

## C1. Entire code in the .tar.gz archive (on Brightspace)

---

## C2.

```
bash-5.1$ vi resnet.py
bash-5.1$ singularity exec --nv --overlay overlay-15GB-500K.ext3:rw /share/apps/images/cuda12.1.1-cudnn8.9.0-devel-ubuntu22.04.2.sif /bin/bash
Singularity> python3 resnet.py --use_cuda --data_path ./data --num_workers 2 --optimizer sgd
Using device: cuda
Files already downloaded and verified
Using Optimizer: sgd

Epoch 1/5
Total Data Loading Time for Epoch 1: 0.527 seconds
Training Time for Epoch 1: 12.956 seconds
Total Running Time for Epoch 1: 14.014 seconds
Epoch 1: Loss: 1.937, Top-1 Accuracy: 29.93%

Epoch 2/5
Total Data Loading Time for Epoch 2: 0.517 seconds
Training Time for Epoch 2: 11.166 seconds
Total Running Time for Epoch 2: 12.177 seconds
Epoch 2: Loss: 1.457, Top-1 Accuracy: 46.05%

Epoch 3/5
Total Data Loading Time for Epoch 3: 0.521 seconds
Training Time for Epoch 3: 11.149 seconds
Total Running Time for Epoch 3: 12.154 seconds
Epoch 3: Loss: 1.202, Top-1 Accuracy: 56.45%

Epoch 4/5
Total Data Loading Time for Epoch 4: 0.540 seconds
Training Time for Epoch 4: 11.151 seconds
Total Running Time for Epoch 4: 12.185 seconds
Epoch 4: Loss: 1.006, Top-1 Accuracy: 64.19%

Epoch 5/5
Total Data Loading Time for Epoch 5: 0.533 seconds
Training Time for Epoch 5: 11.154 seconds
Total Running Time for Epoch 5: 12.174 seconds
Epoch 5: Loss: 0.857, Top-1 Accuracy: 69.62%

Total Data Loading Time over five epochs: 2.638 seconds
Average Training Time over 5 epochs: 11.515 seconds
Average Loss over 5 epochs: 1.292
Average Top-1 Accuracy over 5 epochs: 53.25%
Singularity>
Singularity> █
```

---

## C3.1

```
import matplotlib.pyplot as plt

# Data
workers = [0, 4, 8, 12, 16]
times = [83.267, 2.715, 2.654, 2.680, 2.727]

# Create the bar plot
```

```

plt.figure(figsize=(10, 6))
plt.bar(workers, times)

# Customize the plot
plt.title('Worker Performance')
plt.xlabel('Number of Workers')
plt.ylabel('Time (seconds)')

# Set x-axis ticks to match exactly with the worker numbers
plt.xticks(workers)

# Add value labels on top of each bar
for i, v in enumerate(times):
    plt.text(workers[i], v, f'{v:.3f}', ha='center', va='bottom')

# Find the best performance
best_time = min(times)
best_workers = [workers[i] for i, time in enumerate(times) if time == best_time]

# Add text annotation for best performance
plt.text(0.5, 0.95, f'Best performance: {best_time:.3f} seconds\nNumber of workers: {best_workers}', transform=plt.gca().transAxes, ha='center', va='top', bbox=dict(facecolor='white', alpha=0.8))

# Show the plot
plt.tight_layout()
plt.show()

# Print the best performance to the console
print(f'Best performance: {best_time:.3f} seconds')
print(f'Number of workers: {best_workers}')

```

Number of Workers	Total Time spent on Dataloader over 5 epochs
0	83.267
4	2.715
8	2.654
12	2.680
16	2.727

## Number of workers: 0

```
gn3228@gn3228: ~$ srun --account=ece_gy_9143-2024fa --partition=n1c24m128-v100-4 --gres=gpu:4 --pty /bin/bash
Last login: Mon Oct 7 23:06:34 2024 from 10.47.6.4
gn3228@gn3228: ~$ srun --account=ece_gy_9143-2024fa --partition=n1c24m128-v100-4 --gres=gpu:4 --pty /bin/bash
bash-5.1$ cd Tab2
bash-5.1$ ls
data overlay-15GB-500K.ext3 resnet2.py resnet.sh resnet.py total_loading_time_vs_workers.png
bash-5.1$ vi resnet.py
bash-5.1$ singularity exec --nv --overlay overlay-15GB-500K.ext3:rw ./share/applications/cudnn12.1.1-cudnn8.9.0-devel-ubuntu22.04.2.sif /bin/bash
Singularity> python3 resnet.py --use_cuda ./data --num_workers 0 --optimizer sgd
Using device: cuda
Files already downloaded and verified

Epoch 1/5
Total Data Loading Time for Epoch 1: 16.806 seconds
Training Time for Epoch 1: 15.180 seconds
Total Running Time for Epoch 1: 32.445 seconds
Epoch 1: Loss: 2.089, Top-1 Accuracy: 27.70%

Epoch 2/5
Total Data Loading Time for Epoch 2: 16.711 seconds
Training Time for Epoch 2: 11.142 seconds
Total Running Time for Epoch 2: 28.141 seconds
Epoch 2: Loss: 1.549, Top-1 Accuracy: 43.10%

Epoch 3/5
Total Data Loading Time for Epoch 3: 16.668 seconds
Training Time for Epoch 3: 11.140 seconds
Total Running Time for Epoch 3: 28.096 seconds
Epoch 3: Loss: 1.318, Top-1 Accuracy: 51.79%

Epoch 4/5
Total Data Loading Time for Epoch 4: 16.614 seconds
Training Time for Epoch 4: 11.142 seconds
Total Running Time for Epoch 4: 28.042 seconds
Epoch 4: Loss: 1.074, Top-1 Accuracy: 61.60%

Epoch 5/5
Total Data Loading Time for Epoch 5: 16.466 seconds
Training Time for Epoch 5: 11.141 seconds
Total Running Time for Epoch 5: 27.893 seconds
Epoch 5: Loss: 0.883, Top-1 Accuracy: 68.75%
Total Data Loading Time over five epochs: 83.267 seconds
Singularity> python3 resnet.py --use_cuda ./data --num_workers 4 --optimizer sgd
Using device: cuda
Files already downloaded and verified

Epoch 1/5
|
```

