Report

Vikram Goud

Introduction:

Image categorization, being the fundamental problem in computer vision, has unlimited spheres of its implementation from medical image analysis to driverless cars. CNNs (convolutional neural networks) have the capability of beating the top performance benchmarks on a lot of datasets, hence becoming the most powerful experts for classifying images. Then, however, the quality of data, sample size and structure of the network are some of the things that may affect how good the convolutional neural network performs. Thus, to lead to conclusion, link of such factors is explored in this expression of research considering the image classification model optimization.

Background:

Deep learning models, particularly CNNs, have shown remarkable success in image classification tasks. Traditionally, training CNNs from scratch requires a large amount of labeled data to prevent overfitting and achieve satisfactory performance. However, in scenarios where labeled data is limited, transfer learning techniques, such as using pretrained networks, offer a viable solution by leveraging knowledge learned from large datasets. Understanding the interplay between training sample size and the choice of network architecture is essential for effectively deploying image classification models in real-world applications.

Methodology:

In this paper we conducted experiments using image classification sample set to know how the shift in the number of sample and network choice affects the network performance. A straightforward CNN architecture in the beginning was the place to start, for it was presumed to be more tractable to train. A tiny dataset with the size of 1,000 images was used for that purpose (a training set of 500 images and a validation set of 500). Techniques that avoid overfitting, such as increasing the dataset and using data augmentation, were employed. Next, we add datapoints to our network and modify it.

Furthermore, we tested out various training datasets to discover the superior input size which gave us high predictions outcomes. We employed VGG16 pretrained version as a base in the experiments to appreciate the approximation of training from the beginning plus pretrained network.

Results and Findings:

Training from Scratch:

The CNN trained from scratch was able to do well with a small dataset that was made up of 1000 photos, where the data augmentation strategies were employed to help in reducing overfitting. Besides, we achieved performance gains by increasing the size of the training data confirming the importance of dataset size during CNNs training from the very beginning.

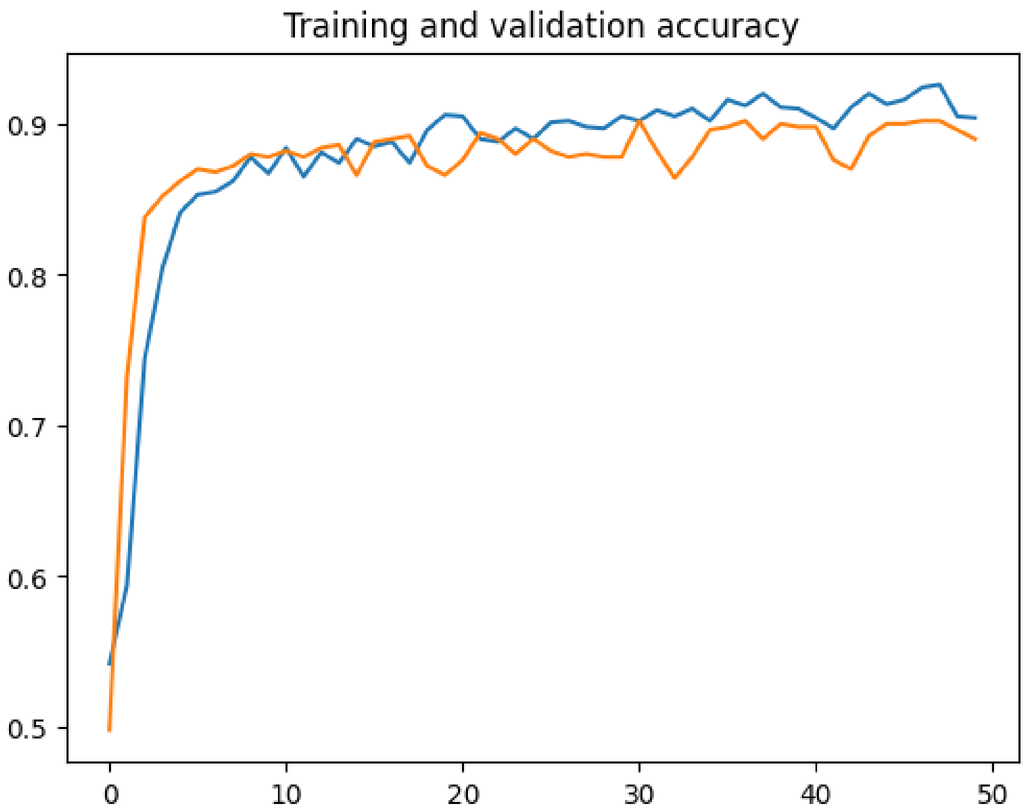
```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #   
=================================================================
 conv2d_4 (Conv2D)           (None, 148, 148, 32)      896       
                                                                 
 max_pooling2d_4 (MaxPoolin  (None, 74, 74, 32)        0         
 g2D)                                                            
                                                                 
 conv2d_5 (Conv2D)           (None, 72, 72, 64)        18496     
                                                                 
 max_pooling2d_5 (MaxPoolin  (None, 36, 36, 64)        0         
 g2D)                                                            
                                                                 
 conv2d_6 (Conv2D)           (None, 34, 34, 128)       73856     
                                                                 
 max_pooling2d_6 (MaxPoolin  (None, 17, 17, 128)       0         
 g2D)                                                            
                                                                 
 conv2d_7 (Conv2D)           (None, 15, 15, 128)       147584    
                                                                 
 max_pooling2d_7 (MaxPoolin  (None, 7, 7, 128)         0         
 g2D)                                                            
                                                                 
 flatten_1 (Flatten)         (None, 6272)              0         
                                                                 
 dense_2 (Dense)             (None, 512)               3211776   
                                                                 
 dense_3 (Dense)             (None, 1)                 513       
                                                                 
=================================================================
Total params: 3453121 (13.17 MB)
Trainable params: 3453121 (13.17 MB)
Non-trainable params: 0 (0.00 Byte)
```

