针对问题：

在十一号机子上，数据读取的时间占了总算法运行时间的一半甚至更多，在提升 dataloader 的 num_works 同时并没有改善。 数据读取占用很久时间的现象理应是不正常的，同时在监控 GPU 运行效率是会发现 GPU 会浪费很多时间在等待读取完的数据输入，即 GPU 利用率会间隔一定时间持续为；同时，这种现象在小规模数据集上不会对总体运行时间带来很大的影响，但在大规模数据集上的每个 batch 运行时间的累加会带来非常大的影响。

怀疑因素：

1. 11 号服务器上的硬盘不是直连的，通过 nas 连接，在每次读取 batch 数据的时候，I/O 口的性能瓶颈；
2. 数据处理的代码段消耗太多时间。

实验：

在 11，7，8 号机子上做了相同的读取数据实验，并分别记录了硬盘读取和数据预处理的时间，并在不同的 num_works(0，1，2，4)的条件下进行实验。实际测算结果如下表所示：

Batchsize：64

服务器	0			1			2			4		
	总数据时间	硬盘读取时间	数据处理时间	总数据时间	硬盘读取时间	数据处理时间	总数据时间	硬盘读取时间	数据处理时间	总数据时间	硬盘读取时间	数据处理时间
11 号机	4.514	0.223	3.048	0.2699	0.0737	0.1419	0.2042	0.0823	0.1535	0.1591	0.0876	0.1662
7 号机	0.2602	0.0532	0.1732	0.3010	0.0550	0.1800	0.2438	0.0617	0.2176	0.2525	0.0654	0.2016
8 号机	0.1196	0.0268	0.0778	0.1493	0.0280	0.0917	0.1345	0.0654	0.1026	0.1212	0.0350	0.0916

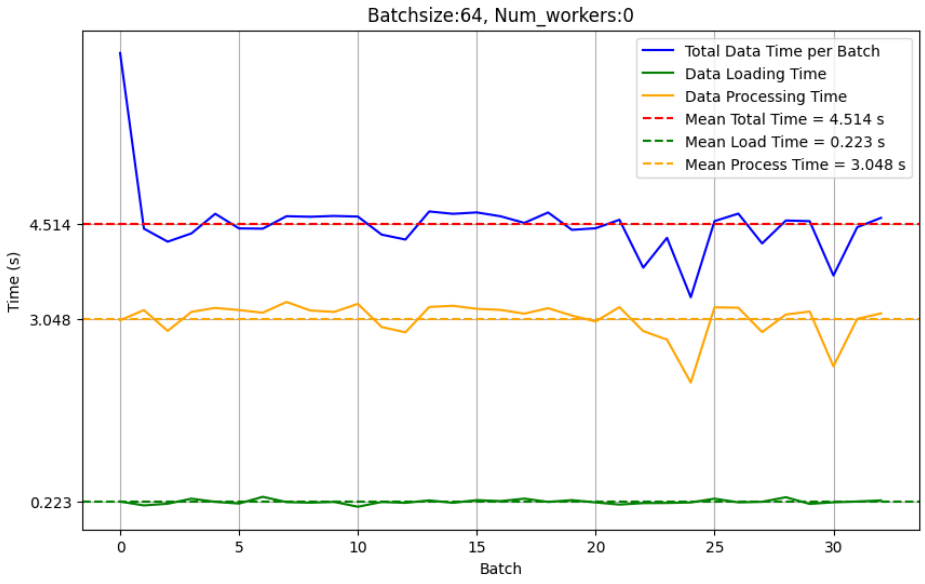
可以看到，8 号机子上的总体偏优，基本比其它两个机子提升两倍之多，我不太了解各服务器的配置，所以暂时不知道具体原因。

