

监督学习和自监督学习在图像分类任务上的性能表现

任务1

神经网络和深度学习

-期末作业-

马超

2024 年 6 月 29 日

目录

1	实验概述	1
2	SimCLR 自监督算法	1
3	监督学习算法	1
4	Linear Classification Protocol	2
5	实验数据集简介	3
5.1	CIFAR-100数据集简介	3
5.2	ImageNet数据集简介	3
5.3	数据集比较	4
6	实验结果	4
6.1	数据集和实验设置	4
6.2	训练过程中的Loss曲线	4
6.3	线性分类器的准确率变化	5
7	github链接及模型权重下载地址	6

1 实验概述

本实验的主要目的是实现SimCLR自监督学习算法并使用该算法在数据集上训练ResNet-18模型。随后在CIFAR-100数据集中使用Linear Classification Protocol对其性能进行评测。将上述结果与在ImageNet数据集上采用监督学习训练得到的表征在相同的协议下进行对比，并比较二者相对于在CIFAR-100数据集上从零开始以监督学习方式进行训练所带来的提升。实验还尝试了不同的超参数组合，探索自监督预训练数据集规模对性能的影响。

2 SimCLR 自监督算法

SimCLR (Simple Framework for Contrastive Learning of Visual Representations) 是一种基于对比学习的自监督算法。其主要思想是通过最大化同一图像不同数据增强视图之间的相似度，并最小化不同图像视图之间的相似度，从而学习图像的表征。SimCLR主要包括以下几个步骤：

1. **数据增强**：对每张输入图像应用一系列数据增强操作，生成两个不同的视图（称为正对）。
2. **特征提取**：使用卷积神经网络（如ResNet）提取每个视图的特征表示。
3. **投影头**：将特征表示通过一个小的全连接网络（投影头）映射到一个低维空间。
4. **对比损失**：使用NT-Xent (Normalized Temperature-Scaled Cross Entropy) 损失函数来最大化正对之间的相似度，并最小化负对（不同图像视图）之间的相似度。

SimCLR的目标是通过对比学习获得高质量的图像表征，这些表征可以用于下游任务，如图像分类。

3 监督学习算法

监督学习是一种机器学习范式，其中模型在带有标注的训练数据上进行训练，从而学习输入数据和目标标签之间的映射关系。在图像分类任务中，常用的监督学习方法包括卷积神经网络（CNN），如ResNet，VGG和Inception等。监督学习主要包括以下几个步骤：

1. **数据准备**：收集并标注数据集，通常包括大量已标注的训练样本。

2. **模型选择**: 选择适合任务的模型架构, 例如ResNet-18。
3. **模型训练**: 在标注数据上训练模型, 通过最小化损失函数 (如交叉熵损失) 来优化模型参数。
4. **模型评估**: 在验证集和测试集上评估模型性能, 常用的指标包括准确率、精确率、召回率和F1分数等。

监督学习的优势在于, 其训练过程直接利用了标注数据, 可以有效学习到输入数据和标签之间的关系, 从而在很多任务上表现出色。

4 Linear Classification Protocol

线性分类协议 (Linear Classification Protocol) 是一种评估自监督学习算法性能的方法。其主要步骤如下:

1. **预训练**: 首先使用自监督学习算法 (如SimCLR) 在大规模无标注数据集上预训练一个神经网络模型, 获取图像的特征表示。
2. **特征提取**: 冻结预训练模型的权重, 在目标数据集 (如CIFAR-100) 上提取图像的特征表示。
3. **线性分类器训练**: 在提取的特征表示上训练一个简单的线性分类器 (如线性层), 并在验证集上调优超参数。
4. **性能评估**: 在测试集上评估线性分类器的性能, 通常使用分类准确率作为指标。

线性分类协议的优势在于, 它能够评估自监督学习算法生成的特征表示的质量, 而无需重新训练整个模型。

5 实验数据集简介

5.1 CIFAR-100数据集简介

CIFAR-100数据集是由加拿大高级研究院（Canadian Institute for Advanced Research）提供的一个用于机器学习和计算机视觉研究的图像数据集。它包含了100个类的60000张彩色图像，每个类包含500张训练图像和100张测试图像。每张图像的分辨率为32x32像素。

- **类别数量：**100个类，每个类包含600张图像（500张训练图像，100张测试图像）。
- **图像尺寸：**32x32像素，彩色图像。
- **数据分布：**每个类的图像数量均匀分布。
- **应用场景：**主要用于图像分类任务的研究，包括监督学习和自监督学习算法的评估。

CIFAR-100数据集是CIFAR-10数据集的扩展版本，CIFAR-10包含10个类的60000张图像，而CIFAR-100则包含更多的类别和相同数量的图像。

5.2 ImageNet数据集简介

ImageNet数据集是一个大规模的视觉数据库，由斯坦福大学视觉实验室创建，并用于图像分类和物体检测等计算机视觉任务。ImageNet包含超过1400万张带有标注的图像，涵盖了20000多个类别。

- **类别数量：**20000多个类，其中ILSVRC（ImageNet Large Scale Visual Recognition Challenge）使用的子集包含1000个类。
- **图像数量：**超过1400万张，其中ILSVRC子集包含120万张训练图像、5万张验证图像和10万张测试图像。
- **图像尺寸：**不固定，通常为较高分辨率的彩色图像。
- **数据分布：**数据分布较为均匀，每个类包含数百到数千张图像。

- **应用场景：**广泛用于图像分类、物体检测、图像分割等计算机视觉任务的研究，特别是在深度学习领域。

ImageNet数据集的引入极大地推动了深度学习在计算机视觉领域的发展。ILSVRC竞赛成为评估图像分类和物体检测算法性能的重要基准。

5.3 数据集比较

特性	CIFAR-100	ImageNet
类别数量	100	20000+（ILSVRC子集为1000）
图像数量	60000	1400万+（ILSVRC子集为120万训练图像）
图像尺寸	32x32像素	不固定（通常为高分辨率）
应用场景	图像分类	图像分类、物体检测、图像分割

表 1: CIFAR-100和ImageNet数据集比较

CIFAR-100数据集由于其较小的图像尺寸和较少的类别数量，适合用于快速原型设计和算法验证。而ImageNet数据集则因其大规模和高分辨率的图像，成为评估图像分类和物体检测算法性能的标准基准。

6 实验结果

6.1 数据集和实验设置

我们使用CIFAR-10数据集进行自监督预训练，并使用CIFAR-100数据集进行线性分类评测。对比实验中，使用ImageNet数据集进行监督学习训练。

6.2 训练过程中的Loss曲线

训练过程中的Loss曲线如图所示：

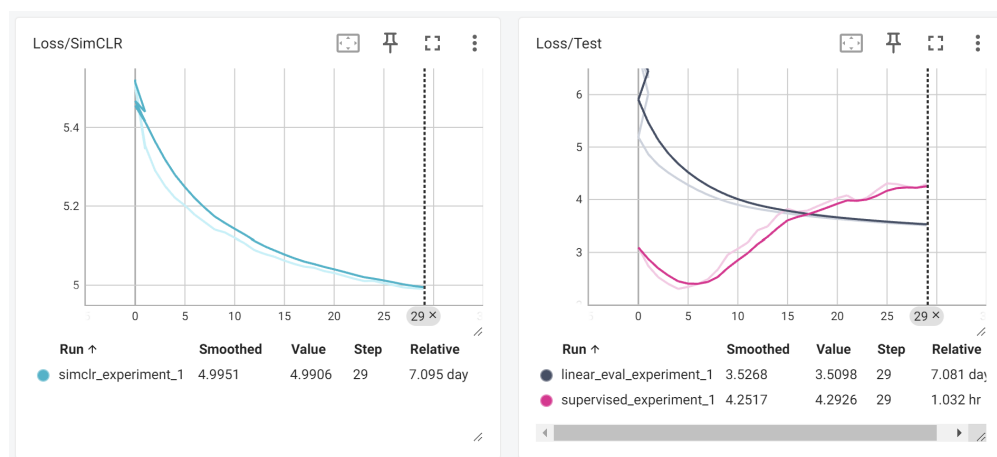


图 1: 训练过程中的Loss曲线

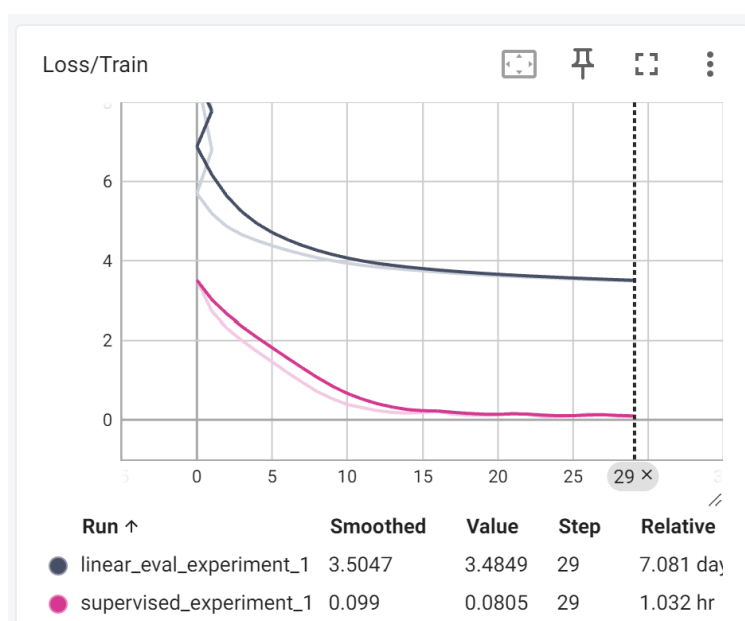


图 2: 训练过程中的Loss曲线

6.3 线性分类器的准确率变化

在训练和测试集上的准确率变化如图所示。可以看到，使用SimCLR自监督预训练的模型在CIFAR-100数据集上的分类准确率显著高于从零开始进行监督学习训练的模型。

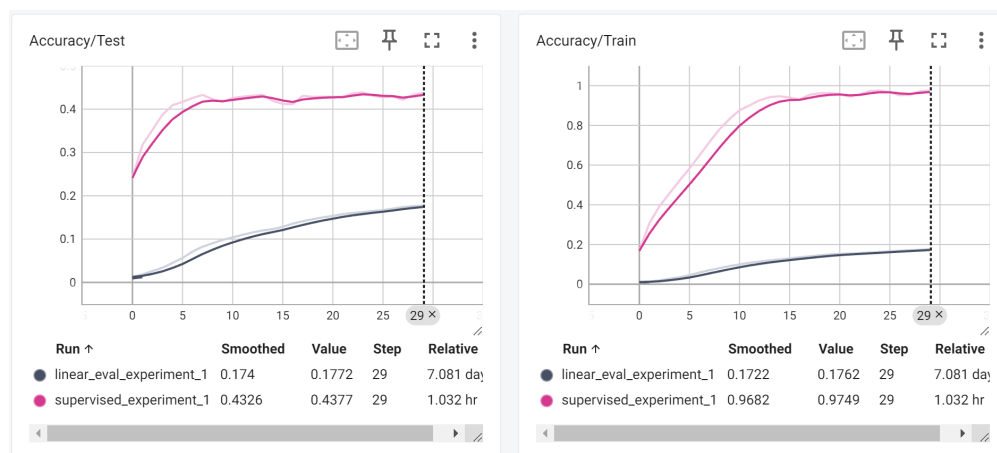


图 3: 线性分类器的准确率变化

7 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/11SLP_1lGeR88k8gcyEe6Zg?pwd=eqvb