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：第3组

小组人员：杜禹欧、张睿诚、曹一凡

完成时间：2024年06月03日

1. 任务目标

实现一个暴力图像的二分类模型，得到一个ViolenceClass，该类含有接口函数classify，输入复数个RGB图片的tensor（是n * 3 * 224 * 224大小），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），预测图片是否包含暴力行为内容。

2. 具体内容

1. 文件结构

- 3-report.pdf 大作业实验报告
- 3-readme.md 接口类说明与调用实例
- 3-classify.py 实现的接口类文件
- 3-SupportingnFiles 其它支持文件和函数
    - train_logs/ 存放训练结果
    - violence_224/ 下载的数据集
    - testjpg/ 存放加入噪声的图片（用于测试集和训练集）
    - lightning_logs 存放Tensorboard可视化文件
    - image/ 存放readme.md的图片
    - noise.py 用于给图片加入噪声
    - dataset.py 数据集加载
    - model.py 神经网络构建
    - train.py 训练模型
    - test.py 用于测试训练结果

2. 实施方案

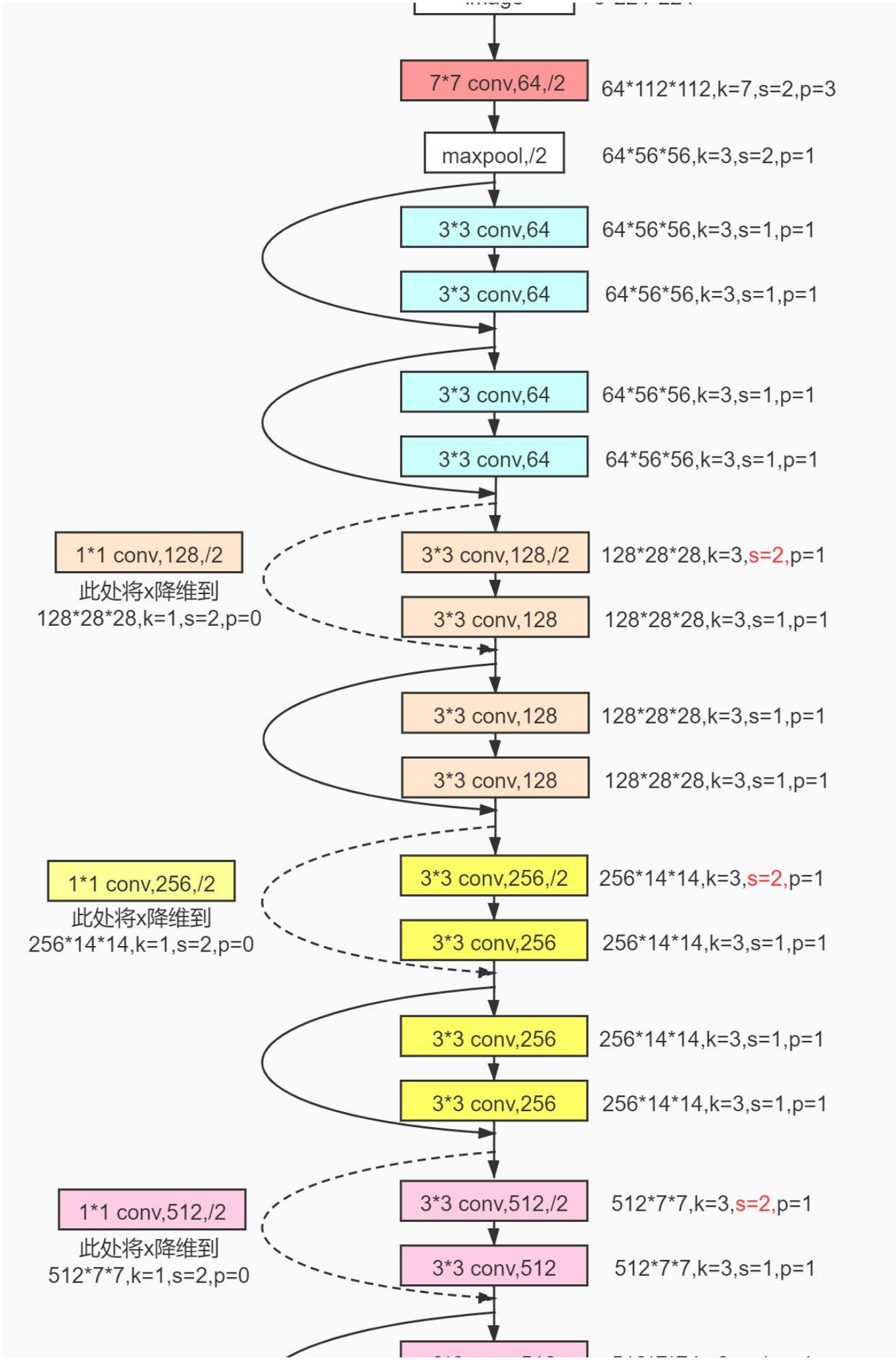
使用数据

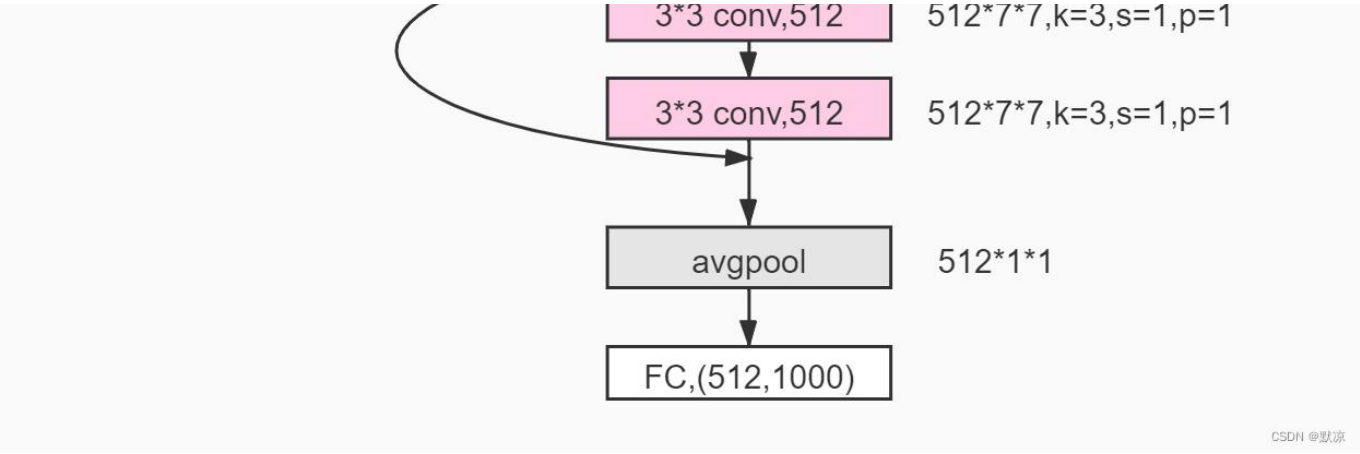
本实验主要使用课程提供的数据集，但对数据集进行了一定的修改添加。课程提供数据集分为train和val两部分，分别为训练数据集和验证数据集，测试数据集需要自行准备。数据集中包含正常行为和暴力行为图像。图像的命名格式皆为：“标签位_四位数字编号”，“_”前的为标签位，正常行为图像以0起始，暴力行为图像以1起始；train部分正常行为图像3660张，暴力行为图像4088张；val部分正常行为图像522张，暴力行为图像583张。为了使得训练出的模型对于噪声干扰图像也具有良好分辨能力，我们在原有数据集的基础上，添加约为总数20%的不同梯度的高斯噪音与椒盐噪音的图像。 同时，由于课程提供的数据集仅包含训练集和验证集，测试集数据需要另外收集。我们收集图片并生成了三个测试集：自然测试集、aigc测试集、噪声测试集。 总结：

