**2025年春季学期《机器学习》实验报告**

**（第三次实验）**

班级： 230617 学号： 23371007 姓名： 刘嘉明

一、在CNN实验中，卷积层和池化层是模型的关键组成部分。请说明它们各自的作用，并比较CNN与MLP在图像处理上的结构差异，解释为何CNN更适合处理具有空间结构的图像数据。

卷积层：通过卷积核在图像局部区域滑动计算，提取边缘、纹理等局部特征

池化层：通过最大池化或平均池化降低特征图尺寸，减少计算量，逐步压缩冗余信息，保留显著特征

结构差异：CNN多层卷积+连接层，MLP直接采用全连接

MLP直接把二维，三维的图像展平为一维，空间结构不易凸显，局部特征学习较为困难，而CNN先是处理二维，三维图像，获取到关键特征后再变成一维，降低了参数量，保留了空间结构特征

二、本实验中在每个卷积层后使用了批归一化（Batch Normalization, BN）。请解释批归一化的基本原理，以及它在训练过程中的具体作用。若移除BN层，模型的训练效果可能会受到哪些影响？

批归一化通过对每一批的输入数据进行标准化处理，使其均值为0、方差为1，从而缓解深层网络中梯度消失/爆炸问题

若删除BN层，训练速度会显著下降，因为需使用更小的学习率以避免梯度不稳定，导致收敛缓慢

三、在训练CNN模型时，我们通常使用翻转、裁剪等数据增强方法。请说明数据增强的原理及其对模型性能的提升作用。

数据增强通过对训练数据进行有意义的变换，生成多样化的新样本，从而扩充数据集

这样以来，可以增加数据多样性，解决小样本场景下的过拟合问题，提升模型的泛化能力

四、请分析RNN在处理音频识别任务时主要有哪些局限性导致其性能不佳？

1. 音频信号的语义通常依赖远距离上下文。RNN依靠循环结构传递隐藏状态，但梯度在时间步上反向传播时会指数级衰减或爆炸，导致模型难以学习长序列的关联

2. 音频信号具有层次化结构（如语音中的音素→音节→单词→句子）。RNN的单一循环结构难以显式建模不同时间尺度的特征

五、请结合模型和音频识别任务特点分析LSTM相较于常规RNN模型有哪些优势？

通过门控机制和细胞状态，可以实现：

1. 遗忘门：决定哪些历史信息需要保留或丢弃
2. 输入门：控制当前输入信息的更新
3. 细胞状态：跨时间步传递关键信息（如语音中的音调、语义上下文），梯度几乎无衰减。

此外，同一单词的持续时间可能因语速不同而拉伸或压缩。而LSTM可以通过遗忘门动态调整对历史信息的记忆强度，适应不同时间尺度的变化。相比之下，常规RNN的固定时间步处理则难以灵活应对时间扭曲。

六、请简述RNN的另一种改进：门控循环单元（Gated Recurrent Unit, GRU）基本原理，并简要比较RNN、LSTM和GRU各自优缺点。

GRU是RNN的另一种改进，通过简化LSTM的结构减少参数量的同时保留核心功能。其核心组件为两个门控机制：更新门和重置门

RNN：结构简单，计算快，但是准确率低

LSTM：门控机制灵活，准确率高，但是参数量大，训练耗费更多资源

GRU：参数量少，训练速度较快，性能接近LSTM，但在复杂情境下能力略差于LSTM