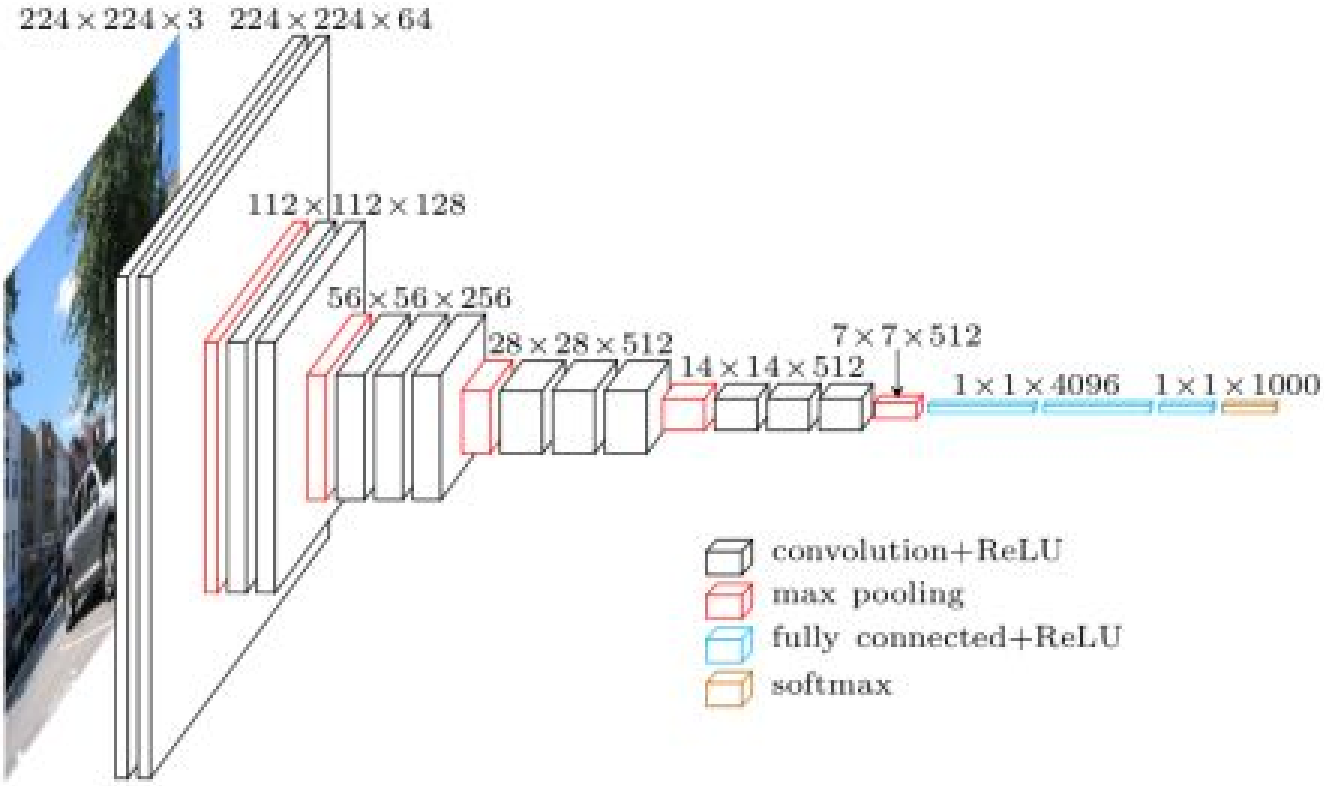


Object Classification

27.02.2019

Bollamreddi Hemanth (150195)
Shah Aalap Anil (150655)

CS783: Visual Recognition
Assignment 2

Method Used

Overview

The dataset consists of images of 5 objects. Each of the class contains varying number of fine classes. The coarser classes are easy to distinguish while some of the finer classes are difficult even for a human to recognize.

Approach

Approach 1 :

CNN's are known for their good feature extraction abilities. Further, ResNets have state of the art performance on image classification tasks.

Coarse Classification :

- We have taken an ResNet-34 (which satisfies the conditions given in the assignment) trained on the ImageNet and have extracted the convolution layers i.e removed the fully connected layers.
- This convolution layer has good starting weights (since it was trained on much bigger dataset).
- We have added a single layer fully connected layer with 5 outputs (no of coarse classes) and trained (fine tuned) it on our dataset. This will give a better feature extractor for our dataset.

