

CNN Performance Estimation

1 Introduction

1.1 Background

In the realm of machine learning, small-scale entities often face the challenge of optimizing CNNs for efficient performance on accessible hardware, balancing the constraints of computational resources and memory demands, a necessity for maintaining competitiveness in a resource-intensive industry.

1.2 Purpose

The aim is to evaluate the performance impact of CNN layers on GPU resources, focusing on the computational time and memory footprint of a Conv2d layer within a neural network handling MNIST data.

1.3 Approach

This entails implementing timed sections within a PyTorch model to measure the execution time and leveraging NVIDIA profiling tools to assess memory utilization, providing insights into resource optimization for deep learning tasks.

2 Pen and Paper Method

Conv2d - 2 [Time and Space Complexity]

Given

$\Rightarrow \text{Conv2d}(32, 64, 3, 1)$

$\Rightarrow \text{Input size} = [64, 32, 13, 13]$

$\Rightarrow \text{batch size} = 64$

$\Rightarrow \text{Output size} = \frac{(\text{input_size} - \text{kernel_size} + 2 \times \text{padding})}{\text{stride} + 1}$

$\therefore \text{Output size} = (64, 64, 11, 11)$

Time Complexity

(Forward Propagation):

Formula: $O(k \times j \times i \dots)$

$\Rightarrow O(k \times (\text{input_channel} \times (\text{kernel_size})^2 \times \text{output_channel} \times \text{output_width} \times \text{output_height}))$

Data:

$$N = 64 \text{ (batch size)}$$

$$\text{input_channel} = 32$$

$$\text{kernel_size} = 3$$

$$\text{output_channels} = 64$$

$$\text{output_width} = \text{output_height} = 11$$

$$\begin{aligned}\text{Time Complexity} &= O(64 \times (32 \times 3^2 \times 64 \times 11 \times 11)) \\ &= O(64 \times (32 \times 9 \times 64 \times 121)) \\ &= O(64 \times 2,211,840) \\ &= O(141,557,760)\end{aligned}$$

$$\therefore \text{Time Complexity} = O(141,557,760)$$

Space Complexity

Parameters,

$$\begin{aligned}\text{Weights} &= (\text{kernel_size})^2 \times \text{input_channel} \times \text{output_channel} \\ &= 3^2 \times 32 \times 64 \\ &= 18,432\end{aligned}$$

$$\Rightarrow \text{Biases} = \text{Output_channels} = 64$$

Output Feature Map:

$$\Rightarrow \text{output_ch} \times \text{output_width} \times \text{output_height} \times \text{batch}$$

