CS783 : Visual Recognition

Rohin Garg
160583

Aaditya Singh
160002

Assignment 2: Image Classification

February 26, 2019

Baseline

For the baseline, we have used pre-trained Resnet18 network. For training, we did not train the resnet layers, only the fully connected layer in the end.
Accuracy on nearly 15% validation data was found to be 89% on fine-grain classification, and 98% on coarse-grain classification.

Improvements on Baseline

Data Augmentation

As the size of the dataset was quite small, we introduced data augmentation techniques like random crops, horizontal and vertical flips, and mean normalization to increase the size of our dataset.

End-to-end Training

Instead of training just the fully connected layers, we also tried training the resnet-18 layers, which resulted in a considerable improvement in accuracy.

Training on hard examples

We found all the examples to which the model was assigning a negative class with a large confidence. Then we trained the model on those examples again.

Reduced number of parameters

We hypothesize that after training the model for fine-grain clasification after using the aforementioned techniques, the model is less likely to make inter-class errors. Hence, we used the model for the fine-grain classification for the coarse-grain classification as well instead of using two separate networks for both the tasks.

By utilizing these approaches and random validation sets, the peak accuracy observed was nearly 93%. Finally, we submitted the final model trained on the entire dataset to avoid loss of data.

Other methods

We also tried replacing the 18-layer resnet by a 121-layer densenet for the coarse-grain classification, which still has a fewer number of parameters than the 34-layer resnet. Further, we used 1 as well as 2 fully connected layers along with data augmentation to improve the accuracy of this network, but this resulted in a decrease of validation accuracy, probably due to overfitting. Hence, we decided to improve upon our 18-layer resnet for the fine-grain classification. In both the approaches, only the fully connected layers were trained. The validation accuracy obtained using this network was 60%.