**基于TensorFlow的MNIST手写数字识别的研究**

摘 要

手写数字识别是计算机视觉领域的经典问题之一。MNIST数据集是一个常用的手写数字识别数据集，包含了大量的手写数字图像及其对应的标签。本项目旨在使用深度学习技术构建一个准确、高效的模型，对MNIST数据集中的手写数字进行识别。项目使用了TensorFlow框架构建了不同模型来实现目标：全连接神经网络和卷积神经网络（CNN）模型。通过对模型在验证集上的准确率进行分析，评估模型的泛化能力，并发现了模型在不同类别上的表现差异。

**关键词**：深度学习，卷积神经网络，TensorFlow

1 算法分析

1.1 算法流程

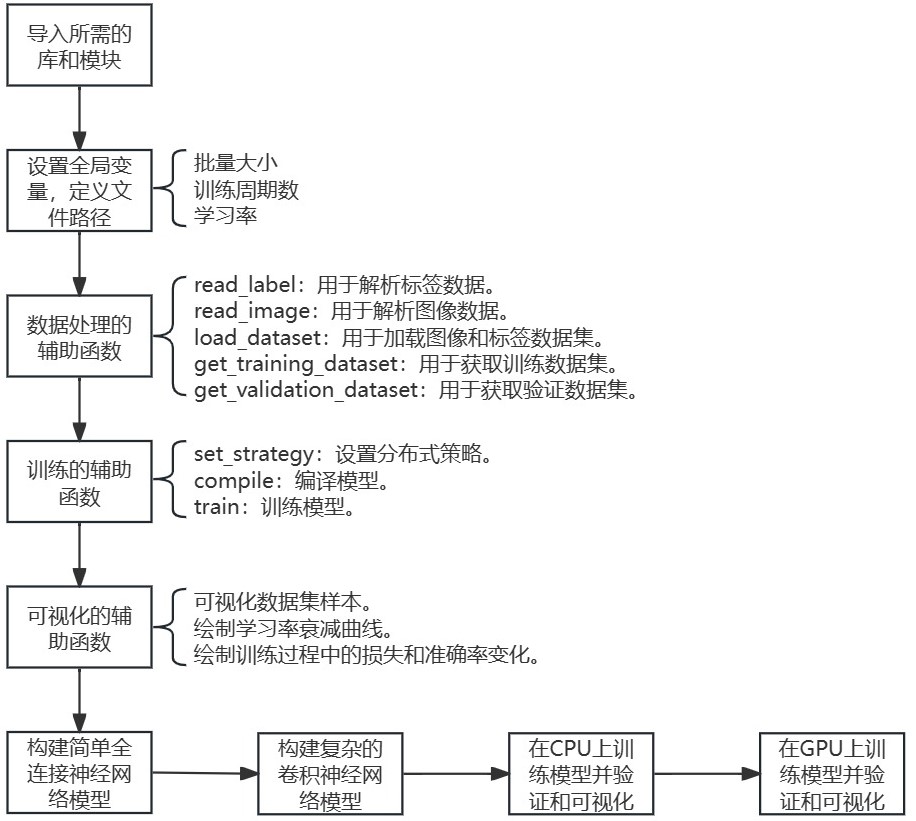


图1.1 总体算法流程图

1.2 主要代码说明

1.2.1 准备数据并展示数据样本

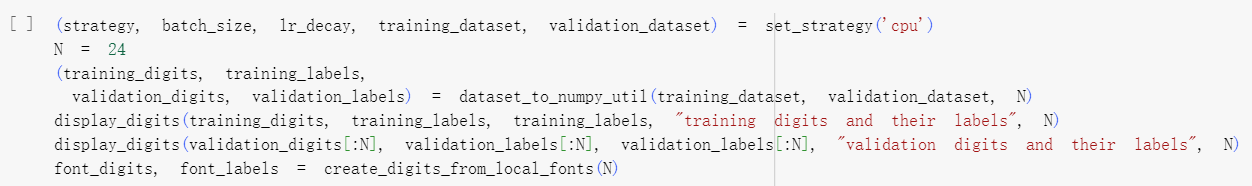


图1.2 准备数据并展示数据样本代码

调用set\_strategy('cpu')函数设置在CPU上进行模型训练，并获取相关参数。调用dataset\_to\_numpy\_util将训练和验证数据集转换为NumPy数组，并获取了指定数量的训练和验证数字样本及其标签。使用display\_digits展示训练数据集和验证数据集中的数字样本及其标签。调用create\_digits\_from\_local\_fonts生成一些来自本地字体文件的数字样本和对应的标签，以了解数据的特点和质量，为后续的模型训练和验证做准备。

1.2.2 简单的全连接神经网络模型



图1.3 全连接神经网络模型代码

首先使用能想到的最简单的网络，只有两层，即输入层和输出层，完成对MNIST数据集的数字识别任务，并观察模型的训练效果以及对新样本的识别能力。具体过程如下：

使用tf.keras.Sequential构建了一个简单的神经网络模型，该模型仅包含两层：

输入层：将输入数据的形状设置为28x28，即MNIST数据集中的图像尺寸。

输出层：包含10个节点，代表数据集中的10个类别（即数字0到9），并使用softmax激活函数进行多类别分类。

使用compile函数编译模型，指定了优化器、损失函数和评估指标。

调用train函数训练模型，包括加载训练数据集、模型的训练和评估。

使用训练好的模型对来自本地字体文件的数字样本和验证数据集中的数字进行识别，并展示未被正确识别的结果。

1.2.3 复杂的卷积神经网络模型



图1.4 卷积神经网络模型代码

首先使用tf.keras.layers.Reshape将输入数据重塑为指定的形状。此处则是将输入形状从一维的28\*28重塑为二维的28x28x1。

此模型使用了三个卷积层用于提取图像特征。每个卷积层都包括一个卷积操作和一个激活函数。卷积操作用于在图像上滑动一个卷积核来提取特征，激活函数则引入了非线性特性。这些层逐渐增加了卷积核的数量和图像的尺寸，以便提取更高级别的特征。

此外还有两个全连接层，用于将卷积层提取的特征映射到类别标签。第一个全连接层包含200个节点，并使用了ReLU激活函数。第二个全连接层则输出10个节点，对应于数据集中的10个类别，并使用softmax激活函数进行多类别分类。

其中，BatchNormalization用于进行批标准化操作，可加速训练过程并增强模型的稳定性。批标准化在每个批次中对数据进行归一化，并进行适当的缩放和平移参数。Dropout用于在训练过程中随机丢弃部分节点，以防止模型过拟合。

