M1.1
```
* Running python /src/m1.1.py
Loading fashion-mnist data... done
Loading model... done
EvalMetric: {'accuracy': 0.8673}
```

M1.2 The GPU spends most of its time in cudaStreamCreateWithFlags and implicit_convolve_sgemm, which create the stream of instructions and perform the convolutions respectively.

```
* Running nvprof python /src/m1.2.py
Loading fashion-mnist data... done
==305== NVPROF is profiling process 305, command: python /src/m1.2.py
Loading model...[22:03:58] src/operator/./../cudnn_algoreg-inl.h:112: Running perfor
done
EvalMetric: {'accuracy': 0.8673}
==305== Profiling application: python /src/m1.2.py
==305== Profiling result:
Time(%)      Time     Calls       Avg       Min       Max  Name
36.73%  50.515ms         1  50.515ms  50.515ms  50.515ms  void cudnn::detail::impli
 28.66%  39.409ms         1  39.409ms  39.409ms  39.409ms  sgemm_sm35_ldg_tn_128x8x
14.10%  19.396ms         2  9.6979ms  461.70us  18.934ms  void cudnn::detail::activ
10.56%  14.516ms         1  14.516ms  14.516ms  14.516ms  void cudnn::detail::pooli
  5.31%  7.2991ms        13  561.47us  1.5360us  5.3801ms  [CUDA memcpy HtoD]
  2.65%  3.6423ms         1  3.6423ms  3.6423ms  3.6423ms  sgemm_sm35_ldg_tn_64x16x
0.82%  1.1255ms         1  1.1255ms  1.1255ms  1.1255ms  void mshadow::cuda::Softma
0.55%  756.61us        12  63.050us  2.0800us  381.98us  void mshadow::cuda::MapPla
0.32%  436.67us         2  218.34us  16.736us  419.94us  void mshadow::cuda::MapPla
  0.28%  391.17us         1  391.17us  391.17us  391.17us  sgemm_sm35_ldg_tn_32x16x
0.02%  22.336us         1  22.336us  22.336us  22.336us  void mshadow::cuda::MapPla
  0.01%  9.8560us         1  9.8560us  9.8560us  9.8560us  [CUDA memcpy DtoH]
==305== API calls:
Time(%)      Time     Calls       Avg       Min       Max  Name
 46.82%  1.85071s        18  102.82ms  17.325us  925.01ms  cudaStreamCreateWithFlag
 28.48%  1.12580s        10  112.58ms     949ns  320.49ms  cudaFree
 20.78%  821.31ms        24  34.221ms  236.10us  814.20ms  cudaMemGetInfo
  3.26%  128.96ms        25  5.1585ms  5.4060us  83.903ms  cudaStreamSynchronize
  0.38%  15.040ms         8  1.8800ms  8.7350us  5.6225ms  cudaMemcpy2DAsync
0.17%  6.7394ms        42  160.46us  12.587us  1.1791ms  cudaMalloc
  0.03%  1.3630ms         4  340.74us  338.56us  343.25us  cuDeviceTotalMem
  0.02%  855.24us       352  2.4290us     246ns  67.187us  cuDeviceGetAttribute
  0.02%  741.66us       114  6.5050us     645ns  303.68us  cudaEventCreateWithFlags
  0.01%  493.21us        23  21.444us  10.378us  85.997us  cudaLaunch
  0.01%  320.83us         6  53.472us  23.284us  80.713us  cudaMemcpy
  0.01%  218.11us         4  54.526us  38.522us  92.078us  cudaStreamCreate
  0.00%  109.43us         4  27.357us  17.599us  31.972us  cuDeviceGetName
  0.00%  94.979us        32  2.9680us     565ns  19.503us  cudaSetDevice
  0.00%  70.858us       110     644ns     419ns  2.3370us  cudaDeviceGetAttribute
  0.00%  60.995us       147     414ns     265ns  1.1010us  cudaSetupArgument
  0.00%  40.094us         2  20.047us  18.193us  21.901us  cudaStreamCreateWithPrio
  0.00%  25.586us        23  1.1120us     440ns  2.0780us  cudaConfigureCall
  0.00%  21.746us        10  2.1740us  1.4410us  6.7240us  cudaGetDevice
  0.00%  8.7270us        16     545ns     368ns     796ns  cudaPeekAtLastError
  0.00%  8.5150us         1  8.5150us  8.5150us  8.5150us  cudaBindTexture
  0.00%  4.3480us         6     724ns     314ns  1.5030us  cuDeviceGetCount
  0.00%  4.2730us         1  4.2730us  4.2730us  4.2730us  cudaStreamGetPriority
```

```
      0.00%  3.9060us         2  1.9530us  1.3920us  2.5140us  cudaStreamWaitEvent
      0.00%  3.5690us         2  1.7840us  1.2320us  2.3370us  cudaEventRecord
      0.00%  3.4930us         6     582ns     378ns     916ns  cuDeviceGet
      0.00%  2.9380us         2  1.4690us  1.3250us  1.6130us  cudaDeviceGetStreamPrior
      0.00%  2.9240us         6     487ns     337ns     661ns  cudaGetLastError
      0.00%  2.8340us         3     944ns     752ns  1.0690us  cuInit
    0.00%  2.1150us         3     705ns     655ns     786ns  cuDriverGetVersion
      0.00%  1.5910us         1  1.5910us  1.5910us  1.5910us  cudaUnbindTexture
      0.00%  1.2040us         1  1.2040us  1.2040us  1.2040us  cudaGetDeviceCount
```

M2.1 * Running python /src/m2.1.py

```
Loading fashion-mnist data... done
Loading model... done
Op Time: 12.146829
Correctness: 0.8562 Model: ece408-high
```