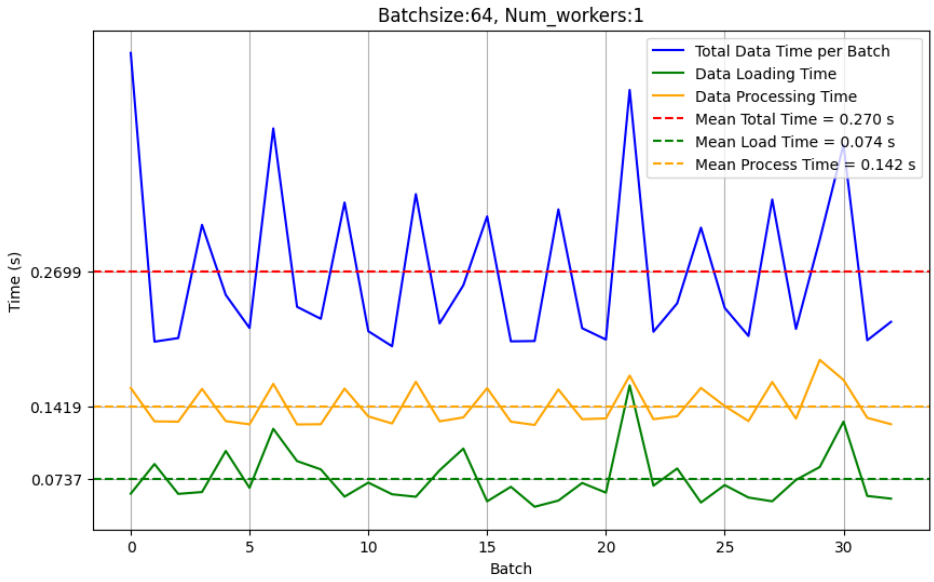
暂时能分析出来：

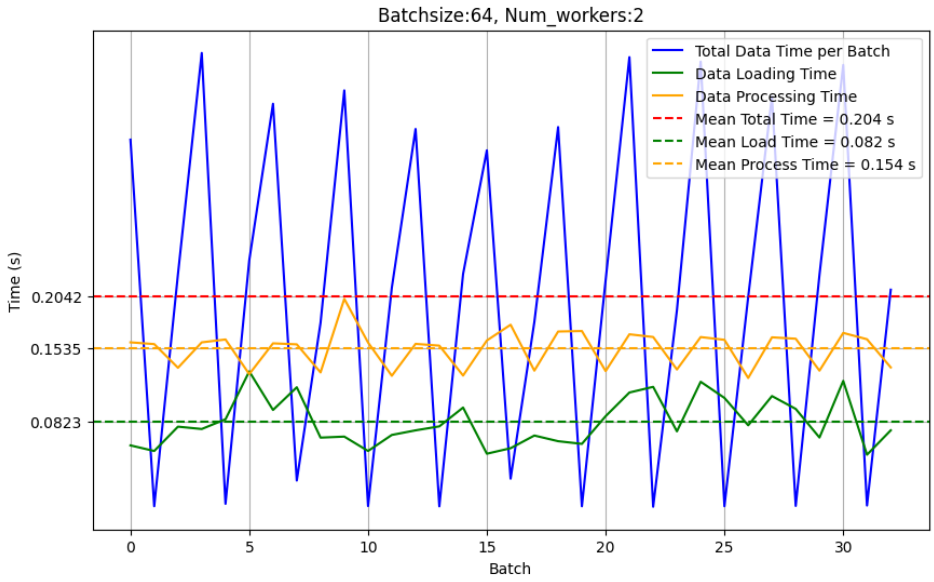
1 同样的数据处理段为什么差别这么大，是因为不同 cpu 的差异吗，毕竟差两倍，在大规模数据集上是 10 小时和 5 小时的区别。

