基于pytorch的mnist图像分类

李洁

1. 引言

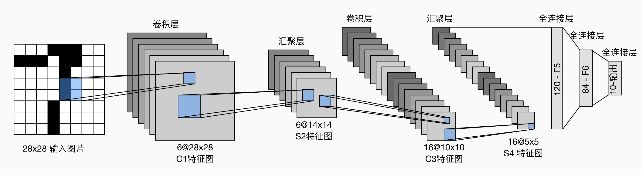
PyTorch框架，通常用于图像识别和语言处理等应用程序的机器学习，实现的模型具有简洁、易于调试等优点。vision transformer（vit）网络是一种基于Transformer架构的深度学习模型，其将图像视为一个序列化的输入，并利用自注意力机制来处理图像中的像素关系。这简化了网络结构，并允许模型捕捉长距离依赖关系。

本实践内容包括以下2个方面：

1、采用lenet、vgg、resnet-18进行mnist图像分类。

2、构建一个超级小的vision transformer网络，测试其在mnist上的准确度。

通过分析比较模型训练时间、测试准确度，深刻理解新添加层对提升网络性能的重要性。

1. 运行结果及分析
   1. lenet

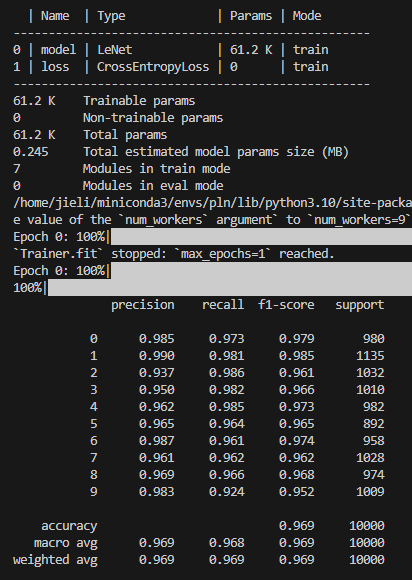
图2-1 LeNet中的数据流

图2-2 LeNet测试结果

训练一轮13s，因网络结构简单。

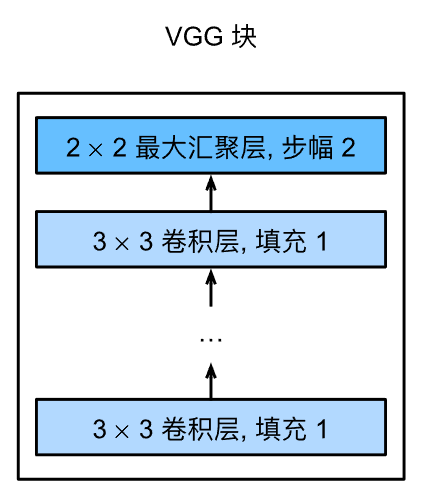
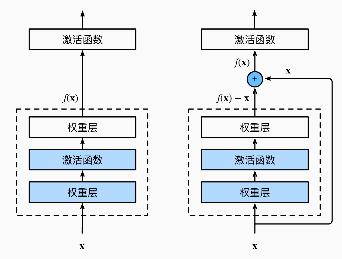
2.2 vgg

图2-3 VGG块模型

使用VGG16,训练一轮7min20s，因网络结构较深，需计算更多的前向传播和反向传播数据。测试准确度较高。

图2-4 VGG测试结果

* 1. resnet-18

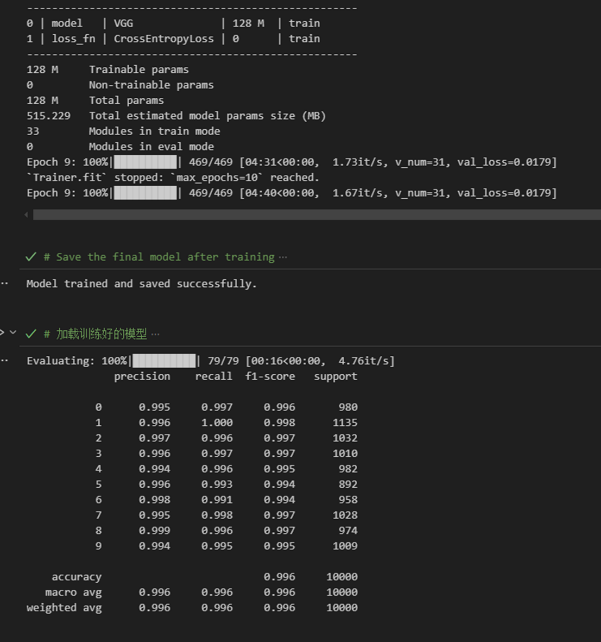
图2-5 左为正常块，右为残差块

图2-6 resnet-18测试结果

使用resnet18，训练一轮27s.在测试集上准确度最高。因使用残差块可直接将输入信息传递至网络深层，加快收敛。

2.4 vit

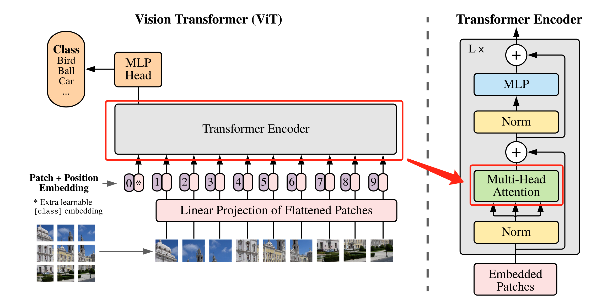
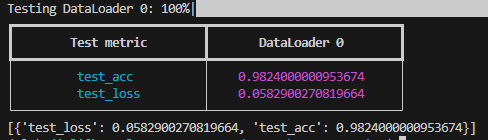
图2-7 多头注意力示意图

图2-8 vit 测试结果

训练一轮2min40s,测试准确度较高。

3.总结

使用框架可使程序易读易改，更清晰。

理论上，随着网络结构的加深，训练时间会增加。但过于复杂会导致过拟合，反而测试精确度很低。应根据任务选模型。