- 扩充了课程组提供的训练集，在其中加入了噪声图像；
- 收集生成图片并建立了自然图像测试集、噪声图像测试集、aigc图像测试集。

使用模型

本实验使用模型基于ResNet18，即为一个深度为18的残差神经网络。ResNet18网络结构流程图如下：


7*7 conv,64,/2
64*112*112,k=7,s=2,p=3
maxpool,/2
64*56*56,k=3,s=2,p=1
3*3 conv,64
64*56*56,k=3,s=1,p=1
3*3 conv,64
64*56*56,k=3,s=1,p=1
3*3 conv,64
64*56*56,k=3,s=1,p=1
3*3 conv,64
64*56*56,k=3,s=1,p=1
1*1 conv,128,/2
此处将x降维到
128*28*28,k=1,s=2,p=0
3*3 conv,128,/2
128*28*28,k=3,s=2,p=1
3*3 conv,128
128*28*28,k=3,s=1,p=1
3*3 conv,128
128*28*28,k=3,s=1,p=1
3*3 conv,128
128*28*28,k=3,s=1,p=1
1*1 conv,256,/2
此处将x降维到
256*14*14,k=1,s=2,p=0
3*3 conv,256,/2
256*14*14,k=3,s=2,p=1
3*3 conv,256
256*14*14,k=3,s=1,p=1
3*3 conv,256
256*14*14,k=3,s=1,p=1
3*3 conv,256
256*14*14,k=3,s=1,p=1
1*1 conv,512,/2
此处将x降维到
512*7*7,k=1,s=2,p=0
3*3 conv,512,/2
512*7*7,k=3,s=2,p=1
3*3 conv,512
512*7*7,k=3,s=1,p=1



该网络通过设置残差结构，能够以较小的计算量实现比较好的图像分类效果。

3. 具体步骤

数据收集与生成：

自然图像：在网络上获取包含人物正常动作的图像与包含暴力行为的图像（包括打架斗殴等），并为图像打上标签，即对图像进行重命名。 aigc图像：由于aigc图像生成比较困难，我们没有在训练集中添加aigc图像，只在测试过程中对aigc图像的分辨能力进行了测试。我们使用了矩阵云平台，在云服务器上部署了stable diffusion。我们使用该模型生成了30张图像，其中有15张为暴力图像，15张为非暴力的任务行为图像。 噪声图像：在原数据集中，随机抽取500个样本，生成第一梯度的高斯噪声（sigma=1）数据集和椒盐噪声数据集(salt_prob=0.02, pepper_prob=0.02)，再随机抽取500个样本，生成第一梯度的高斯噪声（sigma=5）数据集和椒盐噪声数据集(salt_prob=0.1, pepper_prob=0.1)；对自然图像的测试集加上上述两种噪声，生成两种噪声测试集。

建立DataLoader：

在进行训练前，需要将图像数据集转换为DataLoader类，便于加载和管理数据集。因此实现CustomDataset类与CustomDataModule类，完成以下任务：1. 由于模型输入为tensor格式的数据，不能直接使用JEPG格式的图像进行训练，需要对于图像进行格式转换；2. 提取图像标签；3. 设置参数；4. 转换为DataLoader类。此部分实现在dataset.py文件中，具体分析如下： `class CustomDataset(Dataset):` `__init__` ：根据提供的路径加载图像数据，将图像转换为tensor，并且如果如果图像为"train"类图像，则对图像进行随机翻转； `__len__` ：返回数据集的大小; `__getitem__` ：提取图像标签。 `class CustomDataModule(LightningDataModule):` `__init__` ：设置DataLoader参数； `setup` ：分割数据集，生成训练、验证和测试数据集； `train_dataloader`,`val_dataloader`,`test_dataloader`：分别返回训练、验证和测试数据对应的DataLoader。

