CNN 算法实现数字识别建模

一、问题描述

以课堂上讲的手写数字数据集 MNIST 为例，利用 CNN 去实现数字识别建模。

要求

1、在给定的数据集上添加至少 10 个自己用手机拍的 0-9 数字，补充到原数据集中，并显示你添加的数据；

2、提交 python 的 project 打包

3、提交实验报告；

二、数据集说明

1.MNIST 数据集

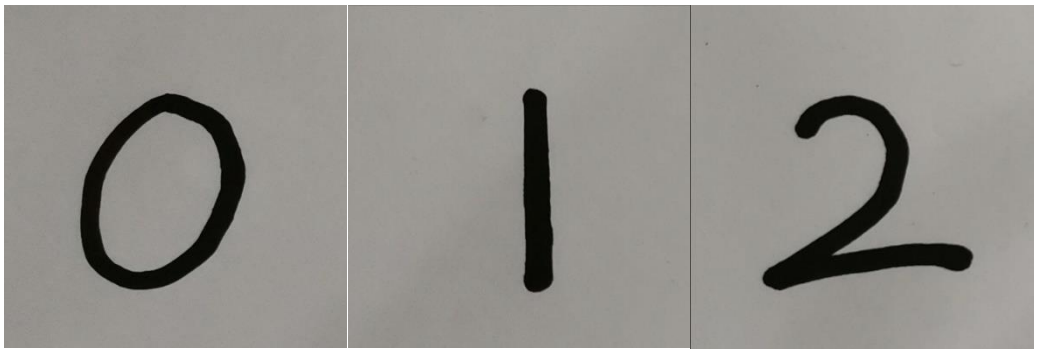
MNIST 数据集是机器学习领域中非常经典的一个数据集，由 60000 个训练样本和 10000 个测试样本组成，每个样本都是一张 28 * 28 像素的灰度手写数字图片。

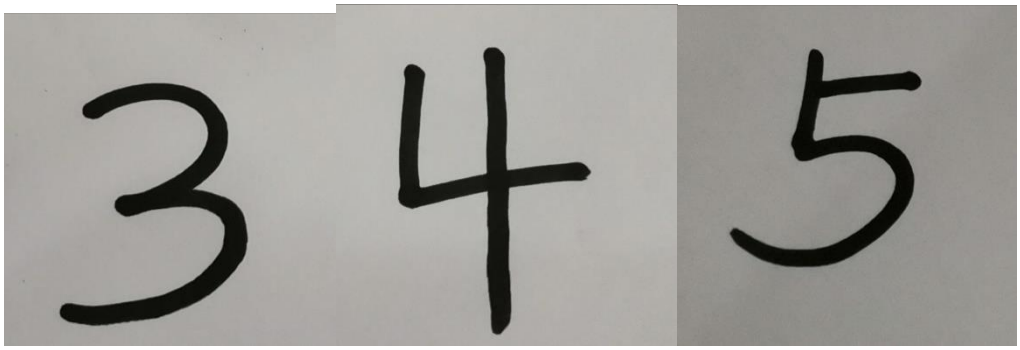
它来自美国国家标准与技术研究所，训练集(training set)由来自 250 个不同人手写的数字构成，其中 50%是高中学生，50%来自人口普查局的工作人员。测试集(test set)也是同样比例的手写数字数据。

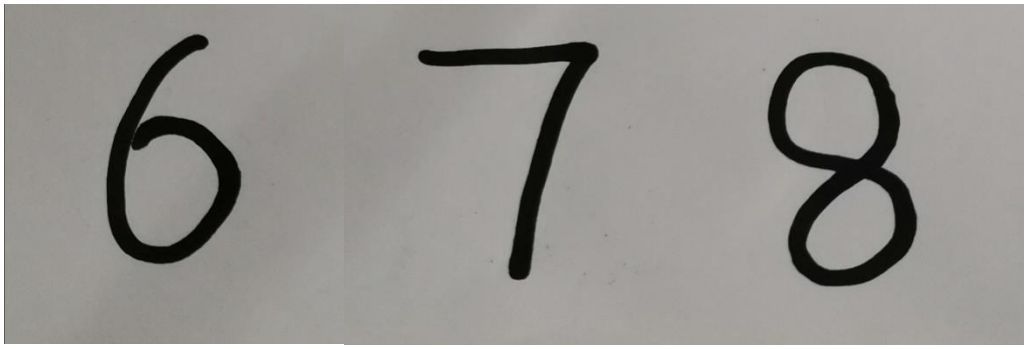
使用 TensorFlow 时，可以使用 tf.keras.datasets.mnist.load_data()来读取数据及标签,使用这种方式时,可以不用事先下载好数据集,它会自动下载并存放到你指定的位置。

