# Report

In this assignment, we trained a deep learning model on the Cats & Dogs dataset using both a network trained from scratch and a pre-trained network (VGG16) to explore the relationship between training sample size and network choice on model performance.

First, we trained a Convnet to classify images of cats and dogs. We started with a small dataset of 1000 training samples, 500 validation samples, and 500 testing samples. We used dropout and data augmentation techniques to reduce overfitting and improve performance. After training the model from scratch, we achieved a training accuracy of 82%, validation accuracy of 80%, training loss of 39%, and validation loss of 40%.
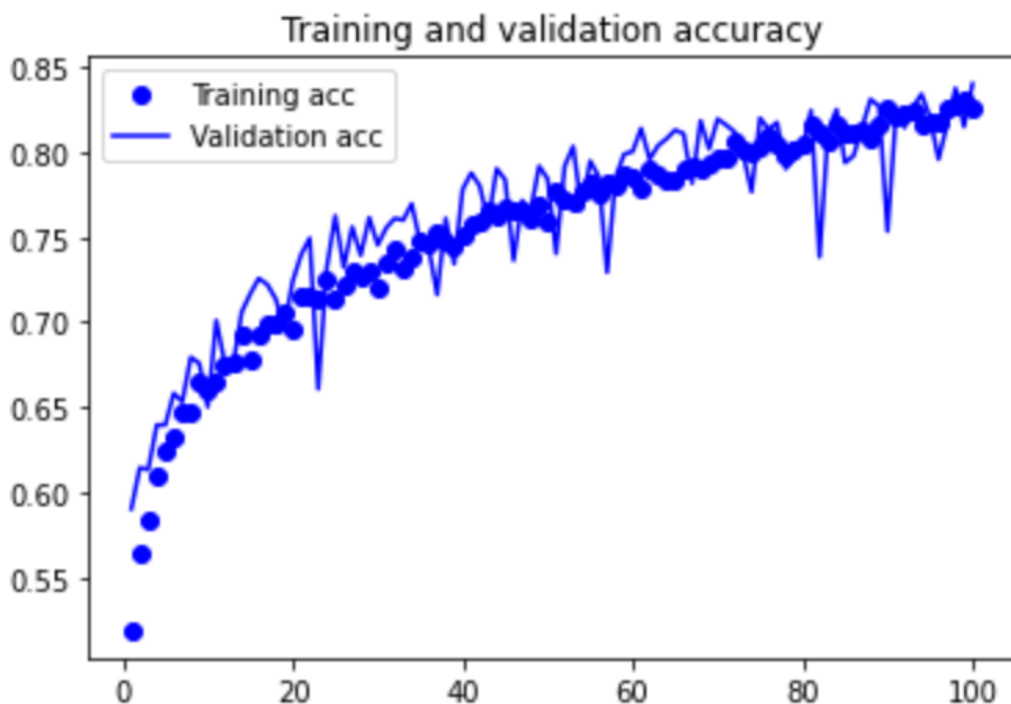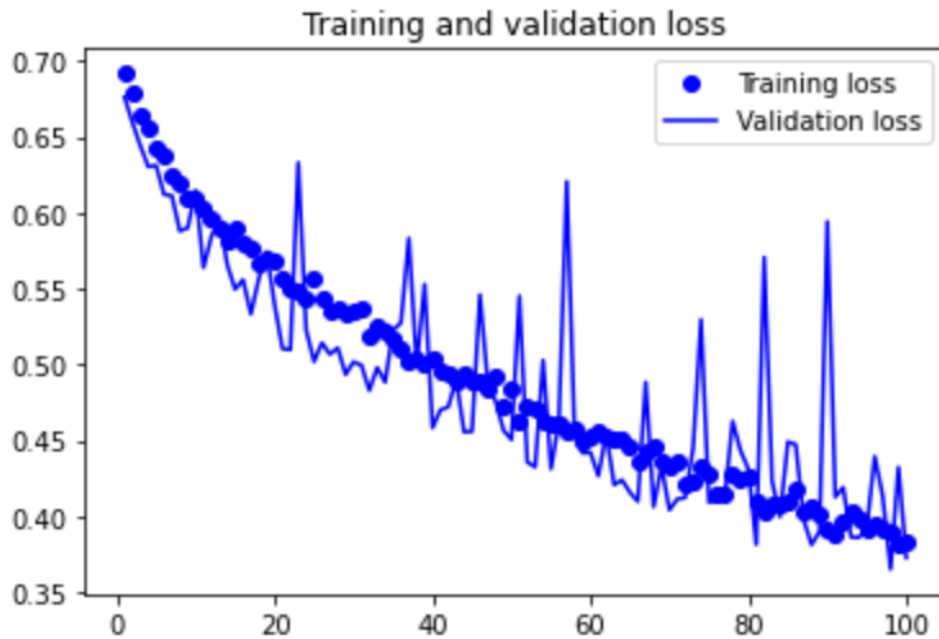