$$\Rightarrow 64 \times 11 \times 11 \times 64$$

$$\Rightarrow \cancel{516,112} \quad 495,616$$

Total space complexity

$$\Rightarrow \text{Parameters} + \text{Output_Feature_Map}$$

$$\Rightarrow 18,432 + 64 + 495,616$$

$$\Rightarrow 514,112$$

$$\therefore \text{Space Complexity} = O(514,112)$$

Memory in Kb

$$\Rightarrow \text{Memory per Output value} = 4 \text{ bytes}$$

$$\Rightarrow \text{Output feature map} \times \text{Memory value}$$

$$\Rightarrow 64 \times 64 \times 11 \times 11 \times 4$$

$$\Rightarrow 1,982,464 \text{ bytes}$$

$$\Rightarrow \underline{1,982,464}$$

$$\Rightarrow 1024$$

$$\Rightarrow \boxed{1936 \text{ Kb}}$$

Summary,

$$\text{Weights} = 72 \text{ Kb}$$

$$\text{Biases} = 0.25 \text{ Kb}$$

$$\text{output_feature_map} = 1936 \text{ Kb}$$

3 Performance Study Using PyTorch Summary

3.1 Output

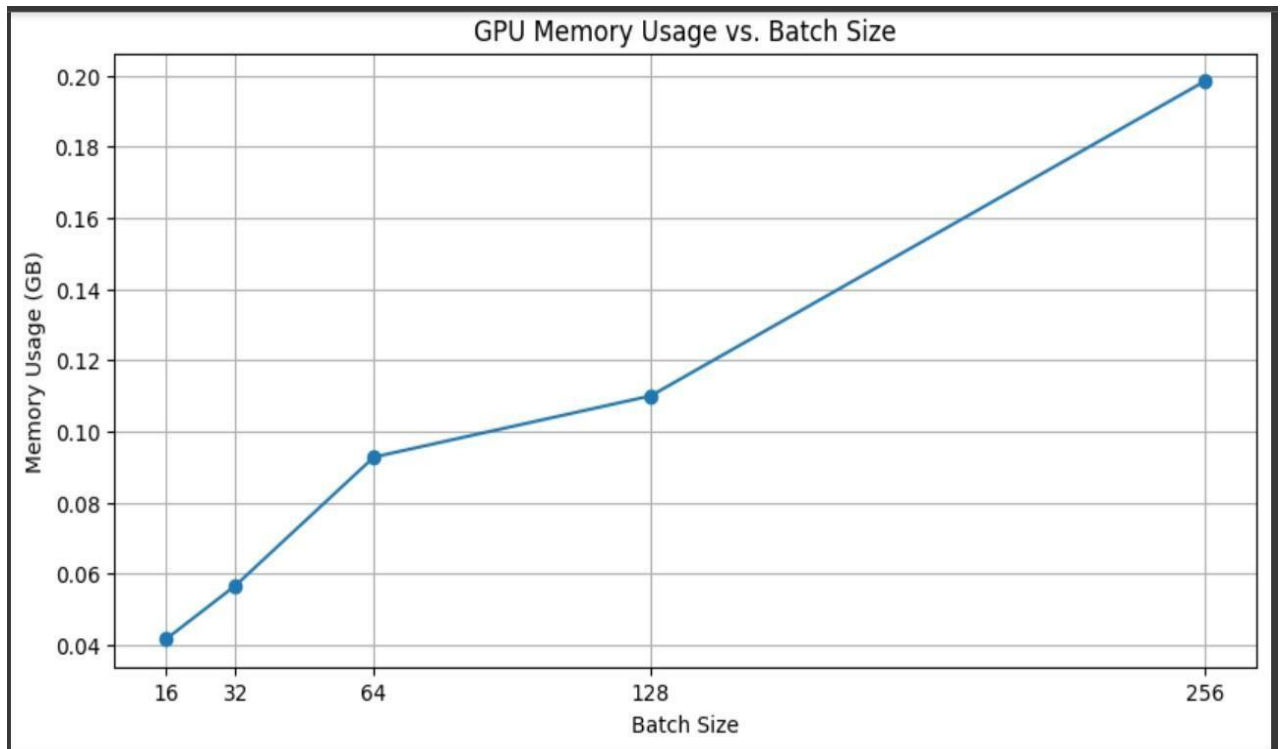
Before

```
pc3427@log-burst:~  
Singularity> vi main.py  
Singularity> python3 main.py  
-----  
Layer (type)          Output Shape          Param #  
-----  
      Conv2d-1         [-1, 32, 26, 26]           320  
      Conv2d-2         [-1, 64, 24, 24]        18,496  
      Dropout-3        [-1, 64, 12, 12]           0  
      Linear-4          [-1, 128]             1,179,776  
      Dropout-5         [-1, 128]               0  
      Linear-6          [-1, 10]                1,290  
-----  
Total params: 1,199,882  
Trainable params: 1,199,882  
Non-trainable params: 0  
-----  
Input size (MB): 0.00  
Forward/backward pass size (MB): 0.52  
Params size (MB): 4.58  
Estimated Total Size (MB): 5.10  
-----  
Memory Usage before training:  
Allocated: 0.00 GB  
Reserved: 0.02 GB  
CPU RAM Free: 84.3 GB  
GPU 0: Memory Free: 38554.0MB / 40960.0MB | Utilization: 4.35791015625%  
Train Epoch: 1 [0/60000 (0%)] Loss: 2.303741  
Train Epoch: 1 [640/60000 (1%)] Loss: 1.250788  
Train Epoch: 1 [1280/60000 (2%)] Loss: 0.813728  
Train Epoch: 1 [1920/60000 (3%)] Loss: 0.426390  
Train Epoch: 1 [2560/60000 (4%)] Loss: 0.556592  
Train Epoch: 1 [3200/60000 (5%)] Loss: 0.172841  
Train Epoch: 1 [3840/60000 (6%)] Loss: 0.326665  
Train Epoch: 1 [4480/60000 (7%)] Loss: 0.387170  
Train Epoch: 1 [5120/60000 (9%)] Loss: 0.309800
```

