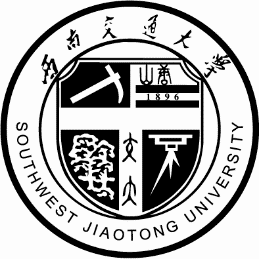
**­­­­­­**



**《机器学习实验》**

**课程设计报告**

**报告题目：基于卷积神经网络的手写数字识别**

**学院名称： 计算机与人工智能学院**

**年 级： 2020级**

**授课教师：**  **廖革元**

**辅导教师： 苟元帅 吴金泽**

**姓 名： 张喆皓**

**学 号： 2020112507**

**成 绩：**

**二〇二三年六月**

**实验要求：**

1. 基于Pytorch深度学习框架，搭建多层卷积神经网络模型，实现手写数字识别，采用精确率、召回率、F1值评测模型表现。
2. 调整学习率，报告不同值对应的模型表现，并加以分析。
3. 调整优化器，报告不同对应的模型表现，并加以分析。
4. 对比方法包括但不限于SVM、决策树，报告其模型表现。
5. 课程报告内容包括但不限于神经网络模型结构介绍，损失函数介绍等。

**课程设计提交形式：**

1. 课程设计报告；
2. 模型代码，包含以下内容：
   1. 训练文件；
   2. 测试文件；
   3. 模型权重；
   4. requirement.txt (环境依赖包中依赖库名称及其版本号，可通过pip或conda导出)
   5. README.md (介绍模型的训练、预测方式及课程设计报告中各实验结果的执行步骤)。

基于卷积神经网络的手写数字识别

**摘 要**：

**关键词**：

1. **相关工作**

手写数字识别作为人工智能的一个重要组成部分，有着极其广泛地应用前景和发展。

首先就国外来讲，阿拉伯数字作为全球通用的数字语言，与各个国家、民族和地区的文化背景无关，是世界上统一使用的符号。数字的识别类别较小，可以方便评估研究方法的有效性和可行性，同时也为字符的识别提供借鉴。最后，数字识别的智能系统，可以应用于财税、金融、邮件分拣等领 域，减少了人工操作带来的不便性和出错率，方便人们生活，为人类向智能领域的发展提供 更大的实用价值。

目前应用于实际生活中为人么所熟知，比如击败李世石的ALPHA GO以及后来青出于蓝而胜于蓝的ALPHA GO ZERO[1] 以及网络春晚大放异彩的钢琴机器人等。 数字识别的方法有MRFSE、运动图像（2）、SVM（3）、BP[4]、聚类分析问等，这些方法在识别之前，需要对样本图像进行统一的预处理，如归一化、平滑去噪、字符 切分、二值化、笔画细化、特征提取、重建模型等，再加上网络训练参数众多使得识别过程 更加耗时、复杂。在实验过程中，大量实践表明，没有一种方法可以获得各种字符特征的100% 识别率，每个研究方法都存在局限性。

卷积神经网络的网络结构使其更接近于生物神经系统，从而可以快速学习并良好记忆。

由牛津大学提出的识别、人脸检测、物体形状等的识别，成为语音分析和图像识别领域的研究热点，并取得 —定研究成果，具有良好的应用前景。

* 1. **基于卷积神经网络的方法**

卷积神经网络（Convolutional Neural Network，CNN）是一种深度学习算法，它在图像识别领域有着广泛的应用。CNN 通过卷积层、池化层和全连接层等组成，其中卷积层和池化层是 CNN 的核心部分。卷积层通过卷积核对输入图像进行卷积操作，提取出图像的特征；池化层则通过对特征图进行降采样，减少特征图的大小，从而减少计算量。CNN 的优点是可以自动提取图像的特征，不需要人工干预，因此在手写数字识别方面有着广泛的应用。

