《人工智能导论》大作业

任务名称：不良内容图像检测

完成组号：13

小组人员：刘李宇轩，王骋昊，马天翊

完成时间：2024.6.19

1. 任务目标

基于暴力图像检测数据集，构建一个检测模型。该模型可以对数据集的图像进行不良内容检测与识别。
数据集由以下数据组成：

- 自然图像；
- 部分 AIGC 生成图像；
- 部分对抗样本图像。

要求：

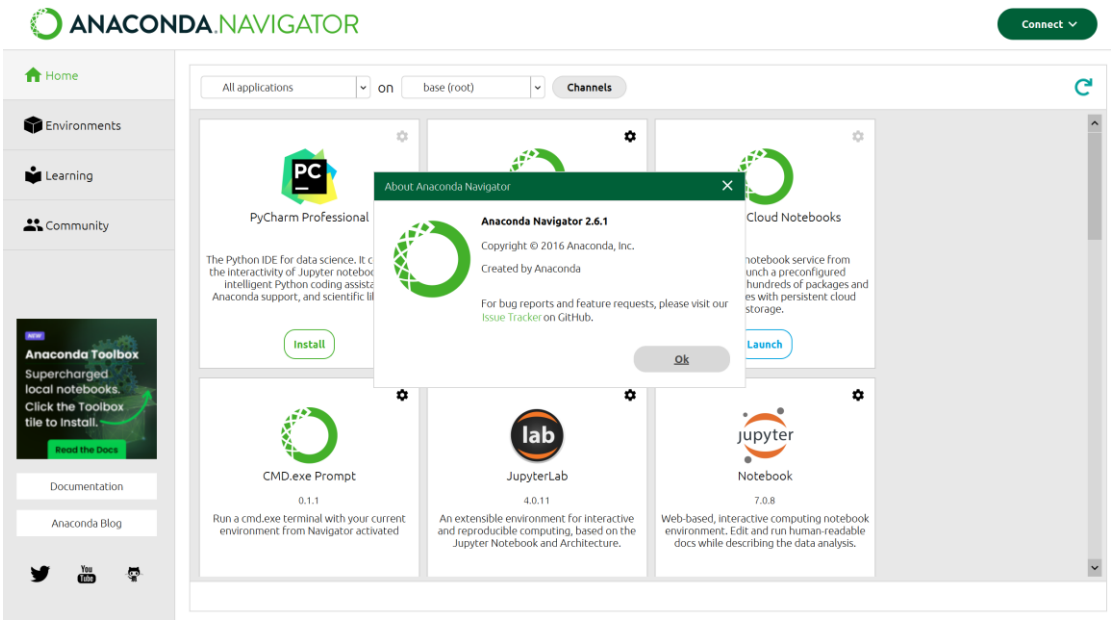
- 模型是 2 分类（0 代表正常图像、1 代表不良图像），分类准确率越高越好；
- 模型具有一定的泛化能力：不仅能够识别与训练集分布类似的图像，对于 AIGC 风格变
- 化、图像噪声、对抗样本等具有一定的鲁棒性；
- 有合理的运行时间。

2. 具体内容

（1）实施方案

软件：

- 环境管理：anaconda/conda 2.6.1
- 语言：python 3.10.14
- 神经框架： pytorch 2.3.1
- 处理图像库：torchvision 0.18.1



在官网安装 anaconda,版本为 2.6.1

安装 python，版本 3.10.14

创建 conda 虚拟环境，安装 pytorch，版本 2.3.1；安装 torchvision，版本 0.18.1

文件结构：

root_dir: ./

sub_dirs: ['_13_other_files_directory']

files: ['classify.py']

root_dir: ./ _13_other_files_directory

sub_dirs: ['lightning_logs', 'violence_224']

files: ['model.py', 'test.py', 'violence_model.ckpt']

设计步骤：

model.py 文件中使用 PIL 库进行图像 io，torchvision 库加载训练模型，torch 库设置数据集，加载数据；pytorch_lightning 库提供数据框架与训练模型框架，加载数据集进行训练；

classify.py 文件中 ViolenceClass 接口类提供接口函数 classify，输入图像转化而来的 n*3*224*224 的 pytorch tensor，输出长度为 n 的 python 列表（每个值为对应的预测类别，即整数 0 或 1），进行图像分类。

test.py 文件加载 ViolenceClass 接口类，加载图像，进行测试。

（2）核心代码分析

model.py：

def load_images_from_folder：

读取指定目录下图片，使用 torchvision 库 transforms 函数调整图像大小为 224x224，转换为 Tensor，并归一化至[0, 1]。

class CustomDataset(Dataset):

以 torch.utils.data 库 Dataset 为父类，__init__函数区分训练集图像与验证集图像，将图像转换为 Tensor，__len__返回各个图像集的图片数量，__getitem__返回图片及其对应的标签（0 代表非暴力，1 代表暴力）。

class TensorDataset(Dataset):

