**Milestone 3 Report**

[Google Drive link to the top-level m3 folder]

1. **Baseline:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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. **Req\_0: \_\_Optimization Name\_\_**

[Google Drive link to subfolder req\_0]

* 1. How does this optimization theoretically optimize your convolution kernel? Expected behavior?
  2. How did you implement your code? Explain thoroughly and show code snippets. Justify the correctness of your implementation with proper profiling results.
  3. Did the performance match your expectation? Explain why or why not, by analyzing profiling results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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>* |

<<Add more profiling analysis here>>

* 1. Does this optimization synergize with any other optimizations? How?
  2. List your references used while implementing this technique. (you must mention textbook pages at the minimum)

1. **Op\_#: \_\_Optimization Name\_\_**

[Google Drive link to subfolder op\_#]

* 1. How does this optimization theoretically optimize your convolution kernel? Expected behavior?
  2. How did you implement your code? Explain thoroughly and show code snippets. Justify the correctness of your implementation with proper profiling results.
  3. Did the performance match your expectation? Explain why or why not, by analyzing profiling results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 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>* |

<<Add more profiling analysis here>>

* 1. Does this optimization synergize with any other optimizations? How?
  2. List your references used while implementing this technique. (you must mention textbook pages at the minimum)