

计算机视觉 lab6 实验报告

实验概述

本次实验实现了 VGG-16、ResNet-34 和 ResNext-50 三个经典的卷积神经网络模型，并在 CIFAR-10 数据集上进行了训练和测试。通过对比不同模型的参数量、训练集和测试集上的表现，分析了它们的泛化能力和性能差异。

数据集与数据增强

数据集

- 使用 CIFAR-10 数据集，包含 10 个类别的 60,000 张 32x32 彩色图像，其中 50,000 张用于训练，10,000 张用于测试。

数据增强

为了提升模型的泛化能力，使用了以下数据增强方法：

1. **随机水平翻转**：以 50% 的概率对图像进行水平翻转。
2. **随机旋转**：对图像进行随机旋转，角度范围为 $\pm 20^\circ$ 。
3. **颜色抖动**：随机调整图像的亮度、对比度、饱和度和色调。
4. **随机裁剪**：对图像进行随机裁剪并缩放到 32x32 大小。
5. **归一化**：将图像像素值归一化到 [0, 1] 范围，并使用 ImageNet 的均值和标准差进行标准化。

```
train_transform = transforms.Compose([
    transforms.RandomHorizontalFlip(),
    transforms.ColorJitter(brightness=0.1, contrast=0.1, saturation=0.1, hue=0.1),
    transforms.RandomRotation(20),
    transforms.RandomResizedCrop(32, scale=(0.8, 1.0)),
    transforms.ToTensor(),
    transforms.Normalize(mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225])
])
```

训练超参

训练超参如下：

- optimizer = AdamW
- lr = 0.001
- wd = 1e-3 (ResNext 1e-4)
- batch_size = 256
- epoch = 50
- scheduler = CosineAnnealingLR

VGG-16

模型结构

- **卷积层**：13 层，卷积核大小为 3x3，池化核大小为 2x2。
- **全连接层**：3 层，神经元个数为 4096。
- **参数量**：约 1.38 亿。

性能

- **训练集准确率**：94.3%
- **测试集准确率**：87.8%

- **分析：**VGG-16 的参数数量最大，但由于其结构简单且深度适中，在 CIFAR-10 数据集上表现优异。

ResNet-34

模型结构

- **模块：**4 个模块，每个模块分别包含 3、4、6、3 个残差块。
- **残差块：**每个残差块包含两个 3x3 卷积层，跳跃连接绕过了卷积层，还包含了 BatchNorm 层。
- **参数量：**约 2,100 万。
- **dropout：**在残差块之间添加了 0.2 的 dropout 层，在全连接层之前添加了 0.5 的 dropout 层。

性能

- **训练集准确率：**85.12%
- **测试集准确率：**82.61%
- **分析：**ResNet-34 通过残差连接缓解了梯度消失问题，性能略低于 VGG-16，但参数量更少，适合中等规模数据集。

