



中国科学院大学

University of Chinese Academy of Sciences

课程作业报告

基于卷积神经网络的手写数字识别

作者姓名: 陈良玉

学科专业: 工业工程与管理

所在单位: 中国科学院大学工程科学学院

2022 年 6 月

摘要

本文基于改版的 AlexNet 经典卷积神经网络针对手写数字识别问题进行实验分析。基础的 AlexNet 是 2012 年 ImageNet 竞赛冠军获得者 Hinton 和他的学生 Alex Krizhevsky 设计的。本文将 AlexNet 进行一定程度的改进，将其适用于处理 MNIST 数据集的训练和测试，模型最终在测试集上的准确率达到 98.6%。

1. 介绍

手写数字识别问题是一种分类问题，即输入手写的 0~9，让机器可识别出每张图片对应什么数字。目前很多经典的卷积神经网络都可以处理这类问题，如 LeNet-5、AlexNet、ResNet 等。本文基于 AlexNet 网络，对其中的卷积层和输出层稍加修改，使其能够很好地处理手写数字识别问题。

2. 解决方案

本次作业采用 AlexNet 作为基础网络框架，在其基础上进行如下改进：AlexNet 原始输入图像尺寸为 227x227x3 大小的 1000 分类，而 MNIST 图像尺寸为 28x28x1 的 10 分类，输入尺寸太小不足以完成网络的下采样过程，故需要对网络进行简单的修改：

卷积层 1：将卷积核步长设为 2；

输出层：输出向量长度设为 10（分类数量）

本文使用的损失函数是交叉熵损失函数，通过 Tensorflow 深度学习框架中的函数 `tf.keras.losses.sparse_categorical_crossentropy` 来调用。具体的网络框架和参数数量见下图。

Model: "Changed_AlexNet"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 14, 14, 96)	11712
max_pooling2d (MaxPooling2D)	(None, 6, 6, 96)	0
conv2d_1 (Conv2D)	(None, 6, 6, 256)	614656
max_pooling2d_1 (MaxPooling2D)	(None, 2, 2, 256)	0
conv2d_2 (Conv2D)	(None, 2, 2, 384)	885120
conv2d_3 (Conv2D)	(None, 2, 2, 384)	1327488
conv2d_4 (Conv2D)	(None, 2, 2, 256)	884992
max_pooling2d_2 (MaxPooling2D)	(None, 1, 1, 256)	0
flatten (Flatten)	(None, 256)	0
dense (Dense)	(None, 2048)	526336
dropout (Dropout)	(None, 2048)	0
dense_1 (Dense)	(None, 2048)	4196352
dropout_1 (Dropout)	(None, 2048)	0
dense_2 (Dense)	(None, 10)	20490

=====
Total params: 8,467,146
Trainable params: 8,467,146
Non-trainable params: 0
=====

图 1 Changed_AlexNet 网络框架和参数数量

3. 实验结果和分析

数据集介绍：MNIST 数据库（Modified National Institute of Standards and Technology database）是一个大型数据库的手写数字是通常用于训练各种图像处理系统。该数据库还广泛用于机器学习领域的培训和测试。它是通过“重新混合”NIST 原始数据集中的样本而创建的。此外，将来自 NIST 的黑白图像归一化以适合 28x28 像素的边界框并进行抗锯齿处理，从而引入了灰度级。MNIST 数据库包含 60,000 个训练图像和 10,000 个测试图像。MNIST 手写数字识别模型的主要任务是：输入一张手写数字的图像，然后识别图像中手写的是哪个数字。MNIST 部分数据如下图所示：