Training and validation loss
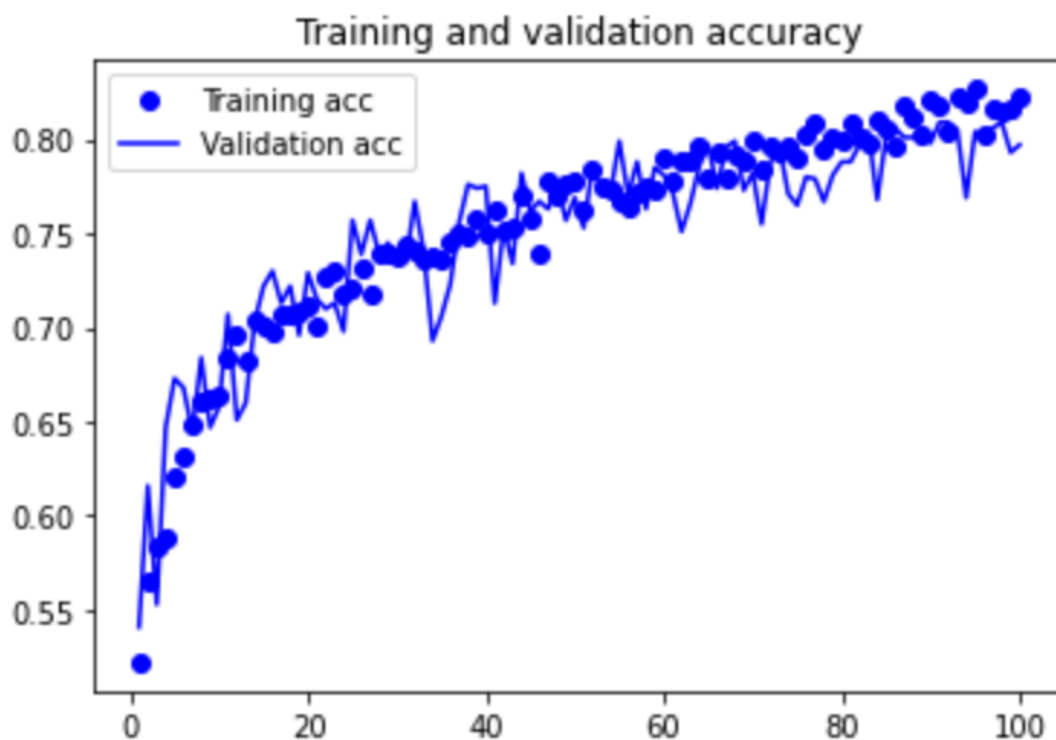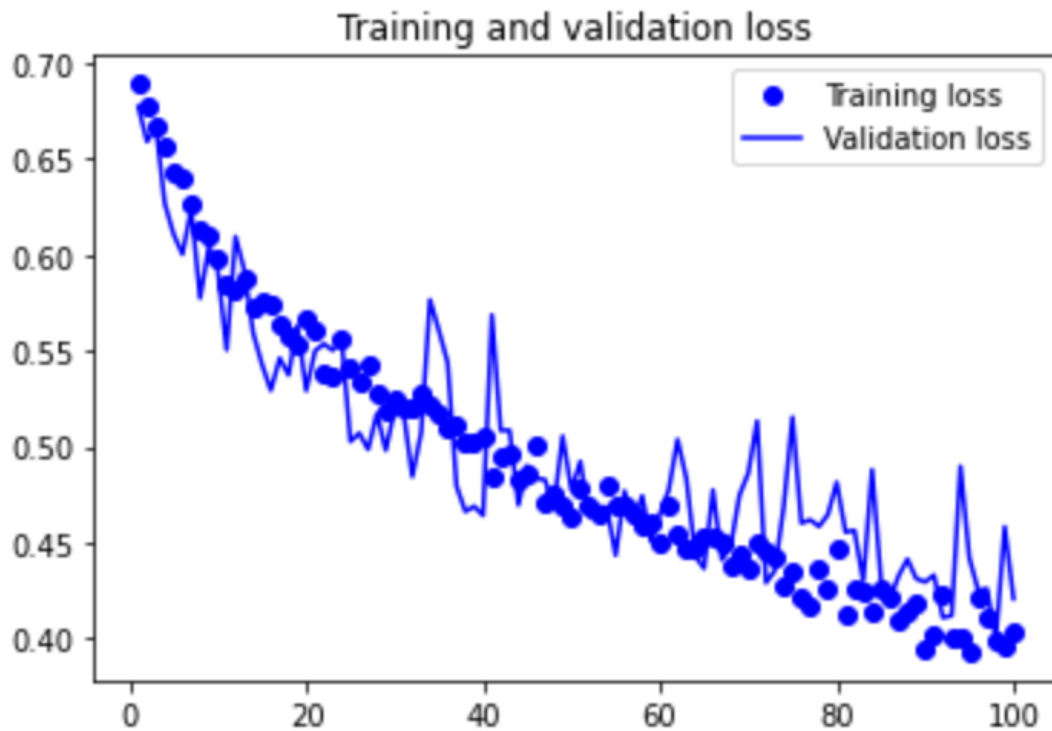
We then increased the training sample size to 2000, retrained the network, and achieved training accuracy of 83%, 84% validation accuracy, 38% training loss, and 36% validation loss. This shows the slight improvement.

Training and validation accuracy

Training and validation loss
