Automated Diagnosis of Pigmented Skin Lesions

TEAM 37

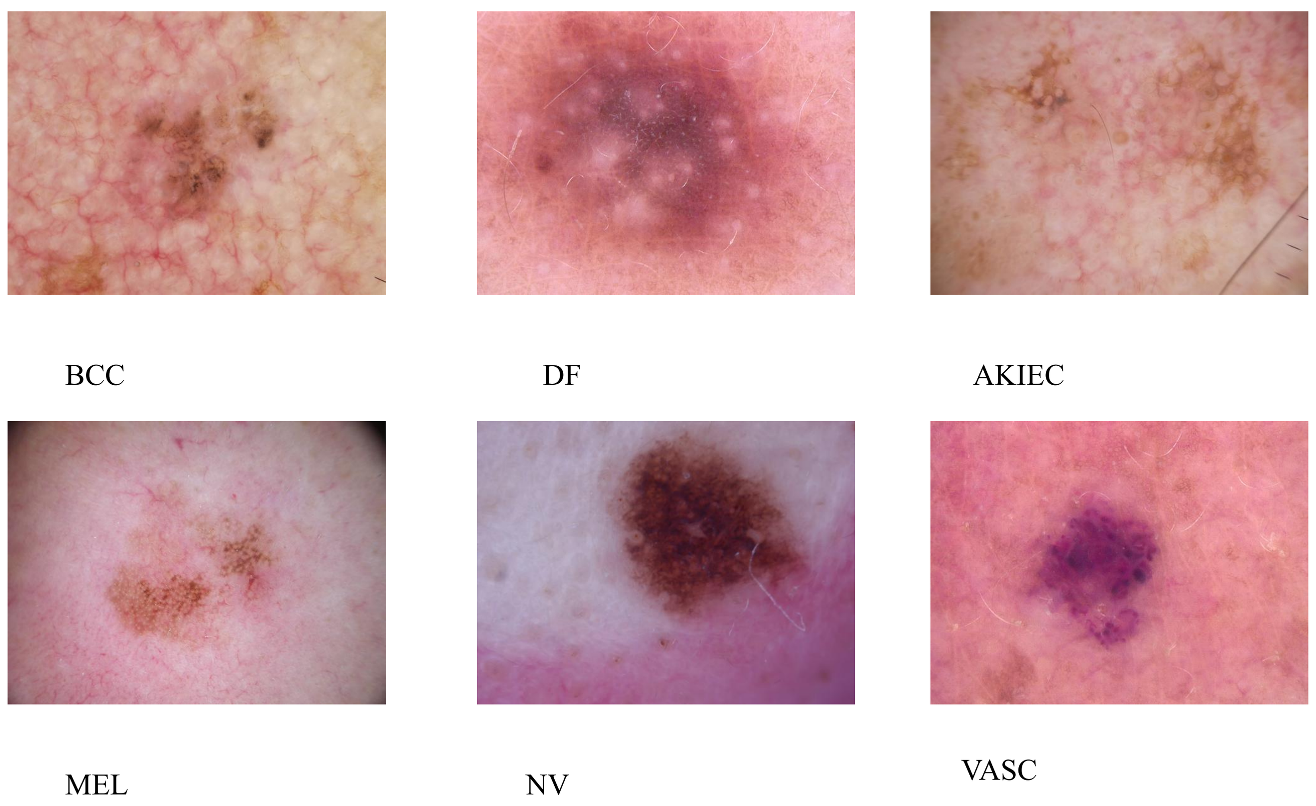
Noopur Khachane, Lucas Lin, Raul Pegan, Louise Xu

