Assignment 2, CNN Implementation from Scratch

Requirements

Please provide a implementation of api for CNN model training and inference. Implementation needs to consider the following requirements:

1, Don't use ai programming frameworks such as PyTorch, Tensorflow and etc.

2, Flexible definition of CNN architecture.

3, Multiple activation options, e.c, Relu, LeakyRelu

4, Classification and regression

5, Weight initialization options

6, SGD optimizers: Momentum, RmsProp, Adam

7,SGD stop criteira

8,Regularization l1, l2, elastic

9,Optimized Convolution layer implementation such as im2col/col2im, FFT(fast fourier transform)

10, Layers need to implement: Conv2d, Pooling(Maxpooling, AvgPooling), Drop-out, Batch Norm, Flatten, FC.

11, Architecture blocks, e.g, Inception Module, Residual block, Depthwise conv/Bottleneck。

12, Implementation of CNN architecture based on your own implementation, recommended such as FaceNet, MobileFaceNet, YOLO V4/5.

13, Bonus implementation: CNN + Transformer

It is highly recommended using different programming languages other than python, and design your own api specification that will satisfy complete deep-learning programming, and also try parallel of training using multi-threads or gpus if available.

Notice

1, Gen code is acceptable but requires to satisfy above.

2, You need to demonstrate you understand and capable of this implementation.

3, Therefore, must run and provide detailed results, such as the model (save in file), confusion matrix and other evaluations.

4, Reports on key points of your design and implementation.

5, Recommended datasets include:

FaceNet/MobileFaceNet:

- **Labeled Faces in the Wild (LFW)**: This dataset is used to evaluate the performance of MobileFaceNet in unconstrained face recognition.

- **MS-Celeb-1M**: This dataset is used to train MobileFaceNet to learn discriminative facial features.
- **CASIA-WebFace**: This is also one of the commonly used training datasets for MobileFaceNet.

Yolo:

- **COCO (Common Objects in Context)**: This is a large - scale object detection dataset with 80 categories and over 330,000 images, commonly used for training and evaluating object detection models like YOLO v5.
- **Pascal VOC**: This dataset contains 20 object categories and over 11,000 images. It's a classic object detection dataset and is also suitable for training and evaluating YOLO v5.
- **OpenImages**: This is a large dataset released by Google, with over 1.7 million training images and 42,000 validation images, covering thousands of categories, providing rich training data for YOLO v5.

6, URLs：

mobileFaceNet: https://arxiv.org/abs/1804.07573

https://github.com/AnyLifeZLB/FaceVerificationSDK

FaceNet: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/http://llcao.net/cu-deeplearning17/pp/class10_FaceNet.pdf,

https://github.com/tbmoon/facenet