

CHOI Jang Hyeon (1155077214)
Supervisors: Professors JIA Jiaya and YU Bei
Mentor: Mr. QI Sun
CSE Department
Term 2, 2020-21
March 14th, 2021

FYP Progress Report: Week 10

The performance difference between the NVIDIA 1080Ti and the Titan Xp is subtle, and considering the small availability of the 1080Ti, the research will be continued on the Titan Xp.

NCHW memory layouts are more suitable to GPUs while the NHWC layout is more suitable for CPUs. A simple if-else code can check the availability of the graphics card with the help of Tensorflow and determine the corresponding memory layout for tuning:

```
import tensorflow as tf

if tf.test.is_gpu_available():
    print("GPU is available. Using the NCHW data layout...")
    memory_layout = "NCHW"
else:
    print("GPU is unavailable. Using the NHWC data layout...")
    memory_layout = "NHWC"
```

However, the other memory layout may be the better choice in some cases, meaning a more comprehensive algorithm needs to be thought of.

Accelerating the configuration exploration is by no means a trivial matter. There are 384,384 possible configurations in the NCHW layout under the current tile settings, and 1,000 configurations already takes 14 hours to tune. I first need to investigate the numerical values of these 1,000 configurations and compare them against the original set of 384,384 configurations. The aim is to include as much variety of configurations as possible and gather a healthy sample suitable for tuning.