Experientments Appendix:

Experiments results for testing GPU activities summary for GPU num = 1,2,4 and Batch size = 8,16,12,128

GPU num = 1, Batch size = 8

```
==13335== NVPROF is profiling process 13335, command: python main.py -a alexnet -b 8 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==13335== Profiling application: python main.py -a alexnet -b 8 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==13335== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:   20.41%  92.639ms       512  180.94us     992ns  1.7844ms  void kernelPointwiseApply2<TensorCAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorCAddOp<floa$
                   19.74%  89.623ms       526  170.39us     992ns  1.8014ms  void kernelPointwiseApply2<TensorAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorAddOp<float>$
                    7.70%  34.978ms        58  603.06us     960ns  17.520ms  [CUDA memcpy HtoD]
                    6.71%  30.469ms       288  105.80us  1.0880us  1.1653ms  void kernelPointwiseApply1<TensorMulConstantOp<float>, float, unsigned int, int=1>(OffsetInfo<TensorMulConstantOp<floa$
                    4.17%  18.930ms       102  185.59us  147.75us  257.58us  maxwell_scudnn_winograd_128x128_ldg1_ldg4_tile228n_nt
                    4.10%  18.616ms        48  387.83us  89.379us  854.27us  sgemm_32x32x32_NT_vec
                    3.81%  17.289ms        48  360.20us  86.275us  700.63us  sgemm_32x32x32_NN_vec
                    3.43%  15.563ms        48  324.23us  70.627us  641.78us  sgemm_128x128x8_TN_vec
                    2.13%  9.6816ms         8  1.2102ms  301.74us  1.8579ms  maxwell_gcgemm_32x32_nt
                    2.12%  9.6186ms        71  135.47us  95.203us  238.25us  maxwell_sgemm_128x64_nt
```

GPU num = 1, Batch size = 16

```
==13339== NVPROF is profiling process 13339, command: python main.py -a alexnet -b 16 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==13339== Profiling application: python main.py -a alexnet -b 16 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==13339== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:   17.33%  92.543ms       512  180.75us     992ns  1.8178ms  void kernelPointwiseApply2<TensorCAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorCAddOp<floa$
                   16.77%  89.523ms       526  170.20us     992ns  1.8179ms  void kernelPointwiseApply2<TensorAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorAddOp<float>$
                    8.03%  42.858ms        58  738.94us  1.0240us  17.874ms  [CUDA memcpy HtoD]
                    5.70%  30.430ms       288  105.66us  1.0880us  1.1652ms  void kernelPointwiseApply1<TensorMulConstantOp<float>, float, unsigned int, int=1>(OffsetInfo<TensorMulConstantOp<floa$
                    3.94%  21.013ms        70  300.18us  265.26us  474.36us  maxwell_scudnn_winograd_128x128_ldg1_ldg4_tile228n_nt
                    3.53%  18.833ms        48  392.36us  92.932us  857.89us  sgemm_32x32x32_NT_vec
                    3.26%  17.387ms        48  362.23us  86.019us  702.49us  sgemm_32x32x32_NN_vec
                    3.15%  16.817ms        87  193.29us  151.08us  244.43us  maxwell_sgemm_128x64_nt
                    3.06%  16.344ms        48  340.49us  76.355us  674.65us  sgemm_128x128x8_TN_vec
                    2.16%  11.550ms        53  217.92us  172.10us  325.65us  maxwell_sgemm_128x64_nn
```