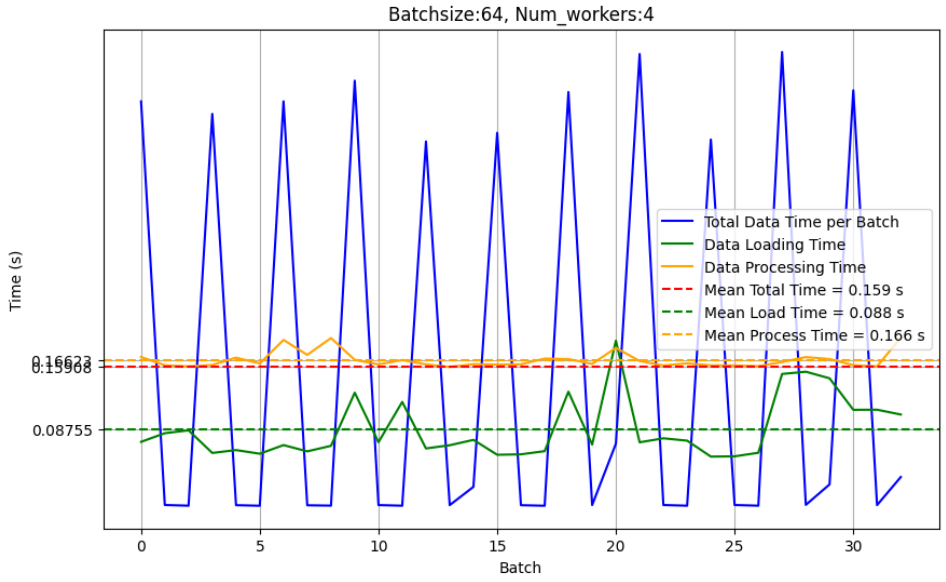
如下是实验截图：

11 号机：

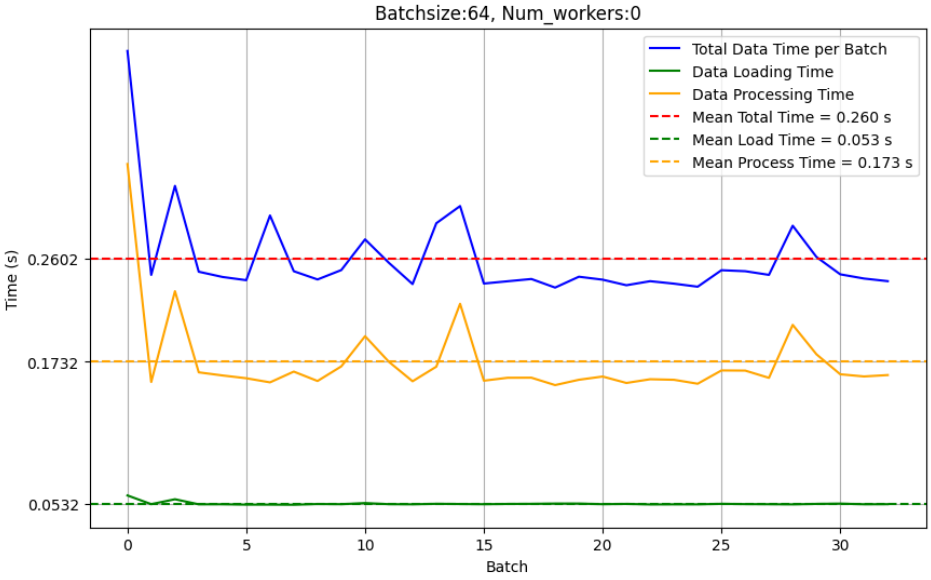


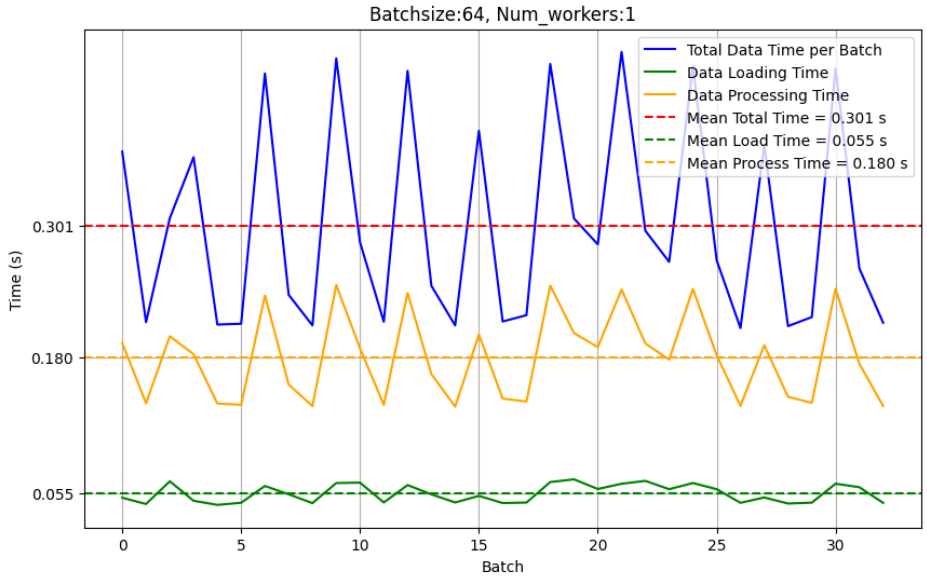

Batchsize:64, Num_workers:1
Total Data Time per Batch
Data Loading Time
Data Processing Time
Mean Total Time = 0.270 s
Mean Load Time = 0.074 s
Mean Process Time = 0.142 s
Time (s)
