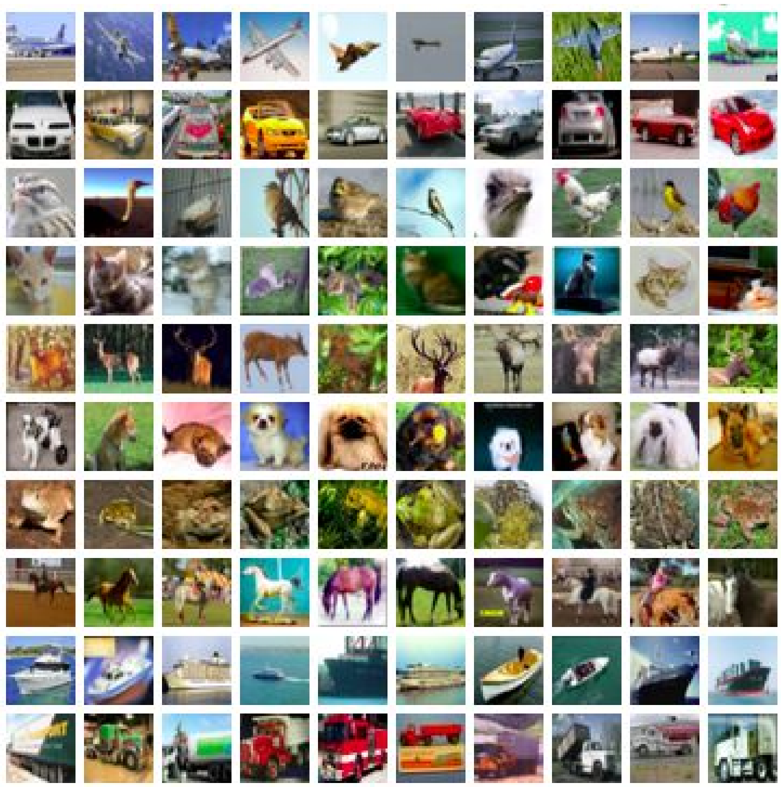


Image-Classification- CIFAR-10

15.04.2019

BNS Vishal(21,160911122), Adarsh R(11,160911090), Vaibhav Dhar(12,160911094)
IT - ‘A’ , Batch-A1
Manipal Institute Of Technology
Manipal

Abstract

In this project, we work on image classification of the CIFAR-10 dataset using the supervised machine learning technique - Convolution Neural Networks. The dataset consists of 60,000 32x32RGB images containing one of 10 object classes, with 6000 images per class. We use cross-validation by splitting the 60,000 data samples into 50,000 training samples and 1,000 validation/test samples to select the optimized hyperparameters for each parametric classifier. Training is done for 125 epochs. The neural network yields performance with a training accuracy above 91.53% and a validation accuracy of 90.210%.

Introduction

The CIFAR-10 dataset is a labeled subset of the 80 million tiny images dataset. They were collected by Alex Krizhevsky, Vinod Nair and Geoffrey Hinton. The CIFAR-10 contains of 60,000 32x32 color images of 10 classes, with 6000 for each class.The 10 different classes represent airplanes, cars, birds, cats, deer, dogs, frogs, horses, ships and trucks. The image on the cover page shows the sample images from the dataset. Over the years, a lot of works have been reported regarding the image classification problem with CIFAR-10 dataset. The highest accuracy so far is achieved by using efficient training of giant neural networks using pipeline parallelism (Nov 2018) that yields 99.0% accuracy on the test set.

In this project, the CIFAR-10 dataset was divided into 50,000 labeled training images and 10,000 validation/ test images to select the best hyperparameters. The classifier trained on the 50,000 samples with the optimized parameters was then used to predict on the testing set and evaluate the prediction accuracy. After carefully designing the network structures and tuning the hyperparameters, we got the best performance with 8 convolution layers along with max-pooling, batch normalization and dropouts between the convolution layers. The network showed a training accuracy of 91.53, and a testing accuracy of 90.210%.

Methodology

The general convolutional neural network (CNN) consists of multiple layers that transform the input image volume into an output volume holding the class scores. The several distinct types of layers are convolutional layer, RELU layer(Activation Layer), MAXPOOL layer and fully-connected layer. For a convolutional neural network, the most important layer is the convolution layer, where each entry in the output volume can be interpreted as an output of a neuron that looks at only a small region in the input and shares parameters with neurons in the same activation map. A lot of breakthroughs for image classification have been made with deep CNN by stacking more and more convolutional layers.

