Version 1:

20K epochs, learning rate = 0.001

Testing Loss:
5K: 1m 31s (- 4m 33s) (5000 25%) 5.1078
10k: 3m 8s (- 3m 8s) (10000 50%) 4.9334
15k: 4m 47s (- 1m 35s) (15000 75%) 4.8771
20k: 6m 29s (- 0m 0s) (20000 100%) 4.8367

Version 2:

Auto Encoder:
5k: 1m 16s (- 3m 50s) (5000 25%) 4.9434
10k: 2m 38s (- 2m 38s) (10000 50%) 4.6610
15k: 4m 4s (- 1m 21s) (15000 75%) 4.5030
20k: 5m 29s (- 0m 0s) (20000 100%) 4.2222

Trained after Auto Encoder Weights:

5k: 1m 29s (- 4m 29s) (5000 25%) 5.1873
10k: 3m 6s (- 3m 6s) (10000 50%) 4.9973
15k: 4m 44s (- 1m 34s) (15000 75%) 4.8067
20k: 6m 23s (- 0m 0s) (20000 100%) 4.6863

Version 3:

5k: 1m 19s (- 4m 19s) (5000 25%) 5.6254
10k: 3m 1s (- 3m 1s) (10000 50%) 4.8564
15k: 4m 14s (- 1m 12s) (15000 75%) 4.7552
20k: 6m 3s (- 0m 0s) (20000 100%) 4.5231