After

```
pc3427@log-burst:~  
Train Epoch: 14 [44160/60000 (74%)] Loss: 0.004841  
Train Epoch: 14 [44800/60000 (75%)] Loss: 0.086831  
Train Epoch: 14 [45440/60000 (76%)] Loss: 0.016361  
Train Epoch: 14 [46080/60000 (77%)] Loss: 0.107713  
Train Epoch: 14 [46720/60000 (78%)] Loss: 0.011860  
Train Epoch: 14 [47360/60000 (79%)] Loss: 0.004715  
Train Epoch: 14 [48000/60000 (80%)] Loss: 0.040008  
Train Epoch: 14 [48640/60000 (81%)] Loss: 0.014918  
Train Epoch: 14 [49280/60000 (82%)] Loss: 0.050742  
Train Epoch: 14 [49920/60000 (83%)] Loss: 0.001852  
Train Epoch: 14 [50560/60000 (84%)] Loss: 0.001311  
Train Epoch: 14 [51200/60000 (85%)] Loss: 0.004660  
Train Epoch: 14 [51840/60000 (86%)] Loss: 0.017244  
Train Epoch: 14 [52480/60000 (87%)] Loss: 0.021238  
Train Epoch: 14 [53120/60000 (88%)] Loss: 0.001500  
Train Epoch: 14 [53760/60000 (90%)] Loss: 0.010422  
Train Epoch: 14 [54400/60000 (91%)] Loss: 0.006629  
Train Epoch: 14 [55040/60000 (92%)] Loss: 0.003442  
Train Epoch: 14 [55680/60000 (93%)] Loss: 0.000459  
Train Epoch: 14 [56320/60000 (94%)] Loss: 0.043133  
Train Epoch: 14 [56960/60000 (95%)] Loss: 0.001035  
Train Epoch: 14 [57600/60000 (96%)] Loss: 0.069848  
Train Epoch: 14 [58240/60000 (97%)] Loss: 0.001160  
Train Epoch: 14 [58880/60000 (98%)] Loss: 0.012721  
Train Epoch: 14 [59520/60000 (99%)] Loss: 0.007149  
Average time per batch for conv2: 0.0001746069573560767 seconds  
  
Test set: Average loss: 0.0254, Accuracy: 9920/10000 (99%)  
  
Memory Usage after training:  
Allocated: 0.02 GB  
Reserved: 0.62 GB  
CPU RAM Free: 84.1 GB  
GPU 0: Memory Free: 37876.0MB / 40960.0MB | Utilization: 6.01318359375%  
Singularity>
```