GPU num = 1, Batch size = 32

```
==13679== NVPROF is profiling process 13679, command: python main.py -a alexnet -b 32 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==13679== Profiling application: python main.py -a alexnet -b 32 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==13679== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:   13.38%  92.581ms       512  180.82us     992ns  1.7835ms  void kernelPointwiseApply2<TensorCAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorCAddOp<floa$
                   12.96%  89.703ms       526  170.54us     992ns  1.8286ms  void kernelPointwiseApply2<TensorAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorAddOp<float>$
                    7.91%  54.762ms        58  944.17us     992ns  17.567ms  [CUDA memcpy HtoD]
                    4.91%  33.962ms       103  329.73us  262.70us  399.15us  maxwell_sgemm_128x64_nt
                    4.39%  30.415ms       288  105.61us     992ns  1.1653ms  void kernelPointwiseApply1<TensorMulConstantOp<float>, float, unsigned int, int=1>(OffsetInfo<TensorMulConstantOp<floa$
                    3.39%  23.476ms        69  340.23us  296.08us  447.18us  maxwell_sgemm_128x64_nn
                    3.00%  20.795ms        38  547.24us  501.17us  750.91us  maxwell_scudnn_winograd_128x128_ldg1_ldg4_tile228n_nt
                    2.76%  19.125ms        48  398.43us  94.115us  889.54us  sgemm_32x32x32_NT_vec
                    2.55%  17.641ms         8  2.2051ms  1.6711ms  2.8028ms  void cudnn::detail::dgrad_engine<float, int=128, int=6, int=8, int=3, int=3, int=5, bool=1>(int, int, int, float const$
                    2.55%  17.614ms        48  366.95us  79.683us  733.15us  sgemm_128x128x8_TN_vec
                    2.54%  17.558ms        48  365.79us  85.827us  700.57us  sgemm_32x32x32_NN_vec
                    2.25%  15.549ms        48  323.94us  130.60us  479.12us  void MaxPoolBackward<float, float>(int, float const *, long const *, int, int, int, int, int, int, int, int, int, int,$
                    2.07%  14.335ms        18  796.41us  782.78us  846.56us  maxwell_scudnn_128x64_stridedB_splitK_medium_nn
                    2.04%  14.108ms         6  2.3513ms  1.4559ms  3.2775ms  void cudnn::detail::wgrad_alg0_engine<float, int=128, int=6, int=8, int=3, int=3, int=5, bool=1, int=512>(int, int, in$
```

GPU num = 1, Batch size = 128

```
==13675== NVPROF is profiling process 13675, command: python main.py -a alexnet -b 128 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==13675== Profiling application: python main.py -a alexnet -b 128 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==13675== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:    8.31%  149.30ms        58  2.5741ms  1.0240us  17.906ms  [CUDA memcpy HtoD]
                    6.94%  124.57ms       117  1.0647ms  158.50us  1.8205ms  maxwell_sgemm_128x64_nn
                    5.34%  95.952ms        71  1.3514ms  862.66us  1.7183ms  maxwell_sgemm_128x64_nt
                    5.16%  92.655ms       512  180.97us     992ns  1.7868ms  void kernelPointwiseApply2<TensorCAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorCAddOp<floa$
                    5.01%  89.937ms       526  170.98us  1.0240us  1.8531ms  void kernelPointwiseApply2<TensorAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorAddOp<float>$
                    4.35%  78.042ms        32  2.4388ms  1.1109ms  4.3302ms  maxwell_sgemm_128x64_tn
                    3.90%  70.064ms         8  8.7579ms  6.7787ms  10.657ms  void cudnn::detail::dgrad_engine<float, int=128, int=6, int=8, int=3, int=3, int=5, bool=1>(int, int, int, float const$
                    3.44%  61.779ms        48  1.2871ms  522.04us  1.8982ms  void MaxPoolBackward<float, float>(int, float const *, long const *, int, int, int, int, int, int, int, int, int, int,$
                    3.14%  56.436ms        18  3.1353ms  3.0893ms  3.2631ms  maxwell_scudnn_128x64_stridedB_splitK_medium_nn
                    3.01%  54.004ms         6  9.0006ms  5.5893ms  12.775ms  void cudnn::detail::wgrad_alg0_engine<float, int=128, int=6, int=8, int=3, int=3, int=5, bool=1, int=512>(int, int, in$
                    2.68%  48.169ms        22  2.1895ms  2.0342ms  2.9911ms  maxwell_scudnn_winograd_128x128_ldg1_ldg4_tile228n_nt
                    2.49%  44.776ms         4  11.194ms  7.7799ms  14.552ms  void cudnn::detail::implicit_convolve_sgemm<float, float, int=512, int=6, int=8, int=3, int=3, int=5, int=1, bool=1, b$
                    2.40%  43.181ms       112  385.54us  26.081us  1.0592ms  void kernelPointwiseApply3<ThresholdUpdateGradInput<float>, float, float, float, unsigned int, int=1, int=1, int=1>(Of$
                    2.27%  40.778ms        23  1.7730ms  879.01us  5.2508ms  cudnn_maxwell_gcgemm_64x64_nt_batched
                    2.16%  38.783ms        17  2.2814ms  2.2312ms  2.6127ms  maxwell_scudnn_128x64_relu_large_nn
                    2.06%  37.079ms        70  529.70us  331.02us  720.96us  void cudnn::winograd_nonfused::winogradForwardOutput4x4<float, float>(cudnn::winograd_nonfused::WinogradOutputParams<f$
```