## Number of Workers: 4

```
Singularity> python3 resnet.py --use_cuda --data_path ./data --num_workers 4 --optimizer sgd
Using device: cuda
Files already downloaded and verified

Epoch 1/5
Total Data Loading Time for Epoch 1: 0.647 seconds
Training Time for Epoch 1: 12.963 seconds
Total Running Time for Epoch 1: 14.226 seconds
Epoch 1: Loss: 2.246, Top-1 Accuracy: 24.80%

Epoch 2/5
Total Data Loading Time for Epoch 2: 0.502 seconds
Training Time for Epoch 2: 11.149 seconds
Total Running Time for Epoch 2: 12.175 seconds
Epoch 2: Loss: 1.613, Top-1 Accuracy: 40.24%

Epoch 3/5
Total Data Loading Time for Epoch 3: 0.539 seconds
Training Time for Epoch 3: 11.156 seconds
Total Running Time for Epoch 3: 12.224 seconds
Epoch 3: Loss: 1.373, Top-1 Accuracy: 49.66%

Epoch 4/5
Total Data Loading Time for Epoch 4: 0.504 seconds
Training Time for Epoch 4: 11.160 seconds
Total Running Time for Epoch 4: 12.211 seconds
Epoch 4: Loss: 1.135, Top-1 Accuracy: 59.20%

Epoch 5/5
Total Data Loading Time for Epoch 5: 0.523 seconds
Training Time for Epoch 5: 11.158 seconds
Total Running Time for Epoch 5: 12.201 seconds
Epoch 5: Loss: 0.985, Top-1 Accuracy: 64.95%
Total Data Loading Time over five epochs: 2.715 seconds
Singularity> █
```

## Number of Workers: 8

```
Singularity> python3 resnet.py --use_cuda --data_path ./data --num_workers 8 --optimizer sgd
Using device: cuda
Files already downloaded and verified

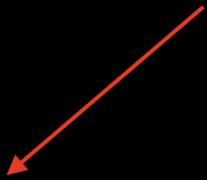
Epoch 1/5
Total Data Loading Time for Epoch 1: 0.555 seconds
Training Time for Epoch 1: 12.939 seconds
Total Running Time for Epoch 1: 14.097 seconds
Epoch 1: Loss: 2.103, Top-1 Accuracy: 25.69%

Epoch 2/5
Total Data Loading Time for Epoch 2: 0.528 seconds
Training Time for Epoch 2: 11.168 seconds
Total Running Time for Epoch 2: 12.269 seconds
Epoch 2: Loss: 1.525, Top-1 Accuracy: 43.41%

Epoch 3/5
Total Data Loading Time for Epoch 3: 0.511 seconds
Training Time for Epoch 3: 11.163 seconds
Total Running Time for Epoch 3: 12.271 seconds
Epoch 3: Loss: 1.272, Top-1 Accuracy: 53.57%

Epoch 4/5
Total Data Loading Time for Epoch 4: 0.520 seconds
Training Time for Epoch 4: 11.160 seconds
Total Running Time for Epoch 4: 12.272 seconds
Epoch 4: Loss: 1.056, Top-1 Accuracy: 61.98%

Epoch 5/5
Total Data Loading Time for Epoch 5: 0.541 seconds
Training Time for Epoch 5: 11.164 seconds
Total Running Time for Epoch 5: 12.321 seconds
Epoch 5: Loss: 0.894, Top-1 Accuracy: 68.30%
Total Data Loading Time over five epochs: 2.654 seconds
```



## Number of Workers: 12

```
Singularity> python3 resnet.py --use_cuda --data_path ./data --num_workers 12 --optimizer sgd
Using device: cuda
Files already downloaded and verified

Epoch 1/5
Total Data Loading Time for Epoch 1: 0.558 seconds
Training Time for Epoch 1: 12.963 seconds
Total Running Time for Epoch 1: 14.124 seconds
Epoch 1: Loss: 1.896, Top-1 Accuracy: 31.21%

Epoch 2/5
Total Data Loading Time for Epoch 2: 0.545 seconds
Training Time for Epoch 2: 11.174 seconds
Total Running Time for Epoch 2: 12.385 seconds
Epoch 2: Loss: 1.428, Top-1 Accuracy: 47.90%

Epoch 3/5
Total Data Loading Time for Epoch 3: 0.536 seconds
Training Time for Epoch 3: 11.158 seconds
Total Running Time for Epoch 3: 12.331 seconds
Epoch 3: Loss: 1.158, Top-1 Accuracy: 58.27%

Epoch 4/5
Total Data Loading Time for Epoch 4: 0.507 seconds
Training Time for Epoch 4: 11.164 seconds
Total Running Time for Epoch 4: 12.339 seconds
Epoch 4: Loss: 0.953, Top-1 Accuracy: 66.02%

Epoch 5/5
Total Data Loading Time for Epoch 5: 0.535 seconds
Training Time for Epoch 5: 11.167 seconds
Total Running Time for Epoch 5: 12.365 seconds
Epoch 5: Loss: 0.805, Top-1 Accuracy: 71.47%
Total Data Loading Time over five epochs: 2.680 seconds
Singularity> █
```

