图像分类实验报告

王松宸 2024201594

2025 年 12 月 13 日

1 实验背景与个人理解

图像分类是计算机视觉中最基础也是最核心的任务之一。其目标是将输入的图像（通常表示为像素矩阵）映射到预定义的类别标签集合中。

在深度学习时代，卷积神经网络（CNN）成为了解决这一任务的主流方法。CNN 通过卷积层提取图像的局部特征（如边缘、纹理），通过池化层降低特征维度并保持平移不变性，最后通过全连接层将提取的高级语义特征映射到类别概率。本次实验中，基于 CIFAR-10 数据集，我设计并实现了模仿 VGG 和 ResNet 架构的两个神经网络模型，并对比了它们的结构特点与运行结果。

2 网络设计与结果分析

2.1 My_VGG_Net

2.1.1 网络结构设计

我设计了一个简化版的 VGG 网络（My_VGG_Net），其核心设计理念源自 VGG 论文中的“堆叠小卷积核”思想。主要特点如下：

- **Block 结构**：网络由两个卷积块（Block）组成。每个 Block 内部包含多个连续的卷积层，最后接一个最大池化层（Max Pooling）。
- **3x3 卷积与 Padding**：大部分卷积层使用了 3×3 的卷积核配合 $padding = 1$。这样做的好处是卷积操作不改变特征图的尺寸，将尺寸减半的任务完全交给池化层，使得网络结构更加清晰。
- **1x1 卷积**：在第二个 Block 中，我引入了一个 1×1 的卷积层（$padding = 0$）。这是模仿 VGG-C 配置的设计，目的是在不改变感受野大小的前提下，增加网络的非线性变换能力，同时调整通道数。
- **全连接层**：最后通过三个全连接层将展平后的特征映射到 10 个分类节点。

2.1.2 网络结构图示

下表展示了 My_VGG_Net 的详细数据流向与张量尺寸变化：

阶段	层类型	配置	输出尺寸
输入	Input	-	$3 \times 32 \times 32$
Block 1	Conv2d $\times 2$	3×3, 64 filters	$64 \times 32 \times 32$
	MaxPool2d	2×2	$64 \times 16 \times 16$
Block 2	Conv2d $\times 2$	3×3, 128 filters	$128 \times 16 \times 16$
	Conv2d (1×1)	1×1, 128 filters	$128 \times 16 \times 16$
	MaxPool2d	2×2	$128 \times 8 \times 8$
分类器	Flatten	-	8192
	Linear	256 units	256
	Linear	60 units	60
	Linear	10 units	10

表 1: My_VGG_Net 网络结构表

2.1.3 运行结果与分析

该网络的运行结果如下图所示：

```
Epoch  5, Step  2000. Loss: 0.605
Epoch  5, Step  4000. Loss: 0.613
Epoch  5, Step  6000. Loss: 0.604
Epoch  5, Step  8000. Loss: 0.617
Epoch  5, Step 10000. Loss: 0.615
Epoch  5, Step 12000. Loss: 0.587
Finished Training
Accuracy of the network on the 10000 test images: 74 %
Accuracy for class: plane is 78.9 %
Accuracy for class: car   is 82.1 %
Accuracy for class: bird  is 48.3 %
Accuracy for class: cat   is 55.6 %
Accuracy for class: deer  is 76.7 %
Accuracy for class: dog   is 78.7 %
Accuracy for class: frog  is 82.0 %
Accuracy for class: horse is 79.6 %
Accuracy for class: ship  is 75.5 %
Accuracy for class: truck is 90.2 %
```

图 1: My_VGG_Net 在 CIFAR-10 上的训练与测试结果

分析:VGG 结构通过加深网络深度提取了更丰富的特征。相比于只有两层卷积的 BaselineNet，My_VGG_Net 拥有 5 个卷积层和 3 个全连接层，具有更强的拟合能力。从结果可以看出，图像分类的准确率提升至 74%。