Optimal Size for a Training Sample:

The optimal dataset was determined by practical experimentation with contrasting training sample sizes to get best results of prediction. A balanced approach with large enough dataset, moderate model complexity and right choice of regularization methods works best to prevent overfitting.

Pre-trained Network:
From a performance point of view, the VGG16 pre-trained model outperforms training from scratch

with little data. The classification accuracy of the pretrained model was also fine-tuned using our own datasets, thus proving the value of transfer learning in situations with very little data.



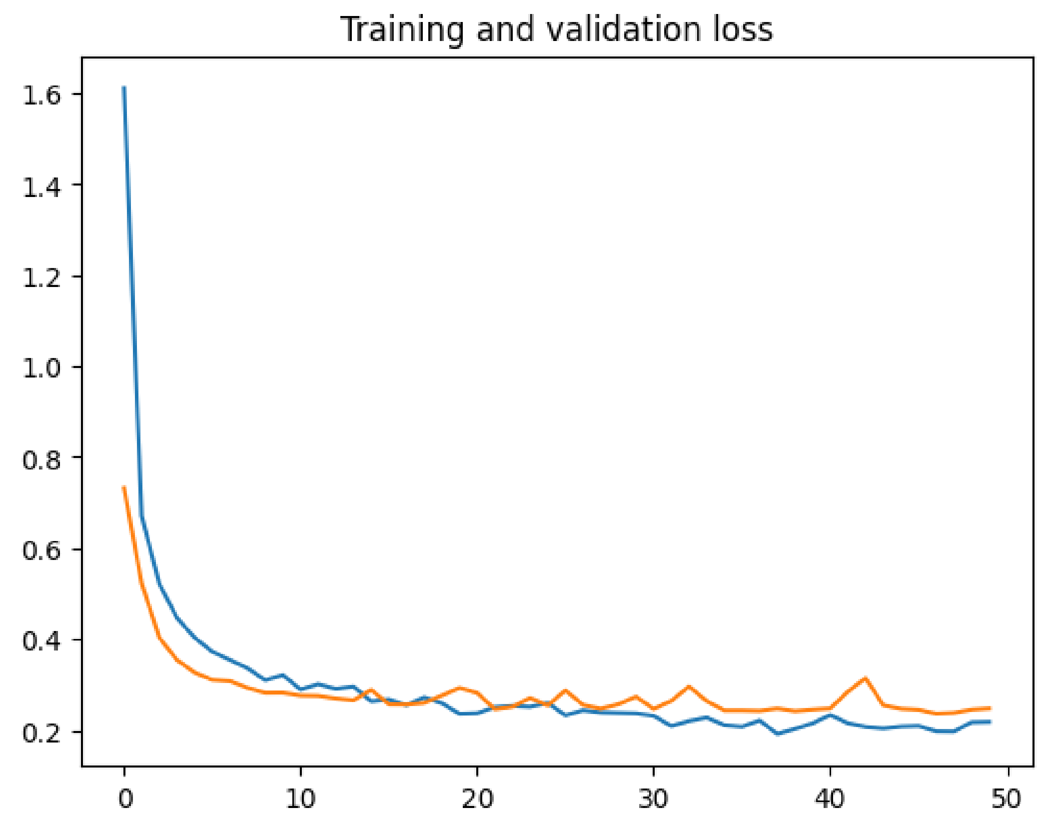


Fig: VGG16 Accuracy and Loss

S No	Method	Training Size	Training Accuracy	Validation Accuracy	Test Accuracy
1	Training from Scratch	1000	73	61	60.3
2	After Augmentation	1000	79	72.2	77.7
3	Optimizing the model	5000	82	82	83.2
	With augmentation	5000	91	90	89.2
		10000	83	91	89
4	VGG16 pre-trained model	1000	91	88	88.5
	With augmentation	1000	90	89	88.9
		3000	89.6	88.4	88.4
		15000	88.45	90	91.7

Conclusion:

This analysis underscores an important note to keep in mind which is how training sample size and the type of network used have great impact on picture classification. By having enough data and proper regularization methods using CNNs can deliver expected results, despite that, transfer learning with pretrained mode is a handy option for when bits of labeled data are available. Via the knowledge on how these components interlace, they may design a precise picture of these pipelines that will be meeting their contexts while balancing the performance of the model with the availability of data as well as computational resources. New optimization strategies as well as architectures of networks could be assessed in the next research studies to improve accuracy of image classification for an array of data and usages.