GPU num = 2, Batch size = 8

```
==11474== NVPROF is profiling process 11474, command: python main.py -a alexnet -b 8 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==11474== Profiling application: python main.py -a alexnet -b 8 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==11474== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:   21.76%  170.12ms        64  2.6582ms  281.80us  11.960ms  ncclBroadcastKernel_copy_i8(ncclColl)
                   11.87%  92.796ms       512  181.24us     992ns  1.7898ms  void kernelPointwiseApply2<TensorCAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorCAddOp<floa$
                   11.50%  89.899ms       526  170.91us     992ns  1.7978ms  void kernelPointwiseApply2<TensorAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorAddOp<float>$
                    9.90%  77.399ms       175  442.28us     960ns  43.732ms  [CUDA memcpy HtoD]
                    4.62%  36.087ms        32  1.1277ms  1.0054ms  1.3389ms  ncclReduceKernel_sum_f32(ncclColl)
                    3.92%  30.655ms       288  106.44us  1.0880us  1.1860ms  void kernelPointwiseApply1<TensorMulConstantOp<float>, float, unsigned int, int=1>(OffsetInfo<TensorMulConstantOp<floa$
                    2.80%  21.899ms       204  107.35us  88.132us  160.30us  maxwell_scudnn_winograd_128x128_ldg1_ldg4_tile228n_nt
                    2.39%  18.717ms        16  1.1698ms  289.96us  1.7950ms  maxwell_gcgemm_32x32_nt
                    2.37%  18.493ms        48  385.27us  88.611us  740.95us  sgemm_32x32x32_NT_vec
                    2.21%  17.294ms        48  360.30us  82.915us  693.18us  sgemm_32x32x32_NN_vec
                    2.05%  16.039ms       142  112.95us  73.123us  238.76us  maxwell_sgemm_128x64_nt
                    2.00%  15.619ms        48  325.40us  71.171us  636.60us  sgemm_128x128x8_TN_vec
```

GPU num = 2, Batch size = 16

```
==11720== NVPROF is profiling process 11720, command: python main.py -a alexnet -b 16 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==11720== Profiling application: python main.py -a alexnet -b 16 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==11720== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:   16.23%  125.08ms        64  1.9544ms  280.62us  11.995ms  ncclBroadcastKernel_copy_i8(ncclColl)
                   12.04%  92.817ms       512  181.28us     992ns  1.7907ms  void kernelPointwiseApply2<TensorCAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorCAddOp<flo
                   11.67%  89.948ms       526  171.00us     992ns  1.8117ms  void kernelPointwiseApply2<TensorAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorAddOp<float
                    5.65%  43.575ms       175  249.00us     960ns  17.521ms  [CUDA memcpy HtoD]
                    4.84%  37.318ms       203  183.83us  146.73us  271.27us  maxwell_scudnn_winograd_128x128_ldg1_ldg4_tile228n_nt
                    4.55%  35.039ms        32  1.0950ms  1.0223ms  1.1535ms  ncclReduceKernel_sum_f32(ncclColl)
                    3.97%  30.584ms       288  106.20us  1.0880us  1.1797ms  void kernelPointwiseApply1<TensorMulConstantOp<float>, float, unsigned int, int=1>(OffsetInfo<TensorMulConstantOp<flo
                    2.46%  18.978ms       141  134.59us  96.196us  238.57us  maxwell_sgemm_128x64_nt
                    2.42%  18.649ms        48  388.53us  92.323us  742.36us  sgemm_32x32x32_NT_vec
                    2.37%  18.237ms        15  1.2158ms  308.08us  1.8494ms  maxwell_gcgemm_32x32_nt
                    2.26%  17.441ms        48  363.35us  86.019us  701.56us  sgemm_32x32x32_NN_vec
                    2.12%  16.377ms        48  341.18us  74.434us  666.58us  sgemm_128x128x8_TN_vec
```

