10,000 samples

GPU	**Time per Batch**	**Time/Epoch**	**10 Epochs**	**20 Epochs**
A100	~0.06 sec	~1.3 minutes	**13 minutes**	**26 minutes**
L40S	~0.08–0.12 sec	~1.6–2.5 minutes	**~16 – 25 minutes**	**~32 – 50 minutes**

1,00,000 Samples

GPU	**Time per Batch**	**Time/Epoch**	**10 Epochs**	**20 Epochs**
A100	**0.03s**	~6.3 minutes	**63 minutes**	**126 minutes**
L40S	0.05–0.08s	~10–17 minutes	**~100 – 170 minutes**	**~200 – 340 minutes**

VRAM

GPU MODEL	**VRAM SIZE**
A100	40GB
LS40	25/40GB
T4- Colab	16GB
MPS	8GB

F32

1) Model weights= 4*0.082= 326MB * 1.2 (buffer)= 390MB
2) Memory = 390 MB

3) Optimizer= 2*4*82= 1.3GB
4) Activation(estimate)= Batch*Length* Hidden Layer*(Encoder+Decoder)* FP

= 8*512*512*(6+6)*4B ~= 1.3GB

5) Caching approx.= 1GB

Total Approx = 5-6 GB

Similarly with using FP16 it is approx. 3- 4 GB