Share Comment Star

Depth

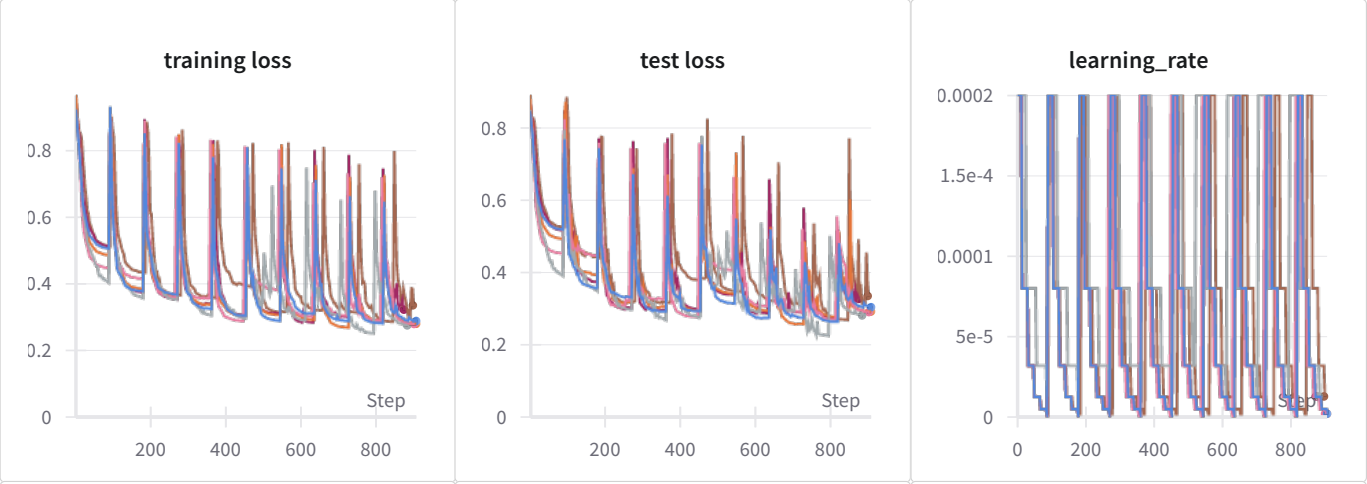
a comparison of performance across increasing depth (ie an increasing number of LSTM layers. Max Depth constrained by GPU memory limitations.

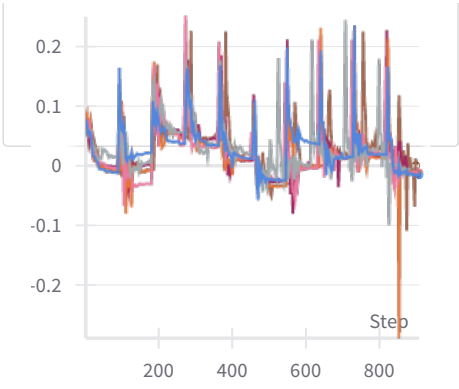
Based on test-train graphs, there are 3 models that are arguably the best performing:
- 4 layers with standard LR
- 4 layers with more aggressive LR
- 6 layers with standard LR

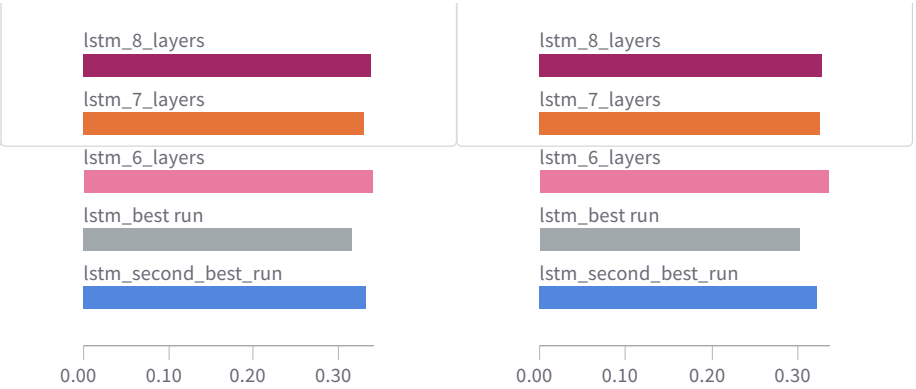
IMO 7 layers is the best because:
- The test train graphs looks the best
- the final MAE is the 2nd lowest
- the average MAE is the 2nd lowest
- the generalisation gap is 2nd lowest

Michael Kingston

Section 1







Fold 9 Loss Curves

lstm... lstm... lstm... lstm... lstm...