GPU num = 2, Batch size = 32

```
==12077== NVPROF is profiling process 12077, command: python main.py -a alexnet -b 32 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==12077== Profiling application: python main.py -a alexnet -b 32 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==12077== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:   15.87%  154.22ms        64  2.4097ms  284.14us  12.026ms  ncclBroadcastKernel_copy_i8(ncclColl)
                    9.55%  92.775ms       512  181.20us     992ns  1.7872ms  void kernelPointwiseApply2<TensorCAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorCAddOp<floa$
                    9.26%  90.014ms       526  171.13us     992ns  1.8204ms  void kernelPointwiseApply2<TensorAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorAddOp<float>$
                    6.19%  60.162ms       175  343.78us     960ns  17.469ms  [CUDA memcpy HtoD]
                    5.49%  53.384ms       171  312.19us  257.58us  478.55us  maxwell_scudnn_winograd_128x128_ldg1_ldg4_tile228n_nt
                    3.70%  35.992ms        32  1.1247ms  1.0158ms  1.3097ms  ncclReduceKernel_sum_f32(ncclColl)
                    3.15%  30.636ms       288  106.37us     992ns  1.1863ms  void kernelPointwiseApply1<TensorMulConstantOp<float>, float, unsigned int, int=1>(OffsetInfo<TensorMulConstantOp<floa$
                    2.93%  28.464ms       143  199.05us  161.93us  242.15us  maxwell_sgemm_128x64_nt
                    2.16%  20.991ms        43  488.17us  429.68us  741.08us  maxwell_scudnn_128x128_stridedB_splitK_small_nn
                    2.13%  20.733ms        16  1.2958ms  333.16us  1.9461ms  maxwell_gcgemm_32x32_nt
                    2.10%  20.398ms        16  1.2749ms  966.47us  1.5450ms  void cudnn::detail::dgrad_engine<float, int=128, int=6, int=8, int=3, int=3, int=5, bool=1>(int, int, int, float const$
```

GPU num = 2, Batch size = 128

```
==12833== NVPROF is profiling process 12833, command: python main.py -a alexnet -b 128 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==12833== Profiling application: python main.py -a alexnet -b 128 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==12833== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:    9.62%  196.66ms       175  1.1237ms     992ns  43.805ms  [CUDA memcpy HtoD]
                    6.88%  140.77ms        64  2.1996ms  280.33us  11.995ms  ncclBroadcastKernel_copy_i8(ncclColl)
                    6.71%  137.18ms       216  635.11us  159.30us  1.4471ms  maxwell_sgemm_128x64_nn
                    5.53%  113.09ms       175  646.21us  462.83us  802.53us  maxwell_sgemm_128x64_nt
                    4.54%  92.804ms       512  181.26us     992ns  1.7900ms  void kernelPointwiseApply2<TensorCAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorCAddOp<floa$
                    4.41%  90.275ms       526  171.63us     992ns  1.8646ms  void kernelPointwiseApply2<TensorAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorAddOp<float>$
                    3.81%  77.858ms        75  1.0381ms  969.03us  1.5376ms  maxwell_scudnn_winograd_128x128_ldg1_ldg4_tile228n_nt
                    3.78%  77.263ms        32  2.4145ms  1.1372ms  3.9375ms  maxwell_sgemm_128x64_tn
                    3.42%  70.003ms        16  4.3752ms  3.1899ms  5.4256ms  void cudnn::detail::dgrad_engine<float, int=128, int=6, int=8, int=3, int=3, int=5, bool=1>(int, int, int, float const$
                    3.02%  61.856ms        96  644.34us  260.59us  952.49us  void MaxPoolBackward<float, float>(int, float const *, long const *, int, int, int, int, int, int, int, int, int, int,$
                    2.77%  56.560ms        36  1.5711ms  1.5424ms  1.6970ms  maxwell_scudnn_128x64_stridedB_splitK_medium_nn
                    2.64%  54.022ms        12  4.5018ms  2.7285ms  6.4100ms  void cudnn::detail::wgrad_alg0_engine<float, int=128, int=6, int=8, int=3, int=3, int=5, bool=1, int=512>(int, int, in$
                    2.12%  43.284ms       192  225.44us  25.889us  538.23us  void kernelPointwiseApply3<ThresholdUpdateGradInput<float>, float, float, float, unsigned int, int=1, int=1, int=1>(Of$
```