0.2699
0.1419
0.0737
0
5
10
15
20
25
30
Batch

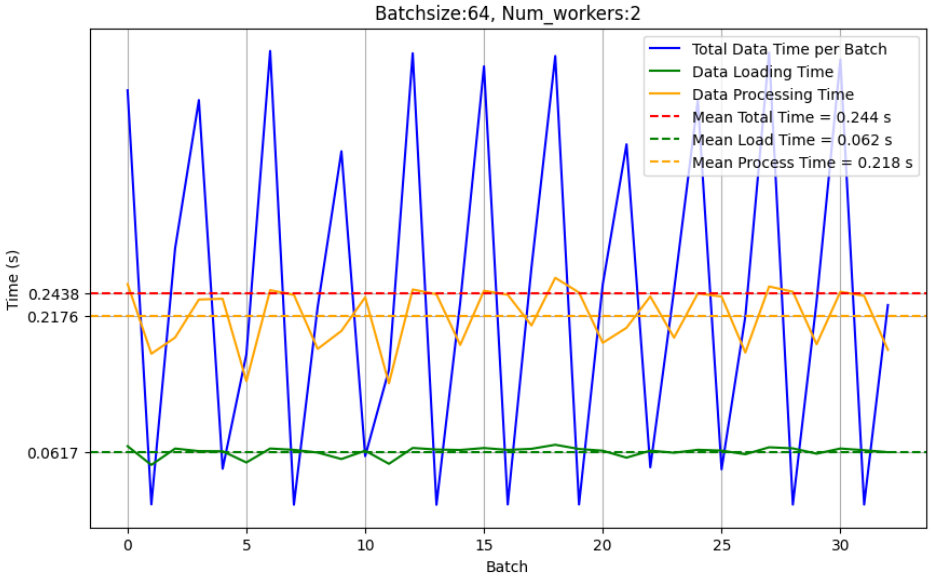

Batchsize:64, Num_workers:2
Total Data Time per Batch
Data Loading Time
Data Processing Time
Mean Total Time = 0.204 s
Mean Load Time = 0.082 s
Mean Process Time = 0.154 s
Time (s)
0.2042
0.1535
0.0823
0
5
10
15
20
25
30
Batch

