

**人工智能软件开发与实践**

**实验报告**

题 目 卷积神经网络CNN

学 院 计算机科学与技术

专 业 人工智能

学 号 2021112845

学 生 张智雄

任 课 教 师 武小荷

哈尔滨工业大学计算机科学与技术学院

2023.9

**实验三:** **卷积神经网络CNN**

1. **实验内容**

搭建Python和Pytorch环境，并在MNIST或CIFAR-10数据集上使用卷积神经网络CNN实现手写体数字识别或图像分类，主要包括三个部分：

1. 下载、预处理数据
2. 使用Pytorch实现CNN分类器，网络结构自行选择（AlexNet，VGG，ResNet等），batch size和epoch按需设置
3. 测试集分类精确度不低于90%
4. **算法简介及其实现细节**

## 2.1 卷积神经网络(CNN)的基本原理

卷积神经网络(Convolutional Neural Network，CNN)是一种深度学习模型，特别设计用于处理和分析具有网格结构的数据，如图像和视频。它能够自动学习图像中的特征并进行高效的图像分类、对象检测、图像生成和分割等任务，其模型结构主要包含以下部分：

1. 卷积层: 卷积层负责从图像中提取特征，如边缘和纹理。它们通过应用过滤器来捕捉这些特征，逐渐形成更复杂的视觉模式。
2. 池化层: 池化层在保留基本信息的同时减小了特征图的大小。最常见的方法是最大池化，它有助于缩小图像，同时保持关键特征并增强鲁棒性。
3. 全连接层: 全连接层结合从前一层提取的特征进行分类和决策。他们将这些特征映射到不同的类别，识别图像中的内容。

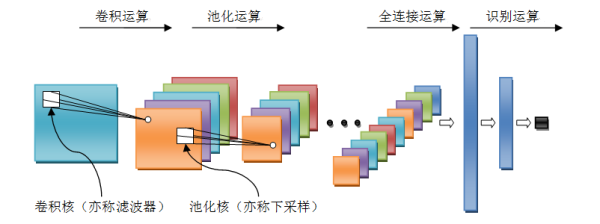


图1 卷积神经网络模型结构

## 2.2 AlexNet的基本结构

AlexNet网络结构相对简单，使用了8层卷积神经网络，前5层是卷积层，剩下的3层是全连接层，具体如下图2所示。

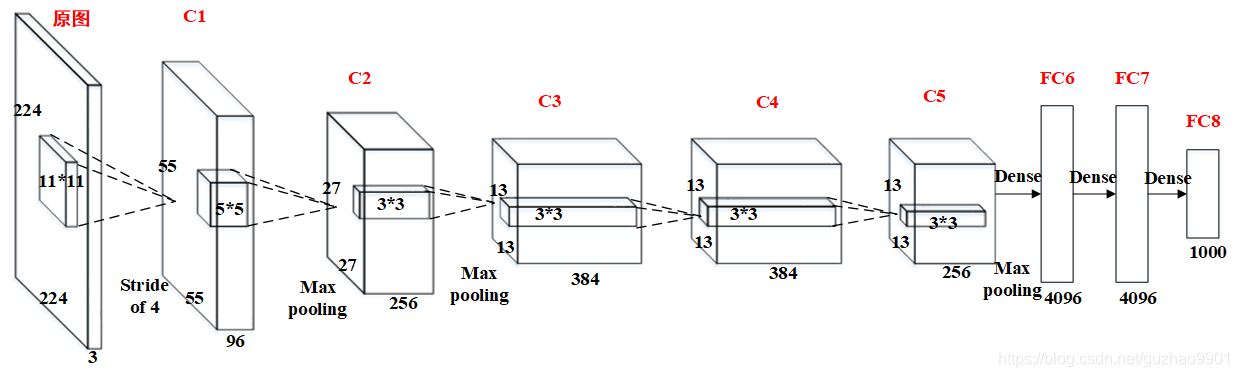


图2 AlexNet判别器结构

与原始的LeNet相比，AlexNet网络结构更深，同时还包括以下特点：

1. ReLU激活函数的引入：采用修正线性单元(ReLU)的深度卷积神经网络能够大幅提高训练速度，同时能够有效防止过拟合现象的出现。
2. 层叠池化操作：AlexNet中池化层采用了层叠池化操作，即池化大小步长，这种类卷积操作可以使相邻像素间产生信息交互和保留必要的联系。
3. Dropout操作：Dropout操作会将概率小于0.5的每个隐层神经元的输出设为0，即去掉一些神经节点，能够有效防止过拟合现象的出现。