CNN 的缺点是需要大量的训练数据和计算资源，同时模型也比较复杂，需要较长的训练时间²。

**1.2 基于SVM的方法**

支持向量机（Support Vector Machine，SVM）是一种常用的分类算法，它在手写数字识别方面也有着广泛的应用。SVM 的工作原理是将输入数据映射到高维空间中，然后在该空间中寻找一个最优的超平面，将不同类别的数据分开。SVM 的优点是可以处理高维数据，对于小样本数据也有较好的分类效果。SVM 的缺点是需要大量的计算资源和时间，同时对于非线性问题需要使用核函数进行处理3。

**1.3 基于决策树的方法**

是哒决策树是一种基于树结构的机器学习算法，可以用于分类和回归。在手写数字识别中，决策树可以通过对每个像素点进行分类来识别数字。例如，如果一个像素点的灰度值大于某个阈值，则将其分类为1，否则将其分类为0。通过这种方式，决策树可以对手写数字进行分类。

决策树的优点是易于理解和解释，可以处理多路输出问题。缺点是容易过拟合，生成的决策树不稳定，微小的数据变化可能导致生成的决策树不同4。

**2. 基础内容**

**2.1 卷积神经网络**

**2.1.1 卷积层**

在最初的卷积层中，成千上万的神经元充当第一组过滤器，搜寻图像中的每个部分和像素，找出模式（pattern）。随着越来越多的图像被处理，每个神经元逐渐学习过滤特定的特征，这提高了准确性。

比如图像是苹果，一个过滤器可能专注于发现“红色”这一颜色，而另一个过滤器可能会寻找圆形边缘，另一个过滤器则会识别细细的茎。如果你要清理混乱的地下室，准备在车库搞个大销售，你就能理解把一切按不同的主题分类是什么意思了（玩具、电子产品、艺术品、衣服等等）。 卷积层就是通过将图像分解成不同的特征来做这件事的。

特别强大的是，神经网络赖以成名的绝招与早期的 AI 方法（比如 Deep Blue 中用到的）不同，这些过滤器不是人工设计的。他们纯粹是通过查看数据来学习和自我完善。

卷积层创建了不同的、细分的图像版本，每个专用于不同的过滤特征——显示其神经元在哪里看到了红色、茎、曲线和各种其他元素的实例（但都是部分的） 。但因为卷积层在识别特征方面相当自由，所以需要额外的一双眼睛，以确保当图片信息在网络中传递时，没有任何有价值的部分被遗漏。

神经网络的一个优点是它们能够以非线性的方式学习。如果不用数学术语解释，它们的意思是能够发现不太明显的图像中的特征——树上的苹果，阳光下的，阴影下的，或厨房柜台的碗里的。这一切都要归功于于激活层，它或多或少地突出了有价值的东西——一些既明了又难以发现的属性5。

**2.1.2 激活层（示例，请删除）**

**2.1.3 损失函数（示例，请删除）**

**2.2 学习率**

**2.3 优化器**

**2.4 SVM**

**2.5 决策树**

**2.6 评估指标**

**2.6.1 查准率(precision)（示例，请删除）**

**2.6.2查全率(recall)（示例，请删除）**

**2.6.3 F1值（示例，请删除）**

**2.7 \*\*\*\***

**2.8 \*\*\*\***

**3. 方法**

**3.1 问题介绍**

手写数字识别是一种计算机视觉任务，目的是让计算机能够识别人类手写的数字。决策树是一种基于树结构的机器学习算法，可以用于分类和回归。在手写数字识别中，决策树可以通过对每个像素点进行分类来识别数字。例如，如果一个像素点的灰度值大于某个阈值，则将其分类为1，否则将其分类为0。通过这种方式，决策树可以对手写数字进行分类。当然，在机器学习中还有其他方法可以进行手写数字识别，例如卷积神经网络、随机森林、支持向量机等。

