Chinese MNIST

Classifying Chinese MNIST Characters

The MNIST database (or Modified National Institute of Standards and Technology database) is a large repository of handwritten digits that represent the "Hello World" of image recognition. In 2017, Dr. K Nazarpour and Dr. M Chen of Newcastle University contributed a Chinese version of the MNIST and a tagged copy was made available on Kaggle website in Sep 2020.

This project attempts to classify Chinese numeral characters into their 15 categories using the Kaggle dataset. The characters were handwritten by 100 volunteers who provided 10 sets of 15 characters each. The dataset contains 15000 scanned images in jpg, each of size 64x64.

Deep Learning Models

Two deep learning models were considered.

1. A Dense model, comprising 2 Dense layers and a final Dense layer for classification output

2. A Convolution-MaxPooling-Dense model, comprising a stack of 4 Convolution and Maxpooling layers, followed by 2 Dense layers and a final Dense layer for classification

Outcomes

Test accuracy results for the two models were as follows.

Model	Test Accuracy
Model 1	83.6%
Model 2	99.1%

Data Sources

Data used in this project was retrieved from Kaggle:
https://www.kaggle.com/gpreda/chinese-mnist

Original data source from Newcastle University:
https://data.ncl.ac.uk/articles/Handwritten_Chinese_Numbers/10280831/1

Original MNIST Dataset: http://yann.lecun.com/exdb/mnist/