## Number of Workers: 16

```
Singularity> python3 resnet.py --use_cuda --data_path ./data --num_workers 16 --optimizer sgd
Using device: cuda
Files already downloaded and verified

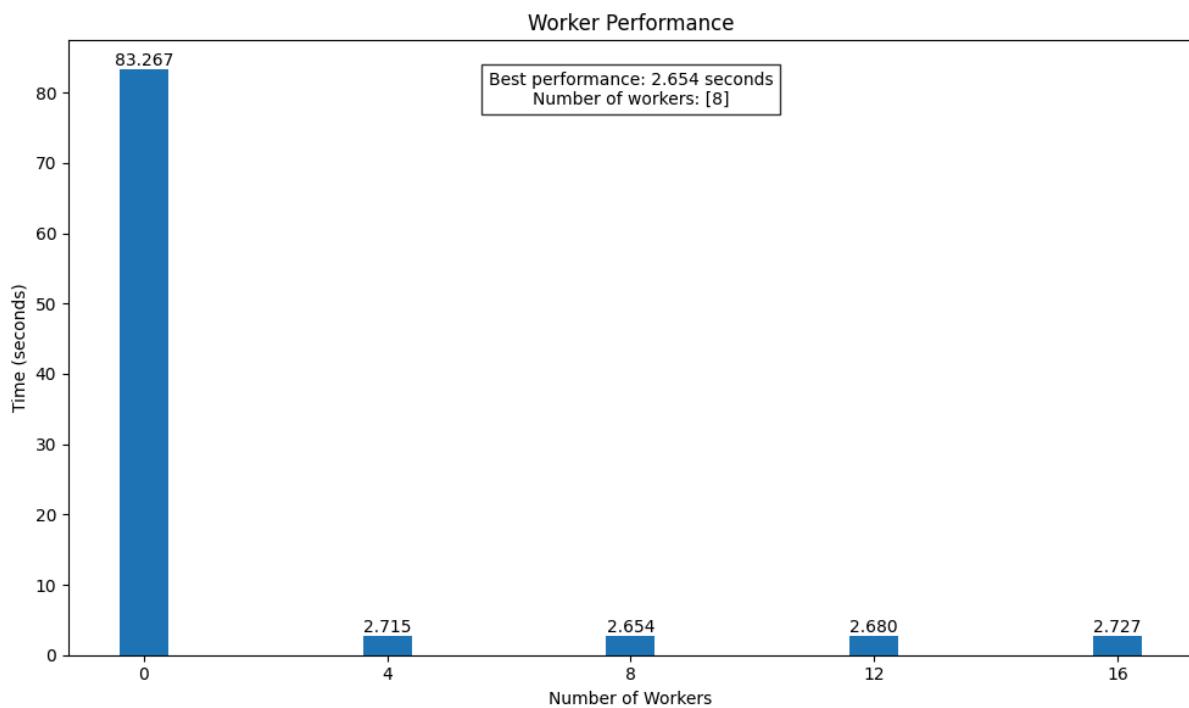
Epoch 1/5
Total Data Loading Time for Epoch 1: 0.580 seconds
Training Time for Epoch 1: 12.930 seconds
Total Running Time for Epoch 1: 14.185 seconds
Epoch 1: Loss: 2.096, Top-1 Accuracy: 26.42%

Epoch 2/5
Total Data Loading Time for Epoch 2: 0.537 seconds
Training Time for Epoch 2: 11.183 seconds
Total Running Time for Epoch 2: 12.466 seconds
Epoch 2: Loss: 1.637, Top-1 Accuracy: 39.35%

Epoch 3/5
Total Data Loading Time for Epoch 3: 0.539 seconds
Training Time for Epoch 3: 11.187 seconds
Total Running Time for Epoch 3: 12.477 seconds
Epoch 3: Loss: 1.426, Top-1 Accuracy: 47.47%

Epoch 4/5
Total Data Loading Time for Epoch 4: 0.521 seconds
Training Time for Epoch 4: 11.175 seconds
Total Running Time for Epoch 4: 12.447 seconds
Epoch 4: Loss: 1.209, Top-1 Accuracy: 56.53%

Epoch 5/5
Total Data Loading Time for Epoch 5: 0.551 seconds
Training Time for Epoch 5: 11.184 seconds
Total Running Time for Epoch 5: 12.479 seconds
Epoch 5: Loss: 1.011, Top-1 Accuracy: 63.86%
Total Data Loading Time over five epochs: 2.727 seconds
Singularity> █
```



## C3.2

Best performance achieved by 8 number of workers

---

## C4: Profiling starting from code in C3

### Number of Workers: 1

```
bash-5.1$ singularity exec --nv --overlay overlay-15GB-500K_ext3:rw /share/apps/images/cuda12.1.1-cudnn8.9.0-devel-ubuntu22.04.2.sif /bin/bash
Singularity> python3 resnet.py --use_cuda --data_path ./data --num_workers 1 --optimizer sgd
Using device: cuda
Files already downloaded and verified

Epoch 1/5
Total Data Loading Time for Epoch 1: 6.664 seconds
Training Time for Epoch 1: 15.366 seconds
Total Running Time for Epoch 1: 22.661 seconds
Epoch 1: Loss: 2.004, Top-1 Accuracy: 27.99%

Epoch 2/5
Total Data Loading Time for Epoch 2: 6.541 seconds
Training Time for Epoch 2: 11.172 seconds
Total Running Time for Epoch 2: 18.145 seconds
Epoch 2: Loss: 1.449, Top-1 Accuracy: 46.43%

Epoch 3/5
Total Data Loading Time for Epoch 3: 6.811 seconds
Training Time for Epoch 3: 11.178 seconds
Total Running Time for Epoch 3: 18.436 seconds
Epoch 3: Loss: 1.196, Top-1 Accuracy: 56.89%

Epoch 4/5
Total Data Loading Time for Epoch 4: 6.740 seconds
Training Time for Epoch 4: 11.182 seconds
Total Running Time for Epoch 4: 18.367 seconds
Epoch 4: Loss: 1.012, Top-1 Accuracy: 64.04%

Epoch 5/5
Total Data Loading Time for Epoch 5: 6.688 seconds
Training Time for Epoch 5: 11.180 seconds
Total Running Time for Epoch 5: 18.321 seconds
Epoch 5: Loss: 0.892, Top-1 Accuracy: 68.45%
Total Data Loading Time over five epochs: 33.445 seconds
Singularity> █
```

