

Image Pre-train Survey

▼ Image Pre-train Survey

▼ 一、Supervised Learning

▼ 预训练方法

- 参考文献
- 数据准备
- 超参设置
- 模型结构
- 训练目标函数
- 评估

▼ 二、Self-Supervised Learning

▼ 基于掩码的预训练方法

- 参考文献
- 数据准备
- 超参设置
- 模型结构
- 训练目标函数
- 评估

▼ 基于对比的预训练方法

- 参考文献
- 数据准备
- 1

一、Supervised Learning

使用有标签的图片来做分类任务以训练模型，强化模型的特征提取能力

预训练方法

参考文献

- ResNet : Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. "Deep Residual Learning for Image Recognition." arXiv:1512.03385 [[cs.CV](#)].

- ViT : Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." arXiv:2010.11929 [[cs.CV](#)].

数据准备

ImageNet等带有标签的数据集

超参设置

- ResNet:
 - 优化器：SGD（随机梯度下降）
 - 学习率：初始学习率为 0.1，当误差平稳时，学习率除以 10
 - 迭代次数：最多训练 60×10^4 次
 - 权重衰减：0.0001
 - 动量：0.9
 - 未使用 dropout
- ViT:
 - 优化器：Adam
 - β_1 : 0.9
 - β_2 : 0.999
 - 批量大小：4096
 - 权重衰减：0.1

模型结构

- 基于CNN的Res-Net50
- 基于transformer的ViT

训练目标函数

训练时做分类任务，损失函数为交叉熵
以降低损失函数为目标

评估

- 在训练结束后，与未训练过的模型同时去适应同一个下游任务，比较二者的训练速度与训练效果。
- 与未预训练而直接适应下游任务的模型进行效果对比

二、Self-Supervised Learning

基于掩码的预训练方法

类似BERT的Mask机制

参考文献

- MAE : Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, Ross Girshick.
"Masked Autoencoders Are Scalable Vision Learners." arXiv:2111.06377 [[cs.CV](#)], Submitted on 11 Nov 2021 (v1), last revised 19 Dec 2021 (this version, v3).

数据准备

准备无标签的图片数据 从ImageNet等较大数据集获取

超参设置

- Masked Patch Rate : 75% (较高的遮盖率能够促使模型去学习读取特征)
- 基于ViT的框架

模型结构

- encoder
 - 基于 Vit
 - Only Unmasked patch will be encoding
 - **masked patch is a shared, learned vector that indicates the presence of a missing patch to be predicted. We add positional embeddings to all tokens**
- decoder
 - 较小的架构 计算开销小
 - 最后一层是线性层 投影到patch像素数量的长度 reshape后得到结果

训练目标函数

decoder输出的内容与被遮蔽的内容进行比较

MSE作为损失函数

反传梯度 达到优化的效果

评估

- 微调适应下游任务, 例如在不同数据集上的分类任务

- 把结果与传统CNN和ViT在不同的数据集的分类任务的效果进行比较

基于对比的预训练方法

参考文献

MoCo系列：动量更新 队列存储

SimCLR系列：end to end 的学习方法，加入project head

SWaV：聚类学习

BYOL：只使用正样本，做一个相互预测的代理任务

SimSiam：预测代理任务 孪生网络

数据准备

任意数据集，对数据进行增强生成正样本与负样本（PS：BYOL Simsiam只使用正样本）

1

对比学习的框架、超参设置，各不相同，又互有借鉴，较难总结统一

评估方式大同小异