创建模型：

在开始训练前，需要创建模型类。我们使用PyTorch Lightning框架定义了一个基于ResNet18的暴力分类器。此部分实现在model.py文件中，具体分析如下： `class ViolenceClassifier(LightningModule):` `__init__`: 加载ResNet18模型（预训练）；替换ResNet最后的全连接层为线性层，使得最后输出为0-1的二分类；设置学习率、损失函数与准确率计算方式； `forward`: 定义前向传播过程; `configure_optimizers`: 配置优化器为Adam优化器； `training_step`,`validation_step`,`test_step`: 定义每个训练、验证和测试的步骤中，需要进行的操作，其中的差别主要在是否需要计算损失与准确率，以及是否需要记录这些数据。

模型训练：

需要实现一个训练脚本，调用以上已实现的类函数进行实例化，完成训练。此部分实现在trian.py文件中，具体实现如下：

1. 配置gpu编号、学习率、batch_size、日志名称等参数；
2. 实例化DataLoader;
3. 设置checkpoint检查点，用于保存验证loss值最小的模型参数；
4. 设置日志记录器，保存训练日志；
5. 实例化训练器；
6. 实例化模型；
7. 进行训练。

模型测试：

实现模型测试脚本，加载训练时checkpoint保存的最佳模型参数，进行测试。此部分实现在test.py文件中，具体实现如下：

1. 配置batch_size等参数；
2. 实例化DataLoader;
3. 从checkpoint加载模型；
4. 实例化训练器；
5. 进行测试

2. 测试结果分析

训练日志：

训练“epouch数-步数”关系图如图1：

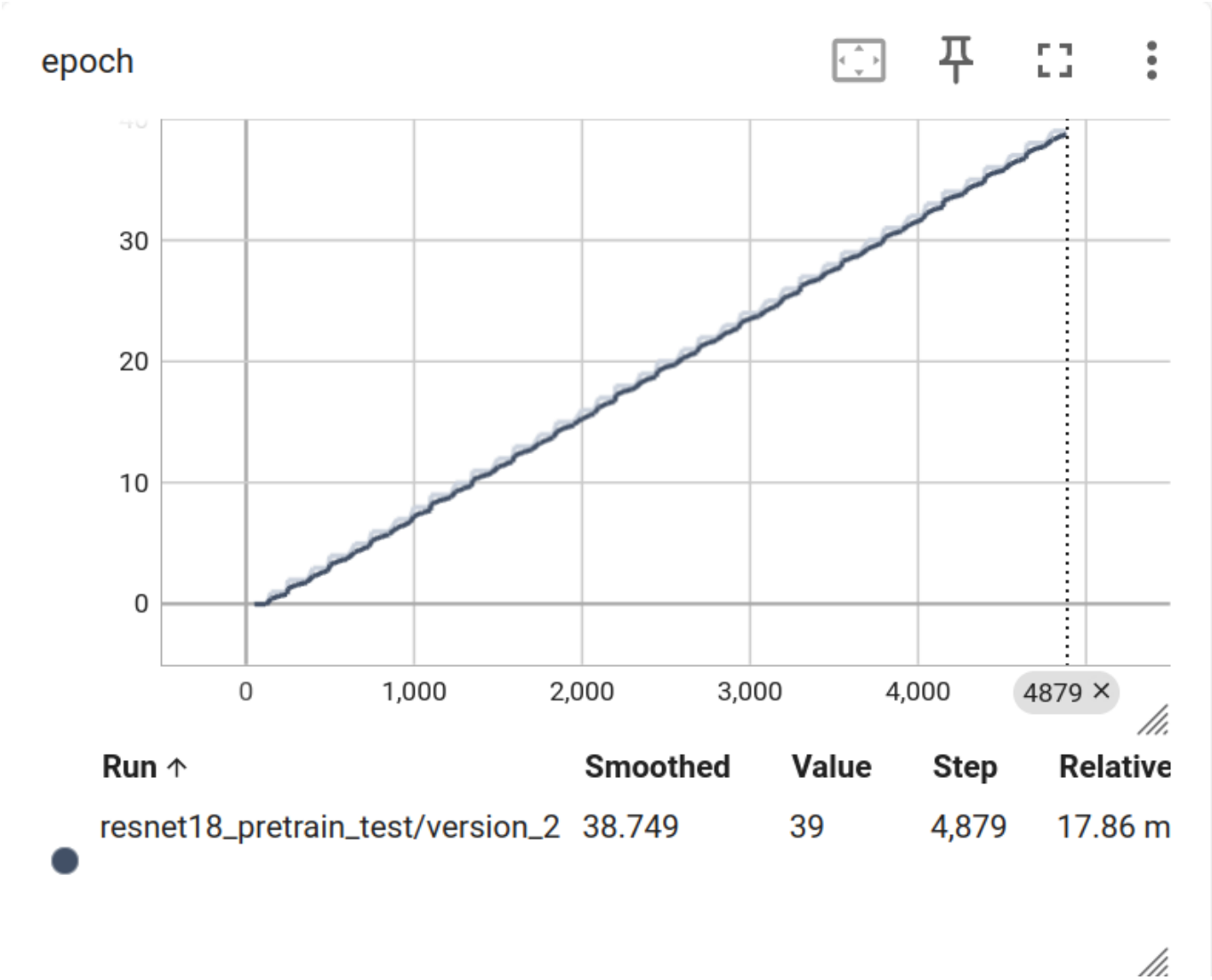


图1

训练过程中训练的“loss值-训练步数”关系图如图2：

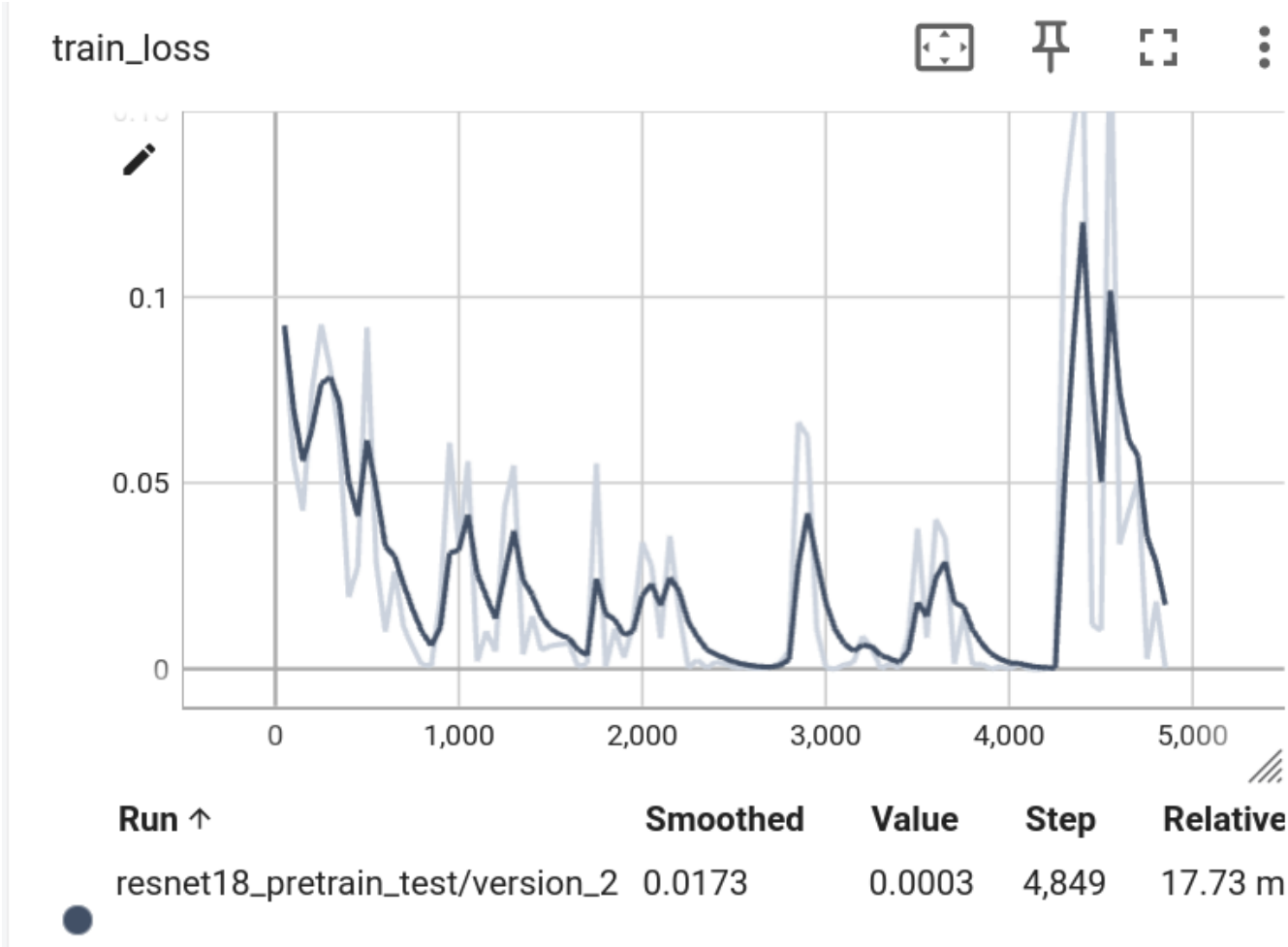


图2