2.2 My_Res_Net

2.2.1 网络结构设计

我还设计了一个基于残差学习的轻量级网络（My_Res_Net），模仿了 ResNet-18 的结构。主要特点如下：

- **残差连接（Shortcut）**：引入了 $F(x)+x$ 的结构。这种跳跃连接允许梯度直接流向浅层，有效缓解了深层网络中的梯度消失问题，使得训练更深的网络成为可能。
- **步长下采样（Stride）**：与 VGG 不同，我在 ResNet 中放弃了最大池化层（Max Pooling），而是在卷积层中设置 $stride = 2$ 来进行下采样。这让网络能够自主学习如何压缩特征图，保留更多信息。
- **全局平均池化（GAP）**：在全连接层之前，我使用了 `AdaptiveAvgPool2d((1, 1))`。这将每个通道的特征图直接压缩为一个数值，极大地减少了全连接层的参数量，降低了过拟合的风险。

2.2.2 网络结构图示

下表展示了 My_Res_Net 的详细数据流向与张量尺寸变化：

阶段	层类型	配置	输出尺寸
输入	Input	-	$3 \times 32 \times 32$
预处理	Conv2d	3×3, 64 filters	$64 \times 32 \times 32$
Block 1	ResBlock	3×3, 64 filters	$64 \times 32 \times 32$
Block 2	ResBlock	3×3, 128 filters, stride 2	$128 \times 16 \times 16$
Block 3	ResBlock	3×3, 256 filters, stride 2	$256 \times 8 \times 8$
分类器	Global AvgPool	-	$256 \times 1 \times 1$
	Flatten	-	256
	Linear	10 units	10

表 2: My_Res_Net 网络结构表

2.2.3 运行结果与分析

该网络的运行结果如下图所示：

```
Epoch  5, Step  2000. Loss: 1.011
Epoch  5, Step  4000. Loss: 0.975
Epoch  5, Step  6000. Loss: 0.989
Epoch  5, Step  8000. Loss: 0.964
Epoch  5, Step 10000. Loss: 0.970
Epoch  5, Step 12000. Loss: 0.937
Finished Training
Accuracy of the network on the 10000 test images: 68 %
Accuracy for class: plane is 78.9 %
Accuracy for class: car   is 81.0 %
Accuracy for class: bird  is 43.9 %
Accuracy for class: cat   is 63.2 %
Accuracy for class: deer  is 57.2 %
Accuracy for class: dog   is 46.5 %
Accuracy for class: frog  is 86.7 %
Accuracy for class: horse is 67.7 %
Accuracy for class: ship  is 76.3 %
Accuracy for class: truck is 80.8 %
```

图 2: My_Res_Net 在 CIFAR-10 上的训练与测试结果

分析：ResNet 的设计使得特征的传递更加顺畅。通过使用 Stride 代替 Pooling，以及引入 GAP，网络能够保持高性能。准确率提升至 68%，低于 My_VGG_Net 的可能原因是 Epoch 数较少，未完全收敛。但为保证不同网络结构的对比公平性，我未增加训练轮数。

3 实验收获

通过本次实验，我不仅熟悉了 PyTorch 框架的基本使用，包括 `Dataset` 加载、模型定义、损失函数与优化器的选择，更深入理解了 CNN 架构演变背后的逻辑：

1. **架构设计的权衡**：VGG 展示了通过堆叠简单 3×3 卷积也能达到很好的效果，但参数量较大；ResNet 则通过残差连接解决了深度带来的退化问题。
2. **细节的重要性**：在实现 ResNet 时，我深刻体会到了维度匹配的重要性。当特征图尺寸或通道数发生变化时，Shortcut 路径也必须进行相应的 1×1 卷积变换，否则无法进行张量相加。
3. **池化策略**：从 VGG 的 Max Pooling 到 ResNet 的 Stride Convolution + Global Average Pooling，我理解了下采样策略的演进及其对特征保留和参数量的影响。