## 2.3 ResNet的基本结构

ResNet的主要创新是引入了残差连接（或称为跳跃连接），以解决深度神经网络训练中的梯度消失和梯度爆炸问题。其核心思想是通过跳跃连接将输入信号直接传递到网络中的后续层，使得网络可以学习到残差信息，从而更轻松地训练非常深的神经网络。这些跳跃连接可以是简单的恒等映射，也可以包含适当的线性变换以匹配维度。这种结构使得神经网络的训练更加稳定，从而能够构建非常深的网络。

下图3为34层ResNet模型的架构图，仿照AlexNet的8层网络结构可以将ResNet划分成8个构建层。

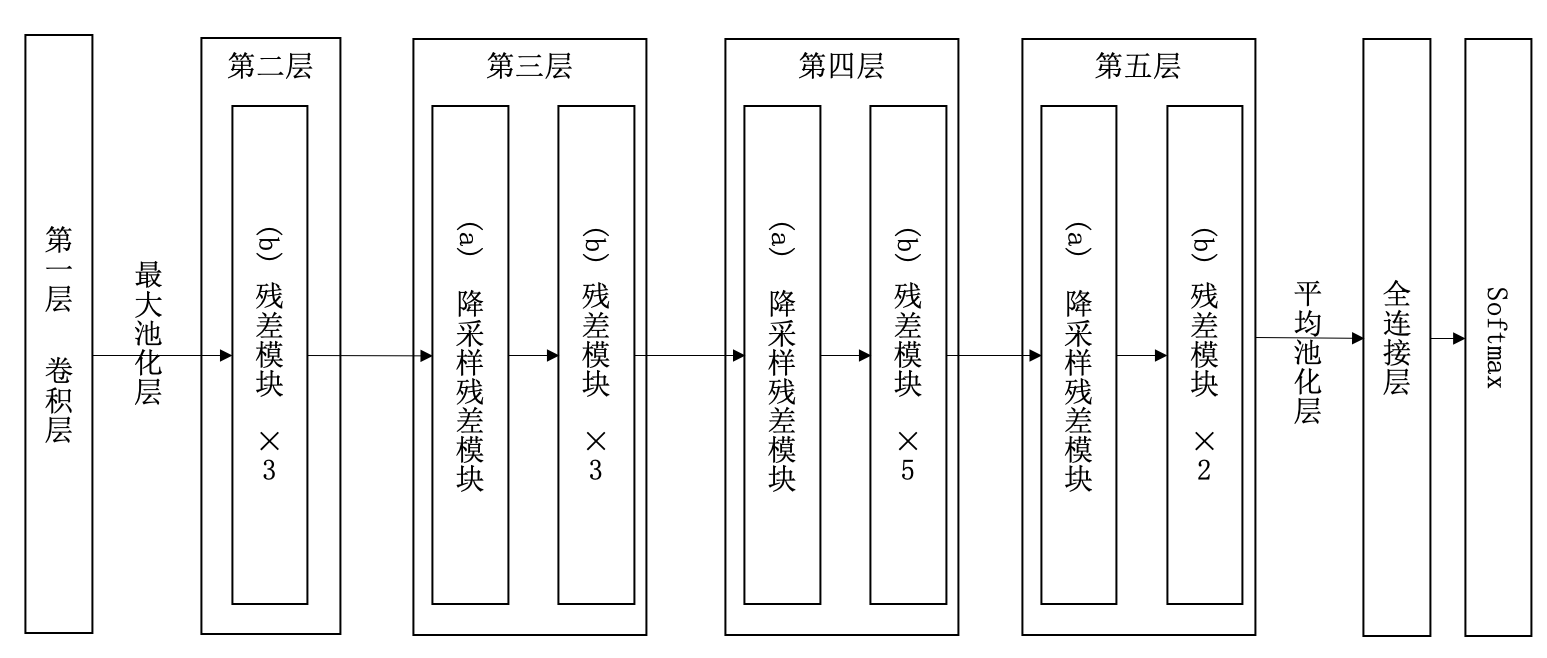


图3 ResNet判别器结构

1. **实验设置及结果分析（包括实验数据集）**

## 3.1 实验数据集及数据预处理

MNIST数据集(Mixed National Institute of Standards and Technology database)是美国国家标准与技术研究院收集整理的大型手写数字数据集，包含60,000个样本的训练集以及10,000个样本的测试集。其中包括0到9的数字。

