Proposal of Image Classification based on CNN Cloud

Tianyang Liu
Electrical and Computer Engineering
University of Florida
Gainesville, US
dotasheepliu@ufl.edu

Yicheng Wang
Electrical and Computer Engineering
University of Florida
Gainesville, US
yichengwang125@ufl.edu

Su Pu
Electrical and Computer Engineering
University of Florida
Gainesville, US
Soupsoup88@ufl.edu

***Abstract*— ImageNet [1] provides huge datasets of images in different categories. While local machines do not have enough computational and storage resource to process them, cloud computing has the power. Convolutional Neural Network (CNN) was proposed in 2012 [2] which receives wide concerns due to its outstanding performance compared to traditional ones, and has potential advantages in image classification. In this project we intend to train a CNN on Amazon AWS to recognize different images. Spark and TensorFlow will be run on the EC2 host and four GPU-enhanced slaves will follow coordination. The host will also be connected to a S3 storage instance as well as a Route 53 to provide web services to the public.**

Keywords—cloud computing; ImageNet; EC2

I. Introduction

Analyzing and classifying pictures is quite a heating topic in nowadays computer vision field. This idea which evolved from artificial network explains how the machine is going to understand and learn the objects from the pictures. The associated database ImageNet which leaded by Feifei Li is the most prestigious and largest research resources. The Large Scale Visual Recognition Challenge (ILSVRC) competition is based on this database and is hold every year since 2010. The result from every year competition is improving very fast and the overall accuracy can reach more than 95%. However, the requirement of the computation resources is very critical, both calculation and storage take up great amount of time and space. Fortunately, the cloud computing is quite mature to help the researchers deal with high performance calculation. Cloud computing relieves the pressure of computing resources when processing big data. By using third party data centers, the users and enterprises from different fields can share the same resources and dynamically change their request. Cloud computing saves money for the companies from buying servers and provides them enough space for higher-demand computation. Based on these technology, our project plans to realize the image classification using CNN neuron network method from a previous paper proposed by Alex Krizhevsky research group. In our proposal, the first part explains some previous work on images processing, and then the explains the platform EC2 from Amazon as well as the Spark framework. The following part describes the CNN in the associated paper [3], [4], [5].

II. Platform

Our cloud would be deployed on Amazon AWS. Specifically, in the CNN training process, a EC2 general purpose host running Spark and TensorFlow should continuously segments and distributes pictures to four GPU-enhanced slaves, and then each slave instance computes, does pooling and sends results back to the host to improve the CNN hyper-parameters. Most CNN is trained on a single node, but the advantage to distribute in this pattern and use MapReduce, is to increase the training speed. As the training completes, we connect EC2 host to Route53 and S3 to make it a webpage-based APP that can be accessed by the public. As users input an image, it goes through the CNN stored in EC2, and the following outputted key words indicate the recognition result. If time permits, we can further store enough poems in the S3. When the cloud recognizes the picture, it does not only tell the user what it is, but also writes an elegant snippet.

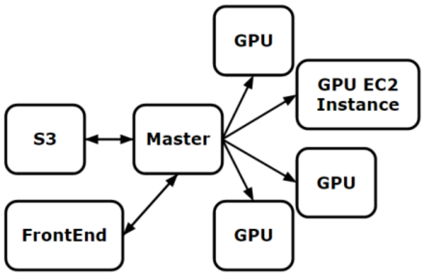

Fig. 1. Structure of the Platform

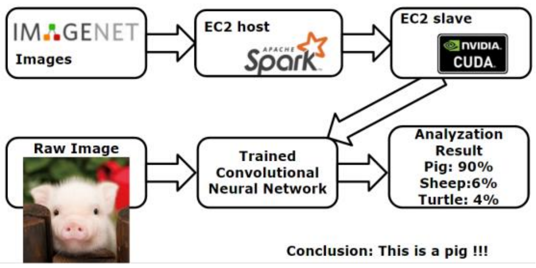

Fig. 2. Execution Flow Chart

As for expenses, most will be cost in the training part. g2.2xlarge provides half the performance of NVIDIA K520 i.e. 1536 cores and 4GB visual memory, as well as whole 8 cores of an Intel Xeon E5-2670 processor, thus four slaves make 2.6 dollars per hour. When the training is done, we discard the slave nodes and concern the response of host server to our clients. The general purpose host could be degraded from M series to T series which deals better with sudden visit but are much more economic.

III. Database

The database we plan to use is ImageNet, which contains 15 million images in 20k categories. However, the total size of the database is 1.2 TB, we would definitely not design a system to classify all these objects as freshmen.

Besides, ImageNet is the official database for the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), which evaluates algorithms for object detection as well as image classification at large scale.

Our purpose is to classify one million images into 1000 categories, and the expect accuracy is 90%.

IV. The Alogrithm

While traditional multilayer perceptron (MLP) shown in Fig.3, models were successfully used for image recognition, the full connection algorithm would lead to the curse of dimensionality. For example, if the size of input image is 200*200*3 (200 columns, 200 rows, 3 color channels), the weights for input layer to first hidden layer would be 120,000*n (n is the number of neurons in the first hidden layer). While as for these kind of input images, the numbers of hidden layers and neurons in each hidden layer are tremendous.

