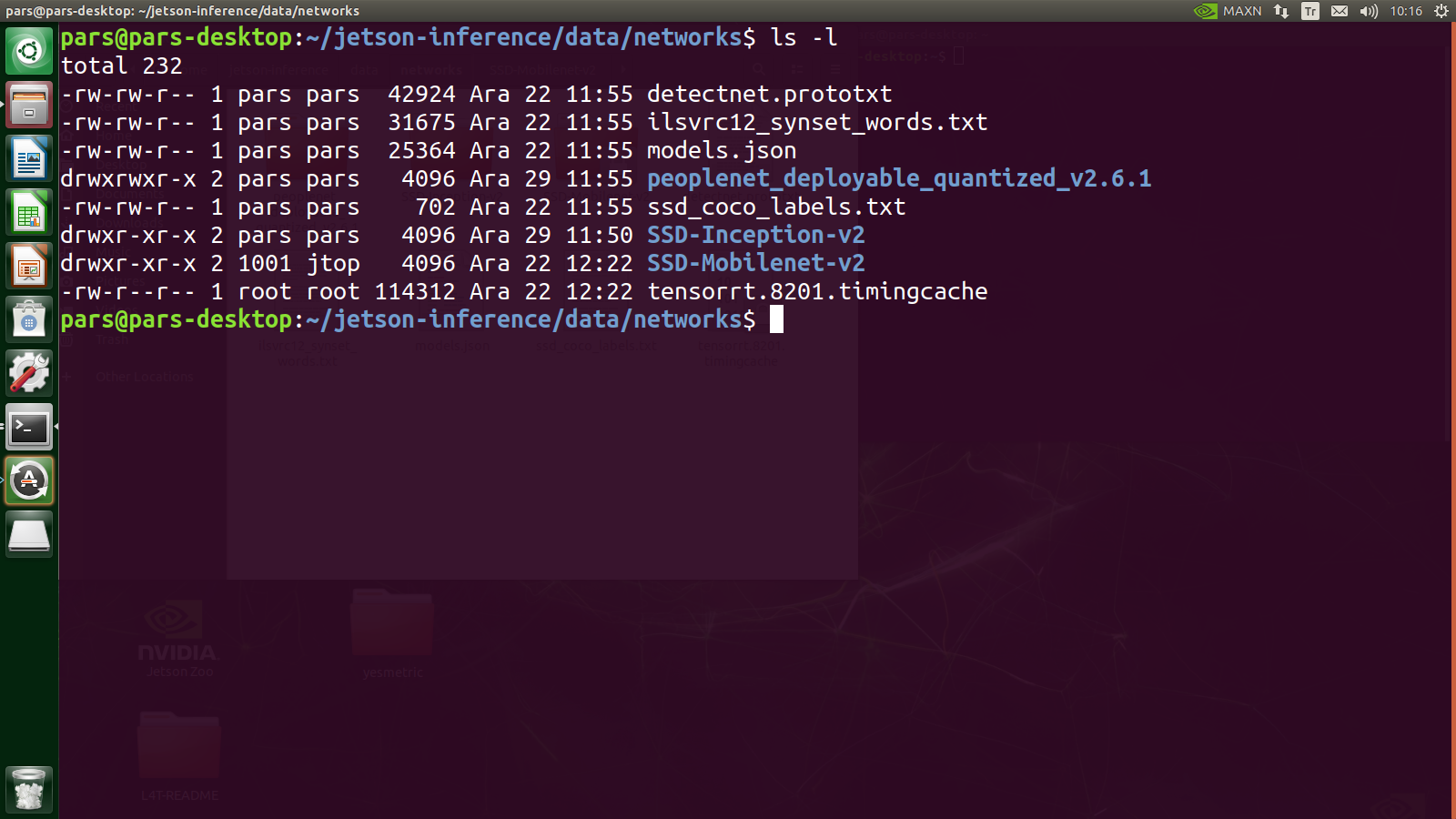
Programming Assignment 4

Group number: 9

Group members:

* Deniz Dönmez
* Yiğit Eren Durmaz
* Gökay Gülsoy

Contents of the data/networks folder after running ssd-mobilenet-v2, ssd-inception-v2, and peoplenet networks with one sample image is as follows:

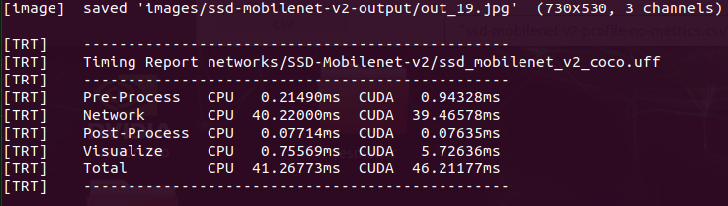


| **Fig 1:** Contents of the data/networks folder after executing each network with one sample image |
| --- |

Profiling commands for ssd-mobilenet-v2, inference time for the last input image, and multiprocessor efficiency metric values for most and least time consuming kernels are as follows:

* sudo /usr/local/bin/cuda/nvprof –csv –log-file images/ssd-mobilnet-v2-output /ssd-moiblenet-v2-profile-no-metrics.csv ./detectnet “images/myimages/\*.jpg” “images/ssd-mobilenet-v2-output\_%i.jpg”
* sudo /usr/local/bin/cuda/nvprof –csv –log-file images/ssd-mobilnet-v2-output /ssd-mobilenet-v2-profile-metrics.csv ./detectnet “images/myimages/\*.jpg” “images/ssd-mobilenet-v2-output\_%i.jpg”

Inference time for the last input image for ssd-mobilenet-v2 is as follows:



| **Fig 2:** Inference time for the last input image fed to ssd-mobilenet-v2 network |
| --- |

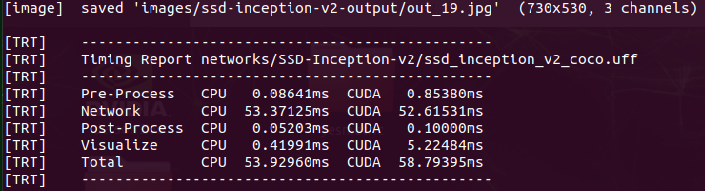
Multiprocessor efficiency metric values for least and most time consuming kernels is as follows:

* Least time consuming kernel: setUniformOffsets => 31.531338%
* Most time consuming kernel: cuDepthWise => 98.481452%

Profiling commands for ssd-mobilenet-v2, inference time for the last input image, and multiprocessor efficiency metric values for the least and most time consuming kernels are as follows:

* sudo /usr/local/bin/cuda/nvprof –csv –log-file images/ssd-inception-v2-output /ssd-inception-v2-profile-no-metrics.csv ./detectnet “images/myimages/\*.jpg” “images/ssd-inception-v2-output\_%i.jpg”
* sudo /usr/local/bin/cuda/nvprof –csv –log-file images/ssd-inception-v2-output /ssd-inception-v2-profile-metrics.csv ./detectnet “images/myimages/\*.jpg” “images/ssd-inception-v2-output\_%i.jpg”

Inference time for the last input image for the ssd-inception-v2 is as follows:



| **Fig 3:** Inference time for the last input image fed to ssd-inception-v2 network |
| --- |

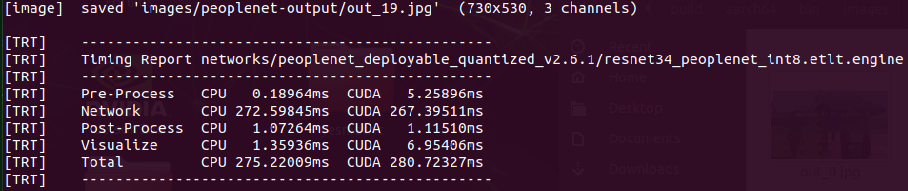
Multiprocessor efficiency metric values for least and most time consuming kernels is as follows:

* Least time consuming kernel: setUniformOffsets => 35.078278%
* Most time consuming kernel: trt\_maxwell\_fp16x2\_hcudnn\_fp16x2\_128x64\_relu\_small\_nn\_v1 => 99.951964%

Profiling commands for peoplenet, inference time for the last input image, and multiprocessor efficiency metric values for the last and most time consuming kernels are as follows:

* sudo /usr/local/bin/cuda/nvprof –csv –log-file images/peoplenet-output /peoplenet-profile-no-metrics.csv ./detectnet “images/myimages/\*.jpg” “images/peoplenet-output\_%i.jpg”
* sudo /usr/local/bin/cuda/nvprof –csv –log-file images/peoplenet-output /peoplenet-profile-metrics.csv ./detectnet “images/myimages/\*.jpg” “images/peoplenet-output\_%i.jpg”

Inference time for the last input image for the peoplenet is as follows:



| **Fig 4:** Inference time for the last input image fed to peoplenet network |
| --- |

Multiprocessor efficiency metric values for least and most time consuming kernels are as follows:

* Least time consuming kernel: generatedNativePointWise => 80.758906%
* Most time consuming kernel: trt\_maxwell\_fp16x2\_hcudnn\_winograd\_fp16x2\_128x128\_ldg1\_ldg4\_relu\_tile148m\_nt\_v1 => 99.543529%

In general for three of these network models multiprocessor efficiency of most time consuming kernels are greater than least time consuming kernels which is plausible because we expect most time consuming kernels to perform highly parallel tasks so that they are expected to utilize all the streaming multiprocessor cores nearly at maximum level. When we compare the network models among themselves highest multiprocessor efficiency for the least time consuming kernel belongs to peoplenet, which indicates that it better utilized the streaming multiprocessor cores for the least time consuming kernels compared to least time consuming kernels of ssd-mobilenet-v2 and ssd-inception-v2 network models. For the most time consuming kernels, multiprocessor efficiencies are close to maximum. Peoplenet has slightly better multiprocessor efficiency compared to ssd-mobilenet-v2 and ssd-inception-v2 according to above analysis results.