

微调在ImageNet上预训练的卷积神经网络实现鸟类识别
任务1

神经网络和深度学习

-期中作业-

马超

2024 年 5 月 30 日

目录

1	引言	1
2	ResNet-18架构简介	1
3	数据集	2
4	模型与训练	2
5	实验与结果	3
6	github链接及模型权重下载地址	3
7	结论	4

摘要

本报告详细描述了微调预训练卷积神经网络以实现鸟类识别的过程。通过使用CUB-200-2011数据集，对比了微调预训练模型与从头训练模型的性能提升。

1 引言

残差网络（Residual Network, ResNet）是由何恺明等人在2015年提出的一种卷积神经网络架构。通过引入残差模块，ResNet有效解决了深层网络中的梯度消失问题，使得网络可以训练得更深更有效。

2 ResNet-18架构简介

ResNet-18是ResNet架构中的一种，具有18层深的卷积神经网络。其结构如下：

- **输入层**：输入图像尺寸为224x224x3。
- **卷积层1**：7x7卷积，64个卷积核，步幅为2，紧接着是3x3最大池化层，步幅为2。
- **残差块1**：包含2个3x3卷积，每个卷积层有64个卷积核。
- **残差块2**：包含2个3x3卷积，每个卷积层有128个卷积核，步幅为2。
- **残差块3**：包含2个3x3卷积，每个卷积层有256个卷积核，步幅为2。
- **残差块4**：包含2个3x3卷积，每个卷积层有512个卷积核，步幅为2。
- **全局平均池化层**：将每个特征图降为1x1。
- **全连接层**：输出层，有1000个神经元（在鸟类识别任务中修改为200个神经元）。

每个残差块包含一个快捷连接（skip connection），将输入直接添加到输出。通过这种方式，ResNet解决了深层网络中的梯度消失问题，使得网络可以更深且更高效地进行训练。

ResNet-18通过引入残差块，使得训练更深层的卷积神经网络成为可能。在鸟类识别任务中，我们将输出层调整为200个神经元，以适应CUB-200-2011数据集中的200个类别。

3 数据集

CUB-200-2011数据集包含200种鸟类，共有11,788张图像。数据集分为训练集和验证集。

数据集特点

- **类别：**数据集包含200种不同的鸟类，每种鸟类都有详细的类别标签。
- **图像数量：**总共包含11788张图像，其中5994张用于训练，5794张用于测试。
- **图像分辨率：**图像分辨率不固定，通常较高，适合进行细粒度的图像分类。
- **标注信息：**每张图像都有类别标签、边界框（bounding box）以及15个关键点（如鸟的嘴、眼睛、翅膀等）标注。此外，还有属性标签（attributes），用于描述鸟类的颜色、形状等特征。

数据集用途

CUB-200-2011数据集广泛用于以下研究领域：

- **细粒度图像分类：**区分相似类别（如不同种类的鸟类）的能力。
- **目标检测：**识别图像中的鸟类并定位其位置。
- **关键点检测：**定位鸟类的关键部位（如眼睛、嘴巴等）。
- **属性预测：**根据图像预测鸟类的属性信息。

4 模型与训练

我们使用在ImageNet上预训练的ResNet-18模型。将输出层修改为200类，以适应CUB-200-2011数据集的需求。使用较小的学习率微调预训练的层，并使用较大的学习率训练新的输出层。

5 实验与结果

我们比较了微调预训练模型和从头训练模型的性能。以下图表展示了损失和准确率曲线。

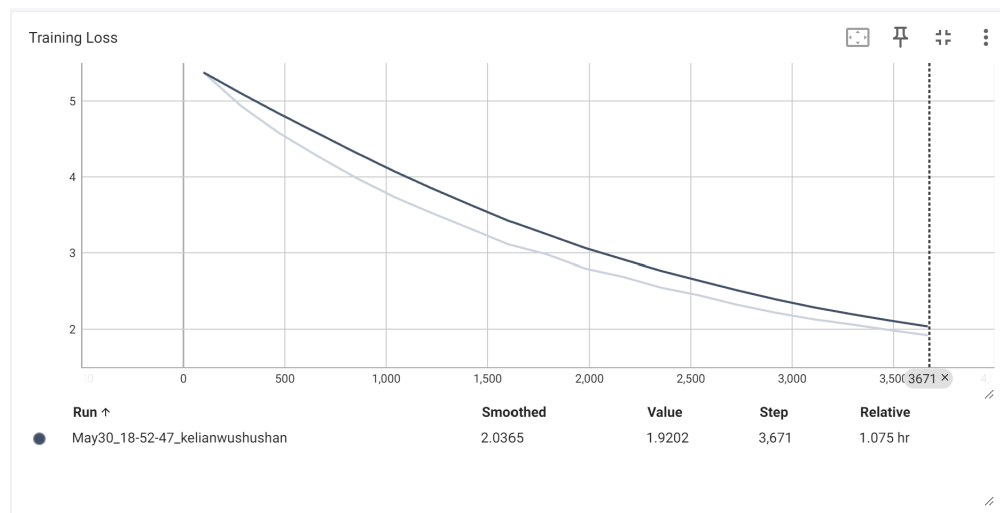


图 1: 训练集和验证集的损失曲线

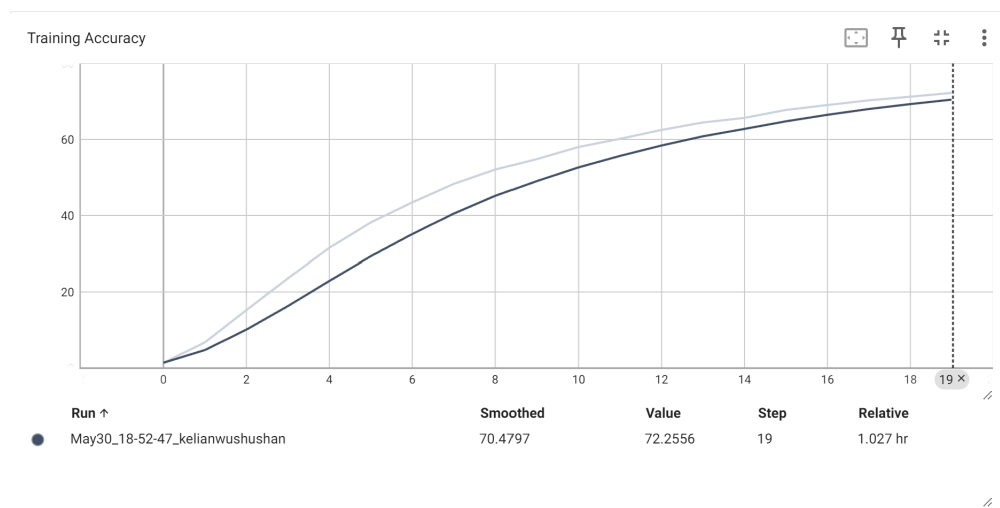


图 2: 训练集和验证集的准确率曲线

6 github链接及模型权重下载地址

repo的public github链接:

<https://github.com/Ethereal-Redolent/hw/tree/master>

模型权重的下载地址:

<https://pan.baidu.com/s/17-gkDYIRdfE16zVFyDsQ7g?pwd=m2hp>

7 结论

微调预训练的卷积神经网络相比于从头训练的模型，显著提升了性能。