# Task 3

The following shows the training execution time variance of different batch sizes, partition settings, models.

**AlexNet**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Batch Size | Number of  Layers in  Stage 0 | Number of  Layers in  Stage 1 | Number of  Layers in  Stage 2 | Number of  Layers in  Stage 3 | Execution time (s) |
| 1 | 2 | 4 | 6 | 8 | 52.1991 |
| 1 | 3 | 3 | 6 | 8 | 51.2575 |
| 1 | 4 | 2 | 4 | 10 | 47.7686 |
| 1 | 4 | 2 | 6 | 8 | 45.952 |
| 1 | 4 | 2 | 8 | 6 | 44.7457 |
| 4 | 2 | 4 | 6 | 8 | 84.0271 |
| 4 | 3 | 3 | 6 | 8 | 83.1174 |
| 4 | 4 | 2 | 4 | 10 | 76.9865 |
| 4 | 4 | 2 | 6 | 8 | 80.4904 |
| 4 | 4 | 2 | 8 | 6 | 79.829 |
| 8 | 2 | 4 | 6 | 8 | 99.2262 |
| 8 | 3 | 3 | 6 | 8 | 97.7753 |
| 8 | 4 | 2 | 4 | 10 | 96.3612 |
| 8 | 4 | 2 | 6 | 8 | 98.1843 |
| 8 | 4 | 2 | 8 | 6 | 92.594 |
| 16 | 2 | 4 | 6 | 8 | 192.997 |
| 16 | 3 | 3 | 6 | 8 | 173.606 |
| 16 | 4 | 2 | 4 | 10 | 199.031 |
| 16 | 4 | 2 | 6 | 8 | 183.451 |
| 16 | 4 | 2 | 8 | 6 | 192.226 |

**Vgg**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Batch Size | Number of  Layers in  Stage 0 | Number of  Layers in  Stage 1 | Number of  Layers in  Stage 2 | Number of  Layers in  Stage 3 | Execution time (s) |
| 1 | 1 | 11 | 10 | 16 | 162.547 |
| 1 | 1 | 11 | 16 | 10 | 164.374 |
| 1 | 2 | 10 | 10 | 16 | 166.763 |
| 1 | 2 | 10 | 16 | 10 | 166.271 |
| 1 | 3 | 9 | 9 | 17 | 214.929 |
| 4 | 1 | 11 | 10 | 16 | 265.898 |
| 4 | 1 | 11 | 16 | 10 | 269.388 |
| 4 | 2 | 10 | 10 | 16 | 263.692 |
| 4 | 2 | 10 | 16 | 10 | 266.719 |
| 4 | 3 | 9 | 9 | 17 | 301.939 |
| 8 | 1 | 11 | 10 | 16 | 372.2 |
| 8 | 1 | 11 | 16 | 10 | 388.67 |
| 8 | 2 | 10 | 10 | 16 | 373.771 |
| 8 | 2 | 10 | 16 | 10 | 372.947 |
| 8 | 3 | 9 | 9 | 17 | 412.178 |
| 16 | 1 | 11 | 10 | 16 | 604.276 |
| 16 | 1 | 11 | 16 | 10 | 601.951 |
| 16 | 2 | 10 | 10 | 16 | 601.33 |
| 16 | 2 | 10 | 16 | 10 | 599.024 |
| 16 | 3 | 9 | 9 | 17 | 647.623 |

**Resnet**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Batch Size | Number of  Layers in  Stage 0 | Number of  Layers in  Stage 1 | Number of  Layers in  Stage 2 | Number of  Layers in  Stage 3 | Execution time (s) |
| 1 | 32 | 67 | 9 | 50 | 529.626 |
| 1 | 52 | 47 | 18 | 41 | 523.017 |
| 1 | 52 | 47 | 27 | 32 | 515.342 |
| 1 | 52 | 47 | 9 | 50 | 505.543 |
| 1 | 62 | 37 | 9 | 50 | 845.541 |
| 4 | 32 | 67 | 9 | 50 | 719.339 |
| 4 | 52 | 47 | 18 | 41 | 722.059 |
| 4 | 52 | 47 | 27 | 32 | 735.306 |
| 4 | 52 | 47 | 9 | 50 | 725.288 |
| 4 | 62 | 37 | 9 | 50 | 1055.59 |
| 8 | 32 | 67 | 9 | 50 | 970.334 |
| 8 | 52 | 47 | 18 | 41 | 1022.76 |
| 8 | 52 | 47 | 27 | 32 | 990.449 |
| 8 | 52 | 47 | 9 | 50 | 969.801 |
| 8 | 62 | 37 | 9 | 50 | 1295.23 |
| 16 | 32 | 67 | 9 | 50 | 1438.66 |
| 16 | 52 | 47 | 18 | 41 | 1441.26 |
| 16 | 52 | 47 | 27 | 32 | 1451.24 |
| 16 | 52 | 47 | 9 | 50 | 1438.01 |
| 16 | 62 | 37 | 9 | 50 | 1729.68 |

# Task 4

**What’s the key factors that affect the training throughput?**

1. Pipelined execution process increases training throughput and vice versa.
2. Avoiding data dependency during execution process increases training throughput and vice versa.
3. Limited memory inhibits execution speed and vice versa.
4. Unevenly distributed partition execution time inhibits execution speed and vice versa.
5. Given different batch size, arrangement of layers which achieve maximum training throughput is different.

**Generally speaking, what’s the best (ideal) partition that can maximize the throughput of training pipeline?**

1. The partition that minimize structural hazard, data hazard and control hazard in pipelining will maximize the throughput. The hazards include structural hazard (Required resource is busy). data hazard (Need to wait for the results of a previous instruction in the pipeline) and control Hazard (Break the pipeline if branch changes the program counter).
2. Evenly distributed partition minimizes the execution time.
3. Enough memory promotes pipeline.

**What’s the distance between your current partition on the three models and the ideal partition?**

1. Some data hazard cannot be eliminated.
2. Some data dependencies make optimal partition impossible.
3. It is impossible to ensure every stage spends the same amount of time. That is because individual layer cannot be perfectly divided. Also, the computer may be executing other programs simultaneously. So, computer resources allocated to execute the model will be less than expected.
4. Operation system has default setting to alter execution partition.