How CNNs perform Sketch Classification?
Sushan Bhattachar (sushanb)
Stanford University

INTRODUCTION
Due to the exponential growth in touch sensitive devices, the number of drawings/sketches/outlines has increased a lot.
People tend to use sketches to express their ideas and emotions quickly and without hassle.
We trained three Convolutional models for doing sketch classification.
We achieved an accuracy of 92% on Resnet model and 35.40% on Baseline model.
Models were implemented in Keras/Tensorflow.

DATASET
The project uses the publicly available Sketchy database, the first large-scale collection of sketch-photo pairs. [1]
It contains 20000 sketches objects across 125 categories.
The dataset is balanced.
There is presence of high intra class variation and low inter class variation which makes classification task challenging.

MODELS
Baseline Model
- It is a 3 layer convolutional network followed by 2 layer of fully connected network.
- The convolutional network consists of actual convolution, ReLU activation, batch normalization and downsampling operation.
- We train the baseline model from scratch for 80 epochs.
- We used multi class cross entropy loss and RMS optimizer for training the model.
- It took 40 minutes to train the model with two NVIDIA K80 GPUs.

Advanced Model 1 (Finetuning Pretrained ResNet)
- ResNet by He et al. to solve vanishing and exploding gradient problem when we use deep convolutional network [2]
- ResNet is only composed of small sized filters of 1x1 and 3x3 with the exception of first convolutional layer of kernel size 7 x 7.
- Every single layer is followed by batch normalization and ReLU activation.
- Due to lack of extensive computation resources, we used the truncated version of ResNet of 50 layers.
- The main difference between ResNet and Standard Neural net is that it contains layers that learn residual functions with the reference to the layer inputs.
- We trained the 50 layer ResNet for 12 epochs using multi class cross entropy loss and Adam Optimizer with 0.001 learning rate.

Models/Results

RESULTS

Figure 1: channel to 3 channel

Advanced Model 2 (DenseNet - Results pending)
- Huang et al. introduced Dense Convolutional Network called as DenseNet where each layer connect to every other layer in a feed-forward fashion [3].
- In traditional convolutional networks with L layers, we have L-1 connections between DenseNet we have L+L(L-1)/2 connections.
- We trained the 121 layer ResNet for 12 epochs using multi class cross entropy loss and Adam Optimizer with 0.001 learning rate.

FUTURE WORK
- Blend/Ensemble numerous and neural methods
- Add Attention mechanism on the top of ResNet and DenseNet
- K-fold validation and Search for fine hyper parameter tuning
- As an experiment, augment the data by zoom, flip, width and height shift
- Continue to tune hyper parameters, schedule learning rates for each epoch.

REFERENCES