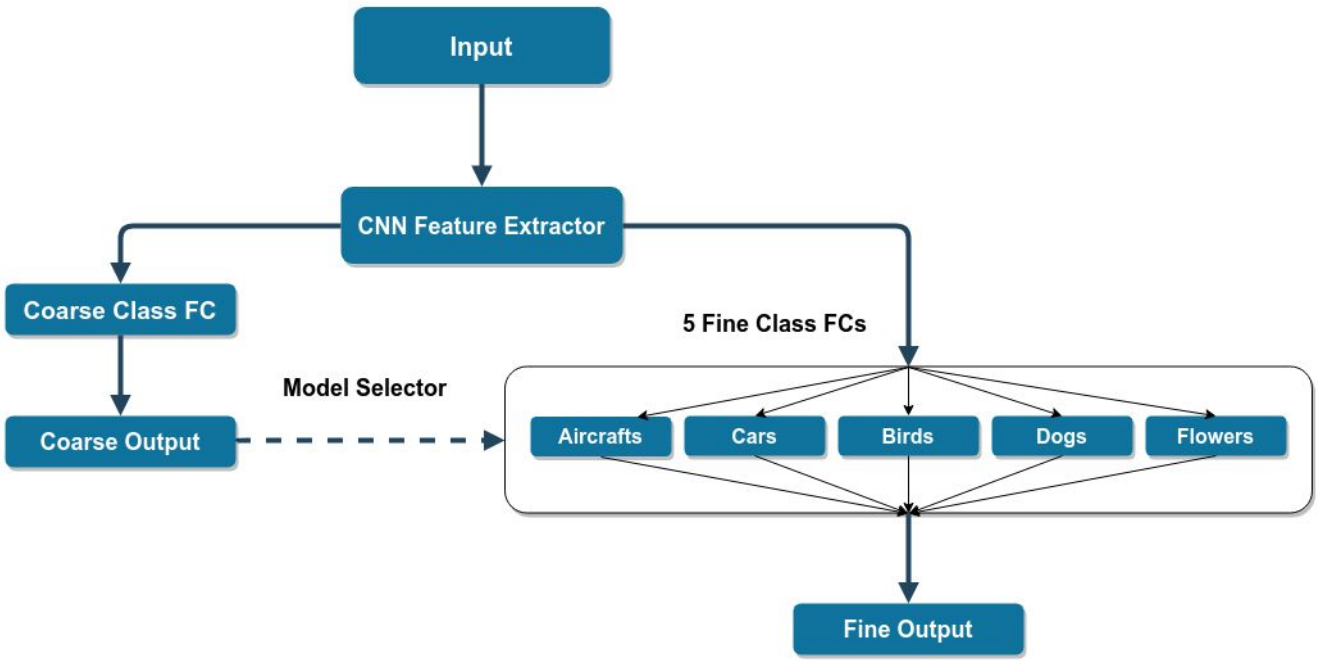
Fine Classification :

- The Convolution layer from the above part is now taken as the feature extractor and the weights are freezed.
- For ‘each’ coarse class, 3 fully connected layers have been added and the weights were trained.

Total Classification :

- An image is first passed through the Conv layer. This extracts the features of the image.
- The coarse class of the image is then known by passing the features to fully connected layer of Coarse Classification Model.
- Depending on the coarse class, from above step, we then pass the features to the appropriate, fine class fully connected layer. Which would give the fine class of the image.

Network :