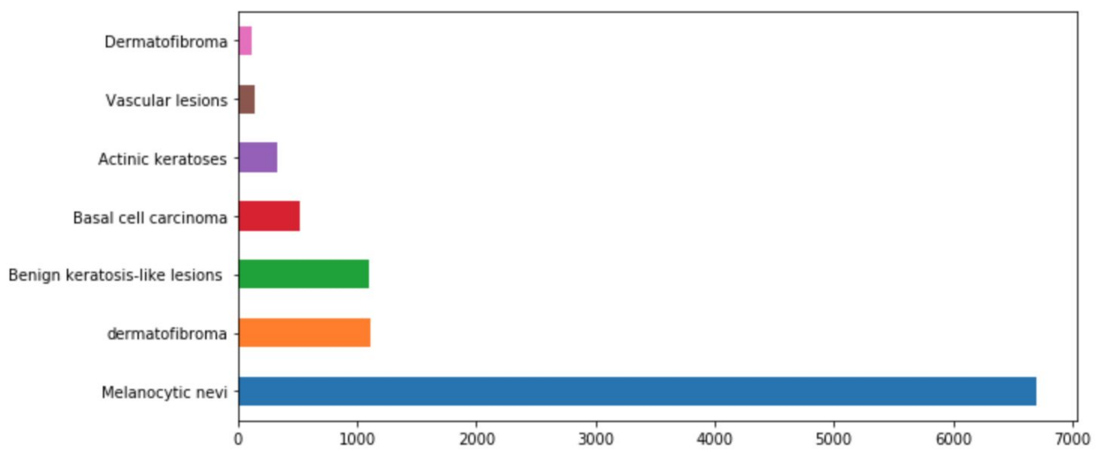
Abstract

Automated diagnosis of skin lesions has become increasingly popular, with many people using mobile applications to determine whether their moles are malignant or benign. CNNs have been found to be quite successful in classifier skin cancer. However, as it is a medical issue, it is important that the reliability of these models be verified. Unfortunately many studies boast high performance classifiers, however the results are incomparable due to lack of transparency. Common methods include using transfer-learning with common architectures such as VGG and Resnet, as well as using CNNs as feature extractors and adding specialized classifiers to the end. We will test the various methods highlighted in the a review paper of skin lesions against a common dataset of skin lesions.

Dataset

The HAM10000 dataset is originally from Harvard Dataverse and contains 10,015 dermatoscopic images. The dataset contains 7 possible labels and other metadata including the body part the lesion is on, how the label was determined, age, etc. The largest issue with the dataset is the incredibly uneven distribution of images across labels.





Preprocessing

The training data for each model was augmented in the same manner:

- RGB values are rescaled to values 0-1
- randomly sheared (intensity= [0-0.2])
- randomly zoomed (x0.8-x1.2)
- randomly flipped horizontally

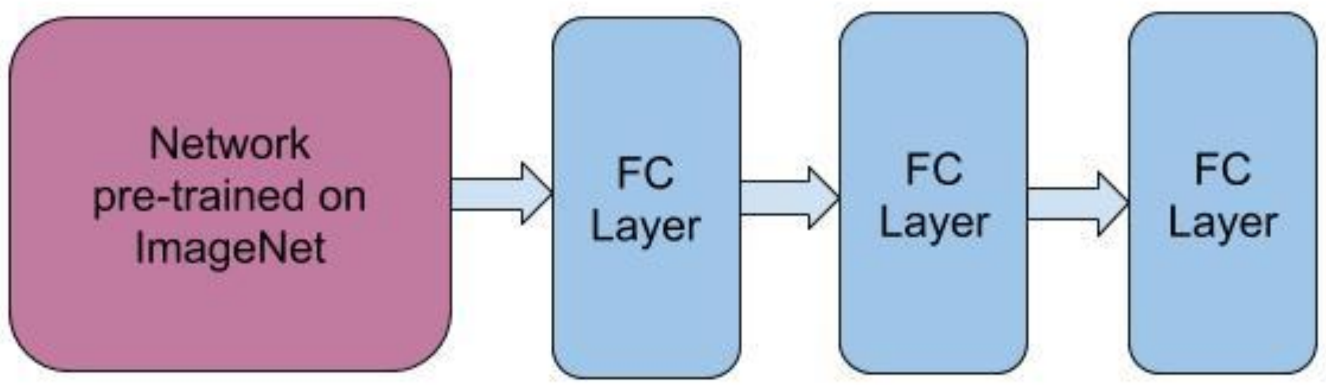
Additionally, all images are resized in accordance to the input dimensions of each model.