图形用户界面, 应用程序

描述已自动生成

图4 MNIST数据集

CIFAR-10数据集(Canadian Institute for Advanced Research-10)是一个由加拿大高级研究所创建和维护的经典计算机视觉数据集，包含60,000张图像。其中分为飞机、汽车、猫、狗等10个不同类别。

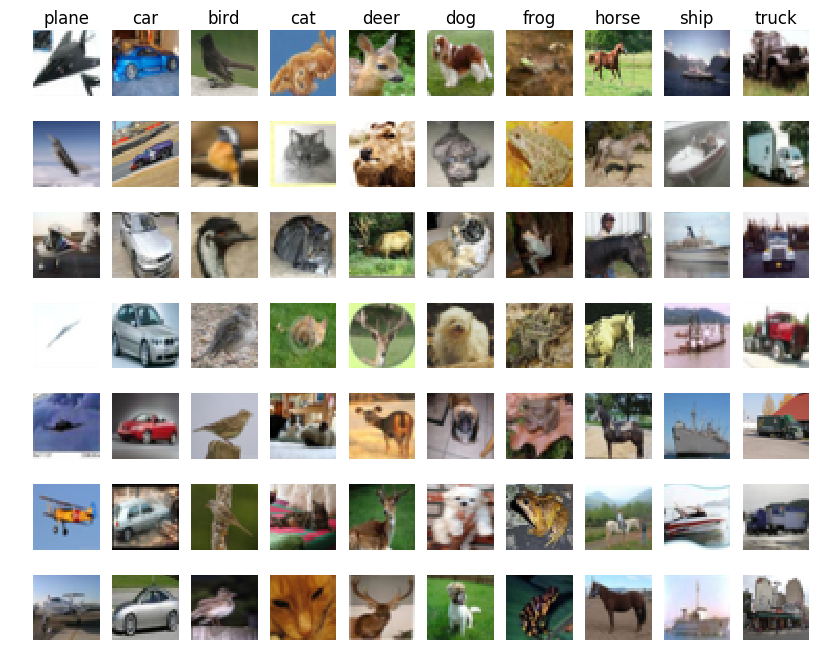


图5 CIFAR-10数据集

在本实验中，使用torchvision自带的数据集加载MNIST和CIFAR-10数据集，并使用transforms.ToTensor方法加载为Tensor张量，最后通过DataLoader加载进GPU进行运算。



图6 数据预处理代码

**3.2 模型设计**

在本次实验中，仿照AlexNet，实现了包含五个卷积层和三个全连接层构建一个深度卷积神经网络，网络的定义是重写nn.Module实现的，卷积层和全连接层之间将数据通过view拉平，同时加入Dropout层防止数据过拟合。

