
Full-attn
Reformer
Performer
Nystrom
Linearized
Linformer
SBM(ours)
Training Loss
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
250
500
750
1000
1250
1500
1750
2000
Epoch
Mask Density
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
250
500
750
1000
1250
1500
1750
2000
Epoch