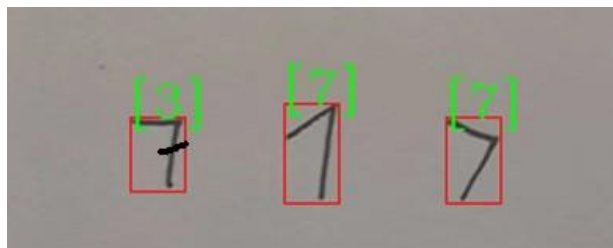


大作业一：自定义数据集实现手写数字识别

一、作业背景

MNIST 数据集是手写数字识别问题中最经典的数据集之一，许多模型在该数据集上取得了良好效果。但由于其来自美国国家标准与技术研究所，数字写法不够丰富，不符合中国人的书写习惯，在某些情况下难以识别出正确效果，如下图。



故丰富 MNIST 数据集使之可以识别更多种写法的手写数字，从而提升识别效果成为了当务之急。

二、作业要求

本次作业的主要任务是使用扩充后的手写数字数据集训练一个全连接神经网络模型 (BP)，使之适应中国人的书写习惯，提升识别效果。

作业应有三部分组成，数据采集与处理，模型训练，模型交叉验证，结果可视化验证。

1、数据采集与处理部分要求：收集自己用手机拍照 0 到 9 的手写数字图片，将其处理为 mnist 格式 (28x28)，添加进训练数据集；

2、模型训练部分要求：使用上述数据集进行模型训练，得到手写数字识别的全连接神经网络；

3、模型交叉验证部分要求：使用混淆矩阵，提升交叉验证的效果，要求在可视化部分展示出添加自己拍照的手写数字前后，（即扩充 mnist 数据集前后）的混淆矩阵，以及 5 折交叉验证效果；

4、结果可视化验证部分要求：使用新的手写数字图片检验模型效果，并进行可视化展示。

三、作业提交

10 月 20 日 23:59 前提交作业到邮箱：coursestj@163.com。作业文件名格式：姓名-学号-第 1 次大作业。