


山东大学计算机科学与技术学院

机器学习与模式识别课程实验报告

学号：201600181073	姓名：唐超	班级：智能 16
实验题目：手势识别		
实验学时：		实验日期：
<p>实验目的：</p> <p style="padding-left: 20px;">输入一张手势图片，输出这张图片所代表的手势数字（0~5）。</p> <div style="display: flex; justify-content: space-around; align-items: center;"><div style="text-align: center;"> 0</div><div style="text-align: center;"> 1</div><div style="text-align: center;"> 2</div><div style="text-align: center;"> 3</div><div style="text-align: center;"> 4</div><div style="text-align: center;"> 5</div></div>		
硬件环境：		
软件环境： Python3.6+tensorflow+keras+jupyternotebook		
<p>实验步骤与内容：</p> <h3>1. 数据集划分</h3> <p>下载 Turkey Ankara Ayrancı Anadolu High School's Sign Language Digits Dataset，对每类手势图片的 80%作为训练集，20%作为验证集。为了检测模型的效果，另外下载 111 张复杂背景手势图片作为测试。</p> <h3>2. 肤色提取</h3> <p>由于数据集中手势的拍摄背景都是简单的白色背景，为了使模型在复杂的背景下仍能对手势进行识别。在训练之前对每张图片进行肤色提取转化为手势（白色）、背景（黑色）的二值图，在每次测试时，也同时将图片进行肤色提取再进行预测。</p> <h3>3. 数据增强</h3> <p>数据集中手势往往正立，位于图片正中心，手的大小也变化不大，这样的数据集容易造成训练结果过拟合，为了增强模型的鲁棒性，在训练时，对每张图片随机旋转一定角度（0~180 度）、拉伸（1~1.2 倍）、水平翻转、平移等操作。部分数据增强结果如下：</p> <div style="display: flex; justify-content: space-around; align-items: flex-end;"></div>		

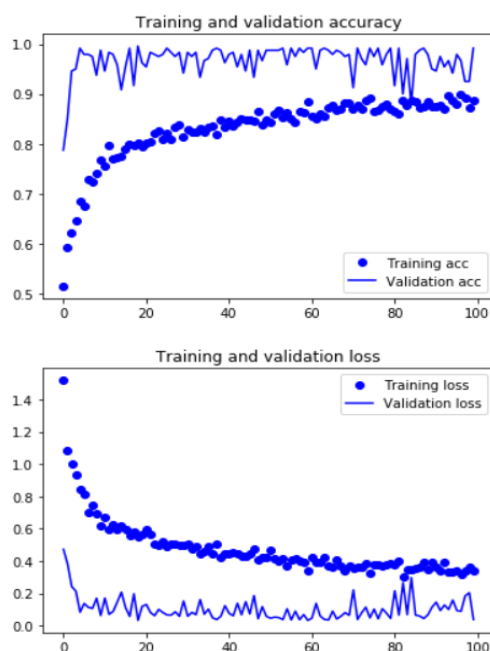
4. 构建网络

4*卷积层、4*池化层、1*flatten、1*dropout、2*全连接。具体如下：

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 198, 198, 32)	896
max_pooling2d_1 (MaxPooling2)	(None, 99, 99, 32)	0
conv2d_2 (Conv2D)	(None, 97, 97, 64)	18496
max_pooling2d_2 (MaxPooling2)	(None, 48, 48, 64)	0
conv2d_3 (Conv2D)	(None, 46, 46, 128)	73856
max_pooling2d_3 (MaxPooling2)	(None, 23, 23, 128)	0
conv2d_4 (Conv2D)	(None, 21, 21, 128)	147584
max_pooling2d_4 (MaxPooling2)	(None, 10, 10, 128)	0
flatten_1 (Flatten)	(None, 12800)	0
dropout_1 (Dropout)	(None, 12800)	0
dense_1 (Dense)	(None, 512)	6554112
dense_2 (Dense)	(None, 6)	3078

5. 训练

将类别交叉熵 (categorical_crossentropy) 作为 loss 函数, batchsize=10, epochs=100, 进行训练, 结果可视化如下:



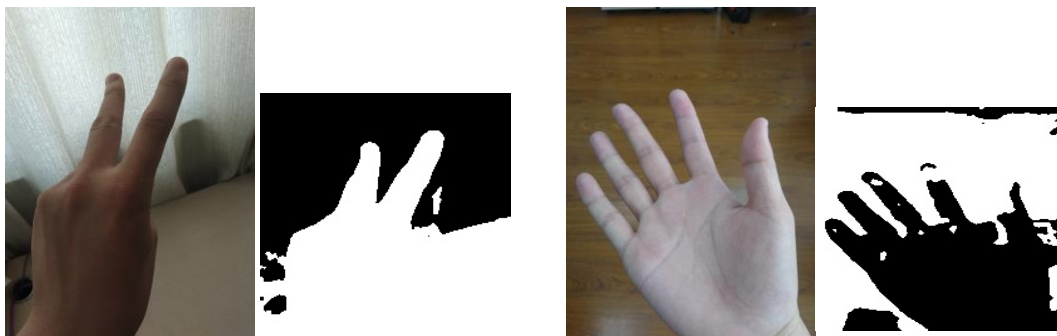
6. 测试

6.1 图片

对另外下载的 111 张图片不进行肤色提取，用没有进行肤色提取的原始数据集训练的模型进行测试，准确率仅为 35%。

采用肤色提取进行训练并测试，准确率提高到 58%。

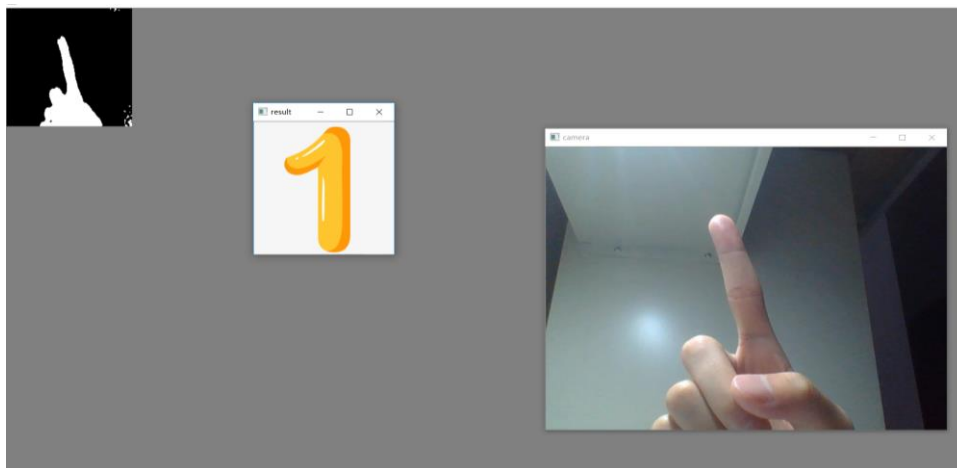
查看测试集时发现，有一部分图片因为背景的颜色与肤色相似，导致肤色提取的结果并不好（如下图所示），删去几张明显由于背景里有肤色相近颜色导致识别结果不好的测试样本，准确率进一步提高到 72%。



可以看出，肤色提取会显著提升复杂背景下的手势识别的准确率，但也有一定的局限，当测试图片中有与肤色颜色相似的背景时，肤色提取的效果非常不好，这样的情况下得到正确结果就更难了。

6.2 电脑摄像头

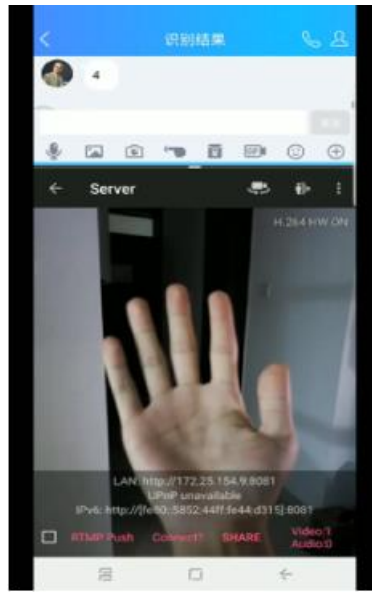
对摄像头接受的每一帧做肤色提取，并对该帧进行手势预测，并实时返回结果。对于没有强光和其他肤色干扰的环境，有着极佳的测试效果。



6.3 手机端

在摄像头识别的基础上，利用 python 调用 win32 的接口与系统交互发送 qq 消息和 ip 摄像头实现了手机端的手势实时识别。

程序调用手机上的 ip 摄像头，对每一帧进行预测，每 10 帧统计预测结果（0~5），返回票数最多的结果，将结果粘贴至剪贴板，并通过 qq 发送至手机。



结论分析与体会：

第一次自己用神经网络完成了一个识别任务，在没有强光和其他肤色干扰的环境，有着极佳的测试效果，但对于这种有干扰的情况下，效果就很差了，另外也没有考虑到测试距离较远的情况，总的来说，还有很大的改进空间。

附录：程序源代码

见邮件其他附件。