**3.2 模型**

本次实验基于Pytorch构建了一个包含2个卷积层（含ReLU激活函数），2个池化层，3个全连接层的卷积神经网络。排列方式为：

下面依次介绍每一层的组成：

1. 卷积层1：

卷积层1输入深度1，输出深度32，卷积核大小为3\*3，移动步长为1。同时，为保证正常完成卷积过程，在输入图像的最外圈补一圈0来占位。

卷积层1在卷积操作完成后，进行标准化和ReLU激活函数的过程。对数据进行标准化是为了规避数据大小对ReLu激活函数的影响，提升准确度。

1. 池化层1：

池化层1使用最大值池化方式，规模为2\*2，即池化后宽高均为原来一半。

1. 卷积层2：

卷积层2输入深度32，输出深度64，卷积核大小为3\*3，移动步长为1。其余参数及后续标准化与ReLU激活过程与卷积层1相同。

1. 池化层2：

池化层2使用最大值池化方式，规模为2\*2，即池化后宽高均为原来一半。

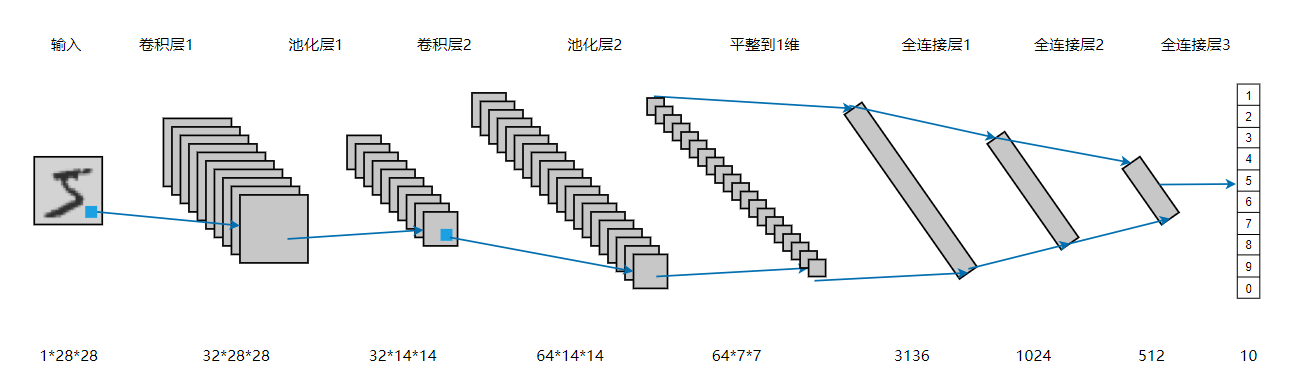
（5）全连接层1-3：

全连接层用于完成最后的分类任务，经过前面两轮卷积+池化的过程，数据的特征点平整得到64\*7\*7=3136个特征点，最终预期的输出分类数为10，因此通过全连接操作完成这个过程。

全连接各层均采用线性模型。全连接层1输入通道数为3136，输出通道数为1024；全连接层2输入通道数为1024，输出通道数为512；全连接层3输入通道数为512，输出通道数为10，分类完成。

此外，每两个全连接层中需要使用ReLU激活函数进行激活，否则再多的层数也相当于只有一层，起不到任何作用。

具体的模型结构如下图所示：

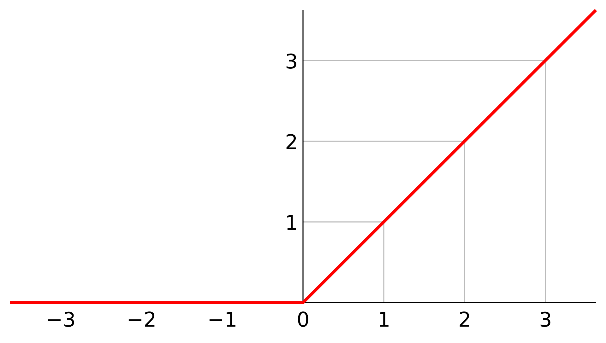


**3.3 损失函数**

本次实验中使用的激活函数均为ReLU线性整流函数。

ReLU的全称为Rectified Linear Unit，是一种人工神经网络中常用的激活函数，通常意义下，其指代数学中的斜坡函数，即

对应图像如下；



而在神经网络中ReLU函数作为神经元的激活函数，为神经元在线性变换之后的非线性输出结果。此时，对于输入的来自上一层神经网络的输入向量，ReLU函数给下一层神经元的输出为