Selected runs without media (1)

Step

909

Fold 8 Loss Curves

lstm ... lstm... lstm... lstm... lstm... lstm...

Step

845

Fold 5 Loss Curves

lstm ... lstm... lstm... lstm... lstm... lstm...

Step

563

Fold 7 Loss Curves

lstm ... lstm... lstm... lstm... lstm... lstm...

Step

751

Fold 6 Loss Curves

lstm ... lstm... lstm... lstm... lstm... lstm...

Step

657

Fold 4 Loss Curves

lstm ... lstm... lstm... lstm... lstm... lstm...

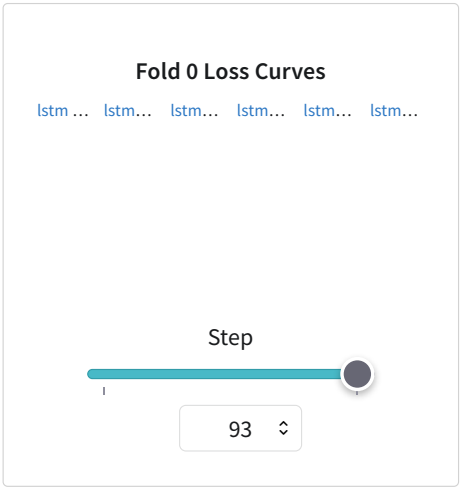
Step

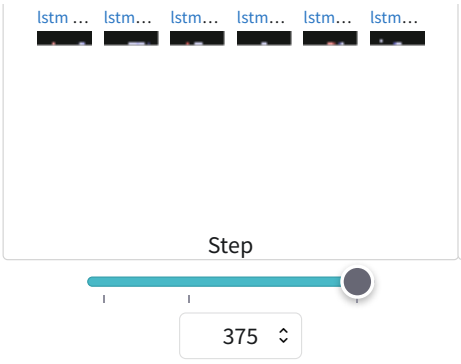
469

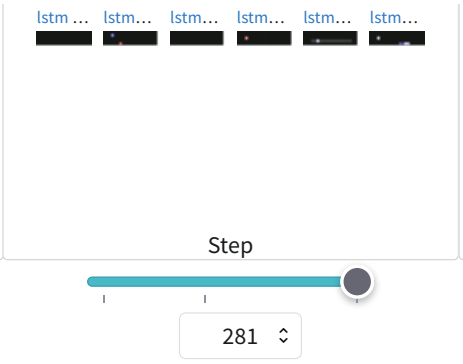
Fold 3 Loss Curves

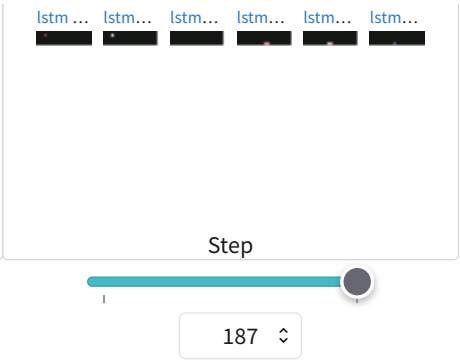
Fold 2 Loss Curves

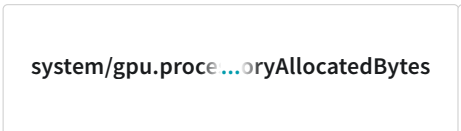