## Number of Workers: 8

```
Singularity> python3 resnet.py --use_cuda --data_path ./data --num_workers 8 --optimizer sgd
Using device: cuda
Files already downloaded and verified

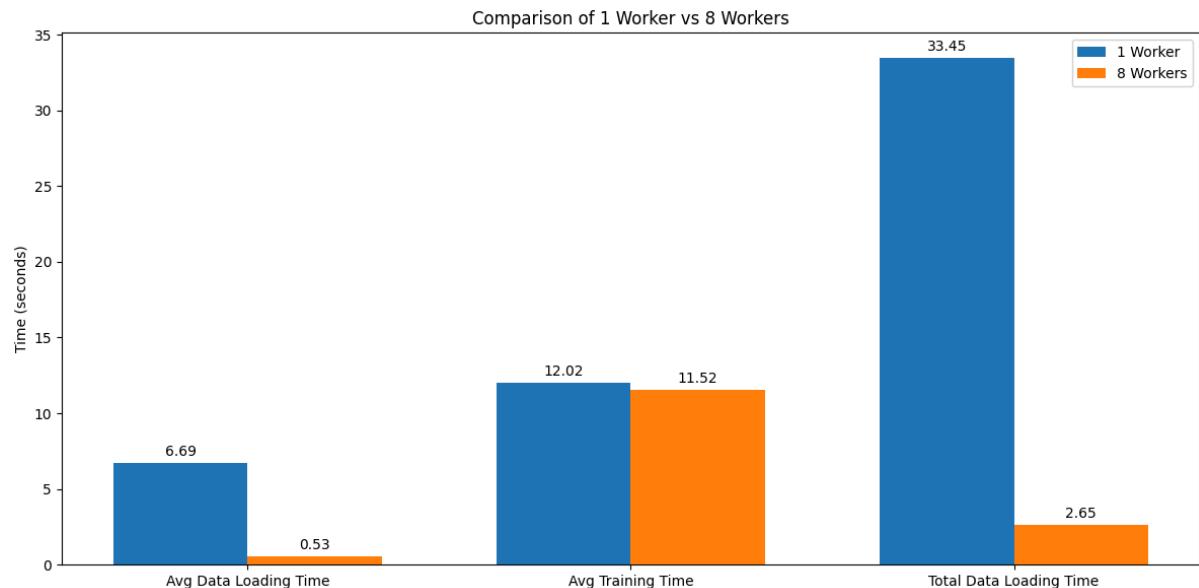
Epoch 1/5
Total Data Loading Time for Epoch 1: 0.555 seconds
Training Time for Epoch 1: 12.939 seconds
Total Running Time for Epoch 1: 14.097 seconds
Epoch 1: Loss: 2.103, Top-1 Accuracy: 25.69%

Epoch 2/5
Total Data Loading Time for Epoch 2: 0.528 seconds
Training Time for Epoch 2: 11.168 seconds
Total Running Time for Epoch 2: 12.269 seconds
Epoch 2: Loss: 1.525, Top-1 Accuracy: 43.41%

Epoch 3/5
Total Data Loading Time for Epoch 3: 0.511 seconds
Training Time for Epoch 3: 11.163 seconds
Total Running Time for Epoch 3: 12.271 seconds
Epoch 3: Loss: 1.272, Top-1 Accuracy: 53.57%

Epoch 4/5
Total Data Loading Time for Epoch 4: 0.520 seconds
Training Time for Epoch 4: 11.160 seconds
Total Running Time for Epoch 4: 12.272 seconds
Epoch 4: Loss: 1.056, Top-1 Accuracy: 61.98%

Epoch 5/5
Total Data Loading Time for Epoch 5: 0.541 seconds
Training Time for Epoch 5: 11.164 seconds
Total Running Time for Epoch 5: 12.321 seconds
Epoch 5: Loss: 0.894, Top-1 Accuracy: 68.30%
Total Data Loading Time over five epochs: 2.654 seconds
```



1. Average Data Loading Time:
  - 1 worker: 6.69 seconds

- 8 workers: 0.531 seconds
  - The 8-worker configuration is significantly faster (about 12.6 times) in data loading.
2. Average Training Time:
- 1 worker: 12.016 seconds
  - 8 workers: 11.519 seconds
  - The training times are quite similar, with the 8-worker configuration being slightly faster.
3. Total Data Loading Time over 5 epochs:
- 1 worker: 33.445 seconds
  - 8 workers: 2.654 seconds
  - Again, the 8-worker configuration shows a dramatic improvement in total data loading time.

Explanation of the differences:

1. Data Loading: The most significant difference is in data loading times. With 8 workers, the data loading process is parallelized, allowing for much faster loading of the dataset. This explains the dramatic reduction in both average and total data loading times.
2. Training Time: The training times are quite similar between the two configurations. The slight improvement with 8 workers could be due to better utilization of system resources.
3. Overall Efficiency: The 8-worker configuration is clearly more efficient, primarily due to the massive improvement in data loading times. This allows for much faster iteration through epochs, which can significantly reduce the overall time required for training a model over many epochs.

