MACHINE LEARNING PROJECT REPORT

Submitted by: Chirag Pabbarju, Kathan Kashiparekh, Panshul Garg

Problem: Identifying number pattern from Google's Street View Housing Numbers data-set.

Baseline paper followed: Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks, Ian J. Goodfellow,Yaroslav Bulatov,Julian Ibarz,Sacha Arnoud,Vinay Shet

Basic Outline of network: The implementation uses an end to end deep learning model using convolution neural networks (CNN). Convolutions are used as feature extractors that try to look for specific features in the images which are then feed into 6 fully connected layers which then use a softmax layer to classify/predict each digit in the image. The first classifier predicts the number of digits in the image and the next 5 predict what those digits are. Obviously,only those number of digits are taken as predicted by the first softmax.

Timeline and an Analysis of the implemented models:

Model 1: The first model implemented had the following features:
- Images were used in the RGB format.
- Filter size was 5*5 with adequate padding. Pooling filter size was 2*2.
- 4 convolution layers were used in the following sequence: Conv+Relu-Conv+Relu+Pool-Conv+Relu-Conv+Relu+Pool.
- Each softmax classifier had one fully connected layer with 1024 nodes in the hidden layer followed by the softmax layer.
- Learning rate in Adam Optimizer was fixed at 1e-4.
- Batch size of 50 was used for training.
- 6 Epochs were used.

Model 2: The second model had the following features:
- Images were used in the grayscale format because we felt that features were being better expressed in grayscale format.
- Each softmax classifier now has 2 fully connected layers,each with 1024 hidden nodes.
- 4 convolution layers were used in the following sequence: Conv+Relu-Conv+Relu+Pool-Conv+Relu-Conv+Relu+Pool.
- Learning rate for Adam Optimizer was fixed at 1e-4.
- Batch size of 50 was used.
- Filter size of 5*5 was used for convolution.
- 6 epochs were used.

Model 3: The third model had the following features:
- 5 convolution layers were used in the following sequence: Conv+Relu-Conv+Relu+Pool-Conv+Relu-Conv+Relu+Pool-Conv+Relu+Pool.
- Filter size of 3*3 was used.
- 12 epochs were used.
- A fully connected layer of 1024 hidden neurons was put after the feature extractors and before branching to the softmax fully connected.

- Filter depth was 32,64,128,256,512 incrementally.
- Dropout was introduced before the softmax with dropout probability 0.5
- Dynamic learning rate was added which decreased as number of epochs increased. It was 1e-4 for 1-6 epochs,1e-5 for 7-10 and 1e-6 for the rest.
- First and third convolution were batch normalized.

Model 4: The fourth had the following features:

- Hard negative mining was added to the above model.
- Batch norm was added to all the conv layers.

Model 5: The fifth had the following features:

- L2 regularization was added to the above model to prevent possible overfitting of data as is seen in the validation accuracy of ~83% of 4th model but on test data it was 72%.
- Test accuracy was ~75%.

MODEL NUMBER	VALIDATION ACCURACY
1	~69%
2	~74%
3	~78%
4	~83%
5	~78%