Fold 1 Loss Curves

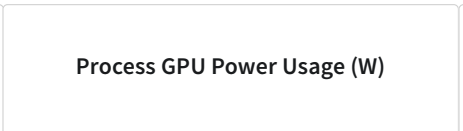

Fold 0 Loss Curves
lstm ... lstm... lstm... lstm... lstm... lstm...
Step
93

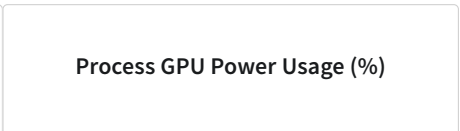

lstm ... lstm... lstm... lstm... lstm... lstm...
Step
375

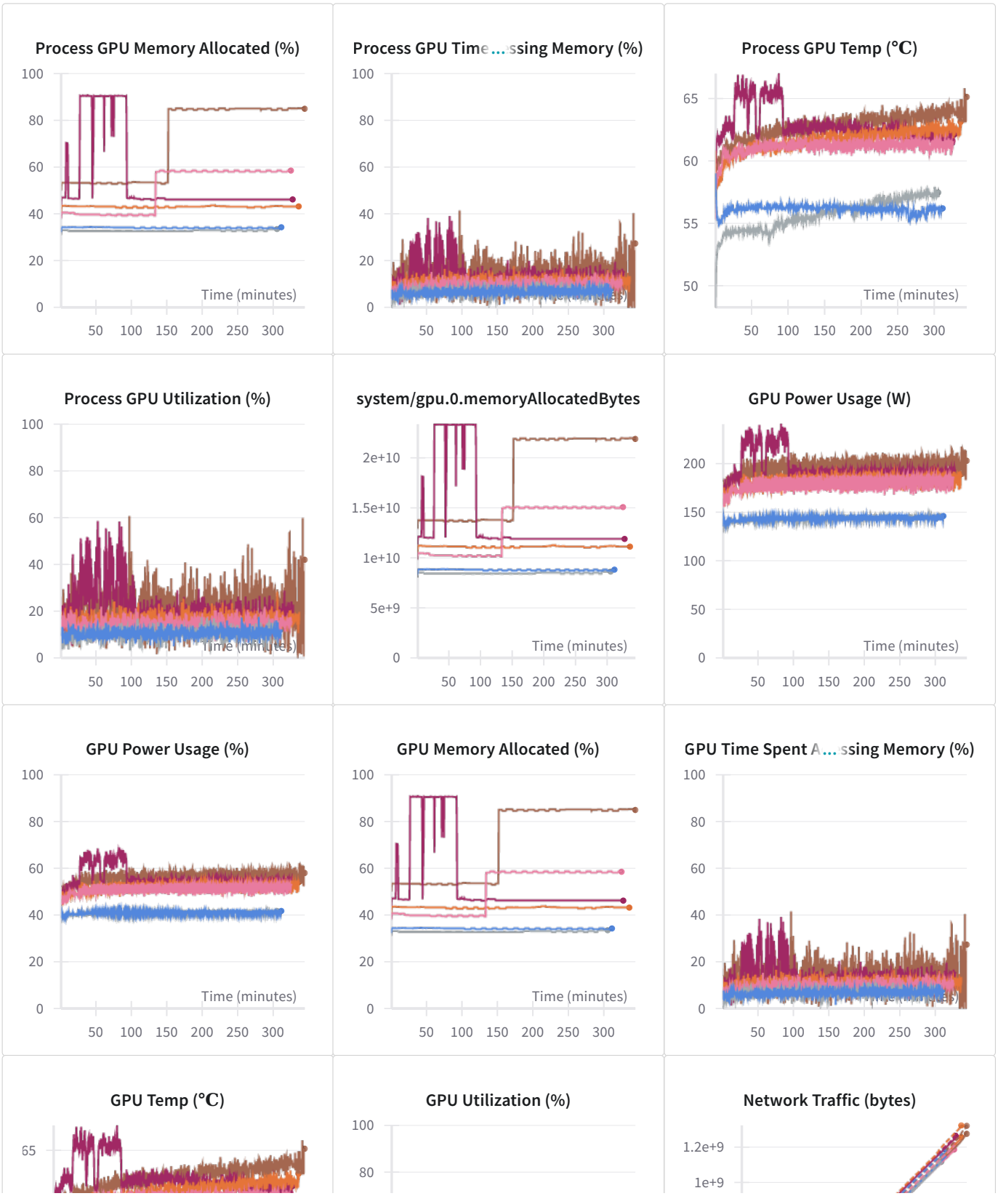

lstm ... lstm... lstm... lstm... lstm... lstm...
Step
281

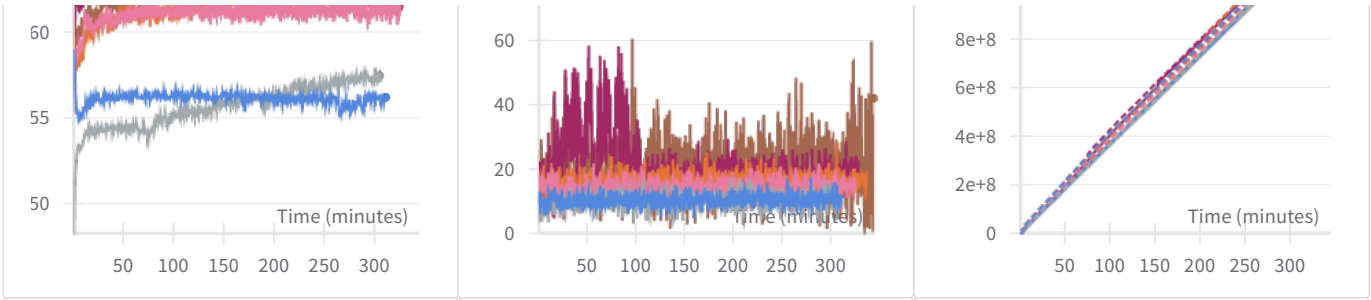

lstm ... lstm... lstm... lstm... lstm... lstm...
Step
187

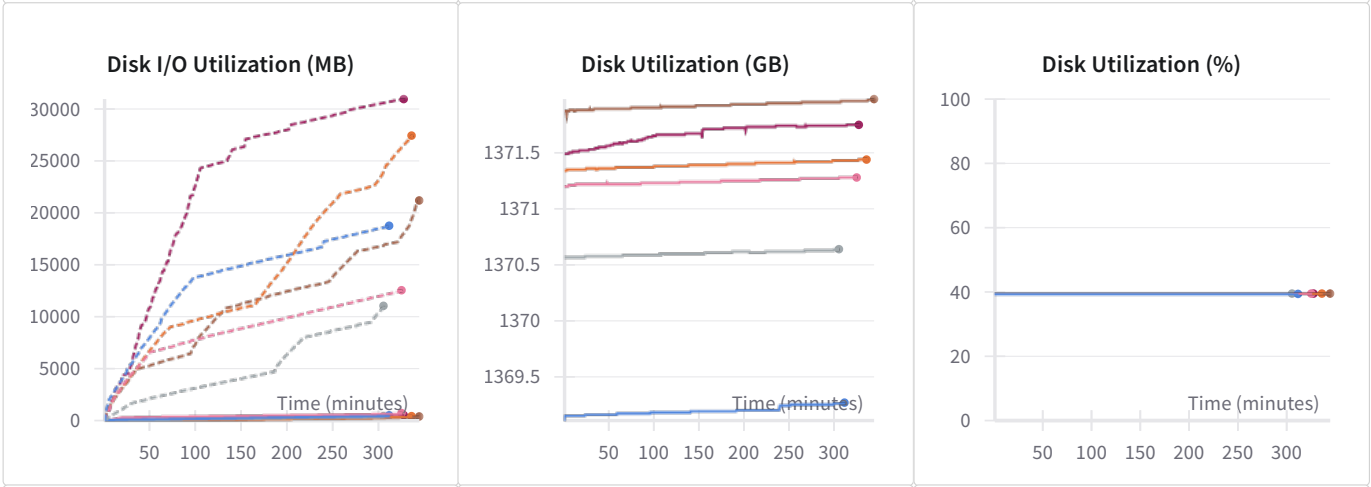

system/gpu.proce ...oryAllocatedBytes

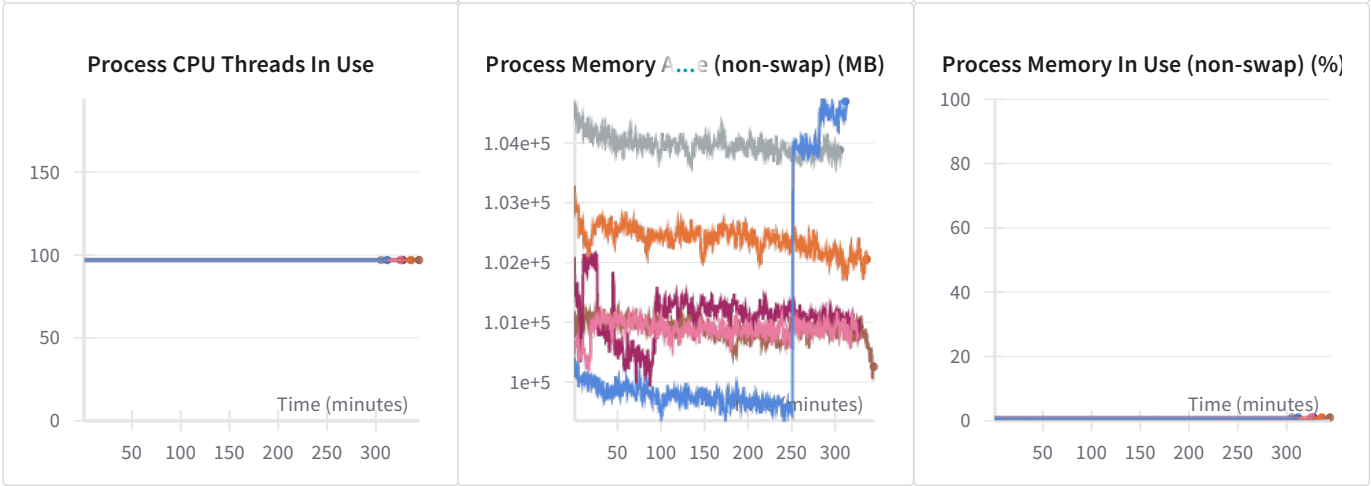

Process GPU Power Usage (W)

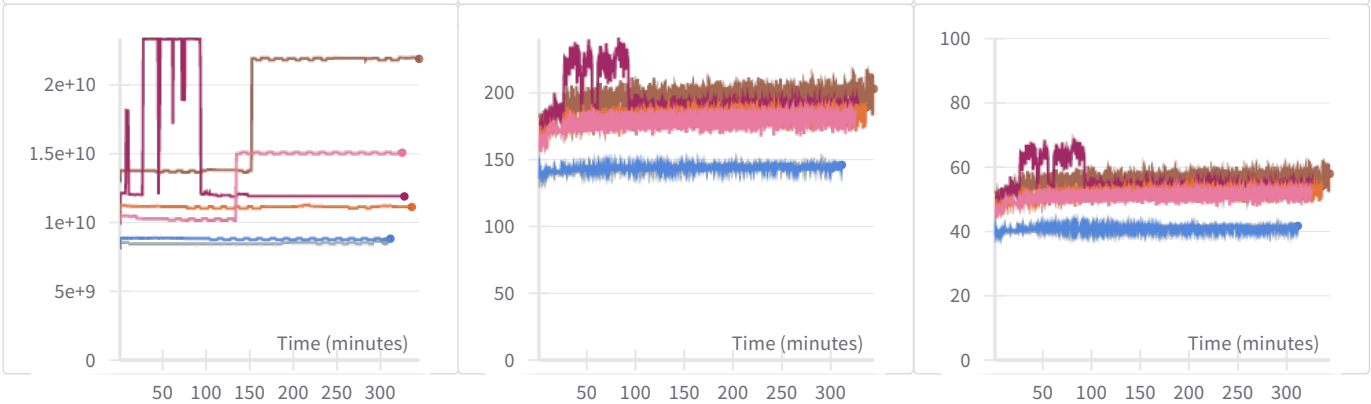

Process GPU Power Usage (%)