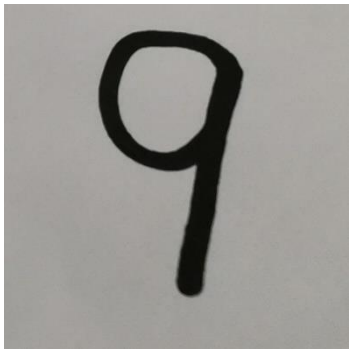
2.自己的数据

通过手机拍摄，得到 0~9 共十张图片如下：









这十张图片作为自己的测试集，用于测试，以检验数字识别建模的成果。

三、算法代码

1.MNIST 数据预处理

在 MNIST 数据预处理中，有以下几步：

1）数据格式转换：保证数据都为 28*28 数据，并且将数据类型转为 float32。

2）规范化：将数据都除 255，使得数据变为 0-1 之间的浮点数，可加速运算。

3）标签转化为 one-hot 形式：便于分类。

代码：

```
# 获取总类别
num_class = len(label_cnt)

# 格式转换
x_train = x_train.reshape((-1,28,28,1))
x_test = x_test.reshape((-1,28,28,1))
x_train = x_train.astype('float32')
x_test = x_test.astype('float32')

# 规范化，将像素值缩至0-1之间
x_train /= 255
x_test /= 255

# 将标签向量转化为one-hot形式的向量
y_train = tf.keras.utils.to_categorical(y_train, num_class)
y_test = tf.keras.utils.to_categorical(y_test, num_class)
```

2.模型搭建

使用改造的 LeNet5 模型，过程如下：

C1 层-卷积层：用 5x5 卷积核，步长为 1 对原图片卷积，一共使用了 8 个卷积核，每一个图中像素是带偏置的卷积再激活的值。

C2 层-池化层：对 C1 的每一个层面，用一个 2x2 窗口进行最大值子采样。

C3 层-卷积层：用 5x5 卷积核，步长为 1 对 C2 层卷积，一共使用 20 个卷积核，对 8 个层面上同一位置像素做了卷积后再求和，用这个和再做一次求偏置，之后再激活。

C4 层-池化层：对 C3 的每一个层面，用一个 2x2 窗口进行最大值子采样。

C5 层-卷积层：用 5x5 卷积核，步长为 1 对 C4 层卷积，用了 120 个卷积核，对 20 个层面上同一位置像素做了卷积后再求和，用这个和再做一次求偏置，之后再激活。

C6 层-全链接层:矩阵压平为向量后,先使用 relu 将向量变为 120 维的,再进行 softmax 计算将向量变为 10 维，得到一个概率值，概率大的位置数即为识别输出的数字。

模型架构如下：

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 28, 28, 8)	208
max_pooling2d (MaxPooling2D)	(None, 14, 14, 8)	0
conv2d_1 (Conv2D)	(None, 10, 10, 20)	4020
max_pooling2d_1 (MaxPooling2	(None, 5, 5, 20)	0
conv2d_2 (Conv2D)	(None, 1, 1, 120)	60120
flatten (Flatten)	(None, 120)	0
dense (Dense)	(None, 120)	14520
dense_1 (Dense)	(None, 10)	1210

代码：

```
# LeNet
model.add(tf.keras.layers.Conv2D(input_shape=(x_train.shape[1], x_train.shape[2], x_train.shape[3]),
    filters=8,kernel_size=(5,5),strides=(1,1),padding='same',activation='relu'))
model.add(tf.keras.layers.MaxPooling2D(pool_size=(2,2)))
model.add(tf.keras.layers.Conv2D(filters=20,kernel_size=(5,5),strides=(1,1),
    padding='valid',activation='relu'))
model.add(tf.keras.layers.MaxPooling2D(pool_size=(2,2)))
model.add(tf.keras.layers.Conv2D(filters=120,kernel_size=(5,5),strides=(1,1),
    padding='valid',activation='relu'))
model.add(tf.keras.layers.Flatten())
model.add(tf.keras.layers.Dense(120,activation='relu'))
model.add(tf.keras.layers.Dense(10,activation='softmax'))
```

