

Name: <name>  
NetID: <netid>  
Section: <class section>

# ECE 408/CS483 Milestone 3

## Report

0. List Op Times, whole program execution time, and accuracy for batch size of 100, 1k, and 10k images from your basic forward convolution kernel in milestone 2. This will act as your baseline this milestone.

Batch Size	Op Time 1	Op Time 2	Total Execution Time	Accuracy
100	<op_time>	<op_time>	<exec_time>	<accuracy>
1000	<op_time>	<op_time>	<exec_time>	<accuracy>
10000	<op_time>	<op_time>	<exec_time>	<accuracy>

1. Optimization 1: <optimization name>

- a. Which optimization did you choose to implement and why did you choose that optimization technique.

<answer here>

- b. How does the optimization work? Did you think the optimization would increase performance of the forward convolution? Why? Does the optimization synergize with any of your previous optimizations?

<answer here>

- c. List the Op Times, whole program execution time, and accuracy for batch size of 100, 1k, and 10k images using this optimization (including any previous optimizations also used).

Batch Size	Op Time 1	Op Time 2	Total Execution Time	Accuracy
100	<op_time>	<op_time>	<exec_time>	<accuracy>
1000	<op_time>	<op_time>	<exec_time>	<accuracy>
10000	<op_time>	<op_time>	<exec_time>	<accuracy>

- d. Was implementing this optimization successful in improving performance? Why or why not? Include profiling results from nsys and Nsight-Compute to justify your answer, directly comparing to your baseline (or the previous optimization this one is built off of).

<answer here>

- e. What references did you use when implementing this technique?

<answer here>

2. Optimization 2: <optimization name>

- a. Which optimization did you choose to implement and why did you choose that optimization technique.

<answer here>

- b. How does the optimization work? Did you think the optimization would increase performance of the forward convolution? Why? Does the optimization synergize with any of your previous optimizations?

<answer here>

- c. List the Op Times, whole program execution time, and accuracy for batch size of 100, 1k, and 10k images using this optimization (including any previous optimizations also used).

Batch Size	Op Time 1	Op Time 2	Total Execution Time	Accuracy
100	<op_time>	<op_time>	<exec_time>	<accuracy>
1000	<op_time>	<op_time>	<exec_time>	<accuracy>
10000	<op_time>	<op_time>	<exec_time>	<accuracy>

- d. Was implementing this optimization successful in improving performance? Why or why not? Include profiling results from nsys and Nsight-Compute to justify your answer, directly comparing to your baseline (or the previous optimization this one is built off of).

<answer here>

- e. What references did you use when implementing this technique?

<answer here>

3. Optimization 3: <optimization name>

(Delete this section blank if you did not implement this many optimizations.)

- a. Which optimization did you choose to implement and why did you choose that optimization technique.

<answer here>

- b. How does the optimization work? Did you think the optimization would increase performance of the forward convolution? Why? Does the optimization synergize with any of your previous optimizations?

<answer here>

- c. List the Op Times, whole program execution time, and accuracy for batch size of 100, 1k, and 10k images using this optimization (including any previous optimizations also used).

Batch Size	Op Time 1	Op Time 2	Total Execution Time	Accuracy
100	<op_time>	<op_time>	<exec_time>	<accuracy>
1000	<op_time>	<op_time>	<exec_time>	<accuracy>
10000	<op_time>	<op_time>	<exec_time>	<accuracy>

- d. Was implementing this optimization successful in improving performance? Why or why not? Include profiling results from nsys and Nsight-Compute to justify your answer, directly comparing to your baseline (or the previous optimization this one is built off of).

<answer here>

- e. What references did you use when implementing this technique?

<answer here>

4. Optimization 4: <optimization name>

(Delete this section blank if you did not implement this many optimizations.)

- a. Which optimization did you choose to implement and why did you choose that optimization technique.

<answer here>

- b. How does the optimization work? Did you think the optimization would increase performance of the forward convolution? Why? Does the optimization synergize with any of your previous optimizations?

<answer here>

- c. List the Op Times, whole program execution time, and accuracy for batch size of 100, 1k, and 10k images using this optimization (including any previous optimizations also used).

Batch Size	Op Time 1	Op Time 2	Total Execution Time	Accuracy
100	<op_time>	<op_time>	<exec_time>	<accuracy>
1000	<op_time>	<op_time>	<exec_time>	<accuracy>
10000	<op_time>	<op_time>	<exec_time>	<accuracy>

- d. Was implementing this optimization successful in improving performance? Why or why not? Include profiling results from nsys and Nsight-Compute to justify your answer, directly comparing to your baseline (or the previous optimization this one is built off of).

<answer here>

- e. What references did you use when implementing this technique?

<answer here>

5. Optimization 5: <optimization name>

(Delete this section if you did not implement this many optimizations.)

- a. Which optimization did you choose to implement and why did you choose that optimization technique.

<answer here>



- b. How does the optimization work? Did you think the optimization would increase performance of the forward convolution? Why? Does the optimization synergize with any of your previous optimizations?

<answer here>

- c. List the Op Times, whole program execution time, and accuracy for batch size of 100, 1k, and 10k images using this optimization (including any previous optimizations also used).

Batch Size	Op Time 1	Op Time 2	Total Execution Time	Accuracy
100	<op_time>	<op_time>	<exec_time>	<accuracy>
1000	<op_time>	<op_time>	<exec_time>	<accuracy>
10000	<op_time>	<op_time>	<exec_time>	<accuracy>

- d. Was implementing this optimization successful in improving performance? Why or why not? Include profiling results from nsys and Nsight-Compute to justify your answer, directly comparing to your baseline (or the previous optimization this one is built off of).

<answer here>

- e. What references did you use when implementing this technique?

<answer here>

6. Optimization 6: <optimization name>

(Delete this section if you did not implement this many optimizations.)

- a. Which optimization did you choose to implement and why did you choose that optimization technique.

<answer here>

- b. How does the optimization work? Did you think the optimization would increase performance of the forward convolution? Why? Does the optimization synergize with any of your previous optimizations?

<answer here>

- c. List the Op Times, whole program execution time, and accuracy for batch size of 100, 1k, and 10k images using this optimization (including any previous optimizations also used).

Batch Size	Op Time 1	Op Time 2	Total Execution Time	Accuracy
100	<op_time>	<op_time>	<exec_time>	<accuracy>
1000	<op_time>	<op_time>	<exec_time>	<accuracy>
10000	<op_time>	<op_time>	<exec_time>	<accuracy>

- d. Was implementing this optimization successful in improving performance? Why or why not? Include profiling results from nsys and Nsight-Compute to justify your answer, directly comparing to your baseline (or the previous optimization this one is built off of).

<answer here>

- e. What references did you use when implementing this technique?

<answer here>