Method

- Transfer learning on ResNet, VGG, Inception
  - We implemented transfer learning using several different popular architectures trained on ImageNet. We then trained a set of dense layers to specify the network for our particular problem.
- Fine-tuned models
  - We then fine tune the entire model by un-freezing the pretrained network and running a full backprop

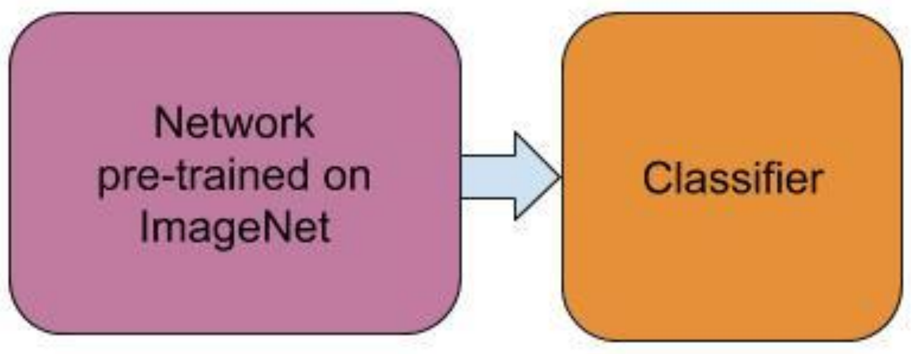
1.) Transfer Learning

- Pre-trained Networks include Resnet, Inception and VGG

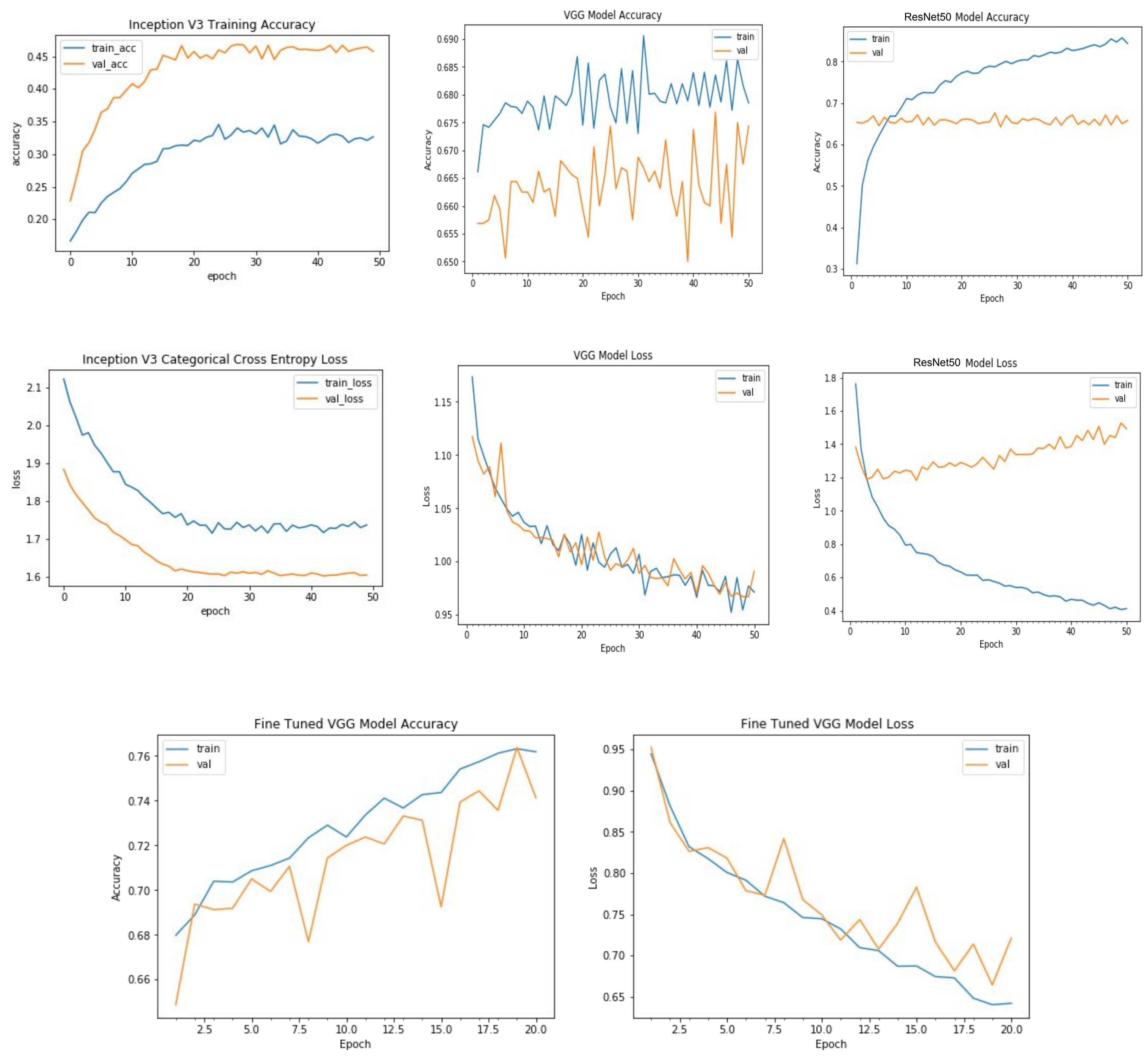


1.) CNN as a Feature Extraction

- Output of CNN is pass into classifier such as SVM or kNN



Results

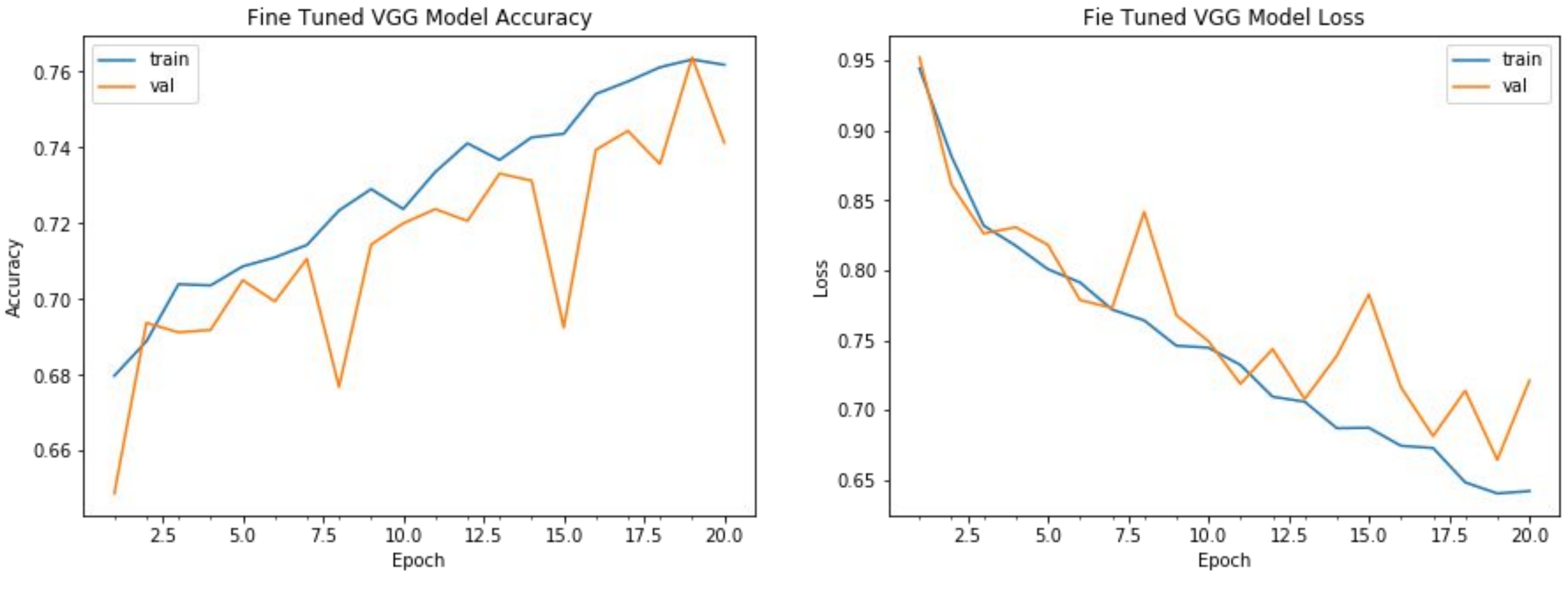


Conclusion

Overall, we were able to achieve decent test accuracies with the three models. The accuracies were not as high as in the papers that we referenced, but that it understandable since our dataset was relatively small compared to the datasets used in these papers. With some further improvements, such as data augmentation or fine-tuning, we may be able to achieve higher accuracies, as we can see from the improved VGG model.

References

[1] Very Deep Convolutional Networks for Large-Scale Image Recognition
https://arxiv.org/abs/1409.1556


Fine Tuned VGG Model Accuracy
train
val
Accuracy
0.76
0.74
0.72
0.70
0.68
0.66
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Epoch
Fie Tuned VGG Model Loss
train
val
Loss
0.95
0.90
0.85
0.80
0.75
0.70
0.65
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Epoch