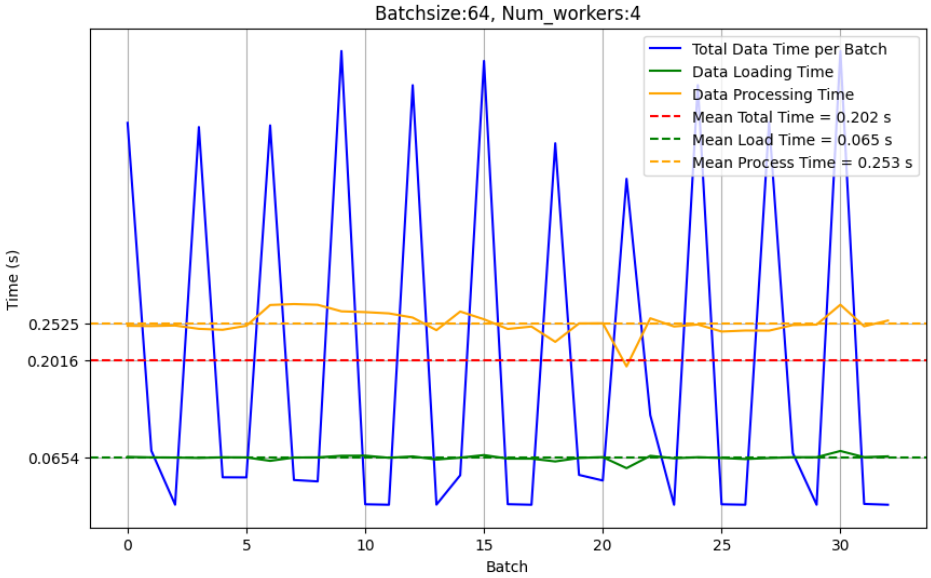

Batchsize:64, Num_workers:4
Total Data Time per Batch
Data Loading Time
Data Processing Time
Mean Total Time = 0.159 s
Mean Load Time = 0.088 s
Mean Process Time = 0.166 s
Time (s)
0.16623
0.15908
0.08755
0
5
10
15
20
25
30
Batch

7 号机：

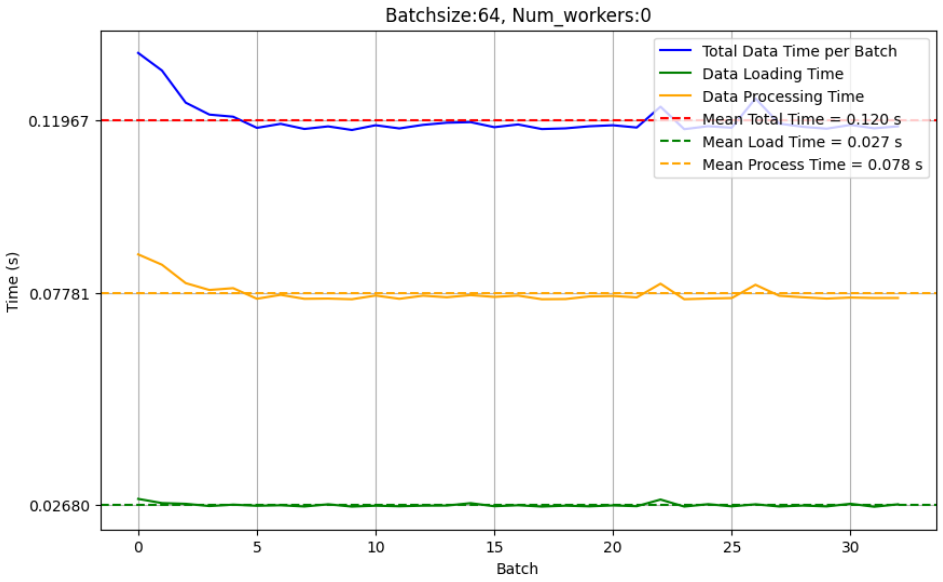

Batchsize:64, Num_workers:0
Total Data Time per Batch
Data Loading Time
Data Processing Time
Mean Total Time = 0.260 s
Mean Load Time = 0.053 s
Mean Process Time = 0.173 s
Time (s)
0.2602
0.1732
0.0532
0
5
10
15
20
25
30
Batch