3.模型训练

模型在训练过程中使用交叉熵损失函数，并使用 Adam 优化器；训练过程中，epochs=5，batch_size=128，验证集占比为 20%。

代码：

```
# 定义模型训练细节，包括交叉熵损失函数，Adam优化器和准确率评价指标
model.compile(loss='categorical_crossentropy',
              optimizer=tf.keras.optimizers.Adam(),
              metrics=['accuracy'])

'''步骤四：训练模型'''
h = model.fit(x_train, y_train, validation_split=0.2 ,batch_size=128, epochs=5)
```

4.自己数据预处理

在处理自己手机拍摄的图片时，有以下算法：

1）图像居中算法：通过获取手写字最左、最右、最上、最下的方位，得到中心点，然后将图像改为正方形，并保留四边各约 20%的背景色边缘，实现手写字居中，且大小适中。

2）先膨胀后腐蚀：消除手写字内部可能出现的背景色小泡。

3）转灰度图并进行高斯滤波：消除可能存在的噪点并使图像平滑。

4）二值化并反色：将图像二值化，并将手写的白底黑字转为黑底白字。

5）骨架提取和膨胀：使手写字粗细相对均匀且适中。

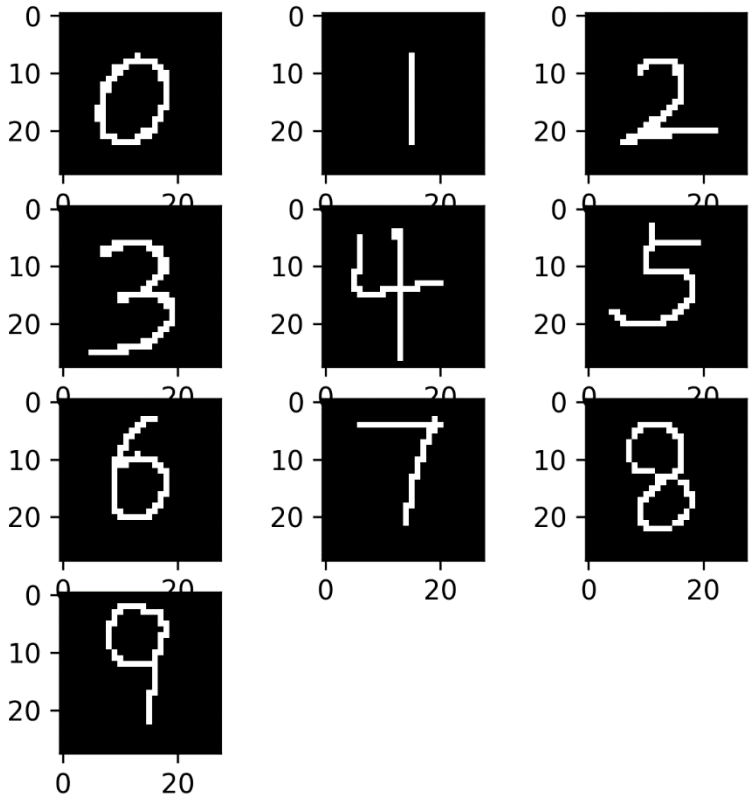
代码：

```
'''获取边界'''
def getbound(img):
    img_binary=img.convert('1')
    img_array = np.array(img_binary)
    shape = img_array.shape
    sumx = shape[0] - np.sum(img_array, axis=0)/255
    sumy = shape[1] - np.sum(img_array, axis=1)/255
    l, r, h, b = 0, 0, 0, 0
    for i in np.arange(shape[1]):
        if sumx[i] >= 3:
            l = i
            break
    for i in range(shape[1]-1, -1, -1):
        if sumx[i] >= 3:
            r = i
            break
    for i in np.arange(shape[0]):
        if sumy[i] >= 3:
            h = i
            break
    for i in range(shape[0]-1, -1, -1):
        if sumy[i] >= 3:
            b = i
            break
    return l, r, h, b
```

```
'''数据预处理'''
def pretreat(filename):
    img = (Image.open(filename).convert('L'))
    l, r, h, b = getbound(img)
    shape = np.array(img).shape
    dx, dy = (r - l)//2, (b - h)//2

    # 获取中心点
    centerx, centery = l+dx, h+dy

    # 改为正方形
    dx = max(dx, dy)
    dy = dx

    # 得到轮廓
    l, r, h, b = max(0, centery-dy*1.2), min(shape[0], centery+dy*1.2), max(0, centerx-dx*1.2),
min(shape[1], centerx+dx*1.2)

    # 切片
    img_array = np.array(img)
    the_img = img_array[l:r, h:b]

    # 先膨胀后腐蚀，消除手写字内部可能出现的小泡
    the_img=morphology.closing(the_img, selem=None, out=None)

    # 转灰度
    the_img=Image.fromarray(cv2.cvtColor(the_img, cv2.COLOR_BGR2RGB))
    img_gray=cv2.cvtColor(np.array(the_img),cv2.COLOR_RGB2GRAY)

    # 高斯滤波
    img_gauss = cv2.GaussianBlur(img_gray,(5,5),0)

    # 二值化，反色
    ret,img_binary=cv2.threshold(img_gauss, 0, 1, cv2.THRESH_BINARY_INV+cv2.THRESH_OTSU)

    # 骨架提取
    img_binary=morphology.skeletonize(img_binary)
```

```
    # 膨胀
    img_binary=morphology.binary_dilation(img_binary, morphology.square(29))

    # 调整大小为(28, 28)
    img_binary=Image.fromarray(img_binary)
    img_binary=img_binary.resize((28,28),Image.ANTIALIAS)

    # 为1的点转为255
    img_array=np.array(img_binary).astype(np.uint8)
    img_array[img_array==1]=255

    # 数据格式规范
    img_array.reshape((-1,28,28,1))
    img_array=img_array[:,:,np.newaxis]

    return img_array
```