1.2.4 在 CPU/GPU 上训练和验证模型

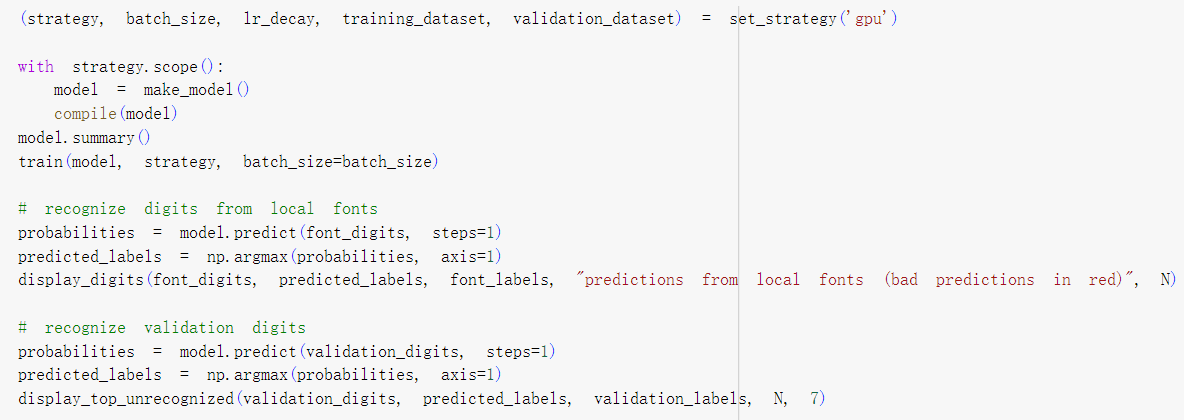


图1.5 在GPU上训练和验证模型代码

目前已有网络的数据结构，接下来需要训练和验证模型，最初在CPU上运行，但消耗过长时间，便改为在GPU上运行。现在需要为Keras提供关键的3个输入：优化器，规定其如何找到最佳解决方案；loss函数，查看每次迭代后模型结果的质量；调用 model.summary()获得结果的文字描述。

2 实验结果分析

2.1 查看训练集和验证集数据



图2.1 部分训练集与验证集的手写数字及标签

2.2 全连接神经网络模型结果

训练10轮后的loss为0.2637，accuracy为0.9279。

损失相对较高，准确率相对较低。这可能是因为全连接神经网络模型在处理图像数据时，没有考虑到图像的空间结构信息，而是将图像展平为一维向量，导致丢失了重要的空间信息。因此，模型在识别复杂图像时表现不佳。

