Static Demo

Run DeepLabV3 server

Semantic image segmentation model

```
./run.sh ./models/new_deeplabv3.pt

python3 tests/static.py images/biker.png
python3 tests/static.py images/seg1.png
python3 tests/static.py images/cat.png
```

- Show making an inference req
- Show status features

Benchmark demo

Run ResNet18 server

```
./bench_run.sh ./models/resnet18.pt
```

Run uniform tests

40 requests per second across 12 threads for 5 seconds

```
./tests/bench.sh ./tests/classification_req.json
```

Dynamic demo

Run ResNet18 server

```
./run.sh ./models/resnet18.pt
```

Open status in another window

This constantly queries the server to ask for how many workers are active

```
python3 tests/status.py
```

Run uniform tests

40 requests per second across 12 threads for 5 seconds

```
./tests/bench.sh ./tests/classification_req.json
```

Then run sine wave tests

Maximum of 8 threads at the peak, varying in a sine wave pattern for 10s Simulates unpredictable waves

```
python3 tests/dynamic.py sine
```

Then run Gaussian tests

Gaussian. Simulates a peak or spike

```
python3 tests/dynamic.py gaussian
```