**3.4 激活函数**

本次实验中使用的损失函数为交叉熵损失函数CrossEntropyLoss。

交叉熵主要用来刻画实际输出与期望输出的距离，交叉熵越小时，两个概率的分布越接近。举个例子来解释，某样本属于第K类，那么在分类训练时，我们期望与第K类相关的所有节点输出值都应该为1，非K类节点输出值都应该为0，此时得到标签[0,0,0,1,0,0,0….,0]，根据标签可以得到样本的分类。交叉熵就是计算当前网络各节点输出与标签的差异，然后利用这种差异进行反向传播，来逐步更新网络参数。

交叉熵的具体计算方法如下：

假设概率分布为期望输出，概率分布为实际输出，为交叉熵，则

在Pytorch中对应的函数方法为nn.CrossEntropyLoss()，内部实现时使用nn.LogSoftmax()和nn.NLLLoss()两个函数进行计算。

**4. 实验**

**4.1 数据集**

**4.2 实验环境**

**4.3 实验结果**

表1 不同学习率对比

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 学习率 | Precision | Recall | F1 |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

表2 不同优化器对比

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 优化器 | Precision | Recall | F1 |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

表3 不同方法对比

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 方法 | Precision | Recall | F1 |  |
| SVM |  |  |  |  |
| 决策树 |  |  |  |  |
| 本文方法 |  |  |  |  |
|  |  |  |  |  |

表4 \*\*\*\*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**4.4 \*\*\***

**5. 总结展望**

（总结：本文主要内容及效果。请删除）

（展望：本文方法的不足之处及对后续改进的思考。请删除）

参考文献

1. <https://blog.csdn.net/qq_52877327/article/details/117161496>
2. 【深度学习系列】手写数字识别--卷积神经网络CNN原理详解(一) - 知乎. <https://zhuanlan.zhihu.com/p/30665319>.
3. 基于SVM技术的手写数字识别 - CSDN博客.<https://blog.csdn.net/qq_40369926/article/details/97687450>.
4. 机器学习入门之：用 scikit-learn 的决策树来实现手写数字识别\_使用sklearn基于决策树的手数字识别\_暖仔会飞的博客-CSDN博客. <https://blog.csdn.net/qq_42902997/article/details/108581656>.
5. <http://t.csdn.cn/nG8qU>

要求：（请删除）

1. 图、表均居中，大小合适、不跨页，图注用五号，位于图的下方；表注用五号，位于表的上方

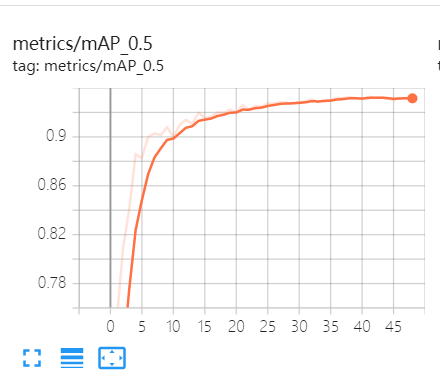


图1 平均精确率曲线图

表1 不同学习率对比表

|  |  |  |  |
| --- | --- | --- | --- |
| 学习率 | Precision | Recall | F1 |
| 0.005 |  |  |  |
| 0.01 |  |  |  |
| 0.05 |  |  |  |
| 0.1 |  |  |  |

1. 公式居中，使用斜体，有编号

(1)

1. 文中须有对图、表、公式的介绍和分析
2. 正文中文用宋体小四号、英文和数字用Times New Roman小四号
3. 参考文献格式符合GB/T 7714—2015
4. 报告中不放代码
5. 严禁抄袭