calculated. The test accuracies are listed below.

ResNet	Test Accuracy
Model R1	95%
Model R2	98%
Model R3	98%
Model R4	98%
Model R5	99%

VGG	Test Accuracy
Model V1	96%
Model V2	99%
Model V3	98%
Model V4	98%
Model V5	96%
Model V6	97%

Figure 2: Training and Validation Loss for ResNet Models

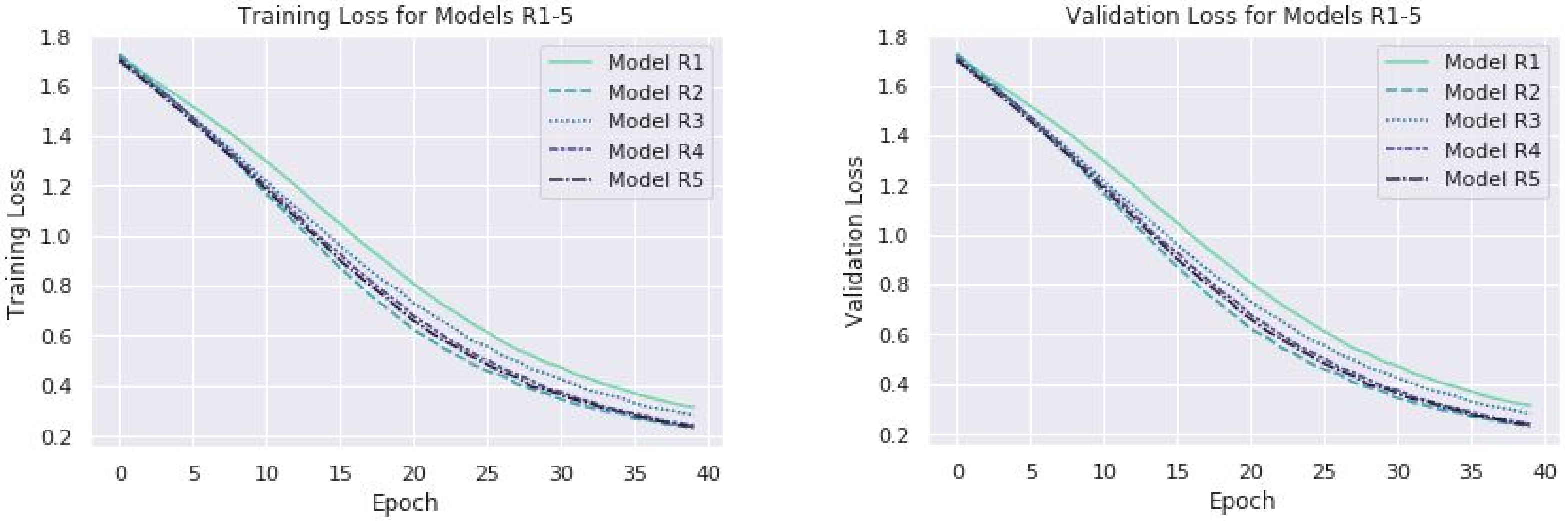
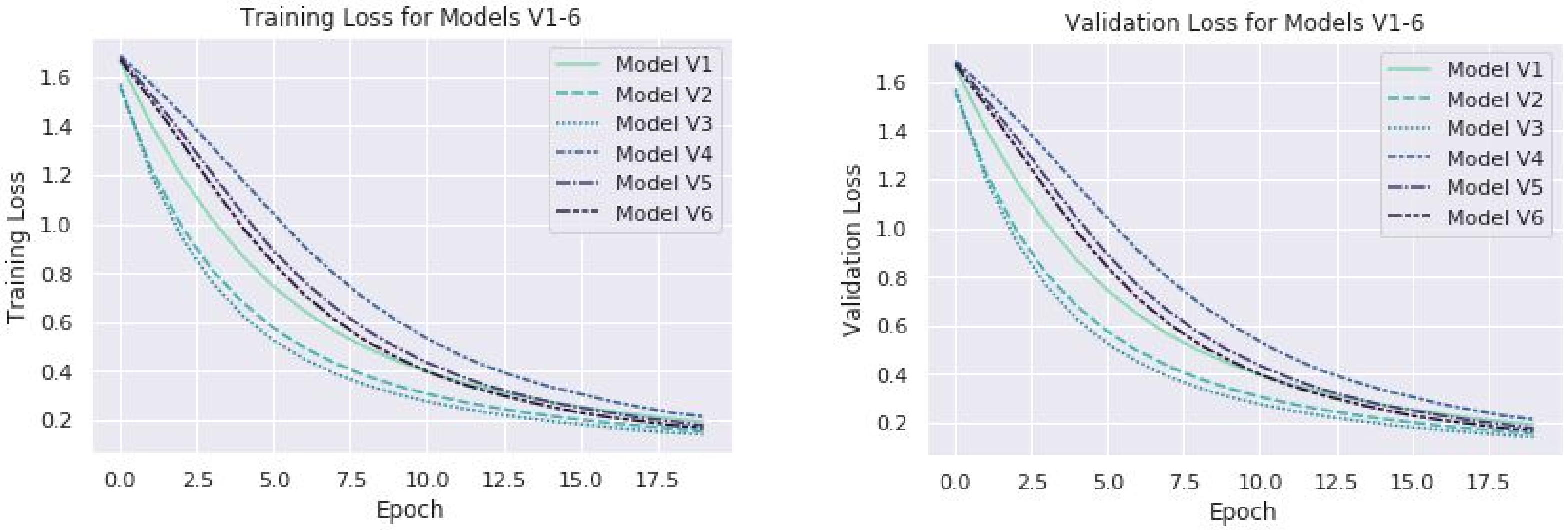


Figure 3: Training and Validation Loss for VGG Models



Conclusions

Given the architectural difference between the VGG network and the ResNet network, it was initially expected that the ResNet networks would perform better at classification tasks. This was not observed in this project. The similar performance between the ResNet and VGG models could be due to the smaller dataset size and the networks all being pretrained, resulting in an essentially minute performance difference.

The VGG networks with batch normalization could be trained with a higher learning rate than the non-batch normalized VGG network. This observation aligns with the effect of batch normalization observed in CNNs that it enables using higher learning rates without accuracy or learning loss.

References

1. ResNet Paper: https://arxiv.org/abs/1512.03385
2. VGG Paper: https://arxiv.org/abs/1409.1556