

Gradient Predictiveness

250
200
150
100
50
0



Standard

Standard + BatchNorm

0

5k

10k

15k

Steps