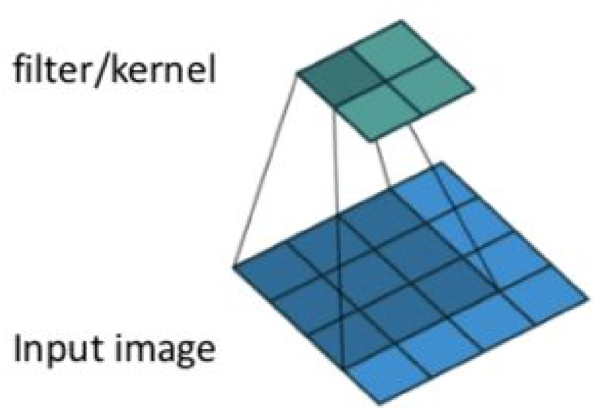
AlexNet is the name of a convolutional neural network, designed by Alex Krizhevsky, and published with Ilya Sutskever and Krizhevsky's PhD advisor Geoffrey Hinton, who was originally resistant to the idea of his student.

AlexNet contained eight layers: the first five were convolution layers, some of them followed by max-pooling layers, and the last three were fully connected layers. It used the non-saturation ReLU action function.

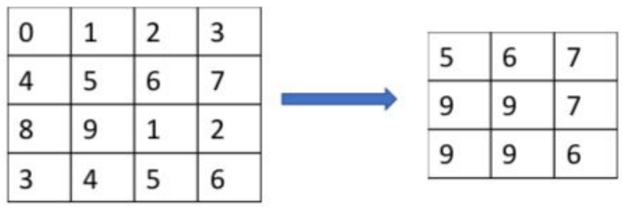
In the network used in this project, it is just an extension of the AlexNet where few more layers are added and optimized. There are eight convolution layers, three MaxPooling, and one fully-connected layers, some of them followed after another. The activation function used in this network is ELU which gave a better accuracy when trained against the ReLU Layer.

Definition and working of each layer:

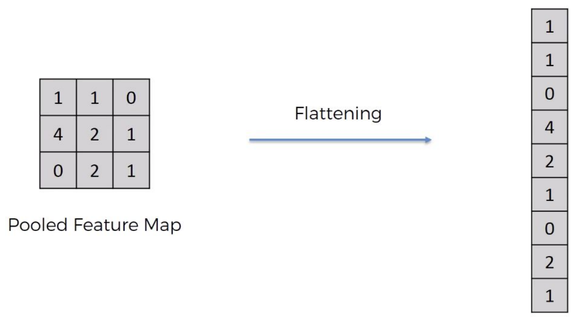
Convolution Layer: Convolution operations extract features of the input image via different filters/ kernels. The training algorithm tunes the kernel values to learn features that are effective for the tasks of interest.



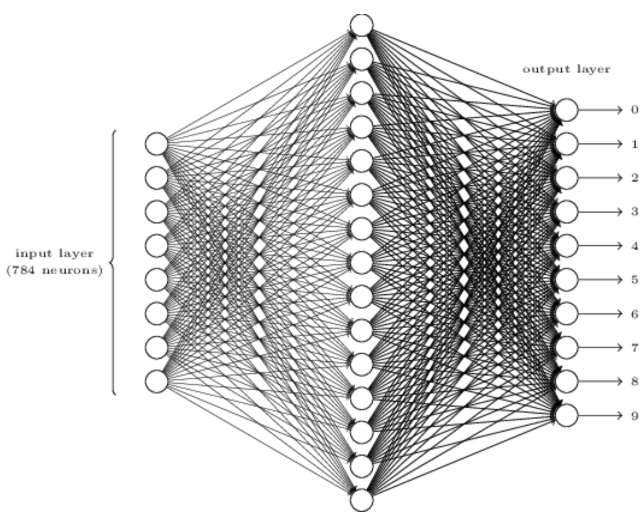
Max-Pooling: It is basically aggregating information. It is similar to the Convolution Layer but without any filter/kernel. We select maximum of each 2X2 matrix in the total image pixel matrix and get a final small max-pooled matrix. Here in the following image, consider a 4X4 pixel matrix, we use a 2X2 pool_size and after max-pooling, we get 3X3 matrix.