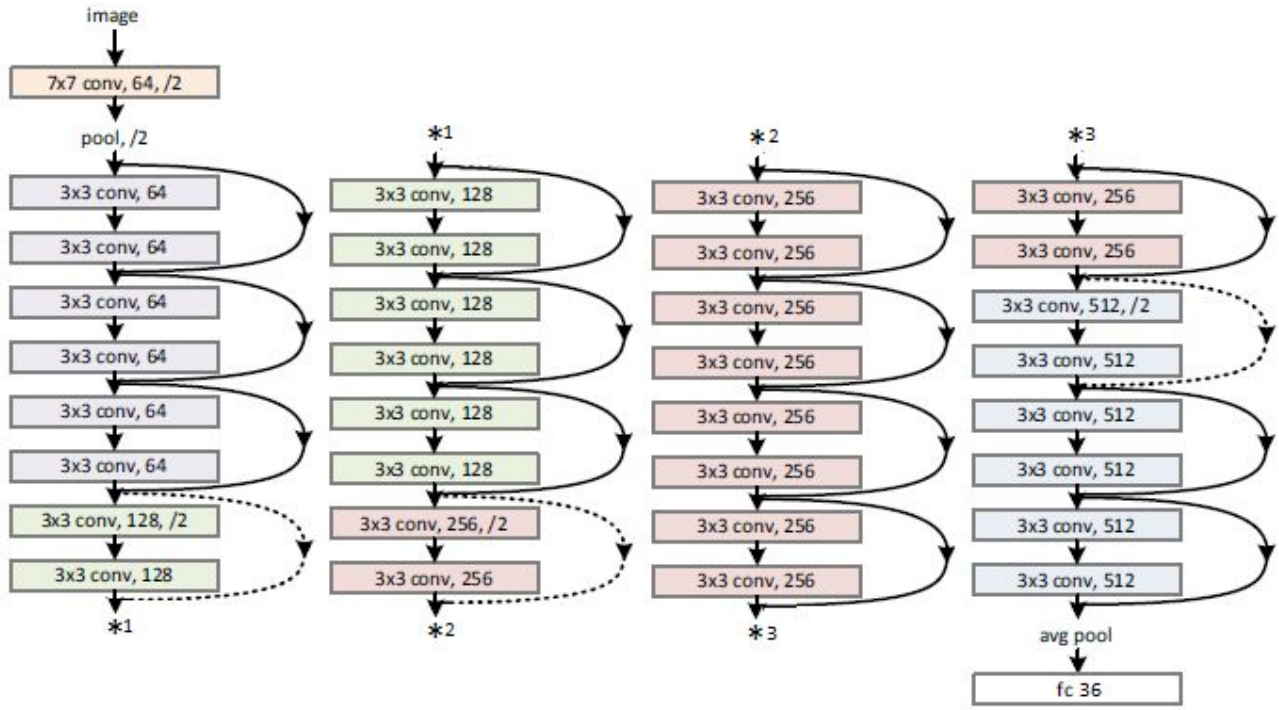


Approach 2 :

In this approach, we have considered the whole problem, as just a single classification problem with number of classes equals to sum of number of fine classes in all coarse classes. I.e It is considered as a '36' class classification problem.

The standard ResNet-34 trained on ImageNet dataset is used (transfer learning) for initial weights and the weights are then fine tuned for our dataset.

Network:



Implementation and Experimentation

Data Augmentation :

We have used several transformations including random rotate (-20, 20) degrees, scaling within (0.8, 1.2) and linear shifts by upto 0.2. This would result in a less biased dataset and therefore would improve accuracy at test time.

Approach 1 :

The first approach is partially inspired by the IEEE paper titled "CNN with coarse-to-fine layer for hierarchical classification".

The coarse accuracy was ~ 100 % for just single fully connected layer. Hence we did not try more fc layers. But for individual fine networks, three fc layers are used. The bottleneck in this approach is the fine networks, as coarse networks were giving almost 100 % accuracy.

Approach 2 :

The second approach is a standard ResNet-34, but with a custom fully connected layer, to reduce the no of parameters to be less than the limit (twice that of ResNet-18).

We have tried ResNet-18 and ResNet-34 for the convolution layers (feature extractors) for both the approaches. ResNet-34 gave better performance. A simple explanation, is that deeper layers give more representative features of that class.

Parameters

We are using a ResNet-34 as reference model. We are removing the original fc layer and added out custom fc layers which is smaller than the original fc layer. This means our model have lesser parameters than the standard ResNet-34 which in itself has less than twice the no of parameters in ResNet-18.

Thus the number of parameters are well within the limit specified in the assignment (twice of standard ResNet-18).

Results

Approach 1:

Coarse	99 %
Aircraft Fine	93 %
Bird Fine	74 %
Car Fine	54 %
Dog Fine	96 %
Flower Fine	98 %
Overall	**85 %**

Approach 2:

Coarse	99 %
Overall	**89 %**