预处理结果：



四、实验分析

1.结果分析

实验结果如下：

对于 MNIST 测试集：测试集损失值: 0.040091145783662796，测试集准确率: 0.9873999953269958。

对于自己的测试集：测试集损失值: 0.3253990411758423，测试集准确率: 0.8999999761581421

混淆矩阵如下:

```
Mnist手写数据集混淆矩阵如下：
[[ 969    0    0    0    0    0    6    2    2    1]
 [   0 1124    2    1    0    0    3    0    5    0]
 [   0    0 1031    0    0    0    0    1    0    0]
 [   0    0    1 1003    0    1    0    3    2    0]
 [   0    0    1    0  979    0    0    0    0    2]
 [   0    0    0   27    0  858    3    1    2    1]
 [   0    1    0    1    1    1  953    0    1    0]
 [   0    1    3    1    0    0    0 1022    1    0]
 [   1    0    8    3    2    0    2    3  954    1]
 [   1    2    2    5    6    2    0    5    5  981]]
我的数据集混淆矩阵如下：
[[1 0 0 0 0 0 0 0 0 0]
 [0 1 0 0 0 0 0 0 0 0]
 [0 0 1 0 0 0 0 0 0 0]
 [0 0 0 1 0 0 0 0 0 0]
 [0 0 0 0 1 0 0 0 0 0]
 [0 0 0 0 0 1 0 0 0 0]
 [0 0 0 0 0 0 1 0 0 0]
 [0 0 0 0 0 0 0 1 0 0]
 [0 0 0 0 0 0 0 0 1 0]
 [0 0 0 0 1 0 0 0 0 0]]
```

MNIST 测试集有近 99%的正确率，自己的测试集有 90%的正确率，可说明整体实验成功。

2.实验过程分析

在本次实验中，有两个难点，一是设计卷积神经网络模型，二是设计关于自己图像的预处理算法。

在模型设计时，需要考虑任务的难度、模型的搭建难度、以及训练所需花费的时间等；由于本次我们使用的是经典的 MNIST 数据集，进行经典的手写字识别，所以这次的任务难度不大，基于以上考虑，我们选择了 LeNet5 模型，而没有使用更为复杂、训练时间更长的 AlexNet、VGGNet、GoogleNet、ResNet 等模型，事实也证明我们使用的模型在测试集上效果很好。

在对自己的图像进行预处理时，我们需要考虑原始图像的方方面面，包括手写字的位置是不是居中、大小是否合适、背景是否有大量的噪点、图像是否相对平滑、线条内部是否有大量的噪点、线条粗细是否均匀适中等。基于以上考虑，我们设计了较为复杂的图像预处理算法。在调试过程中，往往会出现预处理之后图像的效果很不好，此时我们通过在之前的每次操作后都输出一次图像，来观察到底是哪里出了问题，从而调整我们的处理。例如对于骨架提取后的膨胀，就多次调整了滤波器的样式和大小。

在调试过程中，我发现如果将所有的步骤写在一个文件里，会导致浪费大量时间训练模型。因为在实验后期，模型往往已经很成熟，需要修改的是自己图像的预处理。因此我将实验的不同部分划分到不同的文件中，并把训练好的模型参数保存下来，每次调整了图像预处理后只需要拿去测试即可。