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.

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

\[
[p_1, p_2, p_3][\alpha_1 \lambda_1, \alpha_2 \lambda_2, \alpha_3 \lambda_3]^T
\]

where \(p_i\) and \(\lambda_i\) are ith eigenvector and eigenvalue of the 3 * 3 covariance matrix of RGB pixel values.

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 + CNN) as there were some awesome work done in both fields and also saved computational time.

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:

<table>
<thead>
<tr>
<th>Model</th>
<th>Training Sample Size</th>
<th>Training Error</th>
<th>Test Sample Size</th>
<th>Test Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>7600</td>
<td>4.5%</td>
<td>213</td>
<td>15.7%</td>
</tr>
</tbody>
</table>