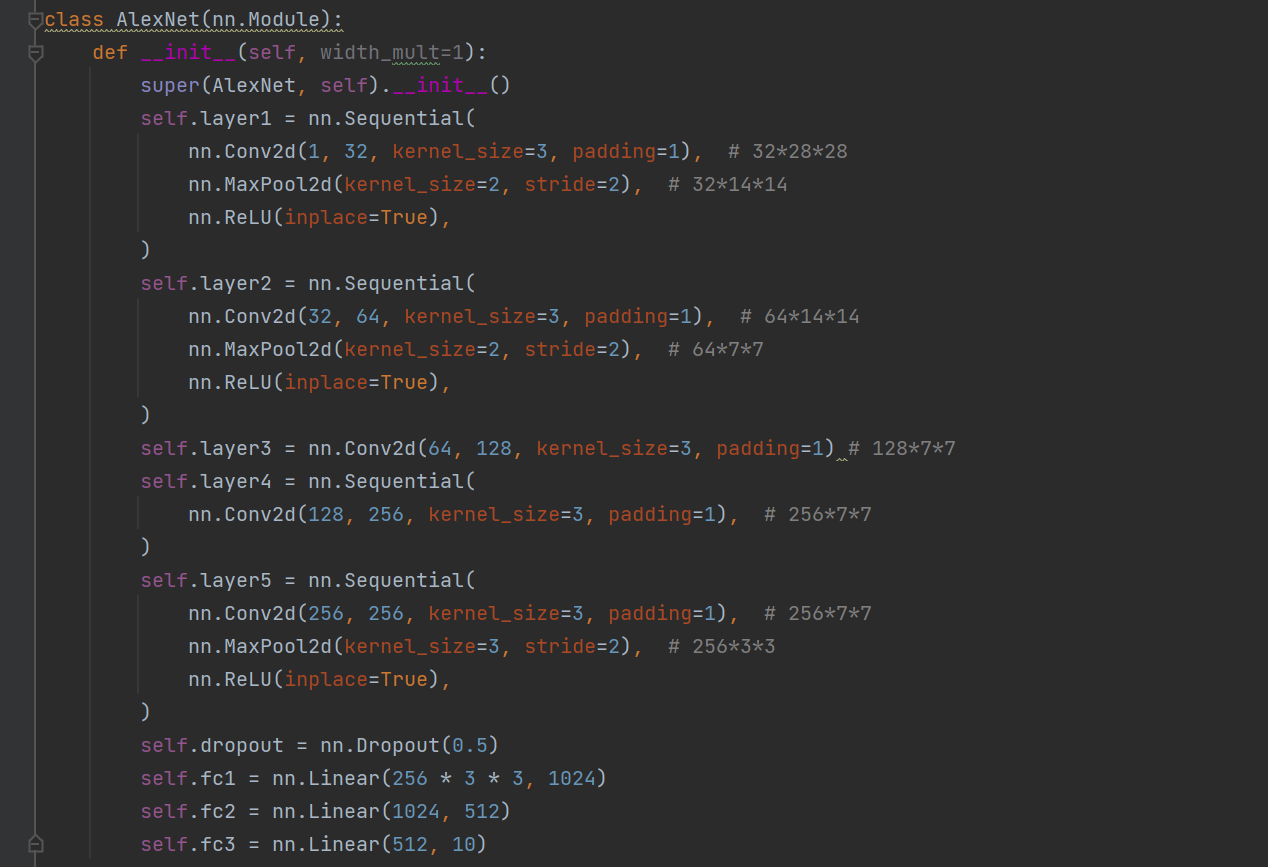


图7 AlexNet模型结构代码

**3.3 实验结果**

在本次实验中，使用交叉熵损失函数和SGD优化器，将模型输入通道根据数据集设为或者，并设置训练超参数为，为，学习率为。训练过程中损失函数的值和在测试集上的准确率变化如下图8所示。

实验发现，随训练过程的进行，损失函数不断降低，在测试集上准确率逐渐升高，最终测试正确率最高能够达到约。损失函数和测试准确率在训练最后阶段呈现波动态，可能原因是在局部最优点附近振荡。