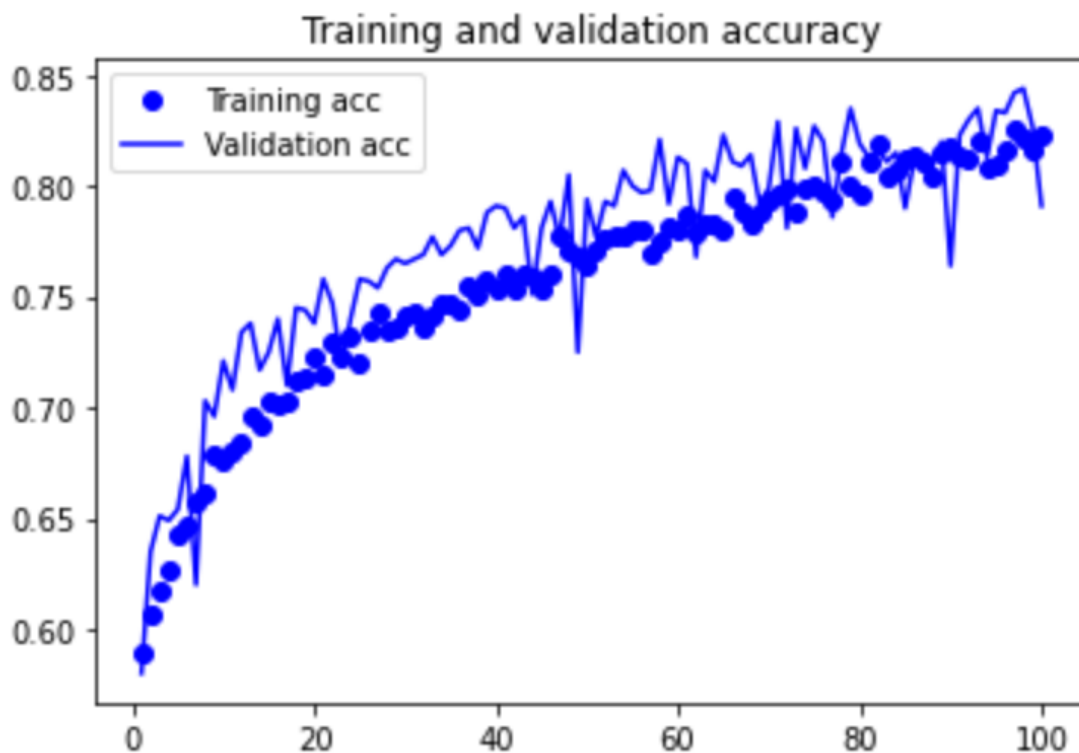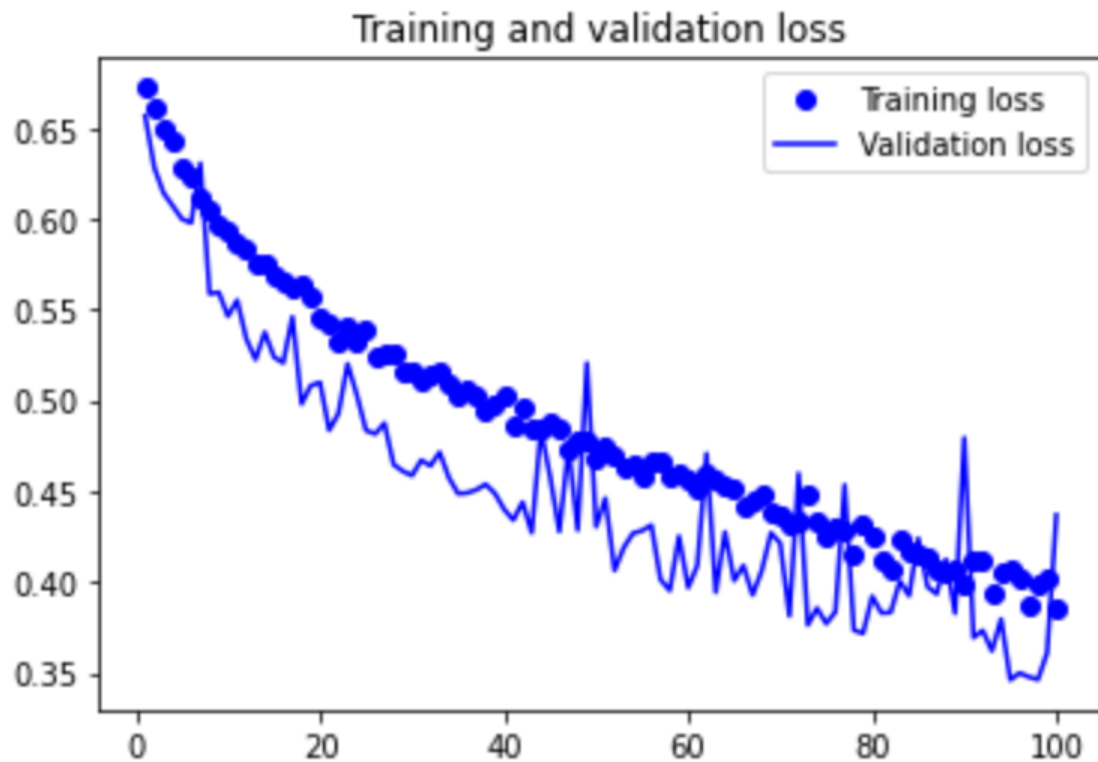
Then we changed the training samples and chose from 3000 to 5000 instead of the first 2000 samples. We got similar results with 83% training accuracy, 84% validation accuracy, 38% training loss, and 36% validation loss. So it suggests that changing the training samples does not necessarily improve the model's performance.

