Large Scale Image Classification

April 29, 2017

Indian Institute Of Information Technology,Allahabad

Members :

IIT2014101 - Yerram Reddy M

IIT2014115 - B. Raju

IIM2014006 - Ashok

Mentor

Dr.Satish Kumar Singh

Candidate's Declaration

We (IIT2014101, IIT2014115, IIM2014006) hereby certify that the work which is being presented in the B.Tech. Mini Project Report entitled "Large Scale Image Classification ", being submitted as a part of End Semester Project Evaluation to the Department of Information Technology of Indian Institute Of Information Technology, Allahabad, is an authentic record of our original work under the guidance and supervision of Dr. Satish Kumar Singh from Jan 17 to May 2017. We have adequately cited and referenced the original sources and have adhered to all principles of academic honesty and integrity.

Supervisor's Certificate

This is to certify that the statement made by the candidates is correct to the best of my knowledge and belief. The project titled "Large Scale Image Classification " is a record of candidates' work carried out by them under my guidance and supervision. I do hereby recommend that it should be accepted in the fulfilment of the requirements of the 6th semester mini project at IIIT Allahabad.

Dr.Satish Kumar Singh

ABSTRACT :

In this work, we have used Convolutional Neural Networks(CNNs) trained on GPUs for classifying images in the Mini Place dataset.We try to train a relatively deep network with a large number of filters per convolutional layer to achieve a high accuracy on the test dataset.

INTRODUCTION :

This section introduces our problem, and the overall plan for approaching our problem.

The goal of this project is to identify the scene category depicted in a photograph. For each image, algorithms will produce a list of at most 5 scene categories in descending order of confidence. The quality of a labeling will be evaluated based on the label that best matches the ground truth label for the image. The idea is to allow an algorithm to identify multiple scene categories in an image given that many environments have multi-labels (e.g. a bar can also be a restaurant) and that humans often describe a place using different words (e.g. forest path, forest, woods). The exact details of the evaluation are available on the Places2 challenge website.

We are planning to train CNN(Convolutional Neural Network) on this dataset and measures the performance of the network. We should improve the classification accuracy of their network models on the validation set of mini places challenge.

PROBLEM STATEMENT :

This section describes our problem precisely specifying the dataset to be used, expected results and evaluation.

The data for this task comes from the Mini Place Dataset from MIT computer vision course project. Specifically, the dataset consisting of 100,000 images for training,

10,000 images for validation and 10,000 images for testing coming from 100 scene categories. The images will be resized to 128*128 to make the data more manageable.

==========
Overview of challenge dataset
==========

There are three types of image data for this competition, all coming from the larger Places2 dataset: training data (TRAINING), validation data (VALIDATION), and test (TEST). There is no overlap in the three sources of data: TRAINING, VALIDATION, and TEST. All three sets of data contain images of 100 categories of scenes.

	Number of images		
Dataset	TRAIN	VALIDATION	TEST
Mini Places	100,000	10,000	10,000

Every image in training, validation and test sets has a single image-level label specifying the presence of one object category. Challenge database statistics:

Training:
- 100,000 images, with 1000 images per category

Validation:
- 10,000 images, with 100 images per category

Test:
- 10,000 images, with 100 images per category

Evaluation Metrics :

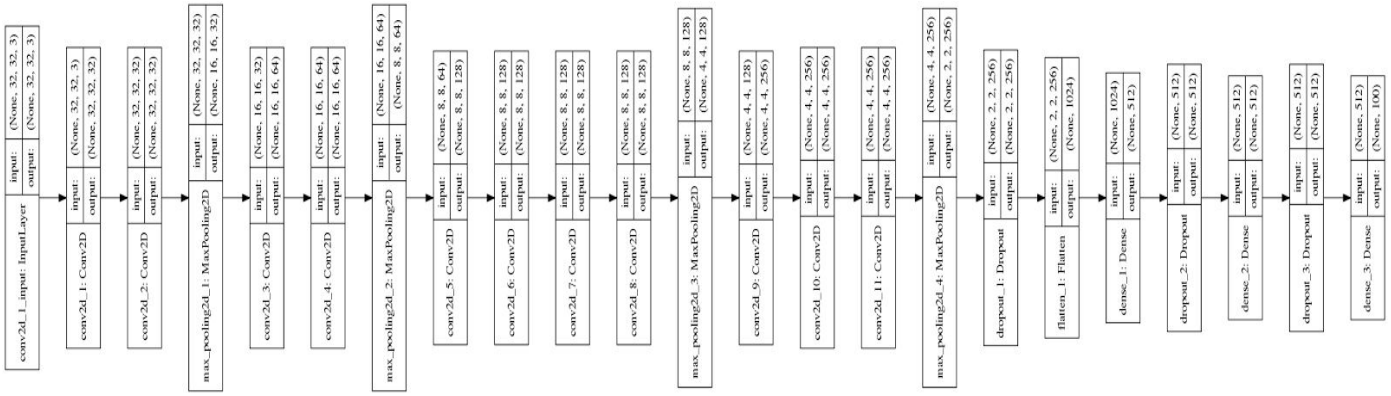
Top-*k* error is the percentage of times the correct scene label is not one of the labels the network considers the *k* best. We calculate Top-5 error on evaluation set because we don't have labels for test set.Top Results from challenge :The winning

network's scores were 41.07% and 13.93% top-1 and top-5 errors respectively.Refer this *http://miniplaces.csail.mit.edu/leaderboard.php* for more information on best results.