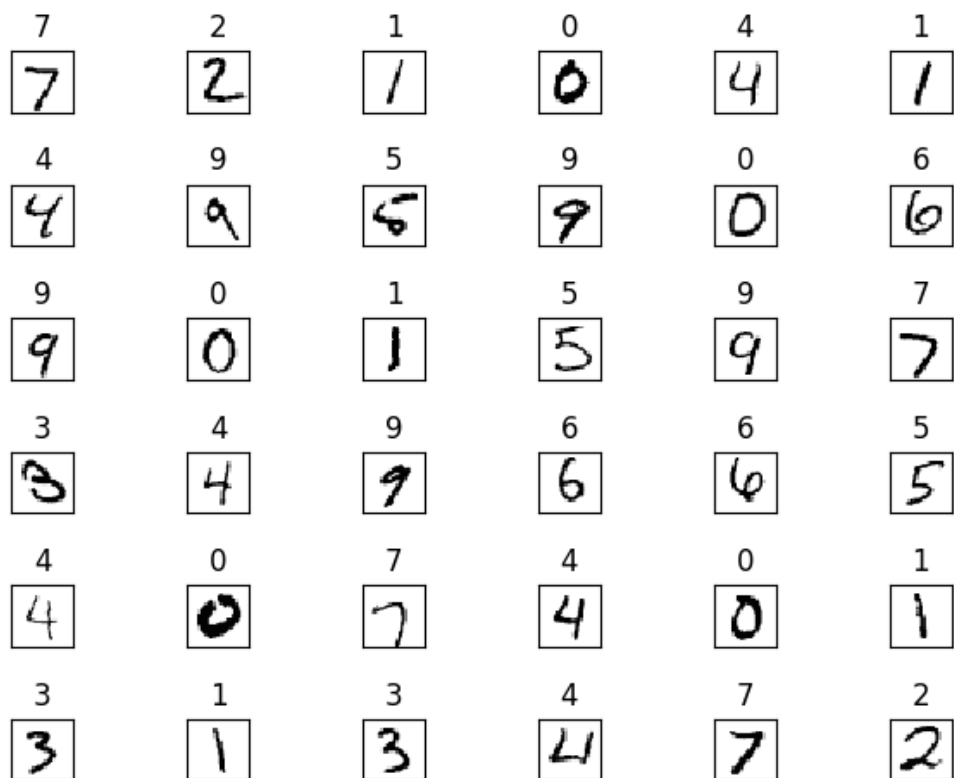


图 2 MNIST 数据集部分数据展示

实验环境如下表所示：

表 1 实验硬件和软件环境

硬件环境	
操作系统	Windows 10
CPU 型号	Intel Core i5 4G@2.50GHz
内存	4 GB
软件环境	
Python	3.8.13
Tensorflow	2.9.1
Keras	2.9.0
Matplotlib	3.5.2

本次作业所选取的 AlexNet 网络本身的拟合能力已经很强大，所以选择了较小的 epoch = 20，训练集和验证集上的 loss 和准确率如下图所示：

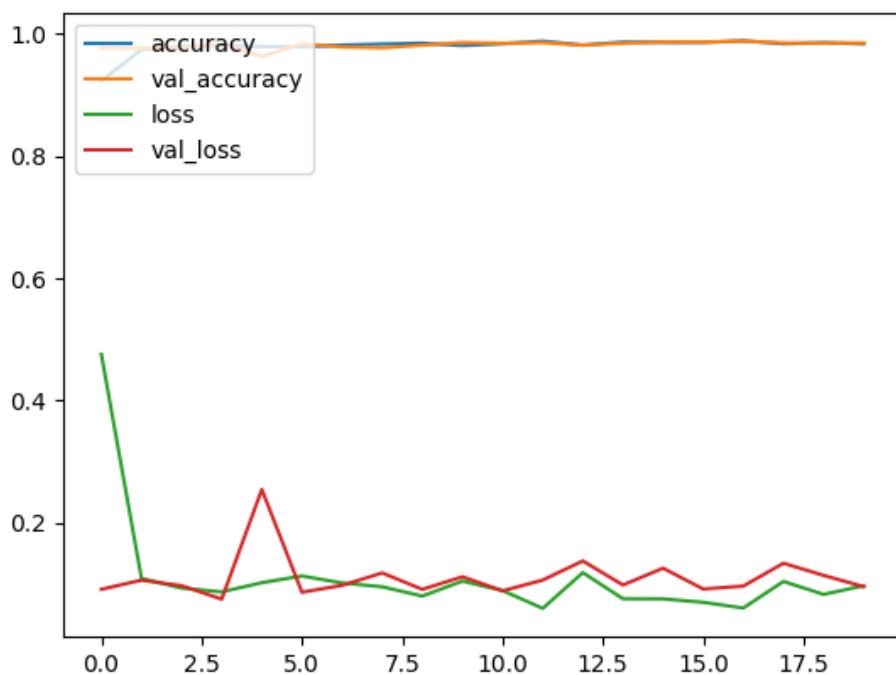


图 3 Changed_AlexNet 基于训练集和验证集的收敛情况

从上图可以看出，模型在训练集上很快收敛，且在验证集上的表现较好，无明显波动情况。故本次模型训练结果令人满意。最终，利用该模型基于测试集的准确率达到 98.6%，如下图所示：

```
844/844 [=====] - 412s 468ms/step - loss: 0.0601 - accuracy: 0.9890 - val_loss: 0.0960 - val_accuracy: 0.9883
Epoch 18/20
844/844 [=====] - 378s 447ms/step - loss: 0.1036 - accuracy: 0.9840 - val_loss: 0.1334 - val_accuracy: 0.9853
Epoch 19/20
844/844 [=====] - 397s 470ms/step - loss: 0.0824 - accuracy: 0.9857 - val_loss: 0.1136 - val_accuracy: 0.9853
Epoch 20/20
844/844 [=====] - 386s 458ms/step - loss: 0.0965 - accuracy: 0.9839 - val_loss: 0.0951 - val_accuracy: 0.9853
313/313 [=====] - 8s 24ms/step - loss: 0.0955 - accuracy: 0.9859
模型在测试集上的评估结果为: [0.09548938274383545, 0.9858999848365784]
```

图 4 模型基于测试集的准确率达到 98.6%

4. 结论

本次作业使用改进的 AlexNet 针对 MNIST 数据集进行分类工作，在对卷积层 1 和输出层的改动后，在测试集上取得了 98.6% 的准确率，效果较好。

参考文献

- [1] LECUN, Yann, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86.11: 2278-2324.
- [2] KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 2012, 25.
- [3] CSDN. TensorFlow2 学习十三、实现 AlexNet[EB/OL]. (2019-11-29)[2022-6-23]. <https://blog.csdn.net/xundh/article/details/103303801>.