

Model

Total params: 8,922,155

Trainable params: 4,512,347

Non-trainable params: 4,409,808

IPUs

1 ipu, batch_size 10

7000/7000 [=====] - 453s 65ms/step - loss: 33335968.0000

Epoch 2/3

7000/7000 [=====] - 22s 3ms/step - loss: 0.1113

Epoch 3/3

7000/7000 [=====] - 24s 3ms/step - loss: 0.0979

training time: 499.7363388000522

2 ipus

7000/7000 [=====] - 456s 65ms/step - loss: 320412900527737600.0000

Epoch 2/3

7000/7000 [=====] - 13s 2ms/step - loss: 941294988954.5288

Epoch 3/3

7000/7000 [=====] - 13s 2ms/step - loss: 258304438665.7449

training time: 482.76148588582873

4 ipus

6992/6992 [=====] - 493s 70ms/step - loss: 527732100064477120.0000

Epoch 2/3

6992/6992 [=====] - 9s 1ms/step - loss: 1744879797116.6504

Epoch 3/3

6992/6992 [=====] - 9s 1ms/step - loss: 972443784371.6722

training time: 511.2401789960917

8 ipus

6976/6976 [=====] - 499s 71ms/step - loss: 32889948279.1756

Epoch 2/3

6976/6976 [=====] - 7s 1ms/step - loss: 0.1105

Epoch 3/3

6976/6976 [=====] - 7s 904us/step - loss: 0.1089

training time: 512.5336864919402

16 ipus

6912/6912 [=====] - 504s 73ms/step - loss: 1103318273.0971

Epoch 2/3

6912/6912 [=====] - 9s 1ms/step - loss: 2105.5997

Epoch 3/3

6912/6912 [=====] - 5s 704us/step - loss: 1005.8742

training time: 518.1304008550011

1 A100-GPU

Batch size 10

Training

Epoch 1/3

7000/7000 [=====] - 287s 41ms/step - loss:
nan

Epoch 2/3

7000/7000 [=====] - 259s 37ms/step - loss:
nan

Epoch 3/3

```
7000/7000 [=====] - 259s 37ms/step - loss:
nan
training time: 815.4172495089006
```

bs 128

```
Training
Epoch 1/3
546/546 [=====] - 39s 71ms/step - loss:
38812012544.0000
Epoch 2/3
546/546 [=====] - 23s 41ms/step - loss:
119075.6172
Epoch 3/3
546/546 [=====] - 23s 41ms/step - loss:
66543.0391
training time: 95.34690408990718
```

bs 256

```
Training
Epoch 1/3
273/273 [=====] - 30s 111ms/step - loss: nan
Epoch 2/3
273/273 [=====] - 14s 50ms/step - loss: nan
Epoch 3/3
273/273 [=====] - 14s 50ms/step - loss: nan
training time: 95.33727809297852
```

bs 512

```
Training
Epoch 1/3
136/136 [=====] - 28s 204ms/step - loss:
1704047804416.0000
Epoch 2/3
136/136 [=====] - 12s 81ms/step - loss:
0.1138
Epoch 3/3
136/136 [=====] - 12s 81ms/step - loss:
0.1138
training time: 54.37938559206668
```

2 A100-GPU

Bs 10

```
Epoch 1/3
3500/3500 [=====] - 299s 85ms/step - loss:
nan
Epoch 2/3
3500/3500 [=====] - 256s 73ms/step - loss:
nan
Epoch 3/3
3500/3500 [=====] - 256s 73ms/step - loss:
nan
training time: 828.7121053638402
```

bs 128

```
Training
Epoch 1/3
272/272 [=====] - 66s 241ms/step - loss: nan
Epoch 2/3
272/272 [=====] - 22s 80ms/step - loss: nan
Epoch 3/3
272/272 [=====] - 22s 80ms/step - loss: nan
training time: 109.71801541000605
```

bs 256

```
Training
Epoch 1/3
136/136 [=====] - 54s 394ms/step - loss:
191244340822016.0000
Epoch 2/3
136/136 [=====] - 12s 86ms/step - loss:
4602008576.0000
Epoch 3/3
136/136 [=====] - 12s 86ms/step - loss:
2362239488.0000
training time: 108.5867206279654
```

bs 512

Epoch 1/3

68/68 [=====] - 51s 749ms/step - loss:

316702.4688

Epoch 2/3

68/68 [=====] - 8s 119ms/step - loss: 15.7828

Epoch 3/3

68/68 [=====] - 8s 119ms/step - loss: 6.7402

training time: 108.58856278494932

TABLE 4: Performance of the representative model on the Mk2 POD-16 IPU's and A100-GPUs

Hardware Platform	Replicas (Devices)	Micro batch size	Time spent on 2nd epoch (seconds)	Throughput (images/second)
A100 GPU	1	512	12.0	5,833
A100 GPU	1	256	13.7	5,109
A100 GPU	1	128	22.5	3,111
A100 GPU	1	10	259.0	270
A100 GPU	2	512	8.1	8,642
A100 GPU	2	256	11.7	5,983
A100 GPU	2	128	21.8	3,211
A100 GPU	2	10	255.5	274
Mk2-IPU	1	10	23.7	2,954
Mk2-IPU	2	10	13.5	5,185
Mk2-IPU	4	10	9.0	7,743
Mk2-IPU	8	10	6.3	11,052
Mk2-IPU	16	10	4.9	21,547