ResNext

模型结构

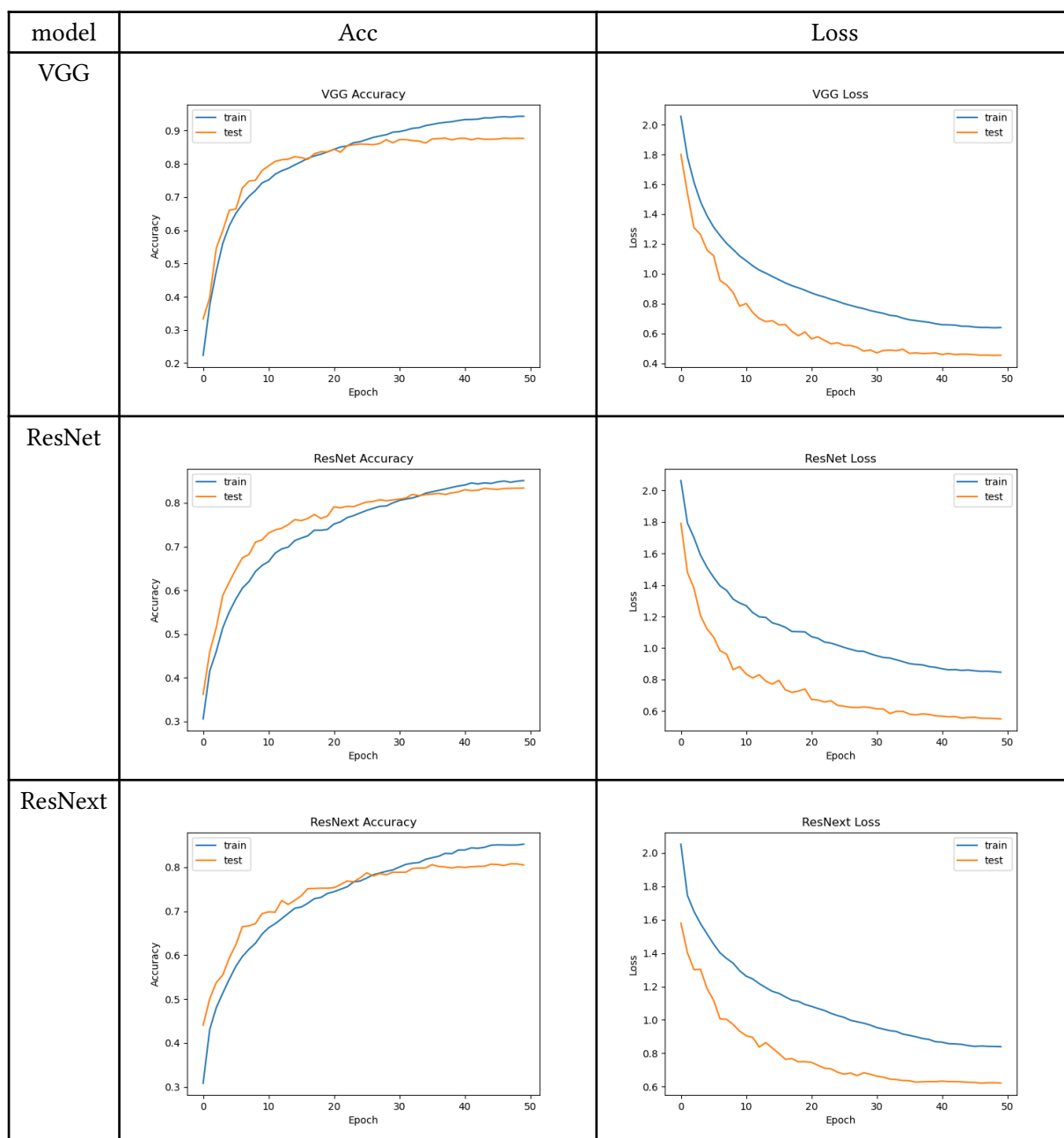
- **模块：**4 个模块，每个模块分别包含 2、2、2、2 个 ResNext 块。
- **ResNext 块：**每个块包含分组卷积（基数 Cardinality=16），瓶颈层宽度为输出通道数的一半。
- **参数量：**约 377 万。

性能

- **训练集准确率：**85.26%
- **测试集准确率：**80.16%
- **分析：**ResNext 通过分组卷积提升了模型的表达能力，但由于其结构复杂且训练难度较大，性能略低于 ResNet-34。

结果展示

以下是各模型的训练和验证曲线：



最终效果均满足要求，且都没有出现较为严重的过拟合现象。

实验总结

1. **VGG-16** 在 CIFAR-10 数据集上表现最佳，但其参数量较大，适合在小规模数据集上使用。
2. **ResNet-34** 通过残差连接缓解了梯度消失问题，性能略低于 VGG-16，但参数量更少，适合中等规模数据集。
3. **ResNext-50** 通过分组卷积提升了表达能力，但由于其结构复杂，训练难度较大，性能略低于 ResNet-34。