---

C5: Training in GPUs vs CPUs

## Device: CUDA

```
Singularity> python3 resnet.py --use_cuda --data_path ./data --num_workers 8 --optimizer sgd
Using device: cuda
Files already downloaded and verified

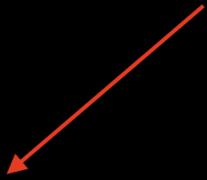
Epoch 1/5
Total Data Loading Time for Epoch 1: 0.555 seconds
Training Time for Epoch 1: 12.939 seconds
Total Running Time for Epoch 1: 14.097 seconds
Epoch 1: Loss: 2.103, Top-1 Accuracy: 25.69%

Epoch 2/5
Total Data Loading Time for Epoch 2: 0.528 seconds
Training Time for Epoch 2: 11.168 seconds
Total Running Time for Epoch 2: 12.269 seconds
Epoch 2: Loss: 1.525, Top-1 Accuracy: 43.41%

Epoch 3/5
Total Data Loading Time for Epoch 3: 0.511 seconds
Training Time for Epoch 3: 11.163 seconds
Total Running Time for Epoch 3: 12.271 seconds
Epoch 3: Loss: 1.272, Top-1 Accuracy: 53.57%

Epoch 4/5
Total Data Loading Time for Epoch 4: 0.520 seconds
Training Time for Epoch 4: 11.160 seconds
Total Running Time for Epoch 4: 12.272 seconds
Epoch 4: Loss: 1.056, Top-1 Accuracy: 61.98%

Epoch 5/5
Total Data Loading Time for Epoch 5: 0.541 seconds
Training Time for Epoch 5: 11.164 seconds
Total Running Time for Epoch 5: 12.321 seconds
Epoch 5: Loss: 0.894, Top-1 Accuracy: 68.30%
Total Data Loading Time over five epochs: 2.654 seconds
```



## Device: CPU

```
gm328     6811  0.0  0.0   3472  1728 pts/0   S+  15:48  0:00 grep python3
Singularity: python3 resnet.py --use_cpu --data_path ./data --num_workers 8 --optimizer sgd
Using device: cpu
Files already downloaded and verified
Using Optimizer: sgd

Epoch 1/5

Total Data Loading Time for Epoch 1: 0.835 seconds
Training Time for Epoch 1: 296.044 seconds
Total Running Time for Epoch 1: 297.111 seconds
Epoch 1: Loss: 1.960, Top-1 Accuracy: 29.76%

Epoch 2/5

Total Data Loading Time for Epoch 2: 0.802 seconds
Training Time for Epoch 2: 295.964 seconds
Total Running Time for Epoch 2: 297.018 seconds
Epoch 2: Loss: 1.465, Top-1 Accuracy: 45.77%

Epoch 3/5

Total Data Loading Time for Epoch 3: 0.781 seconds
Training Time for Epoch 3: 293.387 seconds
Total Running Time for Epoch 3: 294.480 seconds
Epoch 3: Loss: 1.225, Top-1 Accuracy: 55.63%

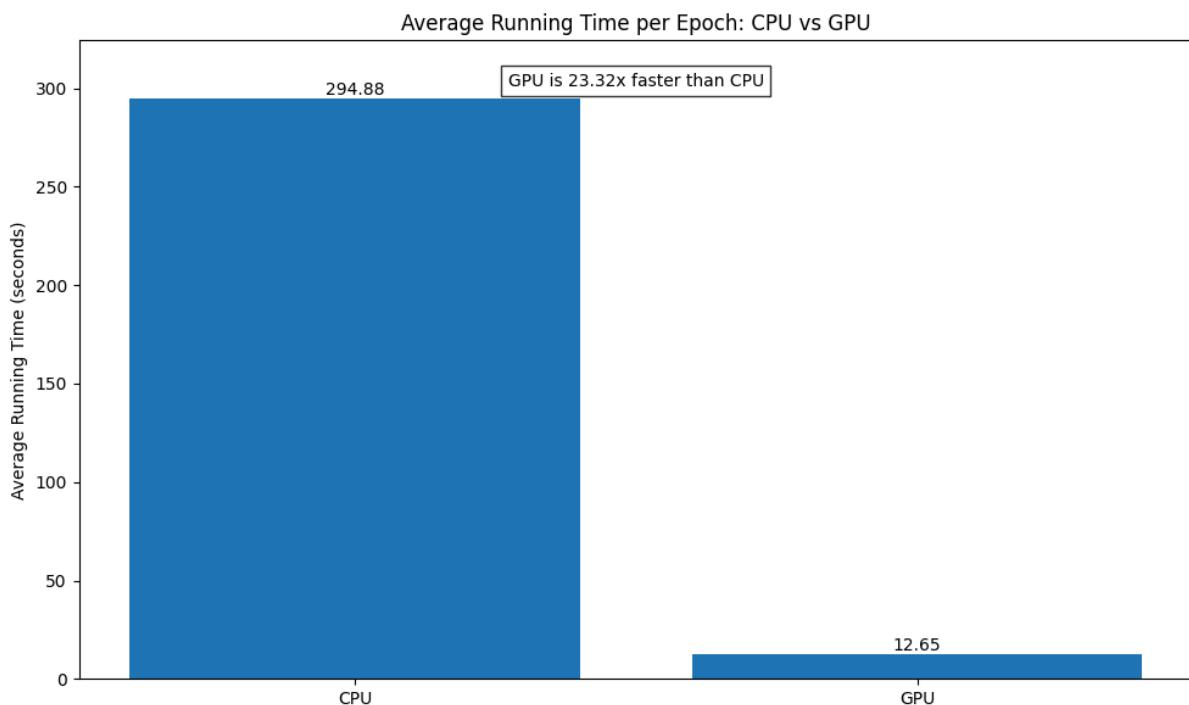
Epoch 4/5

Total Data Loading Time for Epoch 4: 0.768 seconds
Training Time for Epoch 4: 291.977 seconds
Total Running Time for Epoch 4: 293.099 seconds
Epoch 4: Loss: 1.018, Top-1 Accuracy: 63.98%

Epoch 5/5

Total Data Loading Time for Epoch 5: 0.760 seconds
Training Time for Epoch 5: 291.544 seconds
Total Running Time for Epoch 5: 292.673 seconds
Epoch 5: Loss: 0.829, Top-1 Accuracy: 70.93%

Total Data Loading Time over five epochs: 3.946 seconds
Average Training Time over 5 epochs: 293.783 seconds
Average Loss over 5 epochs: 1.299
Average Top-1 Accuracy over 5 epochs: 53.20%
Singularity>
```



## CPU Performance:

- Average running time per epoch: 294.88 seconds (approximately 4 minutes and 55 seconds)

## GPU Performance:

- Average running time per epoch: 12.65 seconds

## Comparison:

- The GPU is approximately 23.32 times faster than the CPU for this specific task.

