**2025年春季学期《机器学习》实验报告**

**（第三次实验）**

班级： 230611 学号： 23373193 姓名： 王天一

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。

卷积层通过滤波器提取图像的局部特征，如边缘和纹理；池化层则对特征图进行降维，减少计算量并赋予模型一定的平移不变性。与将图像展平处理导致空间信息丢失的MLP不同，CNN凭借其局部连接、权值共享和层级结构，能够有效捕捉并利用图像固有的空间层次特性，因此在处理具有空间结构的图像数据时表现更优越。

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小批数据进行标准化处理，使其均值为0、方差为1，并引入可学习参数恢复表达能力。其能缓解内部协变量偏移，加速模型收敛，允许使用更高学习率，并具有一定正则化效果，减少对初始化的敏感性。如果移除BN层，模型训练可能变慢且不稳定，对参数选择更敏感，更易过拟合，深层网络训练难度也会增加。

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。

数据增强的原理是对现有训练数据施加随机变换（如翻转、裁剪）以扩充数据集，让模型学习对这些变换的不变性，而标签保持不变。这些方法能提升模型的泛化能力，即使其更好地适应未见过的数据。同时，通过增加样本多样性，我们能够有效减少过拟合风险，并增强了模型对输入扰动的鲁棒性，从而提升模型性能 。

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？

RNN在处理音频识别任务时，由于其固有结构导致的梯度消失或梯度爆炸问题，难以有效学习音频信号中普遍存在的长程依赖关系。这意味着模型的记忆有限，无法充分捕捉相隔较远但语义相关的音频特征。此外，RNN对序列信息的编码效率不高，且其顺序处理机制限制了并行计算能力，导致处理长音频时效率低下，这些都制约了其在音频识别中的性能表现。

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？

相较于常规RNN，LSTM通过其的门控机制和记忆单元，能有效缓解梯度消失问题，从而出色地捕捉音频信号中的长程依赖关系。这使其能够更精细地控制信息流，记住关键的远距离上下文信息，并遗忘不相关内容。针对音频任务中复杂的时序动态特性，LSTM的这种长时记忆与选择性更新能力，使其在语音识别等任务中表现远超传统RNN。

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的一种改进，它使用重置门和更新门来控制信息流动，以解决长程依赖问题。重置门决定忽略多少过去信息，更新门控制新旧信息的融合程度。相比于RNN，LSTM虽能学长程依赖但参数多计算大，GRU结构更简单，参数较少，计算效率通常优于LSTM，常能在许多任务上达到与LSTM相近的性能，尤其在数据量不大时表现更佳，但LSTM在极复杂任务上可能略有优势 。