CS534 : Traffic Sign Recognition using Convolutional Neural Network

Chaitanya Mitash (cm1074)

The project involves using CNNs for Traffic sign recognition.

I. PART I: TESTING WITH THE BASE MODEL

The base model is a 6-layer CNN with one fully connected layer. On training this CNN with the 31367 images, an accuracy of **97.48%** was achieved.

II. PART II: VGG ARCHITECTURE

The 16 layer network with 2 fully connected layers, when trained completely using the given training images gives an accuracy of **96.10%**. On loading the pre-trained weights from Imagenet, the accuracy changes to **94.97%**

III. PART III: BATCH NORMALIZATION

On using batch normalization before every non-linear activation, the accuracy increases to **98.66%**.

IV. PART IV: QUESTIONS

- Epoch is one pass through the entire dataset. Iteration is a single update, computed over a batch of data. Epoch contains several iterations, depending on the batch size.
- Learning Rate is essentially the size of the step by which the weights are altered based on the computed gradient (corresponding the loss function). Increasing the learning rate can help the network quickly converge/learn quickly. But in case there is a lot of variability in the dataset, it might not learn well or at all, because it will be guided drastically by the latest update.
- Batch size is the number of data that is considered for computing the loss and thus the gradient in each iteration. If the memory is unlimited, the ideal batch size would be the size of the training set. In that case each epoch will have just one batch.
- the feature map shrinks to half the dimension when using a 2X2 max pooling layer. The depth of the feature map does not change. If the input feature map is 48X48, the output feature map after 3 2X2 max-pooling layers would be 6X6. This is assuming, the convolution layers use padding such that the size of the input does not change. In reality, for example in our baseline model, the size change occurs as:
CONV2D (32 filters) - 48,48,32
CONV2D (32 filter, no padding) - 46,46,32
MAXPOOL - 23,23,32
CONV2D (64 filters) - 23,23,64
CONV2D (64 filters, no padding) - 21,21,64
MAXPOOL - 10,10,64
CONV2D (128 filters) - 10,10,128
CONV2D (128 filters, no padding) - 8,8,128
MAXPOOL - 4,4,128

- dropout means randomly disabling certain nodes of the neural network, to learn multiple independent representations. This helps avoid over-fitting and thus generalizes better.

V. HOW TO RUN

- Download data from https://drive.google.com/file/d/0B6tuOE-7NjZATzFtZmsxeFk5WlU/view?usp=sharing
- extract the data inside the project folder `TrafficSignRecognition`
- Run *python datagen.py*
- to test baseline model run command:
python test_baseline.py
- to test VGG model with no pre-training run command:
python test_enhancement_1_fullTrain.py
- to test VGG model with pre-training run command:
python test_enhancement_1_PreTrain.py
- to test baseline model with batch-normalization run command:
python test_enhancement_2.py