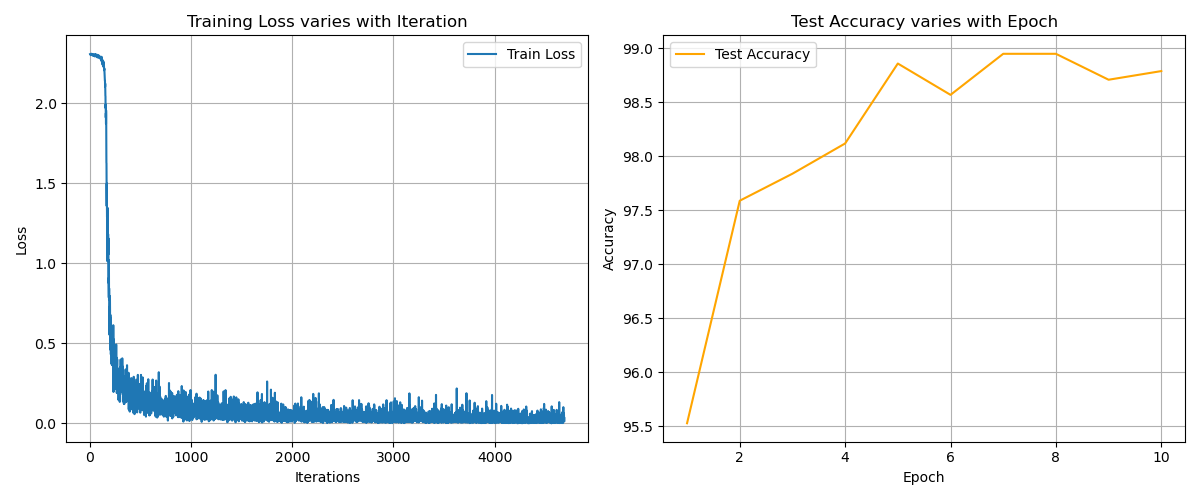


图8 MNIST实验结果（左为损失变化，右为测试集上准确率）

而后通过torch.load方法加载模型对测试集进行直观展示，模型能够对手写数字作出较为准确的分类，具有一定的泛化能力。

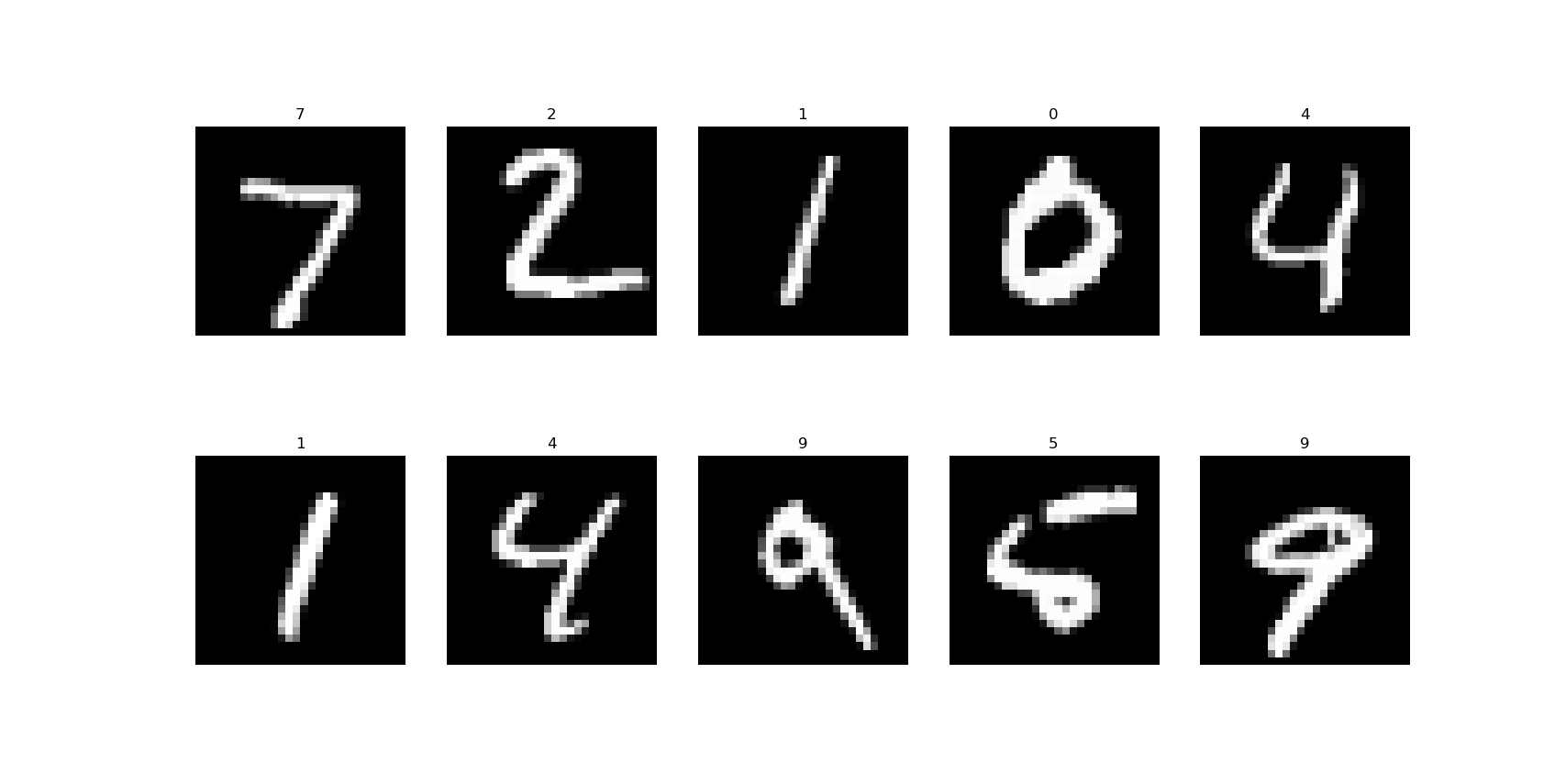


图9 测试集上分类效果

将此模型应用在CIFAR-10数据集上进行训练，分类效果大幅降低，经过30个后准确率仅达到。

图表, 折线图

描述已自动生成

图10 CIFAR-10实验结果（左为损失变化，右为测试集上准确率）

另外使用torch.model中封装好的ResNet18和ResNeXt50模型在CIFAR数据集上进行对比训练，经过足够的训练后，发现测试准确率略有提升，但最高只能达到85%左右，分析可能有以下原因：

1. 模型的表达能力较差，难以在3通道较为复杂的彩色图像中实现较为准确的识别分类。
2. 数据集图像分辨率较低，且种类间具有较为相似的特征，包含信息较少。
3. 数据集缺乏预处理和增强，模型可能很难捕捉到不同类别间的关键特征。
4. 模型训练过程中可能存在过拟合现象，需要对它们进行适当调整或加入正则化约束其参数的更新。
5. **实验结论**

卷积神经网络使用卷积操作，相较于全连接，其网络层与层之间的连接是稀疏的。其次同一层的卷积的参数是共享的，且每一层的卷积运算可以并行处理，具有较快的学习和推理速度，同时也具有较强的表示和学习能力，在图像分类领域具有较为广泛的应用。此外，CNN 还可用作其他任务的基础模型，如生成对抗网络（GAN），作为其backbone模型来辅助生成高质量的图像。

1. **实验收获**

在本次实验中，通过实际使用调试进一步了解了现有深度学习框架Pytorch，在自行搭建CNN模型框架的过程中，熟悉了torch.nn中Conv2d、MaxPool2d、Dropout、Linear等常用网络层的接口使用，锻炼提高了我们的代码能力。

同时，也较为深入的学习了如AlexNet、ResNet、ResNeXt等卷积网络模型，通过实际在训练集上的测试作了对比。

在代码实际编写过程中，会因为不注意维度的变化和处理、输入图像的通道数导致程序出错，同时卷积核的大小、步长以及填充也会影响训练的结果。此外，数据存放的位置，是否加载进GPU进行运算，设置合理的学习率也是我们编写代码中需要考虑的，在逐步调试代码的过程中也是对我们能力的锻炼。

1. **参考文献**

[1] 李航. 统计学习方法[M]. 清华大学出版社, 2012.

[2] 周志华. 机器学习[M]. 清华大学出版社, 2016.

[3] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolute onal neural networks[J]. Advances in neural information processing systems, 2012, 25.

[4] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.