以 torch.utils.data 库 Dataset 为父类，__init__函数初始化

tensor_list，__len__返回 tensor_list 长度，__getitem__返回 tensor 及其对应的标签（测试集图像标签未知，设置默认值 0）。

class CustomDataModule(LightningDataModule):

以 pytorch_lightning 库 LightningDataModule 为父类，__init__ 函数初始化 tensor_list，batch_size，num_workers，setup 函数设置训练集，验证集的 CustomDataset 类实例与测试集的 TensorDataset 类实例，train_dataloader，val_dataloader，test_dataloader 函数分别通过 torch.utils.data 库 DataLoader 函数加载图像，其中 train_dataloader 函数 shuffle 置真，其余置假。

class ViolenceClassifier(LightningModule):

以 pytorch_lightning 库 LightningModule 为父类，__init__函数加载预训练 resnet18 模型，设置二分类，学习率 1e-3， fc 层，交叉熵，准确度，forward 函数进行前向传播，configure_optimizers 函数定义优化器，training_step，validation_step，test_step 函数分别接收 batch，计算 loss 与 accuarcy。

main：

测试模型，进行训练。

classify.py：

class ViolenceClass:

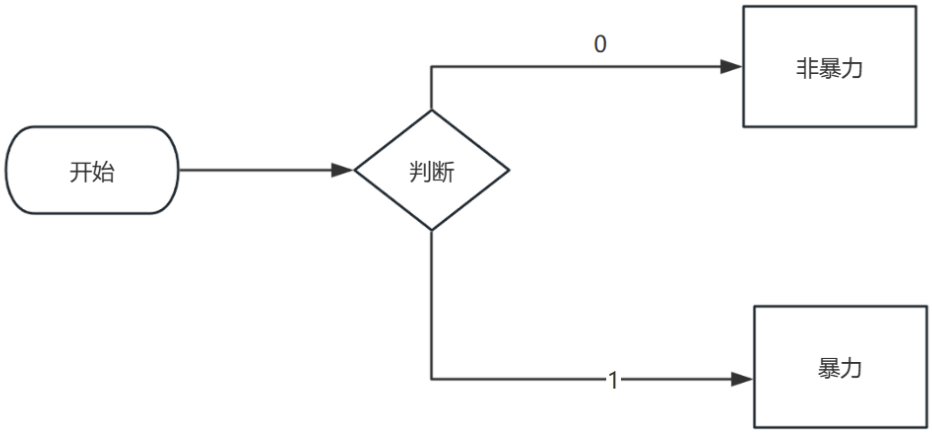
__init__函数加载模型、初始化 trainer；classify 函数实例化 CustomDataModule，给出预测值。

test.py：

读取图片路径，实例化 ViolenceClass，输出预测值。

输出结果：

```
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1,
```



0 即非暴力，1 即暴力。

3．工作总结

（1）收获、心得

王骋昊：本次大作业完成过程中，我们在理解作业要求方面花了比较长的时间，通过网上查找相关知识初步了解模型的构建及训练过程。我负责训练并测试模型，选取 resnet18 模型并使用 pytorch_lightning 库设计数据框架与模型框架，经过 10 个 epoch 完成训练，学到了如何应用机器学习技术来解决图像处理问题。

刘李宇轩：在本次大作业的完成过程中，我与其他组员一同努力，收获颇多。我负责训练并测试模型，在开始阶段，主要困扰我的是对于作业要求，已有代码的解释。我通过详细分析，明白了我们的目的

主要是实现接口 classify.py 中加载数据集，加载模型并调整参数，传入 tensor 并输出预测标签的功能，并因此需要额外建立模型文件，测试文件。我通过与两位同学合作，提高了团队协作能力，知道了如何有效沟通，这在我未来的职业生涯中非常重要。

马天翊：在这一次的人工智能大作业中，我负责 github 代码上传与报告编写。在这一次大作业的完成中，最令我印象深刻的就是我们对于实现这个模型的方法探索。我们试过很多方法，但是最后都因为方法不太合适而舍弃。最终我们经过讨论得出了已知的最好的方法，就是我们展示在大作业中的办法。我觉得这一次大作业不但让我们学到了知识，更让我们有了一次探索的机会，一次通过自己的研究解决难题的机会，这对于我们以后的学习，乃至科研工作，都是有很多锻炼的。

（2）遇到问题及解决思路

我们遇到的主要问题是应该选择哪种类型的模型来进行不良内容图像检测，如何设计合适的模型架构。经过讨论，我们选择了基于深度学习的卷积神经网络（CNN）模型 resnet18 预训练模型进行学习。

4．课程建议

课程可以再安排一些有关神经网络方面的内容理解，大作业布置时间也可以再提前一点，从零理解神经网络仍然有点吃力。