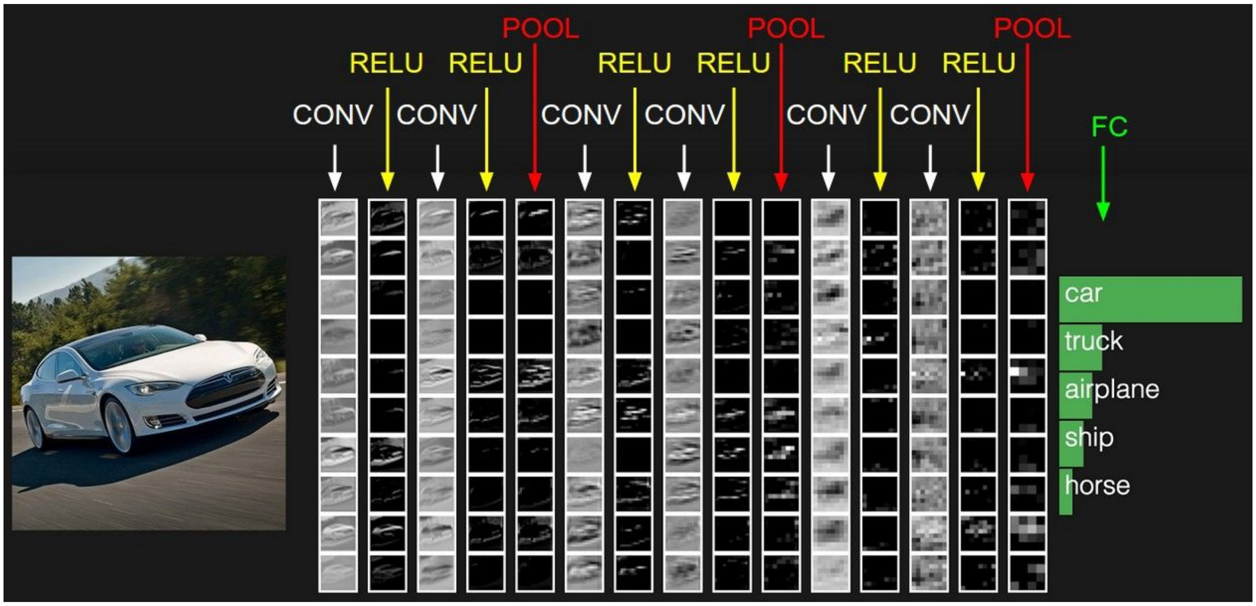
Flattening: The pooled image is then flattened i.e. made into a (NXN, 1) vector which is the input to the Fully Connected Artificial Neural Network.



Fully Connected Layer/ Dense: Here, the input is the flattened pixel matrices and is passed through a hidden layer called the Fully Connected Layer that is used to assign weights to each feature/input value. In Artificial Neural network, the weights are tuned to decide the matrix values' dependency on classifying the image as one of the 10 object class.



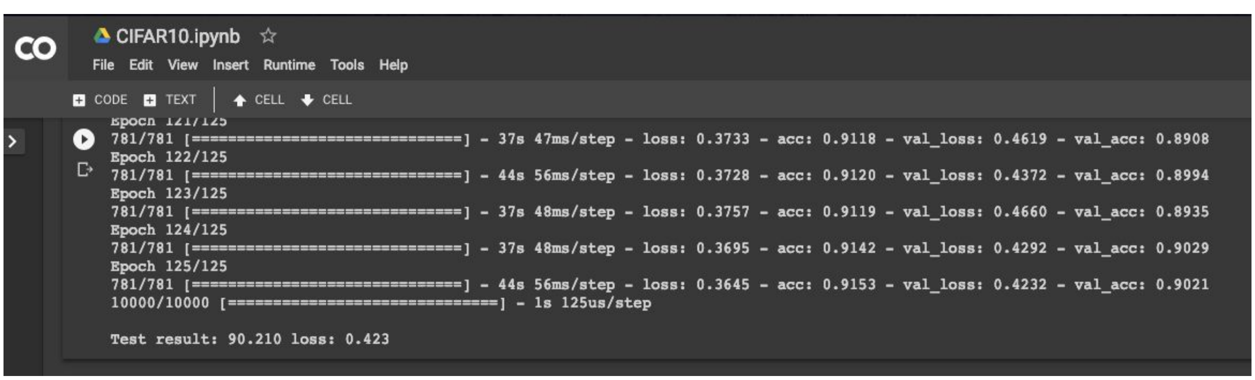
The final Architecture of the model can be visualized from the following image:



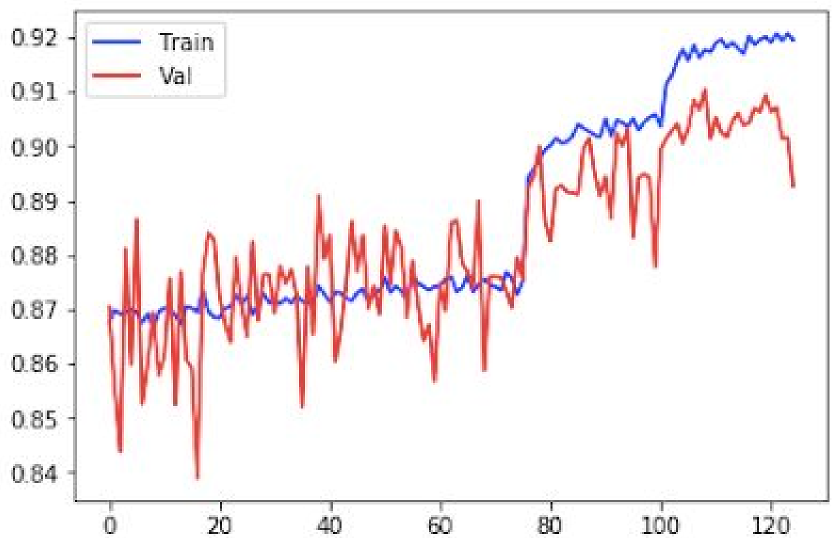
Results and Graphs

Upon training it for 125 epochs, we get training accuracy of 91.53% and test accuracy of 90.210%. As we can see from the following images, it can be interpreted that the error decreases after every epoch and the training and test accuracy improve after each epoch.

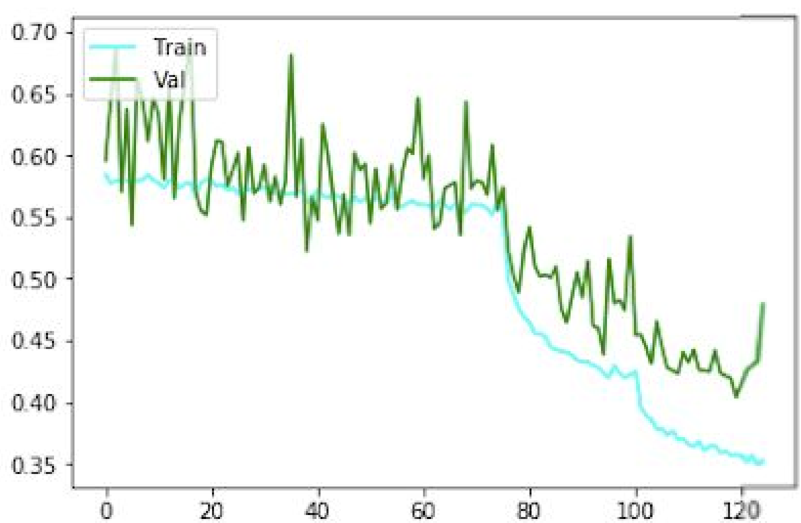
Final Accuracy output:



GRAPH-1: Training and Test Accuracy per Epoch:



GRAPH-2: Error per Epoch:



Conclusion

In this project, we achieved the best classification accuracy of the CIFAR-10 dataset with eight deep convolution neural network. Modified CNN, the network we made with batch normalization, max-pooling and dropouts gave us an accuracy of 90.21%.

Besides, we learned that the control of overfitting is important for classifiers with large set of parameters. To control overfitting, we have used dropout and batch normalization between the convolution layers.

This Image classification has great scope and utilizations in many fields. Currently, Image Classification used for live object detection is utilized for self driving cars, and is also used by Google in their Google Lens.

Literature Survey/ References

[1]Krizhevsky, Alex, Vinod Nair, and Geoffrey Hinton. "The CIFAR-10 dataset." online: http://www.cs.toronto.edu/kriz/cifar. html (2014).

[2]https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

[3] Convolution Neural Networks - deeplearning.ai (Coursera)

[4] https://machinelearningmastery.com/crash-course-convolutional-neural-networks/

[5] Super Data Science - Convolution Neural Networks using Python

[6] https://ieeexplore.ieee.org/document/8308186