

微调 ImageNet 预训练 CNN 实现 Caltech-101 分类

1 实验简介

本实验旨在通过微调在 ImageNet 上预训练的卷积神经网络 (CNN) 来实现 Caltech-101 数据集的分类任务。我们采用了迁移学习的方法, 比较了微调预训练模型 ResNet-18 与从零开始训练模型在 Caltech-101 数据集上的性能差异。

2 数据集

我们使用标准的Caltech-101 数据集[1], 该数据集包含 101 个类别的物体图像, 外加一个背景类别 (共 102 类, 但通常使用 101 类)。数据集示例如图1所示。数据集划分按照如下方式进行: 将每个类别的 80% 划入训练集, 20% 划入测试集。

3 模型架构

如图2所示, 我们选择了 ResNet-18[2] 这一经典的 CNN 架构进行实验: 对于预训练模型, 我们修改了原始网络的最后一层全连接层, 将其输出维度调整为 101 以匹配 Caltech-101 的类别数。

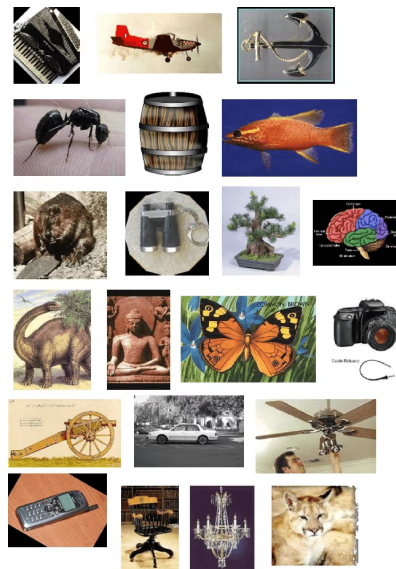


图 1: Caltech-101 数据集示例图像

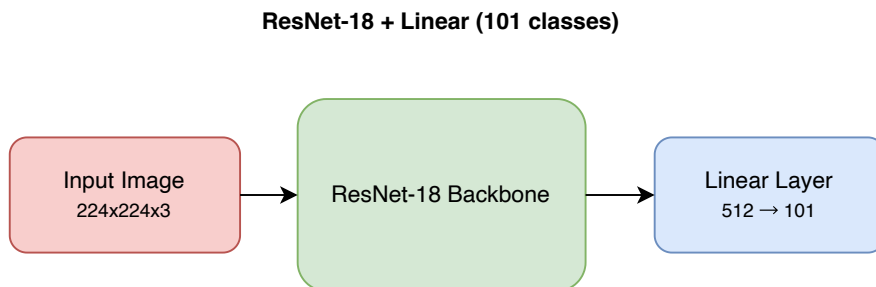


图 2: 微调 ImageNet 所使用的网络架构

4 实验设置

4.1 数据预处理

所有图像被调整为 224×224 像素，并进行了归一化处理：

- 均值 = [0.485, 0.456, 0.406];
- 标准差 = [0.229, 0.224, 0.225].

4.2 训练策略

我们比较了两种训练方式：

1. **微调预训练模型**: 为最后一层全链接层设置较大的学习率，其余层设置较小的学习率。
2. **从零开始训练**: 为所有层设置统一的学习率。

优化器：我们采用了 SGD 优化器，设定动量项的系数为 0.8.

4.3 超参数探索

针对两种不同的训练方式，我们选择的超参数组合和验证集准确率见表1和2，最优的参数选择用加粗表示。

表 1: 微调 Resnet-18 时的不同超参数配置

$lr_{\text{new layer}}$	$lr_{\text{pretrained layer}}$	batch size	epoch	val accuracy(%)
0.05	0.0005	32	50	93.95
0.05	0.0009	32	50	93.67
0.03	0.0007	32	50	94.01
0.05	0.0007	32	50	94.12
0.07	0.0007	32	50	93.89
0.05	0.0007	64	50	93.33
0.05	0.0007	32	25	93.73

5 实验结果

5.1 训练曲线

图??展示了 ResNet-18 模型在最佳超参数设置下的训练曲线。

表 2: 从零训练时的不同超参数配置

lr	batch size	epoch	val accuracy(%)
0.06	32	50	59.83
0.04	32	50	64.07
0.02	32	50	64.58
0.01	32	50	60.28
0.02	64	50	59.20
0.02	32	25	58.25

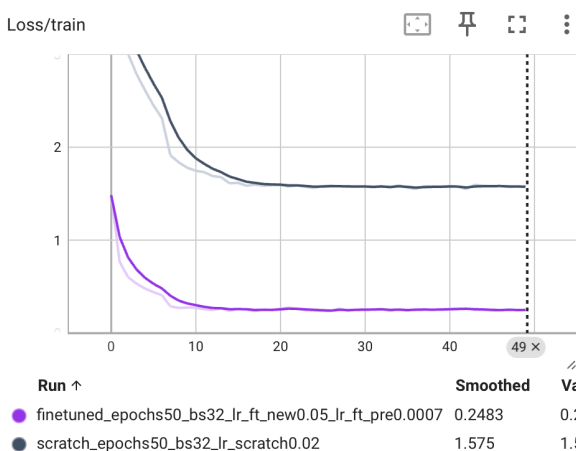


图 3: 最佳超参数设置下训练的 loss 曲线

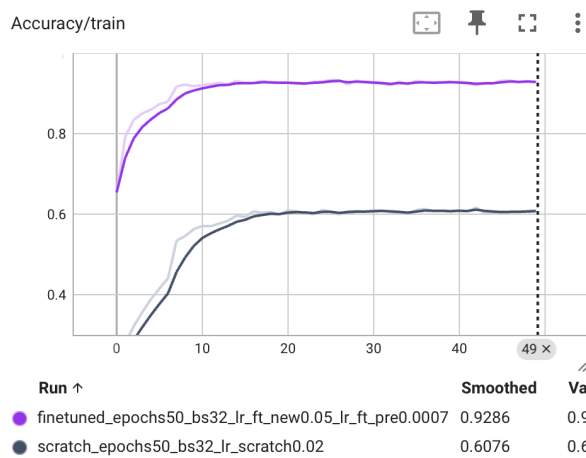


图 4: 最佳超参数设置下训练的 Accuracy 曲线

5.2 模型性能比较

表3比较了模型在不同训练方式在全部数据集下的性能。

表 3: 模型在 Caltech-101 测试集上的性能比较

模型	微调预训练模型 (%)	从零开始训练 (%)
ResNet-18	98.03	70.74

6 分析与讨论

1. **预训练的优势**: 实验结果表明, 使用 ImageNet 预训练模型进行微调比从零开始训练能获得显著更好的性能, 这验证了迁移学习的有效性。
2. **超参数影响**: 微调时, 对新的全链接层设置较大的学习率 (0.05), 其余层设置较小的学习率 (0.0007) 配合较长的训练周期 epoch(50) 获得了最佳性能, 而从零开始训练则需要中等的学习 (0.02) 才能

达到最佳的性能。

7 代码与模型

- 代码仓库: [GitHub Repo](#);
- 预训练模型权重: [Google Drive](#).

参考文献

- [1] Li Fei-Fei, Rob Fergus, and Pietro Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Pattern Recognition Workshop*, 2004.
- [2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.