GPU num = 4, Batch size = 8

```
==152503== NVPROF is profiling process 152503, command: python main.py -a alexnet -b 8 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==152503== Profiling application: python main.py -a alexnet -b 8 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==152503== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:   37.69%  469.58ms       128  3.6686ms  302.00us  12.040ms  ncclBroadcastKernel_copy_i8(ncclColl)
                    7.45%  92.840ms       512  181.33us  1.0240us  1.7894ms  void kernelPointwiseApply2<TensorCAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorCAddOp<float>, f$
                    7.21%  89.817ms       526  170.76us  1.0240us  1.7886ms  void kernelPointwiseApply2<TensorAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorAddOp<float>, flo$
                    6.71%  83.596ms        64  1.3062ms  1.1862ms  1.5517ms  ncclReduceKernel_sum_f32(ncclColl)
                    3.26%  40.616ms       277  146.63us     992ns  16.991ms  [CUDA memcpy HtoD]
                    2.95%  36.724ms        32  1.1476ms  276.52us  1.7655ms  maxwell_gcgemm_32x32_nt
                    2.46%  30.668ms       288  106.49us  1.1200us  1.1819ms  void kernelPointwiseApply1<TensorMulConstantOp<float>, float, unsigned int, int=1>(OffsetInfo<TensorMulConstantOp<float>, f$
                    2.31%  28.770ms       407  70.687us  53.666us  104.42us  maxwell_scudnn_winograd_128x128_ldg1_ldg4_tile228n_nt
```

GPU num = 4, Batch size = 16

```
==153128== NVPROF is profiling process 153128, command: python main.py -a alexnet -b 16 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==153128== Profiling application: python main.py -a alexnet -b 16 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==153128== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:   30.60%  399.10ms       128  3.1180ms  300.59us  12.078ms  ncclBroadcastKernel_copy_i8(ncclColl)
                    7.11%  92.792ms       512  181.23us     992ns  1.7899ms  void kernelPointwiseApply2<TensorCAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorCAddOp<float>, f$
                    6.98%  91.101ms       277  328.88us     960ns  40.822ms  [CUDA memcpy HtoD]
                    6.88%  89.745ms       526  170.62us  1.0240us  1.7879ms  void kernelPointwiseApply2<TensorAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorAddOp<float>, flo$
                    6.48%  84.578ms        64  1.3215ms  1.1720ms  1.6426ms  ncclReduceKernel_sum_f32(ncclColl)
                    3.53%  45.989ms       408  112.72us  89.187us  161.64us  maxwell_scudnn_winograd_128x128_ldg1_ldg4_tile228n_nt
                    2.87%  37.436ms        32  1.1699ms  284.81us  1.7988ms  maxwell_gcgemm_32x32_nt
                    2.50%  32.605ms       282  115.62us  72.931us  237.26us  maxwell_sgemm_128x64_nt
                    2.35%  30.593ms       288  106.22us  1.0880us  1.1817ms  void kernelPointwiseApply1<TensorMulConstantOp<float>, float, unsigned int, int=1>(OffsetInfo<TensorMulConstantOp<float>, f$
```

