**ASSIGNMENT 2: CONVOLUTION REPORT**

Two approaches are used to build a convolutional neural network: training from scratch and using a pretrained convnet to reduce overfitting, which leads to more difficult prediction during training. These techniques are referred to as data augmentation and regularization. The example of cats and dogs is taken into consideration.

I created a new dataset with training, validation, and testing samples using the cats vs. dogs approach data from Kaggle with 1000 samples in the training set and 500 sample validation set.  
and 500 samples in the training set.

Let’s see how the model performs with different sample sizes by reducing overfitting.

Using regularization techniques and fine tuning.

**Q1. Consider the Cats & Dogs example. Start initially with a training sample of 1000, a validation sample of 500, and a test sample of 500 (like in the text). Use any technique to reduce overfitting and improve performance in developing a network that you train from scratch. What performance did you achieve?**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

The graphs shown above are instances of overfitting. Validation accuracy reaches 70-72%, whereas training accuracy rises linearly over time to nearly 100%.

And the validation loss stalls after 5 epochs after decreasing steading ,whereas the training loss keeps decreasing.

Our primary problem will be overfitting because there aren't many training samples. A number of strategies, including dropout, regularization, and data augmentation, can be used to reduce overfitting.  
  
To enhance the model's performance, I used dropout regularization technique to reduce overfitting and improve performance.

|  |  |  |
| --- | --- | --- |
| **MODEL** | **TEST ACCURACY** | **VALIDATION ACCURACY** |
| Without regularization | 70% | 72% |
| With dropout regularization | 73% | 76% |

According to the results, the performance of the model on test set increases after regularizing the model.

Let’s now increase the sample size by 1000 and see the change in performance of the model.

**Q2. Increase your training sample size. You may pick any amount. Keep the validation and test samples the same as above. Optimize your network (again training from scratch). What performance did you achieve?**

**Now, I have increased the sample size of training by 1000, the new size is 2000 and used data augmentation technique for regularization.**

|  |  |  |
| --- | --- | --- |
| **MODEL** | **TEST ACCURACY** | **VALIDATION ACCURACY** |
| With 1000 samples | 73% | 76% |
| With 2000 samples | 85% | 88% |

We can see that the performance of model has now greatly increased by increasing the training sample size and using data augmentation as regularization technique, the test accuracy has changed from 73% to 85%.

**Q3.** **Now change your training sample so that you achieve better performance than those from Steps1 and 2. This sample size may be larger, or smaller than those in the previous steps. The objective is to find the ideal training sample size to get best prediction results?**

**Now, I have increased the training sample size from 2000 to 5000.Let’s evaluate the results**.

|  |  |  |
| --- | --- | --- |
| **MODEL** | **TEST** **ACCURACY** | **VALIDATION ACCURACY** |
| With 1000 samples | 73% | 76% |
| With 2000 samples | 85% | 88% |
| With 5000 samples | 91% | 92% |

**There is an increase in test accuracy of 91% from 85% with increase in training samples**  from 2000 to 5000 which means our model performance has been improving with increase in the training sample data size.

Therefore, the regularized model with 5000 samples is best for getting best predictions, it can be considered as the idea training dataset size.

**Q4. Repeat Steps 1-3, but now using a pretrained network. The sample sizes you use in Steps 2 and 3 for the pretrained network may be the same or different from those using the network where you trained from scratch. Again, use any and all optimization techniques to get best performance.**

I am using the VGG16 pretrained convolution network as the base network in fine tuning process.

Let’s evaluate the results obtained by fine tuning the models with different sample size of training data sets.

|  |  |  |
| --- | --- | --- |
| **FINE TUNED MODEL** | **TEST ACCURACY** | **VALIDATION ACCURACY** |
| With 1000 samples | 97% | 97% |
| With 2000 samples | 99% | 98% |
| With 5000 samples | 98% | 99% |

From the above table we can see that the testing and validation accuracy tend to improve as the sample size increases. They are even better when regularized and fine-tuned when compared to only regularized models.

In Summary, When the training data set is small, overfitting becomes the main concern. To reduce overfitting and improve models performance we can regularize the models with different techniques and by fine-tuning the model we can improve the performance of the model where data augmentation plays the major role by being one of the powerful method of reducing overfitting. Moreover, by increasing the size of the trainingdata set we can reduce overfitting, so it is ideal to take more data for training rather than validation and testing.