

机器学习实验 - 2023 春

实验五：深度学习

实验内容：深度学习

作业提交截止时间：2023 年 5 月 27 日 23:59:59

环境要求：

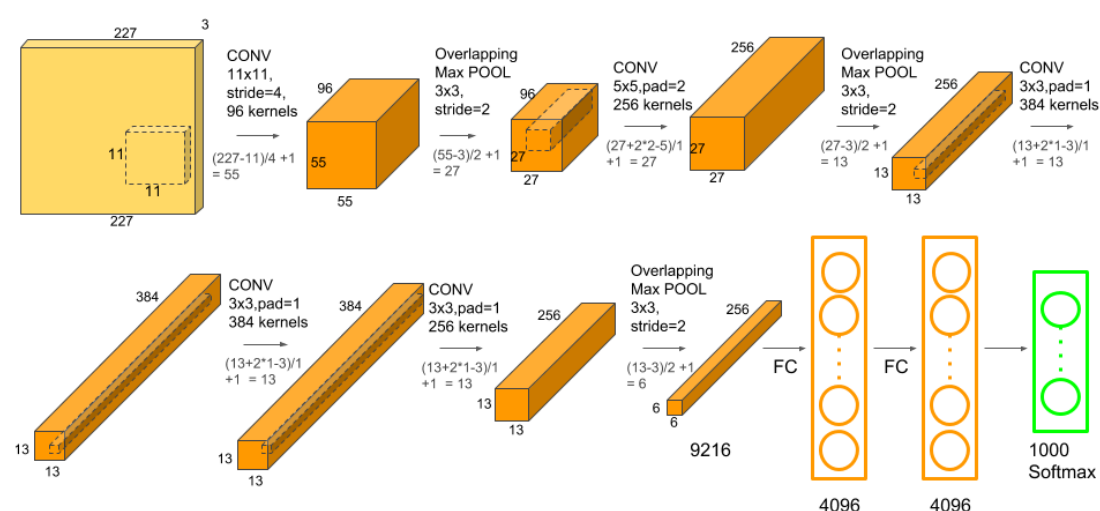
Python, numpy 支持多维度的数组和矩阵运算, pandas 数据处理和分析工具, Matplotlib 图形化工具, pytorch 深度学习库, tensorflow 深度学习库, 中文分词工具 jieba。

任务一：卷积神经网络

本任务中你将使用深度学习库 pytorch 完成图像分类任务, 搭建卷积神经网络的经典模型 AlexNet, 完成对 cifar-10 数据的分类任务, 记录训练过程中的损失和准确率以及测试集的损失和准确率, 并将其可视化, 分析结果。

文件夹 cifar-10-batches-py 包含我们的图像分类问题的数据集, data_batch_1, 2..5 为训练数据, test_batch 为测试数据, 为了方便图片转成张量, 使用 torchvision 加载和处理数据集。图像加载预处理示例代码见 ex5-1.py。

模型细化结构如下图所示



任务二：循环神经网络

本任务中你将深度学习工具 tensorflow.keras 完成中文文本的情感分析,

搭建循环神经网络中的长短期记忆网络 LSTM 模型，使用 `keras` 提供的文本嵌入分词处理工具，完成对中文评论数据集的情感分析。将数据集留出 30% 作为测试集，记录训练过程的损失和准确率以及测试集的损失和准确率，并将其可视化，分析结果。

文件 `comments.csv` 包含我们的中文评论的数据集，`comments` 代表一条评论，`label` 代表文本情感（好评/差评）。以及文本预处理的中文停等词文件 `chineseStopWord.txt`。文本预处理示例代码见 `ex5-2.py`。

作业提交内容：

1. 作业代码截图
2. 实验结果及分析
3. 请将以上结果保存在实验报告（pdf 或者 word 格式）中，命名为 学号+姓名+第几次实验，邮件发送到 facanhe@163.com