

Benchmark

Berechnung der Inferenz-Fps für unterschiedliche Anzahl an Inferenz-Requests

Verwendung

- In *models_dir* pfad zu konvertierten OpenVino Modellen angeben. Die Modellordner haben folgende struktur:

```
└─ model_folder
   └─ frozen_inference_graph.bin
   └─ frozen_inference_graph.xml
   └─ exported_model (optional)
```

Falls vorhanden, wird *exported_model* verwendet, ansonsten wird *exec_model* aus den *.xml* und *.bin* files erzeugt und exportiert.

- In *test_image* Pfad zu inferierender Bilddatei angeben
- *iterationen* gibt an wie oft das Bild inferiert werden soll

benchmark.py

starten über

```
python3 benchmark.py
```

- Modell über Nummer auswählen
- Anzahl der Inferenz-Requests eingeben

benchmark_auto.py

In *models*-Liste mit zu inferierenden Modellen (in *models_dir* enthalten sind) festlegen.

zB:

```
models = ['ssd_mobilenet_v2',
          'ssd_inception_v2',
          'faster_rcnn_inception_v2']
```

starten über

```
python3 benchmark_auto.py
```

werden dann jeweils für 1 bis 4 inferenz-requests automatisch inferiert.

`infer_async.py`

wird von *benchmark.py* und *benchmark_auto.py* für die asynchrone inferenz verwendet.