训练过程中测试的“loss值-训练步数”关系图如图3：

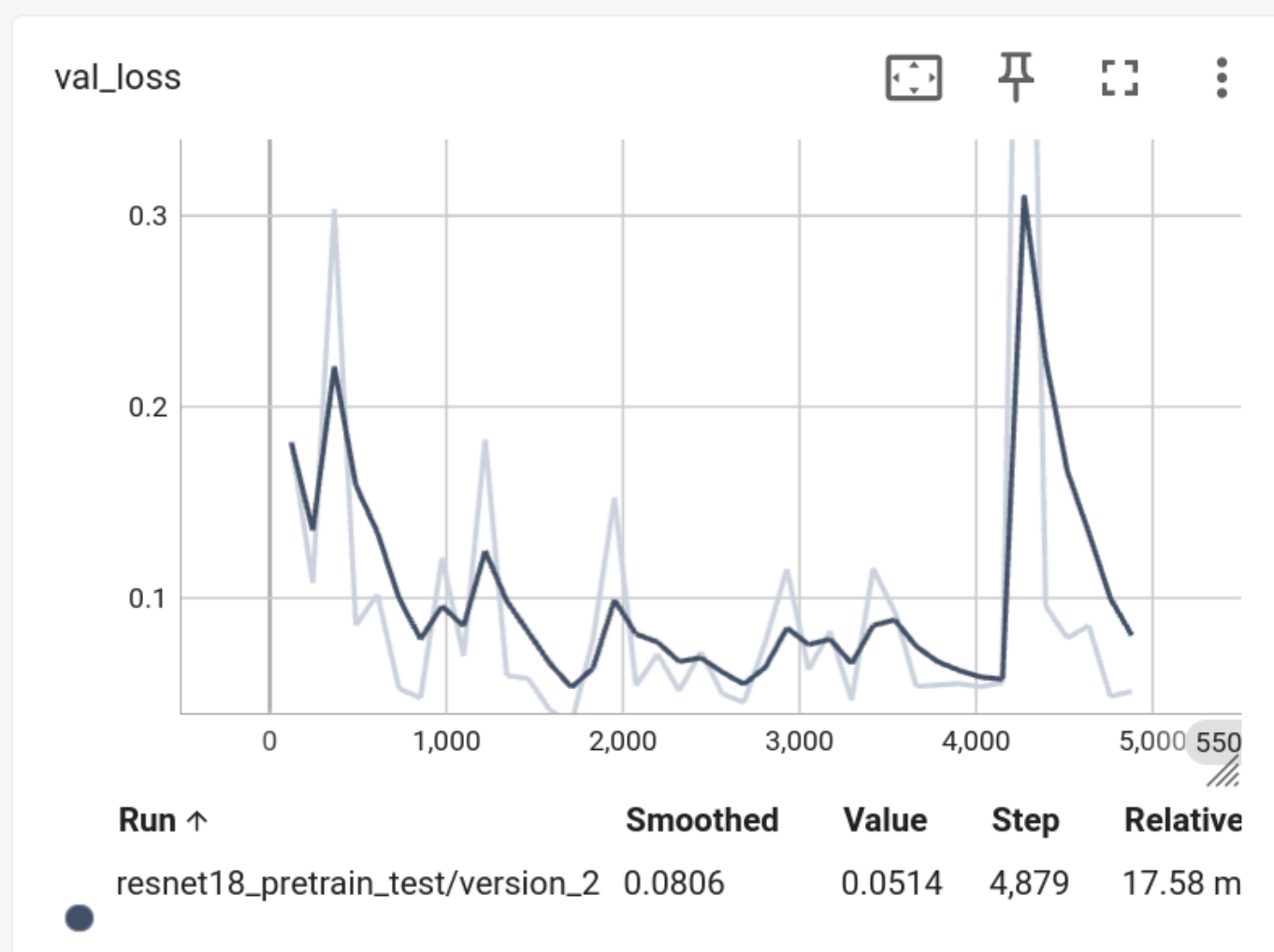


图3

训练40个epouch后停止，其中checkpoint保存在第13个epouch，对比图1与图3，发现val_loss确实在step约为1800时达到最小值（约为第13个epouch的位置）。

在不同测试集上的测试结果：

我们在不同数据集上训练了两个模型，模型0在课程组提供的数据集上训练，模型1在我们扩充后的数据集上训练（加入了噪声图像）。我们分别在自然图像测试集、高斯噪声测试集、椒盐噪声测试集、aigc测试集上对两个模型进行了测试，准确率（accuracy）结果如下：

使用模型	自然图像	高斯噪声	椒盐噪声	aigc图像
模型0	1.00	0.42	0.88	\
模型1	1.00	0.93	0.92	0.7

可以发现：

- 模型对于自然图像的分别能力极强，能很好地分辨生活中实拍的暴力与非暴力图像；
- 我们对于数据集的扩充，使模型在抗噪声方面有了很大进步，模型能够较好地分辨添加了噪声的暴力图像；
- 模型对于aigc图像有一定分辨能力。

3. 工作总结

本实验的主要内容为，基于ResNet18，训练一个对于暴力图像的二分类模型。其中，我们完成的主要工作为：

1. 扩充数据集：我们对于课程提供的数据集进行了扩充，在里面添加了aigc图像和人工添加的图像，以增强模型对于aigc图像的分别能力和抗噪声能力；
2. 训练模型：我们使用本地计算资源，对模型进行了训练，获得了对暴力图像的二分类模型及最佳参数；
3. 测试模型：我们分别使用自制的自然数据集、aigc数据集、噪声图像数据集分别对模型进行了测试；
4. 结果分析：我们对模型训练过程中的日志进行了分析，并对于模型在三种不同测试集下的测试结果进行了分析；
5. 模型部署：我们实现了接口函数classify，输入图像的tensor，输出分类结果。

4. 课程建议

对于本课程，结合实际的课程学习情况，我们有些许微小的建议，如下：

1. 适当增添实践作业；
2. 使用更新的模型库；
3. 希望能够提供更详细的大作业指导文档。