TECHNICAL APPROACH :

This section describe the methods we intend to apply to solve this problem.

Recent days deep learning outperforming traditional computer vision method if they have more data to train.So, we decided to use CNN for our task.We have designed a convolutional network for our dataset which is inspired from classical VGG16 model.Below is our network. First 2 conv layers use 32 3x3 filters , next 2 conv layers use 64 3x3 filters , next 4 conv layers uses 128 3x3 filters and next 3 conv layers uses 256 3x3 filters and in all pooling layers we have used 2x2 max pooling and last 2 dense layers are having 512 neurons and at last we have used softmax layer.Training carefullly with proper value for dropout layers we achieved following results.



(100000, 32, 32, 3) train samples,(10000, 32, 32, 3) test samples

(100000,) train samples labels,(10000,) test samples labels

Total params: 2,896,260.0

Trainable params: 2,896,260.0

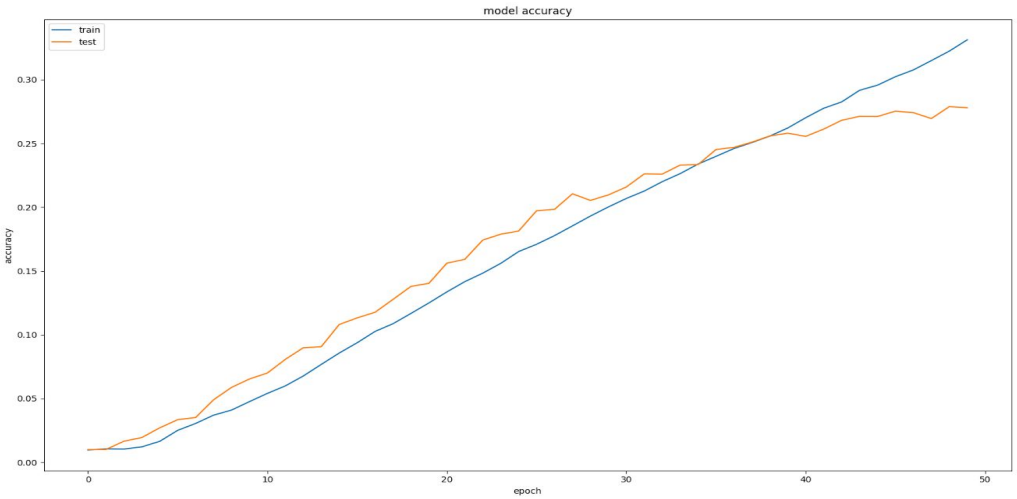
We have implemented the above CNN in keras with tensorflow as backend.During preprocessing we have subtracted R G B planes with 103.939, 116.779, 123.68 in every image.We trained it using stochastic gradient descent with 0.01 learning rate and 0.5

momentum on our GPU which is having 4GB NVIDIA 940MX graphic card with 349 core(8GB RAM) , it takes 1 hour 30 mins per epoch. Since to train network effectively , one should at least train it for 50-200 epochs. That means we need to train our days on our system which may cause the system to die if we don't take proper care. So we decided to resize each image to 32x32x3 so that network parameters decreases to 2.9M from 33M (if we use 128x128x3) which can be manageable on our system . Then we trained network with augmented dataset for 50 epochs which has taken 3 hours 30 mins on our system.

RESULTS :

This section states and evaluate our results up to the end semester.

The Final Result we got using above network 0.3814 top-1, 0.7013 top-5 accuracies on train set and 0.2780 top-1 and 0.5633 top-5 accuracies on test set. Below one is the accuracy vs epoch graph .But the best results on this challenge were 0.5893 top-1 and 0.8607 on test set. Since the result was poor , we continued training upto 150 epochs which leads to 0.8213 top-1 and 0.9796 top-5 on train . 0.2358 top-1 and 0.4894 top-5 on validation set . That means model overfitting the train set . So to evaluate the the model we trained it on cifar100 and cifar10 its giving fairly good results compared to state of art. At final we concluded that resizing 128x128x3 to 32x32x3 leads to information loss in image.



CONCLUSION / FUTURE WORK :

In this work our model have not given good results .So, in future we would improve these results and will meet the state of art .

REFERENCES :

1). http://image-net.org/challenges/LSVRC/2016/

2). http://www.di.ens.fr/willow/pdfs/cvpr06b.pdf

3). http://6.869.csail.mit.edu/fa15/challenge/data.tar.gz

4). https://silp.iiita.ac.in/wordpress/?page_id=449

5). http://cs231n.stanford.edu/

6).https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks