

image Pre-train

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

数据准备

ImageNet等带有标签的数据集

超参设置

- learning rate :
- batch-size:

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。
- 与未见过预训练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

参考文献：MAE-何恺明

类似BERT的Mask机制

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)
- Learning Rate:
- 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

基于对比的预训练训练方法

参考文献： MoCo系列

数据准备

超参设置

模型结构

- ViT
- ResNet-50

训练目标函数

评估