**2025年春季学期《机器学习》实验报告**

**（第三次实验）**

班级： 230617 学号： 23373179 姓名： 周子强

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。

**（1）卷积层的作用：**

1.**特征提取**：通过卷积核（滤波器）的滑动窗口操作，提取图像的局部特征（边缘、纹理、颜色等）。

2.**局部感知**：每个卷积核仅关注输入图像的局部区域，而非全局，符合图像中相邻像素关联性强的特点。

3.**权值共享**：同一卷积核在整个输入图像上共享参数，大幅减少参数量，降低过拟合风险。

**（2）池化层的作用：**

1.**下采样（降维）**：通过最大池化或平均池化，减少特征图的空间尺寸，降低计算复杂度。

2.**保留主要特征**：池化操作能保留局部区域的显著特征（如最大池化保留最明显的特征），增强模型对微小平移的鲁棒性。

3.**防止过拟合**：通过降低特征维度，间接减少后续层参数，抑制过拟合。

**（3）CNN与MLP的结构差异：**

|  |  |  |
| --- | --- | --- |
| **特征** | **CNN** | **MLP** |
| **连接方式** | 局部连接（卷积核） | 全连接（每个神经元连接所有输入） |
| **参数共享** | 卷积核权值共享 | 无参数共享 |
| **空间信息处理** | 保留二维空间结构 | 输入展平为一维向量 |
| **层次化特征提取** | 底层特征->高级语义 | 直接学习全局特征 |

**（4）为什么CNN更适合处理具有空间结构的图像数据：**

1.局部相关性：图像的相邻像素具有强相关性，CNN的卷积核通过局部感知捕捉这种关联，而MLP全连接忽略了这种空间信息。

2.平移不变性：池化层使模型对目标位置变化不敏感。

3.参数效率：权值共享和局部连接大幅减少参数量。

4.层次化特征学习：浅层卷积提取特征边缘/纹理，深层卷积组合为复杂语义，更贴合图像的多层次特性。

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？

**（1）批归一化的基本原理：**

1.标准化输入分布：对每个批次的输入数据，按通道计算均值和方差，并进行标准化：

其中，和分别为当前批次均值和方差，是极小值防止除零。

2.可学习的缩放与偏移：引入两个可训练参数（缩放因子）和（偏移因子），恢复数据表达能力：

**（2）BN在训练过程中的具体作用：**

1.加速模型收敛：标准化输入分布，缓解内部协变量偏移，使每层输入的分布稳定，降低梯度更新的震荡；允许使用更大的学习率，加快训练速度。

2.缓解梯度问题：减轻梯度消失或梯度爆炸的问题，适用于深层网络。

3.正则化效果：每个批次的均值和方差引入噪声，抑制过拟合。

4.降低对初始化的敏感性：减少网络对权重初始化的依赖，提升训练稳定性。

**（3）移除BN层的影响：**

1.训练速度显著下降。

2.模型稳定性降低。

3.过拟合风险增加。

4.对初始化和参数敏感。

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。

**（1）数据增强的原理：**通过对训练数据进行有意义的变换，生成多样化的新样本，如几何变换（翻转、裁剪）、颜色变换（亮度调整）、噪声添加、随机裁剪、不同方向旋转、颜色抖动。

**（2）对模型性能的提升作用：**增强泛化能力、减少过拟合、提升鲁棒性、加速收敛。

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.**长时依赖处理能力不足：**RNN的隐藏状态通过简单递归更新（

），随着时间步增加，梯度在反向传播时容易消失或爆炸，导致无法有效捕捉长距离依赖。语音任务中语义依赖上下文多个音素，普通RNN难以建模远距离时间步的关联，导致关键信息丢失。

2.**隐藏状态表达能力有限：**隐藏状态仅通过线性变换和非线性激活（tanh）更新，无法灵活控制信息流动，难以捕捉复杂时序模式。

3.**计算效率低下：**RNN按照时间步顺序计算，无法并行化处理序列。

4.**对输入扰动敏感：**RNN缺乏对噪声、语速变化等扰动的鲁棒性设计。

5.**参数共享的局限性：**RNN的权值共享虽减少参数量，但固定权重难以适应多尺度时序模式。

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？

1.**长时依赖处理能力：**LSTM引入细胞状态作为记忆通道，通过门控机制（遗忘门、输入门）长期保留关键信息。其中，遗忘门决定从细胞状态中丢弃哪些信息（无关噪声），输入门筛选当前输入中需要存储到细胞状态的新信息（关键音素），输出门控制细胞状态对当前输出的影响。

2.**动态信息流控制：**LSTM输入门和遗忘门根据当前输入和前一状态动态调整权重，例如语速变化时遗忘门可决定保留更多历史信息；输出门控制当前隐藏状态的表达强度，适应不同时间步的特征重要性差异。

3.**梯度稳定性与收敛速度：**LSTM细胞状态的更新公式允许梯度直接流过时间步，环节梯度消失问题；遗忘门和输入门的Sigmoid函数输出值在0~1，能够稳定梯度传播。

4.**噪声鲁棒性：**LSTM的遗忘门可以过滤噪声，保留关键语音特征。

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。

**（1）GRU基本原理：**GRU是RNN的改进结构，通过简化LSTM的门控机制，结合更新门和重置门来管理时序信息流，核心公式如下：

1.重置门：

控制前一时刻隐藏状态对当前候选状态的影响程度，用于捕捉短期依赖。

2.更新门：

决定保留多少旧状态，并引入多少新候选状态。

3.候选隐藏状态：

结合当前输入和经重置门筛选的历史状态，生成临时状态。

4.最终隐藏状态：

通过更新门融合旧状态和新候选状态，完成信息更新。

**（2）优缺点对比：**

|  |  |  |
| --- | --- | --- |
| **模型** | **优点** | **缺点** |
| **RNN** | 结构简单，参数量少 | 无法处理长时依赖，梯度不稳定，隐藏状态表达能力有限 |
| **LSTM** | 引入遗忘门、输入门、输出门和细胞状态，有效捕捉长时依赖，鲁棒性强 | 结构复杂，参数量大，计算成本高，训练速度慢 |
| **GRU** | 合并更新门和重置门，简化结构，训练更快，资源消耗低 | 对极长序列建模能力弱于LSTM，某些任务可能牺牲部分性能换取效率 |