---

## C6: Experimenting with different optimizers

### (i) SGD:

```
bash-5.1$ singularity exec --nv --overlay overlay-15GB-500K.ext3:rw /share/apps/images/cudnn12.1.1-cudnn8.9.0-devel-ubuntu22.04.2.sif /bin/bash
Singularity: python3 resnet.py --use_cuda --data_path ./data --num_workers 8 --optimizer sgd
Using device: cuda
Files already downloaded and verified

Epoch 1/5
Total Data Loading Time for Epoch 1: 0.641 seconds
Training Time for Epoch 1: 15.480 seconds
Total Running Time for Epoch 1: 16.948 seconds
Epoch 1: Loss: 1.874, Top-1 Accuracy: 51.75%

Epoch 2/5
Total Data Loading Time for Epoch 2: 0.509 seconds
Training Time for Epoch 2: 11.157 seconds
Total Running Time for Epoch 2: 12.239 seconds
Epoch 2: Loss: 1.427, Top-1 Accuracy: 47.42%

Epoch 3/5
Total Data Loading Time for Epoch 3: 0.535 seconds
Training Time for Epoch 3: 11.148 seconds
Total Running Time for Epoch 3: 12.265 seconds
Epoch 3: Loss: 1.161, Top-1 Accuracy: 58.04%

Epoch 4/5
Total Data Loading Time for Epoch 4: 0.541 seconds
Training Time for Epoch 4: 11.153 seconds
Total Running Time for Epoch 4: 12.280 seconds
Epoch 4: Loss: 0.954, Top-1 Accuracy: 66.22%

Epoch 5/5
Total Data Loading Time for Epoch 5: 0.554 seconds
Training Time for Epoch 5: 11.181 seconds
Total Running Time for Epoch 5: 12.385 seconds
Epoch 5: Loss: 0.792, Top-1 Accuracy: 72.27%

Total Data Loading Time over five epochs: 2.780 seconds
Average Training Time over 5 epochs: 12.024 seconds
Average Loss over 5 epochs: 1.242
Average Top-1 Accuracy over 5 epochs: 55.14%
Singularity> █
```

## (ii) SGD with nesterov:

```
bash-5.1$ singularity exec --nv --overlay overlay-15GB-500K.ext3:rw /share/apps/images/cuda12.1.1-cudnn8.9.0-devel-ubuntu22.04.2.sif /bin/bash
Singularity> python3 resnet.py --use_cuda --data_path ./data --num_workers 8 --optimizer sgd_nesterov
Using device: cuda
Files already downloaded and verified
Using Optimizer: sgd_nesterov

Epoch 1/5
Total Data Loading Time for Epoch 1: 0.550 seconds
Training Time for Epoch 1: 13.029 seconds
Total Running Time for Epoch 1: 14.181 seconds
Epoch 1: Loss: 1.947, Top-1 Accuracy: 32.50%

Epoch 2/5
Total Data Loading Time for Epoch 2: 0.513 seconds
Training Time for Epoch 2: 11.229 seconds
Total Running Time for Epoch 2: 12.353 seconds
Epoch 2: Loss: 1.410, Top-1 Accuracy: 48.86%

Epoch 3/5
Total Data Loading Time for Epoch 3: 0.533 seconds
Training Time for Epoch 3: 11.217 seconds
Total Running Time for Epoch 3: 12.341 seconds
Epoch 3: Loss: 1.125, Top-1 Accuracy: 59.88%

Epoch 4/5
Total Data Loading Time for Epoch 4: 0.547 seconds
Training Time for Epoch 4: 11.244 seconds
Total Running Time for Epoch 4: 12.432 seconds
Epoch 4: Loss: 0.895, Top-1 Accuracy: 68.32%

Epoch 5/5
Total Data Loading Time for Epoch 5: 0.555 seconds
Training Time for Epoch 5: 11.231 seconds
Total Running Time for Epoch 5: 12.383 seconds
Epoch 5: Loss: 0.722, Top-1 Accuracy: 74.87%

Total Data Loading Time over five epochs: 2.697 seconds
Average Training Time over 5 epochs: 11.590 seconds
Average Loss over 5 epochs: 1.220
Average Top-1 Accuracy over 5 epochs: 56.89%
Singularity>
```

## (iii) Adagrad:

```
Singularity> python3 resnet.py --use_cuda --data_path ./data --num_workers 8 --optimizer adagrad
Using device: cuda
Files already downloaded and verified
Using Optimizer: adagrad

Epoch 1/5
Total Data Loading Time for Epoch 1: 0.551 seconds
Training Time for Epoch 1: 13.299 seconds
Total Running Time for Epoch 1: 14.453 seconds
Epoch 1: Loss: 2.181, Top-1 Accuracy: 25.22%

Epoch 2/5
Total Data Loading Time for Epoch 2: 0.549 seconds
Training Time for Epoch 2: 11.293 seconds
Total Running Time for Epoch 2: 12.430 seconds
Epoch 2: Loss: 1.652, Top-1 Accuracy: 37.41%

Epoch 3/5
Total Data Loading Time for Epoch 3: 0.544 seconds
Training Time for Epoch 3: 11.289 seconds
Total Running Time for Epoch 3: 12.433 seconds
Epoch 3: Loss: 1.390, Top-1 Accuracy: 48.69%

Epoch 4/5
Total Data Loading Time for Epoch 4: 0.533 seconds
Training Time for Epoch 4: 11.292 seconds
Total Running Time for Epoch 4: 12.437 seconds
Epoch 4: Loss: 1.162, Top-1 Accuracy: 58.18%

Epoch 5/5
Total Data Loading Time for Epoch 5: 0.526 seconds
Training Time for Epoch 5: 11.290 seconds
Total Running Time for Epoch 5: 12.421 seconds
Epoch 5: Loss: 1.010, Top-1 Accuracy: 64.03%

Total Data Loading Time over five epochs: 2.704 seconds
Average Training Time over 5 epochs: 11.693 seconds
Average Loss over 5 epochs: 1.479
Average Top-1 Accuracy over 5 epochs: 46.71%
Singularity>
Singularity>
Singularity>
```

