智能计算系统分组实验

1. 实验目的

掌握深度学习部署方法，使用开源工具链进行模型转换，考虑算子实现、转换工具完善等，完成模型部署全流程的体验与改进。

2. 实验环境

硬件环境：CPU

软件环境：Python3.7+、Pytorch 1.13.1、ONNXRuntime 1.14.1、Caffe 1.0.0、Darknet https://github.com/AlexeyAB/darknet

数据集：ImageNet 2012

3. 实验内容和步骤

任务一：利用工具链进行模型的部署流程学习体验（60%）

目前主流工具链均包含模型转换函数，例如 pytorch2onnx，tensorflow2tensorflow lite 等。此外，还有许多基于预训练模型转换的工具可供选择，例如 MMDNN，MMDeploy 等。任选一项或多项主流图像分类网络（MobileNet 系列、ResNet 系列、MobileOne 等），通过工具或内置函数进行模型转换，转换后的目标格式可以为 ONNXRuntime、Caffe、NCNN、Tensorflow lite 等的任意一种。

任务二：使用模型转换工具将预训练模型转换至 Darknet（需改进转换工具），考虑实现新算子（40%）

Darknet 是一个用 C 编写的开源神经网络框架。它速度快，易于安装，并支持 CPU 计算。目前已有的工具链支持 Caffe2darknet 的模型转换，模型转换的过程中会遇到转换工具不能正常转换模型、框架间算子不兼容、不同框架下算子实现的功能一致但名称不一致等问题。建议使用的工具链有：

Caffe2Darknet： https://github.com/KerwinKai/Caffe2Darknet.git

Caffe_model_zoo： https://github.com/KerwinKai/Caffe_model_zoo

建议可选择进行转换的模型有（**仅选一个即可**）：MobileNet-v2、ResNet18、ShufflenetV2、SqueezeNet。

Task

使用Caffe2DarkNet，将已有的Caffe框架下的预训练模型转换至Darknet框架。提供四种轻量级网络以供选择，有以下挑战需要解决克服。

MobileNet_v2

转换工具提示 `assert (i + 1 < layer_num and layers[i + 1]['type'] == 'ReLU')`，转换后推理未测试

ResNet18

转换成功，推理时需要解决 `Flatten` 在DarkNet上实现的问题

shufflenet_v2

转换工具提示 `unknown type Concat`，可调研DarkNet上该层的实现方法（能否实现、有无类似功能的层），更新转换工具，转换后推理未测试

squeezenet

转换工具提示 `unknown type Concat`，可调研DarkNet上该层的实现方法（能否实现、有无类似功能的层），更新转换工具，转换后推理未测试

4. 评分标准

60 分标准：完成任务一，利用工具链进行模型的部署流程学习体验

60-100 分标准：完成任务二，包括但不限制于，提交改进模型转换工具的 pr、实现模型转换后 DarkNet 上缺失的算子、选择一个预训练模型转换至 DarkNet 框架并实现对单张图片推理的精度相差不大。

备注：以上为基本的评分准则，最终评分会根据代码实现和相应的结果有一定的附加分数。

5. 文件提交格式

任务一：提交 ipynb 文件，文件中需记载模型转换的代码及过程，保留运行结果。并测试对于同一张图片，转换前与转换后的推理结果置信度需相差不大。命名为：任务一.ipynb

任务二：提交 ipynb 文件，加入新算子后编译好的 DarkNet 文件夹、及修改后的 caffe2darknet.py 文件，压缩为任务二.zip。Ipynb 文件中需记录实现的算子，并证明在某一轻量网络上的推理能力，可参考 **Caffe_model_zoo** 中的 how2get.ipynb 文件。caffe2darknet.py 文件需加入对应新算子的映射转换。

6.【课程大作业格式内容要求（注明小组成员姓名，学号，专业等信息）】

一、**背景调研**（主要调研模型转换部署相关的论文，如 MobiSys 等会议内的科研论文等，插入引用的参考文献）

二、**模型介绍**(主要介绍本组使用的模型网络，模型详细的分块说明，如数据集读取、激活函数、损失函数等，可以截图贴出来核心代码片段）

三、**实验分析**（分析一下模型转换的实验结果，准确率、吞吐量等评测指标，还可调研图像预处理对模型推理准备度的影响）

四、**组内分工**（详细介绍小组内分工情况，每位成员分工内容，贡献程度）