4 Comparison Graph



5 Performance Study Using NVIDIA Profiling

5.1 Output

Batch 16

```
pc3427@log-burst:~  
Train Epoch: 1 [59040/60000 (98%)] Loss: 0.106755  
Train Epoch: 1 [59200/60000 (99%)] Loss: 0.084940  
Train Epoch: 1 [59360/60000 (99%)] Loss: 0.004357  
Train Epoch: 1 [59520/60000 (99%)] Loss: 0.010249  
Train Epoch: 1 [59680/60000 (99%)] Loss: 0.001311  
Train Epoch: 1 [59840/60000 (100%)] Loss: 0.058322  
Average time per batch for conv2: 0.0003705306088000002 seconds  
==PROF== Target process 125038 terminated before first instrumented API call.  
pc3427 1  
==PROF== Profiling "computeOffsetsKernel" - 1: 0%...50%...100% - 1 pass  
==PROF== Profiling "volta_scudnn_128x64_relu_xreg..." - 2: 0%...50%...100% - 1 pass  
==PROF== Profiling "elementwise_kernel" - 3: 0%...50%...100% - 1 pass  
pc3427 2  
==PROF== Target process 135122 terminated before first instrumented API call.  
==PROF== Disconnected from process 123060  
[123060] python3.9@127.0.0.1  
void cask_cudnn::computeOffsetsKernel<(bool)0, (bool)0>(cask_cudnn::ComputeOffsetsParams), 2024-Mar-25 12:32:21, Context 1, Stream 7  
Section: Command line profiler metrics  
-----  
smssp_sass_thread_inst_executed_op_fadd_pred_on.avg          inst          0  
smssp_sass_thread_inst_executed_op_fadd_pred_on.max          inst          0  
smssp_sass_thread_inst_executed_op_fadd_pred_on.min          inst          0  
smssp_sass_thread_inst_executed_op_fadd_pred_on.sum          inst          0  
-----  
volta_scudnn_128x64_relu_xregs_large_nn_v1, 2024-Mar-25 12:32:22, Context 1, Stream 7  
Section: Command line profiler metrics  
-----  
smssp_sass_thread_inst_executed_op_fadd_pred_on.avg          inst          115200  
smssp_sass_thread_inst_executed_op_fadd_pred_on.max          inst          120832  
smssp_sass_thread_inst_executed_op_fadd_pred_on.min          inst          110592  
smssp_sass_thread_inst_executed_op_fadd_pred_on.sum          inst          36864000  
-----  
void at::native::elementwise_kernel<(int)128, (int)2, void at::native::gpu_kernel_impl<at::native::CUDataFuctor_add<Float>>(at::TensorIteratorBase &, const  
T1 &):[lambda(int) (instance 1)]>(int, T3), 2024-Mar-25 12:32:22, Context 1, Stream 7  
Section: Command line profiler metrics  
-----  
smssp_sass_thread_inst_executed_op_fadd_pred_on.avg          inst          0  
smssp_sass_thread_inst_executed_op_fadd_pred_on.max          inst          0  
smssp_sass_thread_inst_executed_op_fadd_pred_on.min          inst          0  
smssp_sass_thread_inst_executed_op_fadd_pred_on.sum          inst          0  
-----  
Singularity>
```


Batch 32

```
pc3427@log-burst- X + v
Train Epoch: 1 [58240/60000 (97%)] Loss: 0.096107
Train Epoch: 1 [58560/60000 (98%)] Loss: 0.038584
Train Epoch: 1 [58880/60000 (98%)] Loss: 0.186977
Train Epoch: 1 [59200/60000 (99%)] Loss: 0.096600
Train Epoch: 1 [59520/60000 (99%)] Loss: 0.076027
Train Epoch: 1 [59840/60000 (100%)] Loss: 0.075420
==PROF== Target process 190669 terminated before first instrumented API call.
Average time per batch for conv2: 0.000382708134933332 seconds
pc3427 1
==PROF== Profiling "computeOffsetsKernel" - 1: 0%...50%...100% - 1 pass
==PROF== Profiling "volta_scudnn_128x64_relu_xreg..." - 2: 0%...50%...100% - 1 pass
==PROF== Profiling "elementwise_kernel" - 3: 0%...50%...100% - 1 pass
pc3427 2
==PROF== Target process 196156 terminated before first instrumented API call.
==PROF== Disconnected from process 188853
[188853] python3.9@127.0.0.1
void cask_cudnn::computeOffsetsKernel<(bool)0, (bool)0>(cask_cudnn::ComputeOffsetsParams), 2024-Mar-25 12:37:01, Context 1, Stream 7
Section: Command line profiler metrics
-----
smsp_sass_thread_inst_executed_op_fadd_pred_on.avg          inst          0
smsp_sass_thread_inst_executed_op_fadd_pred_on.max          inst          0
smsp_sass_thread_inst_executed_op_fadd_pred_on.min          inst          0
smsp_sass_thread_inst_executed_op_fadd_pred_on.sum          inst          0
-----

volta_scudnn_128x64_relu_xregs_large_nn_v1, 2024-Mar-25 12:37:01, Context 1, Stream 7
Section: Command line profiler metrics
-----
smsp_sass_thread_inst_executed_op_fadd_pred_on.avg          inst          115200
smsp_sass_thread_inst_executed_op_fadd_pred_on.max          inst          118784
smsp_sass_thread_inst_executed_op_fadd_pred_on.min          inst          110592
smsp_sass_thread_inst_executed_op_fadd_pred_on.sum          inst          36864000
-----

void at::native::elementwise_kernel<(int)128, (int)2, void at::native::gpu_kernel_impl<at::native::CUDataFuctor_add<float>>(at::TensorIteratorBase &, const
T1 &):[lambda(int) (instance 1)]>(int, T3), 2024-Mar-25 12:37:01, Context 1, Stream 7
Section: Command line profiler metrics
-----
smsp_sass_thread_inst_executed_op_fadd_pred_on.avg          inst          0
smsp_sass_thread_inst_executed_op_fadd_pred_on.max          inst          0
smsp_sass_thread_inst_executed_op_fadd_pred_on.min          inst          0
smsp_sass_thread_inst_executed_op_fadd_pred_on.sum          inst          0
-----

Singularity>
```