GPU num = 4, Batch size = 32

```
==153528== NVPROF is profiling process 153528, command: python main.py -a alexnet -b 32 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==153528== Profiling application: python main.py -a alexnet -b 32 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==153528== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:   26.68%  377.18ms       128  2.9467ms  300.75us  12.073ms  ncclBroadcastKernel_copy_i8(ncclColl)
                    8.44%  119.74ms       277  432.29us     960ns  41.336ms  [CUDA memcpy HtoD]
                    6.55%  92.903ms       512  181.45us  1.0240us  1.7931ms  void kernelPointwiseApply2<TensorCAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorCAddOp<float>, f$
                    6.33%  89.852ms       526  170.82us     992ns  1.7996ms  void kernelPointwiseApply2<TensorAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorAddOp<float>, flo$
                    6.14%  87.106ms        64  1.3610ms  1.1599ms  2.1832ms  ncclReduceKernel_sum_f32(ncclColl)
                    4.56%  64.699ms       341  189.73us  145.51us  263.75us  maxwell_scudnn_winograd_128x128_ldg1_ldg4_tile228n_nt
                    3.44%  48.745ms       341  142.95us  96.964us  238.38us  maxwell_sgemm_128x64_nt
                    2.16%  30.606ms       288  106.27us     992ns  1.1835ms  void kernelPointwiseApply1<TensorMulConstantOp<float>, float, unsigned int, int=1>(OffsetInfo<TensorMulConstantOp<float>, f$
```

GPU num = 4, Batch size = 128

```
==153812== NVPROF is profiling process 153812, command: python main.py -a alexnet -b 128 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==153812== Profiling application: python main.py -a alexnet -b 128 --epochs 1 --lr 0.02 /beegfs/work/courses/2019-CSCI-GA-3033-025/imagenet_pytorch_small
==153812== Profiling result:
            Type  Time(%)      Time     Calls       Avg       Min       Max  Name
 GPU activities:   16.70%  446.90ms       128  3.4914ms  303.95us  12.323ms  ncclBroadcastKernel_copy_i8(ncclColl)
                   10.89%  291.27ms       277  1.0515ms     992ns  40.913ms  [CUDA memcpy HtoD]
                    5.49%  146.80ms       384  382.29us  159.46us  1.5006ms  maxwell_sgemm_128x64_nn
                    4.23%  113.24ms       350  323.53us  261.83us  449.04us  maxwell_sgemm_128x64_nt
                    3.75%  100.45ms        87  1.1546ms  791.17us  1.3668ms  maxwell_scudnn_128x128_stridedB_splitK_small_nn
                    3.64%  97.338ms        64  1.5209ms  1.0894ms  2.8580ms  ncclReduceKernel_sum_f32(ncclColl)
                    3.47%  92.901ms       512  181.45us     992ns  1.7905ms  void kernelPointwiseApply2<TensorCAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorCAddOp<float>, f$
                    3.37%  90.108ms       526  171.31us     992ns  1.8157ms  void kernelPointwiseApply2<TensorAddOp<float>, float, float, unsigned int, int=1, int=1>(OffsetInfo<TensorAddOp<float>, flo$
                    2.89%  77.406ms        32  2.4189ms  1.1322ms  3.8342ms  maxwell_sgemm_128x64_tn
                    2.83%  75.800ms        32  2.3687ms  1.7539ms  2.9376ms  void cudnn::detail::dgrad_engine<float, int=128, int=6, int=8, int=3, int=3, int=5, bool=1>(int, int, int, float const *, i$
                    2.33%  62.408ms       192  325.04us  130.15us  481.62us  void MaxPoolBackward<float, float>(int, float const *, long const *, int, int, int, int, int, int, int, int, int, int, int,$
                    2.16%  57.706ms        24  2.4044ms  1.5040ms  3.4394ms  void cudnn::detail::wgrad_alg0_engine<float, int=128, int=6, int=8, int=3, int=3, int=5, bool=1, int=512>(int, int, int, fl$
                    2.14%  57.342ms        72  796.42us  779.36us  851.33us  maxwell_scudnn_128x64_stridedB_splitK_medium_nn
```