Nirbhay Sharma (B19CSE114)

Dlops -assignment - 5

Code Files

the code contains the following files

- **cnn_model.py** - to return the models i.e. (vgg, resnet)
- **config.py** - contains all the configurations like batch_size, lr, epochs etc
- **custom_dataset.py** - contains the data loader for the dataset and return the split of dataset into train, test and validation
- **tb_gpu.py** - contains the training code for normal gpu without optimizations
- **tb_gpu_optimized.py** - contains the training code for optimized version (includes optimizations including AMP, pin_memory, n_workers, cudnn.benchmark etc)
- **main.py** - contains the final training code for optimize and unoptimized, depending on the command line argument passed
- **pytorch_inference.py** - contains the inferencing code for .pt model i.e. normal pytorch inference
- **onnx_convert.py** - contains the script for converting the .pt model to .onnx format
- **onnx_inference.py** - contains the script for inferencing the .onnx model

How to run

- just change the config files for epochs, batch_size etc
- run main.py file as

  to run the model in normal mode change optimized = False in config.py file and run the following command

  ```
  python3 main.py --mode normal
  ```

  to run the model in optimized form change optimized = True in config.py file and run the following command

  ```
  python3 main.py --mode optimize
  ```

Results

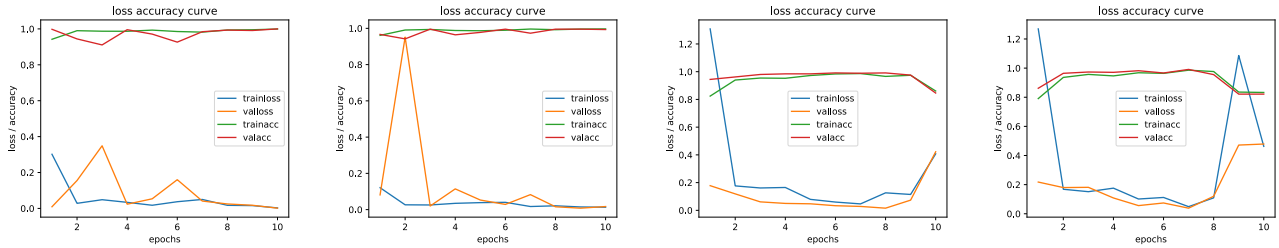
hyperparameters

- batch_size=32
- lr= 0.001
- epochs=10

- Image_size = (224,224,3)

Model	normal train time	optimize train time	accuracy	optimized accuracy
vgg16	49 Minutes	8 Minutes	0.85	0.83
resnet50	51 Minutes	4 Minutes	0.99	0.99

loss accuracy curves for resnet, resnet_optimized, vgg, vgg_optimized repectively



the model is converted to onnx model and the model can be found here

the models are converted to onnx using onnx_convert.py script and tested using onnx_inference.py

---