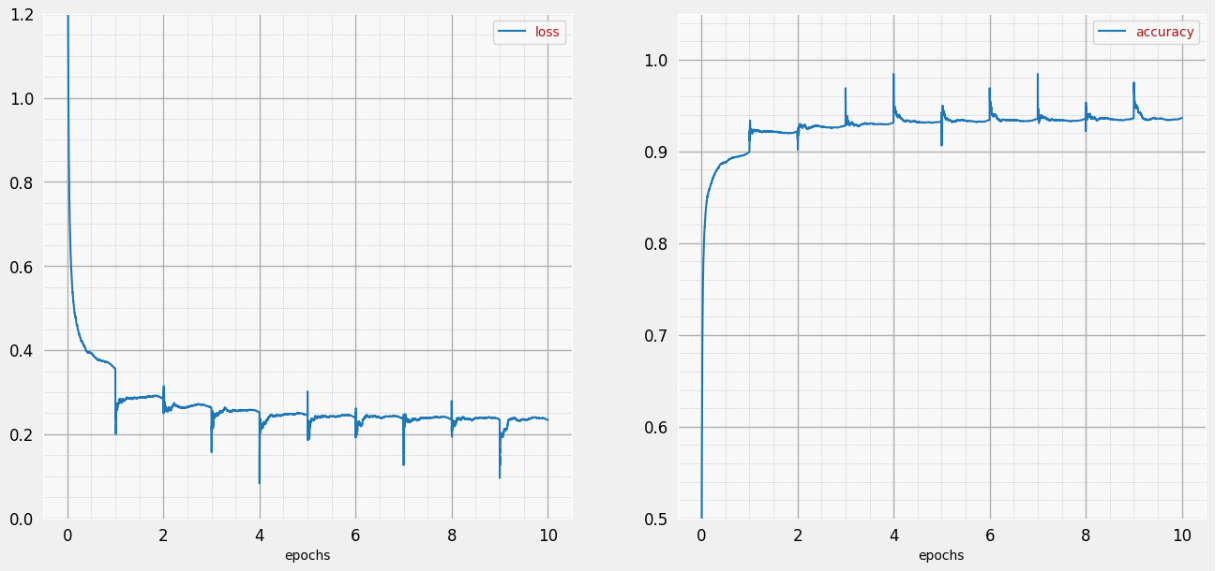


图2.2 全连接神经网络模型训练后的loss-epochs与accuracy-epochs



图2.3 本地字体识别结果与识别错误的手写数字及对应标签

2.3 卷积神经网络模型结果

训练10轮后的loss为0.0172，accuracy为0.9945。

损失相对较低，准确率相对较高。卷积神经网络模型能够有效地捕获图像的空间结构信息，并通过卷积和池化操作提取图像的局部特征，因此在识别图像时表现更好。

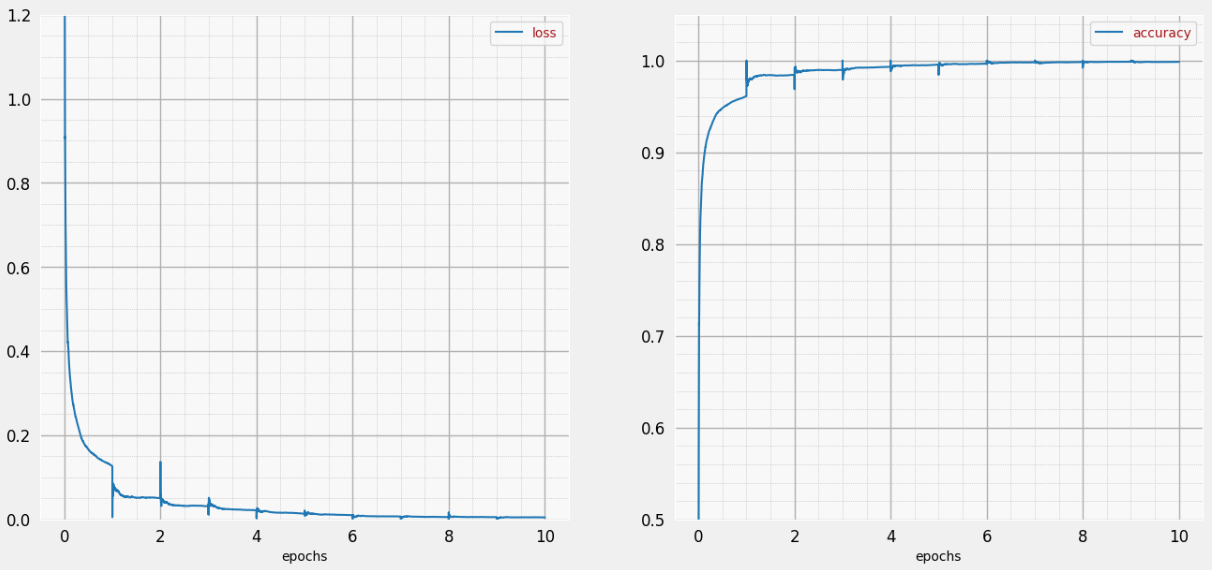


图2.4 卷积神经网络模型训练后的loss-epochs与accuracy-epochs

2.4 分析比较

卷积神经网络模型相对于全连接神经网络模型在性能上有明显提升。这表明在处理图像数据时，利用卷积操作可以更好地捕获图像的空间特征，从而提高模型的性能。

全连接神经网络模型在处理图像数据时存在一定的局限性，容易出现过拟合等问题。因此，在处理图像等具有空间结构的数据时，卷积神经网络模型是更为合适和有效的选择。

3 总 结

本文针对MNIST手写数字识别任务，探讨了2种不同类型的神经网络模型：全连接神经网络模型和卷积神经网络模型。全连接神经网络模型使用了最简单的结构，但在处理图像数据时存在一定的局限性。由于其无法捕获图像的空间结构信息，导致模型在识别复杂图像时表现不佳，在验证数据上可能会有较高的损失和相对较低的准确率。与之相反，卷积神经网络模型通过卷积和池化操作，能更有效捕获图像的空间特征，从而提高模型的性能。在相同的训练轮数下，卷积神经网络模型取得了更低的损失和更高的准确率。这表明在处理复杂图像数据时，卷积神经网络模型是更为合适和有效的选择。

<https://colab.research.google.com/drive/1aK7_4aOJt4Wh0zAvdyYBSRtqZdZgJj2t?usp=sharing>（本文代码链接）

参考文献

[1] Zhu W. Classification of MNIST handwritten digit database using neural network[J]. Proceedings of the research school of computer science. Australian National University, Acton, ACT, 2018, 2601.

[2] Ertam F, Aydın G. Data classification with deep learning using Tensorflow[C]//2017 international conference on computer science and engineering (UBMK). IEEE, 2017: 755-758.