Batch 64

```
pc3427@log-burst- X + v
Train Epoch: 1 [57600/60000 (96%)] Loss: 0.189804
Train Epoch: 1 [58240/60000 (97%)] Loss: 0.063211
Train Epoch: 1 [58880/60000 (98%)] Loss: 0.113528
Train Epoch: 1 [59520/60000 (99%)] Loss: 0.028631
Average time per batch for conv2: 0.00037083920895522333 seconds
==PROF== Target process 46597 terminated before first instrumented API call.
pc3427 1
==PROF== Profiling "computeOffsetsKernel" - 1: 0%...50%...100% - 1 pass
==PROF== Profiling "volta_scudnn_128x64_relu_xreg..." - 2: 0%...50%...100% - 1 pass
==PROF== Profiling "elementwise_kernel" - 3: 0%...50%...100% - 1 pass
pc3427 2
==PROF== Target process 49446 terminated before first instrumented API call.
==PROF== Disconnected from process 44508
[44508] python3.9@127.0.0.1
void cask_cudnn::computeOffsetsKernel<(bool)0, (bool)0>(cask_cudnn::ComputeOffsetsParams), 2024-Mar-25 12:25:50, Context 1, Stream 7
Section: Command line profiler metrics
-----
smsp_sass_thread_inst_executed_op_fadd_pred_on.avg          inst          0
smsp_sass_thread_inst_executed_op_fadd_pred_on.max          inst          0
smsp_sass_thread_inst_executed_op_fadd_pred_on.min          inst          0
smsp_sass_thread_inst_executed_op_fadd_pred_on.sum          inst          0
-----

volta_scudnn_128x64_relu_xregs_large_nn_v1, 2024-Mar-25 12:25:51, Context 1, Stream 7
Section: Command line profiler metrics
-----
smsp_sass_thread_inst_executed_op_fadd_pred_on.avg          inst          115200
smsp_sass_thread_inst_executed_op_fadd_pred_on.max          inst          120832
smsp_sass_thread_inst_executed_op_fadd_pred_on.min          inst          108544
smsp_sass_thread_inst_executed_op_fadd_pred_on.sum          inst          36864000
-----

void at::native::elementwise_kernel<(int)128, (int)2, void at::native::gpu_kernel_impl<at::native::CUDataFuctor_add<float>>(at::TensorIteratorBase &, const
T1 &):[lambda(int) (instance 1)]>(int, T3), 2024-Mar-25 12:25:51, Context 1, Stream 7
Section: Command line profiler metrics
-----
smsp_sass_thread_inst_executed_op_fadd_pred_on.avg          inst          0
smsp_sass_thread_inst_executed_op_fadd_pred_on.max          inst          0
smsp_sass_thread_inst_executed_op_fadd_pred_on.min          inst          0
smsp_sass_thread_inst_executed_op_fadd_pred_on.sum          inst          0
-----

Singularity>
```