#### (iv) Adadelta:

```
Singularity>
Singularity> python3 resnet.py --use_cuda --data_path ./data --num_workers 8 --optimizer adadelta
Using device: cuda
Files already downloaded and verified
Using Optimizer: adadelta

Epoch 1/5
Total Data Loading Time for Epoch 1: 0.551 seconds
Training Time for Epoch 1: 13.443 seconds
Total Running Time for Epoch 1: 14.578 seconds
Epoch 1: Loss: 1.381, Top-1 Accuracy: 49.26%

Epoch 2/5
Total Data Loading Time for Epoch 2: 0.518 seconds
Training Time for Epoch 2: 11.622 seconds
Total Running Time for Epoch 2: 12.744 seconds
Epoch 2: Loss: 0.892, Top-1 Accuracy: 68.37%

Epoch 3/5
Total Data Loading Time for Epoch 3: 0.556 seconds
Training Time for Epoch 3: 11.605 seconds
Total Running Time for Epoch 3: 12.753 seconds
Epoch 3: Loss: 0.688, Top-1 Accuracy: 75.86%

Epoch 4/5
Total Data Loading Time for Epoch 4: 0.545 seconds
Training Time for Epoch 4: 11.602 seconds
Total Running Time for Epoch 4: 12.743 seconds
Epoch 4: Loss: 0.580, Top-1 Accuracy: 79.68%

Epoch 5/5
Total Data Loading Time for Epoch 5: 0.545 seconds
Training Time for Epoch 5: 11.604 seconds
Total Running Time for Epoch 5: 12.753 seconds
Epoch 5: Loss: 0.503, Top-1 Accuracy: 82.47%

Total Data Loading Time over five epochs: 2.716 seconds
Average Training Time over 5 epochs: 11.975 seconds
Average Loss over 5 epochs: 0.809
Average Top-1 Accuracy over 5 epochs: 71.13%
Singularity>
Singularity>
Singularity> █
```

## (v) Adam:

```
Singularity> python3 resnet.py --use_cuda --data_path ./data --num_workers 8 --optimizer adam
Using device: cuda
Files already downloaded and verified
Using Optimizer: adam

Epoch 1/5
Total Data Loading Time for Epoch 1: 0.550 seconds
Training Time for Epoch 1: 13.215 seconds
Total Running Time for Epoch 1: 14.351 seconds
Epoch 1: Loss: 1.459, Top-1 Accuracy: 46.60%

Epoch 2/5
Total Data Loading Time for Epoch 2: 0.531 seconds
Training Time for Epoch 2: 11.391 seconds
Total Running Time for Epoch 2: 12.515 seconds
Epoch 2: Loss: 1.031, Top-1 Accuracy: 63.14%

Epoch 3/5
Total Data Loading Time for Epoch 3: 0.539 seconds
Training Time for Epoch 3: 11.382 seconds
Total Running Time for Epoch 3: 12.524 seconds
Epoch 3: Loss: 0.869, Top-1 Accuracy: 69.35%

Epoch 4/5
Total Data Loading Time for Epoch 4: 0.507 seconds
Training Time for Epoch 4: 11.381 seconds
Total Running Time for Epoch 4: 12.473 seconds
Epoch 4: Loss: 0.747, Top-1 Accuracy: 74.19%

Epoch 5/5
Total Data Loading Time for Epoch 5: 0.530 seconds
Training Time for Epoch 5: 11.389 seconds
Total Running Time for Epoch 5: 12.507 seconds
Epoch 5: Loss: 0.659, Top-1 Accuracy: 77.35%
```

Total Data Loading Time over five epochs: 2.657 seconds  
Average Training Time over 5 epochs: 11.752 seconds  
Average Loss over 5 epochs: 0.953  
Average Top-1 Accuracy over 5 epochs: 66.13%

```
Singularity>
Singularity>
Singularity> █
```

## C7: Experimenting without Batch Norm

```
bash-5.1$ singularity exec --nv --overlay overlay-15GB-500K.ext3:rw /share/apps/images/cuda12.1.1-cudnn8.9.0-devel-ubuntu22.04.2.sif /bin/bash
Singularity> python3 resnet.py --use_cuda --data_path ./data --num_workers 8 --optimizer sgd
Using device: cuda
Files already downloaded and verified
Using Optimizer: sgd

Epoch 1/5
Total Data Loading Time for Epoch 1: 0.513 seconds
Training Time for Epoch 1: 11.391 seconds
Total Running Time for Epoch 1: 12.495 seconds
Epoch 1: Loss: 2.025, Top-1 Accuracy: 23.88%

Epoch 2/5
Total Data Loading Time for Epoch 2: 0.539 seconds
Training Time for Epoch 2: 9.665 seconds
Total Running Time for Epoch 2: 10.831 seconds
Epoch 2: Loss: 1.589, Top-1 Accuracy: 41.40%

Epoch 3/5
Total Data Loading Time for Epoch 3: 0.506 seconds
Training Time for Epoch 3: 9.653 seconds
Total Running Time for Epoch 3: 10.754 seconds
Epoch 3: Loss: 1.359, Top-1 Accuracy: 51.06%

Epoch 4/5
Total Data Loading Time for Epoch 4: 0.532 seconds
Training Time for Epoch 4: 9.661 seconds
Total Running Time for Epoch 4: 10.765 seconds
Epoch 4: Loss: 1.172, Top-1 Accuracy: 58.75%

Epoch 5/5
Total Data Loading Time for Epoch 5: 0.510 seconds
Training Time for Epoch 5: 9.656 seconds
Total Running Time for Epoch 5: 10.755 seconds
Epoch 5: Loss: 1.007, Top-1 Accuracy: 64.72%
```

Total Data Loading Time over five epochs: 2.599 seconds  
Average Training Time over 5 epochs: 10.005 seconds  
Average Loss over 5 epochs: 1.431  
Average Top-1 Accuracy over 5 epochs: 47.96%

```
Singularity>
Singularity>
```

---

Q1. How many convolutional layers are in the ResNet-18 model?

