Report

Vineeth Rudraksha

In this study, we investigate the relation between the number of training samples and the network architecture selection for images, with the aim of classification. Taking the Cats & Dogs dataset as an illustrative example, we concretely look at the differing performances utilizing spawned Convolutional Neural Networks (CNNs) as well as pretrained networks for training. The task is that the changes of sample size should be studied for the optimization purpose of each strategy.

Problem Statement:

The challenge is to use CNNs to build an image classification model that can discriminate between cats and dogs. Our goal is to determine whether using a pretrained network offers any benefit over starting from scratch and how the size of the training sample affects the model's performance.

Methodology:

We use two models; one is called CNN into the scratch and the other is a CNN pre-trained (MobileNetV2 with ImageNet's weight implementation).

CNN Trained from Scratch:

Convolutional, pooling, and fully connected layers place early in the design of the CNN. The use of regularization and data augmentation are two strategies common to reduce the overfitting and pour efficiency of the process. We evaluate the model performance on fixed validation and test sets and optimize the size of training sample, if necessary.

CNN with Pretrained Network (ImageNet):

To attack Cats & Dogs organization (classification) problem we pre-trained the MobileNetV2 structure on ImageNet and fine-tuned the model's parameters. Convolution layers of MobileNetV2 are decisively frozen and Fully Connected Layers are added for the classification. We vary the arrangement once over again and assess the results accordingly by applying the same test and validation sets.

Results and Findings:

CNN's Initial Training:

Those approaches concerning data augmentation and regularization make the model finally achieve the accuracy of 71% on the validation set and 69.8% on the test set with the initial training sample size of 1000.

Fig: Data Augmentation

Performance can be improved by the increased size of samples such as a validation accuracy of 87% and 84% on the test set is obtained with a set of 3000.

Though improvement can be observed with the growth of training set size, its benefits are lowering.

<u>Optimal Size for a Training Sample:</u>

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.

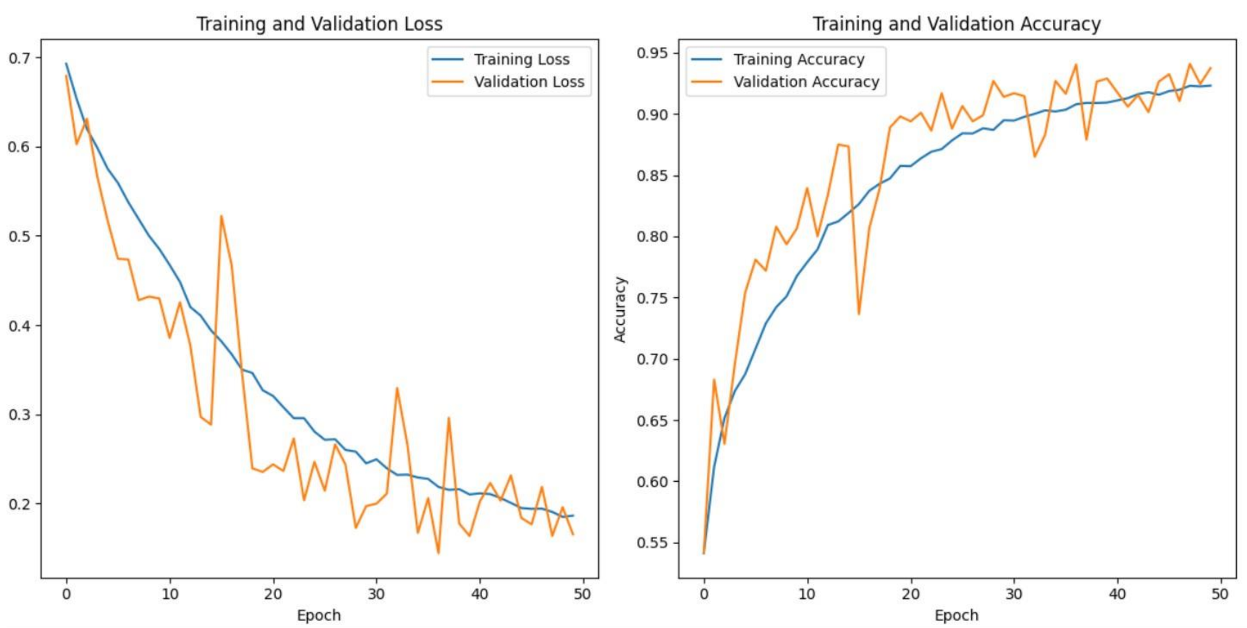

Fig: training, validation accuracy and loss for 15000 samples **Pre-trained**

<u>Network MobileNetV2:</u>

The images that are contained in the ImageNet dataset are not merely supplied but are accompanied by proper annotations including the classes that they belong to for the purpose of training and benchmarking computer vision models, particularly those involved in image classification. The millions of photos are grouped into tens of thousands of pictures that could be searched under different themes.

Notwithstanding the situation, the project was implemented in order to boost research on objects recognition, situation understanding and image interpretation as well as to strengthen visual computer technology development.
By adapting the MobileNetV2 model, an accuracy of 97.8% is reached on the validation set and 97.6% on the test set, using the same initial training sample of 1000.
An accuracy of 98.1% on the test set and 97.4% on the validation set is also obtained when the training set size is 3000.
As the training sample size is enlarged, the performance will be diminished at some point, similar to the CNN that initially had been trained from scratch.

TABLE FOR METRICS

Method	Train Size	Validation and Test Sample Size	Test Accuracy	Validation Accuracy	Training Accuracy
Training from Scratch	1000	500,500	0.6980	0.7160	0.7060
After Augmentation	1000	500,500	0.6959	0.6960	0.7180
Optimizing the model	3000	500,500	0.8519	0.8380	0.8550
With augmentation	3000 15000	500,500	0.9284	0.9375	0.9232
Image Net pre-trained Model	1000	500,500	0.9760	0.9780	0.9930
With augmentation	1000	500,500	0.9819	0.9720	0.9847

Conclusion:

The network architecture choice is normally a main point that affects the performance of the model, either training from scratch or with a pre-trained network. Contrary to the original training, ImageNet's transfer learning property makes pretrained network deal better with fewer training samples. On the one hand, the difference in performance between the two methods shines when the sample size is small, but with an increasing sample size, the gap becomes smaller. Both strategies will arrive at similar performance levels eventually, where the extensive use of transfer learning does not necessarily perform better than training on larger datasets. Besides this, when training from zero-shot, regularization approaches and data augmentation for instance, are of crucial importance in mitigating overfitting. Generally, the interconnection between the variation of the training sample size and the type of neural network illustrates that choosing the right techniques becomes the important thing which is based on the data and the computer resources.