Batch 128

```
pc3427@log-burst:~  
Train Epoch: 1 [52480/60000 (87%)] Loss: 0.121056  
Train Epoch: 1 [53760/60000 (90%)] Loss: 0.099359  
Train Epoch: 1 [55040/60000 (92%)] Loss: 0.160793  
Train Epoch: 1 [56320/60000 (94%)] Loss: 0.149049  
Train Epoch: 1 [57600/60000 (96%)] Loss: 0.221626  
Train Epoch: 1 [58880/60000 (98%)] Loss: 0.144202  
Average time per batch for conv2: 0.0003538357889125798 seconds  
==PROF== Target process 23762 terminated before first instrumented API call.  
pc3427 1  
==PROF== Profiling "computeOffsetsKernel" - 1: 0%...50%...100% - 1 pass  
==PROF== Profiling "volta_scudnn_128x64_relu_xreg..." - 2: 0%...50%...100% - 1 pass  
==PROF== Profiling "elementwise_kernel" - 3: 0%...50%...100% - 1 pass  
pc3427 2  
==PROF== Target process 26716 terminated before first instrumented API call.  
==PROF== Disconnected from process 14551  
[14551] python3.9@127.0.0.1  
void cask_cudnn::computeOffsetsKernel<(bool)0, (bool)0>(cask_cudnn::ComputeOffsetsParams), 2024-Mar-25 12:43:03, Context 1, Stream 7  
Section: Command line profiler metrics  
-----  
smsp__sass_thread_inst_executed_op_fadd_pred_on.avg inst 0  
smsp__sass_thread_inst_executed_op_fadd_pred_on.max inst 0  
smsp__sass_thread_inst_executed_op_fadd_pred_on.min inst 0  
smsp__sass_thread_inst_executed_op_fadd_pred_on.sum inst 0  
-----  
volta_scudnn_128x64_relu_xregs_large_nn_v1, 2024-Mar-25 12:43:03, Context 1, Stream 7  
Section: Command line profiler metrics  
-----  
smsp__sass_thread_inst_executed_op_fadd_pred_on.avg inst 115200  
smsp__sass_thread_inst_executed_op_fadd_pred_on.max inst 120832  
smsp__sass_thread_inst_executed_op_fadd_pred_on.min inst 108544  
smsp__sass_thread_inst_executed_op_fadd_pred_on.sum inst 36864000  
-----  
void at::native::elementwise_kernel<(int)128, (int)2, void at::native::gpu_kernel_impl<at::native::CUDAFunction_add<float>>(at::TensorIteratorBase &, const  
T1 &)::[lambda(int) (instance 1)]>(int, T3), 2024-Mar-25 12:43:03, Context 1, Stream 7  
Section: Command line profiler metrics  
-----  
smsp__sass_thread_inst_executed_op_fadd_pred_on.avg inst 0  
smsp__sass_thread_inst_executed_op_fadd_pred_on.max inst 0  
smsp__sass_thread_inst_executed_op_fadd_pred_on.min inst 0  
smsp__sass_thread_inst_executed_op_fadd_pred_on.sum inst 0  
-----  
Singularity>
```