- 1 initial convolutional layer in the main ResNet class
- 4 layer groups, each containing 2 BasicBlocks
- Each BasicBlock has 2 convolutional layers

The total number of convolutional layers is:  $1 + (4 * 2 * 2) = 17$  convolutional layers

---

Q2. What is the input dimension of the last linear layer?

The input size of the last linear layer is 512. This comes from the final part of the model, which produces 512 outputs (or channels). These outputs are averaged by a layer that reduces everything to a simpler form before being sent into the final fully connected layer.

---

Q3.

Trainable Parameters: 11173962, Gradients: 11173962

```
def count_parameters(model):
    return sum(p.numel() for p in model.parameters() if p.requires_grad)

def count_gradients(model):
    return sum(p.numel() for p in model.parameters() if p.requires_grad and p.grad is
not None)

# In the main function, after initializing the model:
model = ResNet18().to(device)
num_params = count_parameters(model)
print(f"Number of trainable parameters: {num_params}")

# After training (at the end of the main function):
num_grads = count_gradients(model)
print(f"Number of gradients: {num_grads}")
```

Using optimizer: SGD

```
bash-5.1$ singularity exec --nv --overlay overlay-15GB-500K.ext3:rw /share/apps/images/cuda12.1.1-cudnn8.9.0-devel-ubuntu22.04.2.sif /bin/bash
Singularity: python3 resnet.py --use_cuda --data_path ./data --num_workers 2 --optimizer sgd
Using device: cuda
Files already downloaded and verified
Number of trainable parameters: 11173962 ←
Using Optimizer: sgd

Epoch 1/5
Total Data Loading Time for Epoch 1: 0.669 seconds
Training Time for Epoch 1: 15.886 seconds
Total Running Time for Epoch 1: 17.253 seconds
Epoch 1: Loss: 1.925, Top-1 Accuracy: 31.31%

Epoch 2/5
Total Data Loading Time for Epoch 2: 0.505 seconds
Training Time for Epoch 2: 11.201 seconds
Total Running Time for Epoch 2: 12.191 seconds
Epoch 2: Loss: 1.489, Top-1 Accuracy: 45.22%

Epoch 3/5
Total Data Loading Time for Epoch 3: 0.514 seconds
Training Time for Epoch 3: 11.207 seconds
Total Running Time for Epoch 3: 12.210 seconds
Epoch 3: Loss: 1.279, Top-1 Accuracy: 53.64%

Epoch 4/5
Total Data Loading Time for Epoch 4: 0.544 seconds
Training Time for Epoch 4: 11.215 seconds
Total Running Time for Epoch 4: 12.243 seconds
Epoch 4: Loss: 1.110, Top-1 Accuracy: 60.31%

Epoch 5/5
Total Data Loading Time for Epoch 5: 0.541 seconds
Training Time for Epoch 5: 11.221 seconds
Total Running Time for Epoch 5: 12.258 seconds
Epoch 5: Loss: 0.967, Top-1 Accuracy: 65.55%

Total Data Loading Time over five epochs: 2.773 seconds
Average Training Time over 5 epochs: 12.146 seconds
Average Loss over 5 epochs: 1.354
Average Top-1 Accuracy over 5 epochs: 51.21%
Number of gradients: 11173962 ←
Singularity> █
```

---

#### Q4. Using Optimizer: Adam

Trainable Parameters: 11173962, Gradients: 11173962

```
Singularity> python3 resnet.py --use_gpu --data_path ./data --num_workers 2 --optimizer adam
usage: resnet.py [-h] [--use_cuda] [--use_cpu] [--data_path DATA_PATH] [--num_workers NUM_WORKERS] [--optimizer {sgd,sgd_nesterov,adagrad,adadelta,adam}]
resnet.py: error: unrecognized arguments: --use_gpu
Singularity> python3 resnet.py --use_cuda --data_path ./data --num_workers 2 --optimizer adam
Using device: cuda
Files already downloaded and verified
Number of trainable parameters: 11173962 ←
Using Optimizer: adam

Epoch 1/5
Total Data Loading Time for Epoch 1: 0.658 seconds
Training Time for Epoch 1: 15.507 seconds
Total Running Time for Epoch 1: 16.849 seconds
Epoch 1: Loss: 1.501, Top-1 Accuracy: 44.73%

Epoch 2/5
Total Data Loading Time for Epoch 2: 0.510 seconds
Training Time for Epoch 2: 11.398 seconds
Total Running Time for Epoch 2: 12.404 seconds
Epoch 2: Loss: 1.034, Top-1 Accuracy: 63.11%

Epoch 3/5
Total Data Loading Time for Epoch 3: 0.536 seconds
Training Time for Epoch 3: 11.399 seconds
Total Running Time for Epoch 3: 12.432 seconds
Epoch 3: Loss: 0.846, Top-1 Accuracy: 70.27%

Epoch 4/5
Total Data Loading Time for Epoch 4: 0.520 seconds
Training Time for Epoch 4: 11.402 seconds
Total Running Time for Epoch 4: 12.425 seconds
Epoch 4: Loss: 0.720, Top-1 Accuracy: 75.14%

Epoch 5/5
Total Data Loading Time for Epoch 5: 0.514 seconds
Training Time for Epoch 5: 11.400 seconds
Total Running Time for Epoch 5: 12.420 seconds
Epoch 5: Loss: 0.633, Top-1 Accuracy: 78.24%

Total Data Loading Time over five epochs: 2.738 seconds
Average Training Time over 5 epochs: 12.222 seconds
Average Loss over 5 epochs: 0.947
Average Top-1 Accuracy over 5 epochs: 66.29%
Number of gradients: 11173962 ←
Singularity>
Singularity>
Singularity> █
```