



Team Name: Fast
ECE408 Final Project Report: Milestone 2

Jiming Chen (jimingc2)
Jiangyan Feng (jf8)
Maxim Korolkov (korolko2)
University of Illinois at Urbana-Champaign
October 11, 2019

1 Milestone 2

1.1 Kernels that collectively consume more than 90% of the program time

1. CUDA memcpy HtoD (30.43%)
2. volta_scudnn_128x64_relu_interior_nn_v1 (17.97%)
3. volta_gcgemm_64x32_nt (17.29%)
4. fft2d_c2r_32x32 (8.80%)
5. volta_sgemm_128x128_tn (7.85%)
6. op_generic_tensor_kernel (6.60%)
7. fft2d_r2c_32x32 (6.48%)

1.2 CUDA API calls that collectively consume more than 90% of the program time

1. cudaStreamCreateWithFlags (41.94%)
2. cudaMemGetInfo (32.96%)
3. cudaFree (21.22%)

1.3 Explanation of the difference between kernels and API calls

Kernels are code that is called from the CPU but is executed on the GPU. API calls on the other hand are called from the CPU and are executed on the CPU but are able to send the required information to the GPU, such as memcpy information.

1.4 Output of rai running MXNet on the CPU

```
Loading fashion-mnist data... done
Loading model... done
New Inference
EvalMetric: {'accuracy': 0.8154}
17.09user 4.49system 0:08.92elapsed 241%CPU (0avgtext+0avgdata 6045876maxresident)k
0inputs+2824outputs (0major+1602890minor)pagefaults 0swaps
```

1.5 Program run time on CPU

17.09user 4.49system 0:08.92elapsed 241%CPU

1.6 Output of rai running MXNet on the GPU

```
Loading fashion-mnist data... done
Loading model... done
New Inference
EvalMetric: {'accuracy': 0.8154}
4.91user 3.48system 0:04.71elapsed 177%CPU
(0avgtext+0avgdata 2981120maxresident)k
0inputs+1712outputs (0major+732138minor)pagefaults 0swaps
```

1.7 Program run time on GPU

4.91user 3.48system 0:04.71elapsed 177%CPU

1.8 Whole program execution time

```
m1.1.py: 17.09user 4.49system 0:08.92elapsed
m1.2.py: 4.91user 3.48system 0:04.71elapsed
m2.1.py: 83.12user 8.17system 1:13.74elapsed
Total: 105.12user, 16.14system, 1:27.37elapsed
```

1.9 Op times

```
Op Time: 10.801294
Op Time: 59.181779
Correctness: 0.7653
```

2 Milestone 1

Team Name: fast

Team Members: Jiming Chen, Jiangyan Feng, Maxim Korolkov