Batch 256

```
pc3427@log-burst:~  
Train Epoch: 1 [46080/60000 (77%)] Loss: 0.153298  
Train Epoch: 1 [48640/60000 (81%)] Loss: 0.184827  
Train Epoch: 1 [51200/60000 (85%)] Loss: 0.123459  
Train Epoch: 1 [53760/60000 (89%)] Loss: 0.122601  
Train Epoch: 1 [56320/60000 (94%)] Loss: 0.199652  
Train Epoch: 1 [58880/60000 (98%)] Loss: 0.101914  
Average time per batch for conv2: 0.00036184534042553186 seconds  
==PROF== Target process 50696 terminated before first instrumented API call.  
pc3427 1  
==PROF== Profiling "computeOffsetsKernel" - 1: 0%...50%...100% - 1 pass  
==PROF== Profiling "volta_scudnn_128x64_relu_xreg..." - 2: 0%...50%...100% - 1 pass  
==PROF== Profiling "elementwise_kernel" - 3: 0%...50%...100% - 1 pass  
pc3427 2  
==PROF== Target process 53056 terminated before first instrumented API call.  
==PROF== Disconnected from process 48988  
[48988] python3.9@127.0.0.1  
void cask_cudnn::computeOffsetsKernel<(bool)0, (bool)0>(cask_cudnn::ComputeOffsetsParams), 2024-Mar-25 12:45:04, Context 1, Stream 7  
Section: Command line profiler metrics  
-----  
smsp__sass_thread_inst_executed_op_fadd_pred_on.avg inst 0  
smsp__sass_thread_inst_executed_op_fadd_pred_on.max inst 0  
smsp__sass_thread_inst_executed_op_fadd_pred_on.min inst 0  
smsp__sass_thread_inst_executed_op_fadd_pred_on.sum inst 0  
-----  
volta_scudnn_128x64_relu_xregs_large_nn_v1, 2024-Mar-25 12:45:04, Context 1, Stream 7  
Section: Command line profiler metrics  
-----  
smsp__sass_thread_inst_executed_op_fadd_pred_on.avg inst 115200  
smsp__sass_thread_inst_executed_op_fadd_pred_on.max inst 118784  
smsp__sass_thread_inst_executed_op_fadd_pred_on.min inst 108544  
smsp__sass_thread_inst_executed_op_fadd_pred_on.sum inst 36864000  
-----  
void at::native::elementwise_kernel<(int)128, (int)2, void at::native::gpu_kernel_impl<at::native::CUDAFunction_add<float>>(at::TensorIteratorBase &, const  
T1 &)::[lambda(int) (instance 1)]>(int, T3), 2024-Mar-25 12:45:04, Context 1, Stream 7  
Section: Command line profiler metrics  
-----  
smsp__sass_thread_inst_executed_op_fadd_pred_on.avg inst 0  
smsp__sass_thread_inst_executed_op_fadd_pred_on.max inst 0  
smsp__sass_thread_inst_executed_op_fadd_pred_on.min inst 0  
smsp__sass_thread_inst_executed_op_fadd_pred_on.sum inst 0  
-----  
Singularity>
```

6 Difference Between Estimated and Measured

6.1 Time Complexity:

- *Estimated:* The pen-and-paper method gave a theoretical time complexity of $O(141,557,760)$ for processing the Conv2d-2 layer.
- *Measured:* The profiling results provided actual execution times, which differ due to factors like hardware efficiency, parallelization, and real-time system load. Average time per batch for conv2: 0.0001746069573560767 seconds

6.2 Space Complexity:

- *Estimated:* Calculations predicted a space complexity of $O(514,112)$ considering weights, biases, and output feature maps.
- *Measured:* Memory profiling revealed the actual memory usage, which are based on the system's memory management, overheads, and potential optimizations employed by the deep learning framework.
- Memory Usage after training: Allocated: 0.02 GB Reserved: 0.62 GB
CPU RAM Free: 84.1 GB
GPU: Memory Free: 37876.0MB/40960.6MB | Utilization: 6.01318359375