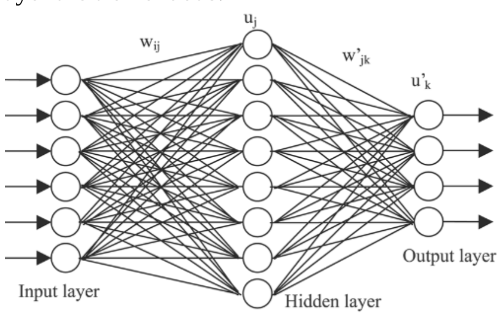

Fig. 3. Structure of MLP

The full connectivity of neurons is wasteful in the framework of image recognition, and the huge number of parameters quickly leads to overfitting. A simple ConvNet mainly contains 3 layers: Convolutional Layer, Pooling Layer, and the Fully-Connected Layer.

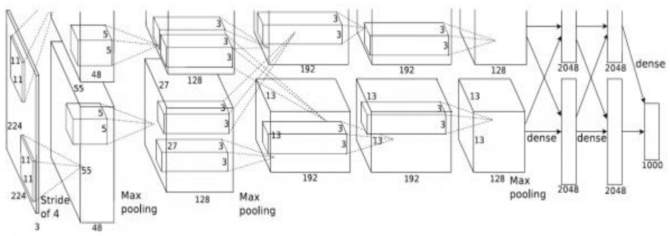

Fig. 4. Structure of CNN

A. Convolutional Layer

We call the layer convolutional because it is related to convolution of two signals:

$$f[x,y] * g[x,y] = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} f[n_1,n_2] \cdot g[x-n_1, y-n_2]$$

which means elementwise multiplication and sum of a filter and the signal.

The layer's parameters consist of a set of kernels that extend through the full depth. So the kernels could be regarded as 3-dimensional filters, and each filter gets a 2-dimensional activation map. For example, in Fig.4, the author applied 96 filters to the first layer, the depth of the output is 96.

B. Pooling Layer

The idea behind this layer is that once the feature is found, the exact location is not as important as the rough location relative to other feature. So this is a non-linear down sample layer, which makes the representations smaller and more manageable. For example, the pooling algorithm in Fig.5 is max pooling, which means to segment the whole image into non-overlapping region, and output of each sub-region is the maximum inside. It operates over each activation map independently, so the depth will not change.

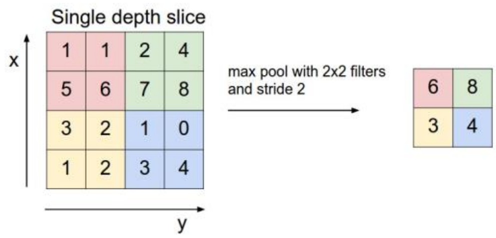

Fig. 5. Max pooling

C. Fully-Connected Layer

Just like the connection method in ordinary NN, the neurons of one layer connect all the neurons in previous layer, which means there are N*M parameters (N is the number of neurons in present layer, and M for the previous layer).

V. Conclusion

This proposal points out the possibility and urgent need to put CNN based image classification process on the cloud. The structure of our system features one EC2 host connected to four GPU-enhanced slaves, as well as a S3 for storage and Route53 for web services. It also gives specifications on three main steps of CNN training: convolution, pooling, and fully connection. We believe the feasibility of this project and if time permits, hope to give it more attractive functions and run it for longer periods.

References

[1] http://image-net.org/

[2] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

[3] Perronnin F, Dance C. Fisher kernels on visual vocabularies for image categorization[C]//2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007: 1-8.

[4] Sánchez J, Perronnin F. High-dimensional signature compression for large-scale image classification[C]//Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011: 1665-1672.

[5] Boser, Bernhard E., Isabelle M. Guyon, and Vladimir N. Vapnik. "A training algorithm for optimal margin classifiers." Proceedings of the fifth annual workshop on Computational learning theory. ACM, 1992.

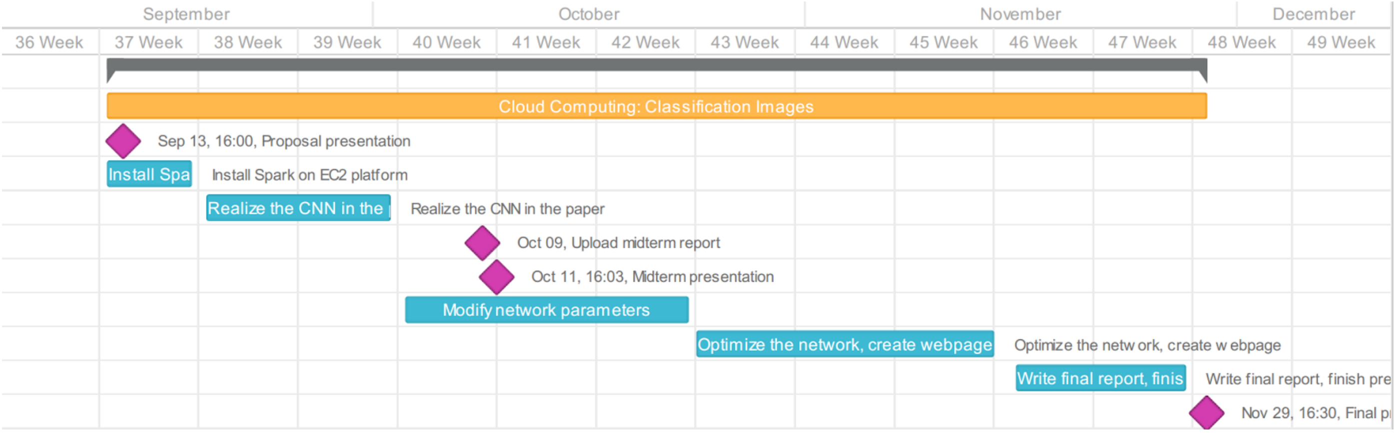

Fig. 6. Time arrangement for the project