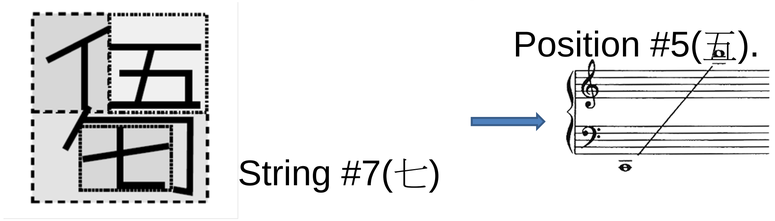
# CS229 Guqin Notation and Music Style Recognition

Chen Shi pupushi@stanford.edu
Stanford University

## Predication

Inspired by *Convolutional Neural Networks* about digit recognition in the class, I did this project on Guqin Notation translation, which converted reduced characters with positioned 1-10 (一 -十) t in handwritten Chinese characters to other music recording method. Previous work used radical extraction for handwritten recognition. [1]




## Data

Handwritten music sheets were read in by text detection, and cropped out an adjusted region and saved as black and white 224*224 images. I also randomly produced (一 -十) characters in different fonts to enlarge the training pool.

## Features

For each black and white image, position of the cropped region was recorded in the file name by text detection.

Now each single image should only contain a single unit of digit information. 96 convolutional kernels were learned by the first convolutional layer on the 224*224*3 input images.

## Models

Basically I built a four-layer convolutional neural network with ReLUs, for 10 classifiers using CaffeNet[2] CNN models.

A neuron's output f as a function of its input x is f(x) = max(0; x) as non-saturating nonlinearity model.

To reduce overfitting, I performed PCA with assigned randomized variables (αi) to alter the intensities of the RGB channels in training images.

$$[\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3][\alpha_1\lambda_1, \alpha_2\lambda_2, \alpha_3\lambda_3]^T$$

where pi and ƛi are ith eigenvector and eigenvalue of the 3 *3 covariance matrix of RGB pixel values.

## Experimental Results

I managed to collect about 760 training images for each classifiers (1-10) and training errors range from (0% for 10 to 12.2% for 3). Below shows the average results from 10 classifiers:

(Accuracy rate per classifier will be discussed in final report.)

| Model | Training Sample Size | Training Error | Test Sample Size | Test Error |
|---|---|---|---|---|
| CNN | 7600 | 4.5% | 213 | 15.7% |

## Discussion

I am quite excited the model does successfully recognize 一(1) to 十(10) with an average 85% accuracy rate. To simplify my experiments, I did adjust the strategy to have a two step approach (text detection[3] + CNN) as there were some awesome work done in both fields and also saved computational time. Originally, I expected to tell the style differences between the schools of Guqin playing from 2 schools of notations I collected. It ended up that the notations do not carry rhythm information and it is more up to each player to interpret the piece. This brought up an interesting direction that we can build a model to read in the recording and compose according to the style of it.

## Future

- Implement position into learning model
- Double chord recognition
- Introducing more notations on playing techniques
- Handwriting feature and App development

1. Enzhi,N et al. "A Radical Cascade Classifier for HandwrittenChinese Character Recognition" JOCCH Volume 3 Issue 3, March 2011
2. Krizhevsky, A et al. "ImageNet Classification with Deep Convolutional Neural Networks" NIPS 2012
3. Chen, Huizhong, et al. "Robust Text Detection in Natural Images with Edge-Enhanced Maximally Stable Extremal Regions." Image Processing (ICIP), 2011 18th IEEE International Conference