7 Bonus diff shape/dimension of Conv2d-2 Layer

```
self.conv2 = nn.Conv2d(32, 128, 5, 1)
```

```
pc3427@log-burst:~  
Singularity> vi main.py  
Singularity> python3 main.py  
-----  
Layer (type)          Output Shape          Param #  
-----  
Conv2d-1              [-1, 32, 26, 26]      320  
Conv2d-2              [-1, 128, 22, 22]     102,528  
Dropout-3             [-1, 128, 11, 11]     0  
Linear-4              [-1, 128]             1,982,592  
Dropout-5             [-1, 128]             0  
Linear-6              [-1, 10]              1,290  
-----  
Total params: 2,086,730  
Trainable params: 2,086,730  
Non-trainable params: 0  
-----  
Input size (MB): 0.00  
Forward/backward pass size (MB): 0.76  
Params size (MB): 7.96  
Estimated Total Size (MB): 8.72  
-----  
Memory Usage before training:  
Allocated: 0.01 GB  
Reserved: 0.06 GB  
CPU RAM Free: 59.1 GB  
GPU 0: Memory Free: 14859.0MB / 16384.0MB | Utilization: 7.879638671875%  
GPU 1: Memory Free: 16147.0MB / 16384.0MB | Utilization: 0.018310546875%  
Train Epoch: 1 [0/60000 (0%)] Loss: 2.286213  
Train Epoch: 1 [640/60000 (1%)] Loss: 1.537523  
Train Epoch: 1 [1280/60000 (2%)] Loss: 0.606463  
Train Epoch: 1 [1920/60000 (3%)] Loss: 0.580142  
Train Epoch: 1 [2560/60000 (4%)] Loss: 0.412750  
Train Epoch: 1 [3200/60000 (5%)] Loss: 0.339578  
Train Epoch: 1 [3840/60000 (6%)] Loss: 0.308050
```



```
pc3427@log-burst:~ × + ∨

Train Epoch: 14 [38400/60000 (64%)] Loss: 0.006117
Train Epoch: 14 [39040/60000 (65%)] Loss: 0.001814
Train Epoch: 14 [39680/60000 (66%)] Loss: 0.028619
Train Epoch: 14 [40320/60000 (67%)] Loss: 0.039198
Train Epoch: 14 [40960/60000 (68%)] Loss: 0.004925
Train Epoch: 14 [41600/60000 (69%)] Loss: 0.001715
Train Epoch: 14 [42240/60000 (70%)] Loss: 0.002213
Train Epoch: 14 [42880/60000 (71%)] Loss: 0.001605
Train Epoch: 14 [43520/60000 (72%)] Loss: 0.005828
Train Epoch: 14 [44160/60000 (74%)] Loss: 0.152756
Train Epoch: 14 [44800/60000 (75%)] Loss: 0.018037
Train Epoch: 14 [45440/60000 (76%)] Loss: 0.036829
Train Epoch: 14 [46080/60000 (77%)] Loss: 0.001293
Train Epoch: 14 [46720/60000 (78%)] Loss: 0.003019
Train Epoch: 14 [47360/60000 (79%)] Loss: 0.194374
Train Epoch: 14 [48000/60000 (80%)] Loss: 0.005047
Train Epoch: 14 [48640/60000 (81%)] Loss: 0.010128
Train Epoch: 14 [49280/60000 (82%)] Loss: 0.026868
Train Epoch: 14 [49920/60000 (83%)] Loss: 0.024038
Train Epoch: 14 [50560/60000 (84%)] Loss: 0.001583
Train Epoch: 14 [51200/60000 (85%)] Loss: 0.005513
Train Epoch: 14 [51840/60000 (86%)] Loss: 0.002177
Train Epoch: 14 [52480/60000 (87%)] Loss: 0.003087
Train Epoch: 14 [53120/60000 (88%)] Loss: 0.009407
Train Epoch: 14 [53760/60000 (90%)] Loss: 0.028581
Train Epoch: 14 [54400/60000 (91%)] Loss: 0.094591
Train Epoch: 14 [55040/60000 (92%)] Loss: 0.006442
Train Epoch: 14 [55680/60000 (93%)] Loss: 0.004611
Train Epoch: 14 [56320/60000 (94%)] Loss: 0.005151
Train Epoch: 14 [56960/60000 (95%)] Loss: 0.000920
Train Epoch: 14 [57600/60000 (96%)] Loss: 0.017100
Train Epoch: 14 [58240/60000 (97%)] Loss: 0.000457
Train Epoch: 14 [58880/60000 (98%)] Loss: 0.022846
Train Epoch: 14 [59520/60000 (99%)] Loss: 0.001342
Average time per batch for conv2: 0.00013037963208359187 seconds

Test set: Average loss: 0.0248, Accuracy: 9933/10000 (99%)

Memory Usage after training:
Allocated: 0.03 GB
Reserved: 1.56 GB
CPU RAM Free: 58.9 GB
GPU 0: Memory Free: 13283.0MB / 16384.0MB | Utilization: 17.498779296875%
GPU 1: Memory Free: 16147.0MB / 16384.0MB | Utilization: 0.018310546875%
Singularity>
```

8 Summary

The comprehensive analysis of the convolutional neural network (CNN) dedicated to MNIST classification demonstrates distinct behaviour in time complexity and memory usage across varying batch sizes. My assessment compares theoretical estimations with measured data using tools such as nsight, providing valuable insights into the relationship between batch size and computational performance. Such an evaluation is instrumental in optimizing CNN training within constrained hardware environments, ensuring effective resource utilization for the neural network's operational parameters.