CS 298 Proposal

Geetika Bansal
San Jose State University
Faculty Advisor: Dr. Chris Pollett

Neural net Captcha Cracker

Neural networks have achieved tremendous success recently. They have achieved state of the art performance in many tasks like object detection, speech recognition.

In the field of computer vision convolutional neural networks have been used for tasks like object detection and segmentation. RNNs have been used in sequence prediction tasks like language translation and handwriting generation (alex et al). Oriol et al recently combined the two networks for image caption generation, to get better than state of the art performance on various datasets like Pascal and Flickr30k, when measured using BLEU-1 score. They used a convolutional neural network to digest an image's content into a vector of floats and then used LSTM (Long Short Term Memory, a popular variant of RNN) to generate captions (sequence of words). The entire network was trained together on Imagenet.

I propose to use a similar approach for captcha recognition, where the task is to predict the sequence of characters in a CAPTCHA image. A convolutional neural network with alternate convolutional and max pool layers will be used to learn features at progressively higher level and to digest all the features to a vector. An LSTM will then be used to generate the sequence of characters terminated by a special character "</S>" predicted by the network. The LSTM will maintain its internal hidden and control states. At every iteration it will be fed with the vector generated from the convolutional neural network.

Since the problem is similar to object detection and simpler than image caption generation where neural networks have shown great performance I expect the network to give good results.

Since neural networks are really powerful, they require a lot of training data to avoid overfitting. I plan on generating training data using libraries like SkimpyGimpy. Furthermore I will enhance the data set by synthetically creating training data for instance by rotating characters and by dropping character pixels.

Project Deliverables:

Training Data:

As discussed above supervised systems like neural networks require a lot of training data to avoid overfitting. I plan on generating this data using libraries like SkimpyGimpy and then would do image manipulation to synthetically create even more data.

Trained Model:

I plan on using Theano framework to train the neural network using the training data as mentioned move. Theano framework has become really popular in the neural networks community because of the flexibility it provides in creating different networks.

Inference Tool For CAPTCHA recognition:

After the model is trained I will develop a tool to load the model for inference for CAPTCHA recognition for demo. If time permits I will wrap this in an http server so that we could use the tool through a web browser.

CS298 Report:

Written project report containing all the details of the project.

Innovative and challenging aspects of Project:

To my knowledge convolutional neural networks with LSTM haven't been tried before for CAPTCHA recognition. Since these are fairly new models there isn't a lot of literature on them. Most of the research is being done in industry and thus the tips and tricks to train these models aren't openly shared. It would be challenging to understand how these models work and then to train these models which so far has been done by PHDs with years of experience in this domain. Also it would be challenging to understand frameworks like Theano to train these models and then use for inference.

Schedule:

September mid : Training Data
October mid : Trained Model
November mid : Inference Tool For CAPTCHA recognition
December : CS298 Report

References:

[2015]Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan ,Show and Tell: A Neural Image Caption Generator

[2014]Alex Graves Generating Sequences With Recurrent Neural Networks

[2014]Ilya Sutskever, Oriol Vinyals, Quoc V Le, Sequence to Sequence Learning WIth Neural Networks