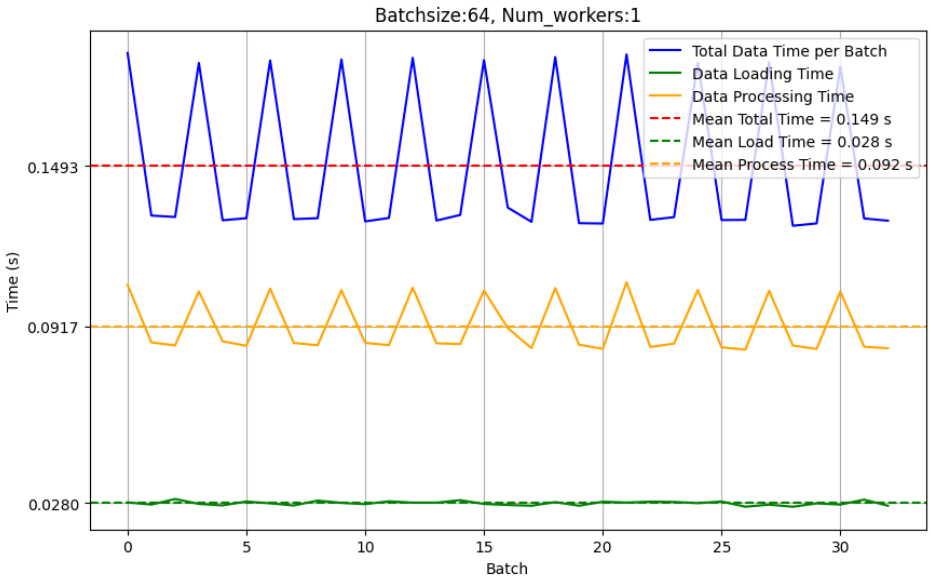

Batchsize:64, Num_workers:1
Total Data Time per Batch
Data Loading Time
Data Processing Time
Mean Total Time = 0.301 s
Mean Load Time = 0.055 s
Mean Process Time = 0.180 s
Time (s)
0.301
0.180
0.055
0
5
10
15
20
25
30
Batch

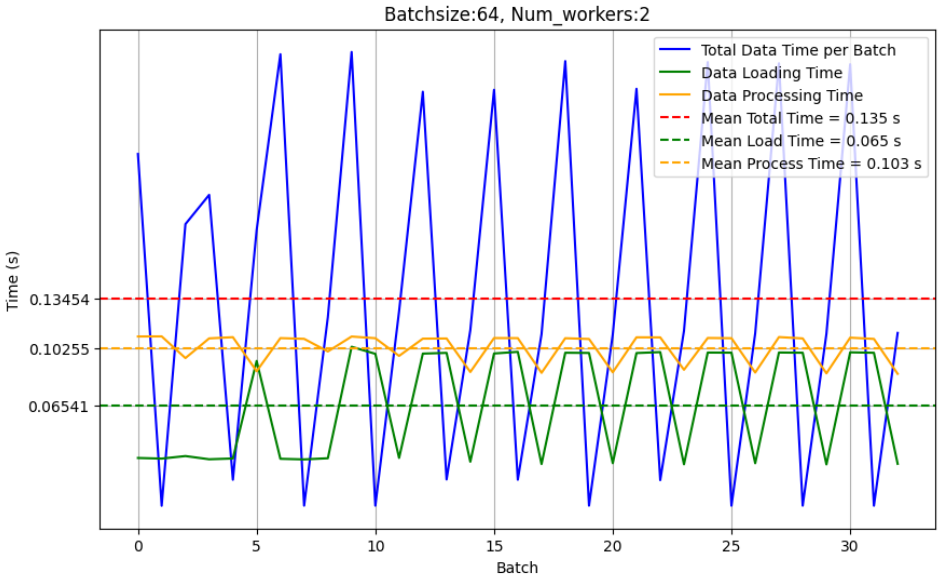

Batchsize:64, Num_workers:2
Total Data Time per Batch
Data Loading Time
Data Processing Time
Mean Total Time = 0.244 s
Mean Load Time = 0.062 s
Mean Process Time = 0.218 s
Time (s)
0.2438
0.2176
0.0617
0
5
10
15
20
25
30
Batch

