Qeios

Peer Review

Review of: "Optimizing Multi-GPU Training with Data Parallelism and Batch Size Selection"

Mingyi Liu[1]

1. Georgia Institute of Technology, United States

The paper is well-written and easy to understand. I have the following comments that may help improve its scientific impact:

1. **Related Work**: In its current form, references are introduced individually without clear connections. I suggest grouping references by topic into subsections. For instance, you could include a subsection specifically for works focusing on batch size investigations.
2. **Experiments and Results**:
    1. The paper reports accuracy and loss results in figures, but overhead and training time are frequently mentioned in discussions. To better support the conclusions, it seems necessary to have a table summarizing various model performance metrics across different batch sizes.
    2. The experimental implementation lacks detail. For example, it is unclear how many runs were conducted for each experiment.
    3. If applicable, merging figures could enhance the visualization of accuracy and loss trends caused by batch size changes.
3. **Contributions**:
    1. The paper notes that existing works also conclude moderate batch sizes achieve high performance. However, it does not discuss any new findings in comparison to prior studies.
    2. With only one architecture and one dataset applied, the scalability claims throughout the paper are not entirely convincing. I recommend toning down these claims.
4. **Minor Fixes**:
    1. Figures 2 and 3 appear identical.

 doi.org/10.32388/MZGJ77

2. In the context of “Batch Size 64” results, Figure 4 should be referenced instead of Figure 5.

3. Figures for “Batch Size 128” results are missing.

Declarations

Potential competing interests: No potential competing interests to declare.