Process GPU Memory Allocated (%)
Process GPU Time ...ssing Memory (%)
Process GPU Temp (°C)
Process GPU Utilization (%)
system/gpu.0.memoryAllocatedBytes
GPU Power Usage (W)
GPU Power Usage (%)
GPU Memory Allocated (%)
GPU Time Spent A...ssing Memory (%)
GPU Temp (°C)
GPU Utilization (%)
Network Traffic (bytes)
Time (minutes)


60
55
50
Time (minutes)
50 100 150 200 250 300
60
40
20
0
Time (minutes)
50 100 150 200 250 300
8e+8
6e+8
4e+8
2e+8
0
Time (minutes)
50 100 150 200 250 300


Disk I/O Utilization (MB)
30000
25000
20000
15000
10000
5000
0
Time (minutes)
50 100 150 200 250 300
Disk Utilization (GB)
1371.5
1371
1370.5
1370
1369.5
Time (minutes)
50 100 150 200 250 300
Disk Utilization (%)
100
80
60
40
20
0
Time (minutes)
50 100 150 200 250 300


Process CPU Threads In Use
150
100
50
0
Time (minutes)
50 100 150 200 250 300
Process Memory A...e (non-swap) (MB)
1.04e+5
1.03e+5
1.02e+5
1.01e+5
1e+5
minutes)
50 100 150 200 250 300
Process Memory In Use (non-swap) (%)
100
80
60
40
20
0
Time (minutes)
50 100 150 200 250 300


2e+10
1.5e+10
1e+10
5e+9
0
Time (minutes)
50 100 150 200 250 300
200
150
100
50
0
Time (minutes)
50 100 150 200 250 300
100
80
60
40
20
0
Time (minutes)
50 100 150 200 250 300

Created with ❤ on Weights & Biases.

https://wandb.ai/michael_kingston/electricity_demand_forecasting/reports/Depth--Vmlldzo3NTM1MTk4