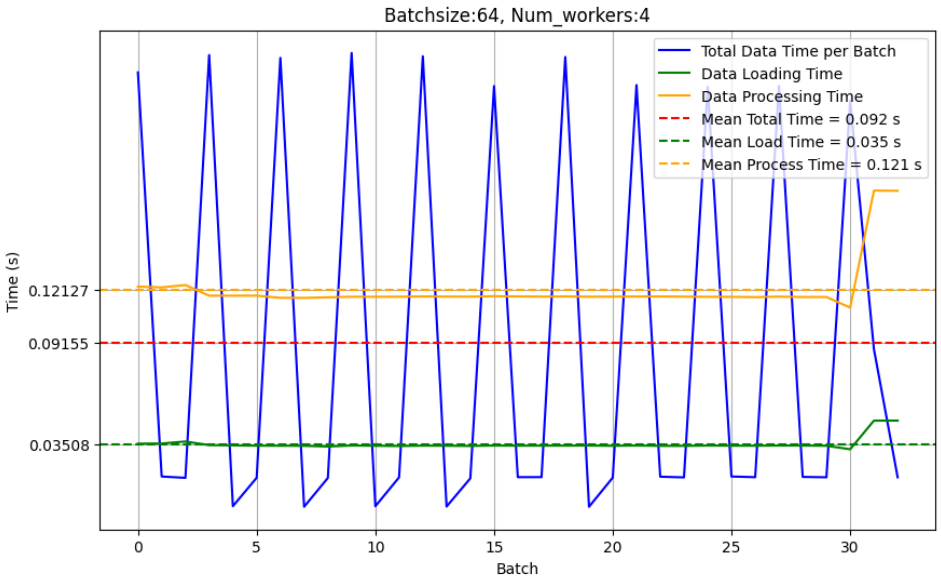

Batchsize:64, Num_workers:4
Total Data Time per Batch
Data Loading Time
Data Processing Time
Mean Total Time = 0.202 s
Mean Load Time = 0.065 s
Mean Process Time = 0.253 s
Time (s)
0.2525
0.2016
0.0654
0
5
10
15
20
25
30
Batch

8 号机：


Batchsize:64, Num_workers:0
Total Data Time per Batch
Data Loading Time
Data Processing Time
Mean Total Time = 0.120 s
Mean Load Time = 0.027 s
Mean Process Time = 0.078 s
0.11967
0.07781
0.02680
Time (s)
0
5
10
15
20
25
30
Batch


Batchsize:64, Num_workers:1
Total Data Time per Batch
Data Loading Time
Data Processing Time
Mean Total Time = 0.149 s
Mean Load Time = 0.028 s
Mean Process Time = 0.092 s
0.1493
0.0917
0.0280
Time (s)
0
5
10
15
20
25
30
Batch


Batchsize:64, Num_workers:2
Total Data Time per Batch
Data Loading Time
Data Processing Time
Mean Total Time = 0.135 s
Mean Load Time = 0.065 s
Mean Process Time = 0.103 s
0.13454
0.10255
0.06541
Time (s)
0
5
10
15
20
25
30
Batch


Batchsize:64, Num_workers:4
Total Data Time per Batch
Data Loading Time
Data Processing Time
Mean Total Time = 0.092 s
Mean Load Time = 0.035 s
Mean Process Time = 0.121 s
Time (s)
0.12127
0.09155
0.03508
0
5
10
15
20
25
30
Batch