Training and validation accuracy

Training and validation loss

Finally, we repeated the experiment with a VGG16 pretrained network. We achieved much higher accuracy with this network, with a training accuracy of 97%, validation accuracy of 90%, training loss of 9%, and validation loss of 23% for the first step samples, and a training accuracy of 97%, validation accuracy of 89%, training loss of 9%, and validation loss of 25% for the third step samples.
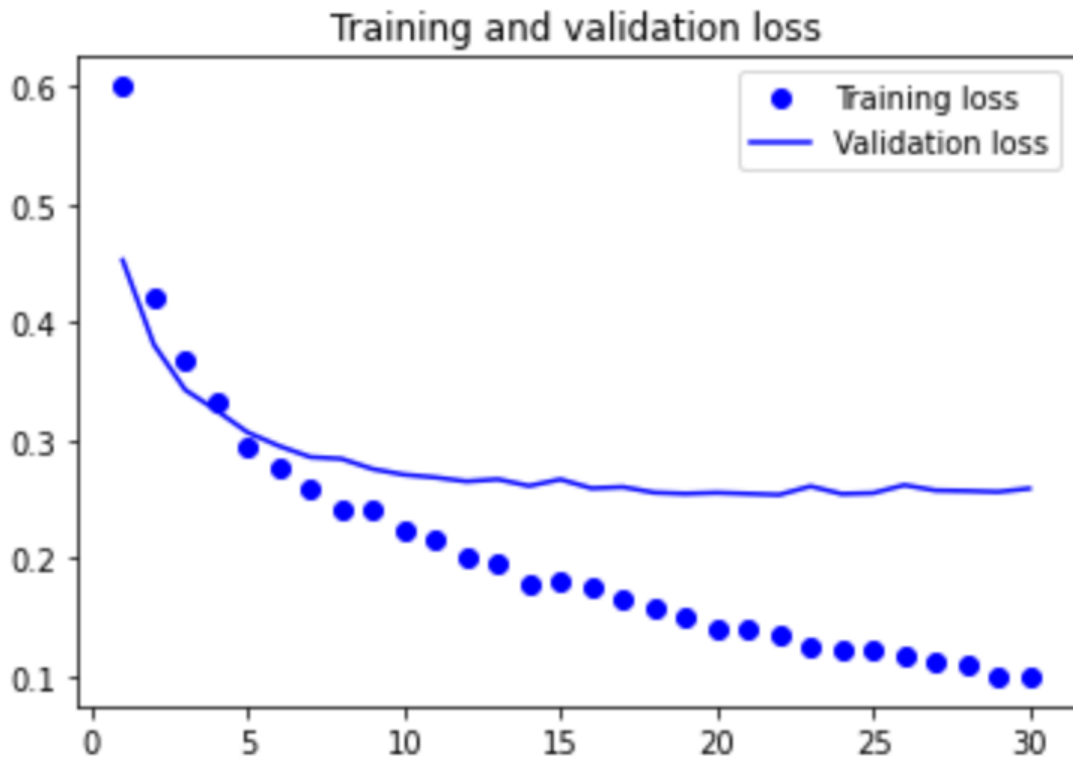
Step:1

Training and validation accuracy

Training and validation loss

Step:3

Training and validation accuracy

Training and validation loss
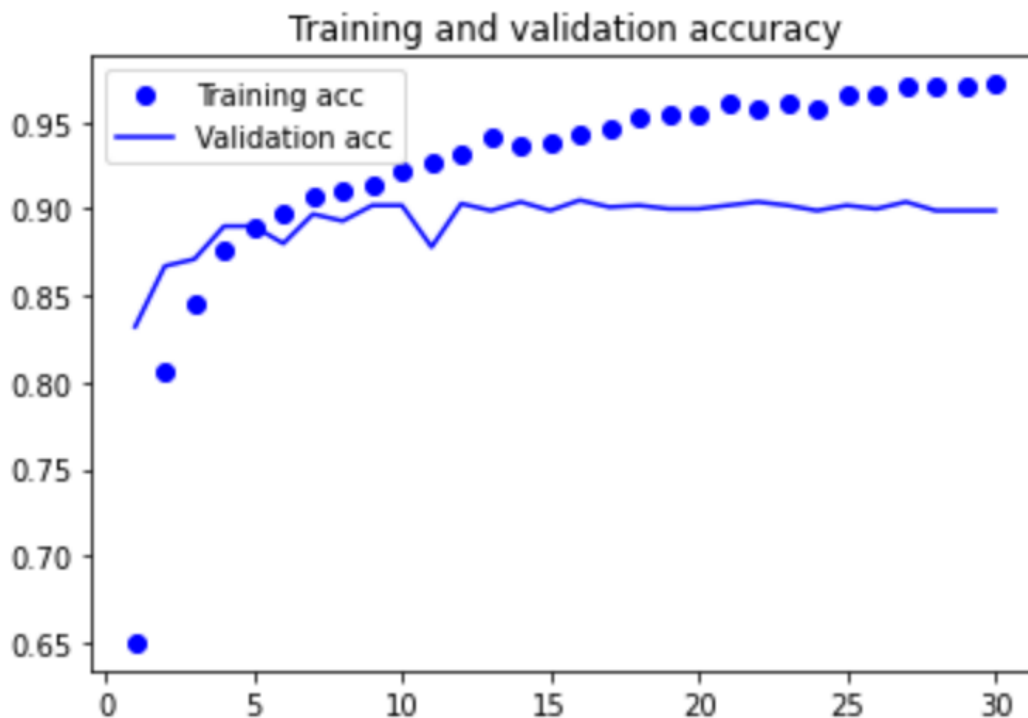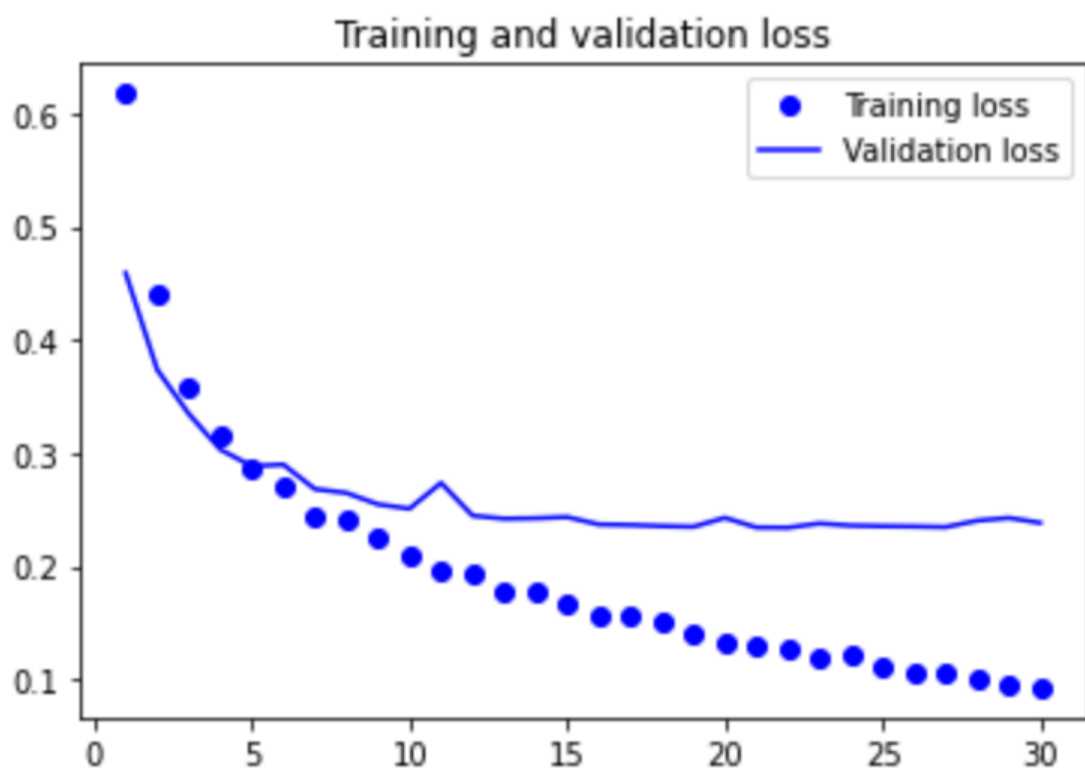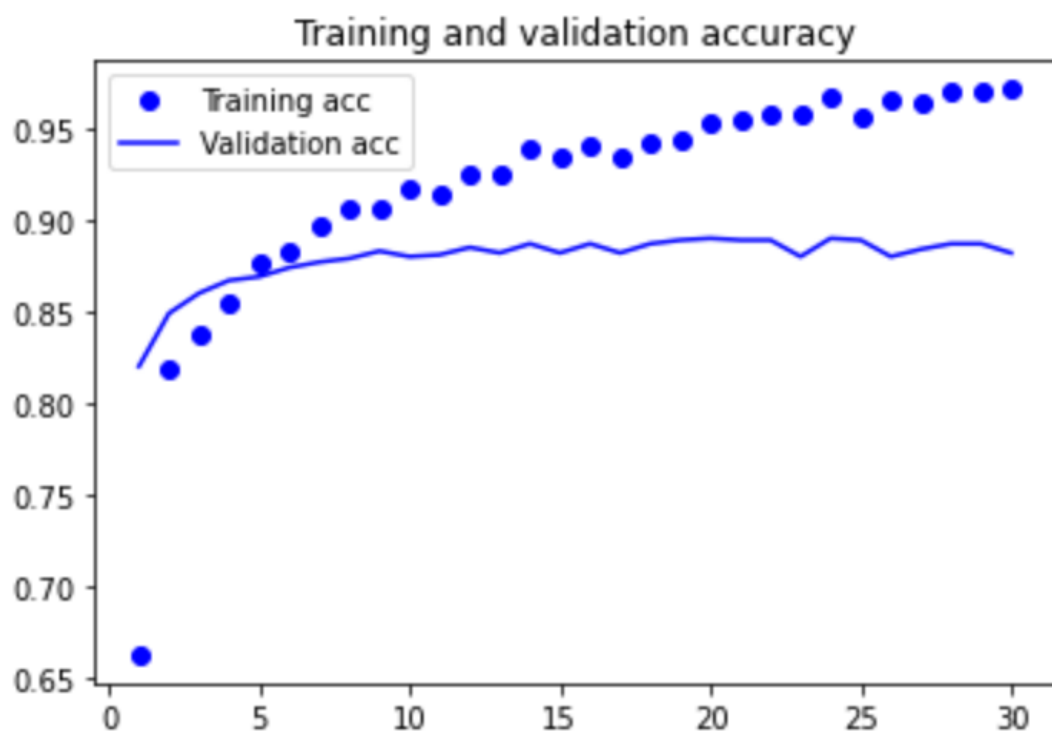
To summarize, the choice of the network has a significant impact on performance.   Using a pretrained network can significantly improve performance. Also, increasing the training sample size can slightly improve performance, but the improvement is limited after a certain point. The pretrained model outperformed the model trained from scratch in this case.