b19cse114-lab2.md 2/3/2022

Nirbhay Sharma (B19CSE114)

DLops lab-assignment - 2

| config | Model Name, Batchsize | Pytorch Throughput | ONNX Throughput | TensorRT FP32 Throughput | TensorRT FP16 Throughput | TensorRT INT8 Throughput |
|----------------------|-----------------------------|-----------------------|--------------------|--------------------------------|--------------------------------|--------------------------------|
| (TF32 = True) | ResNet50, 32 | 1933 | 1196 | 3168 | 4780 | 5881 |
| (TF32 = True) | ResNet18, 32 | 5848 | 3006 | 5171 | 6653 | 7430 |
| (TF32 = False) | ResNet18,32 | 4842 | 3158 | 5282 | 6387 | 8203 |

[commands]

- 1. mode -> train
- 2. mode -> convert
- 3. mode -> pytorch
- 4. mode -> onnx
- 5. mode -> tensorrt32
- 6. mode -> tensorrt16
- 7. mode -> tensorrt8