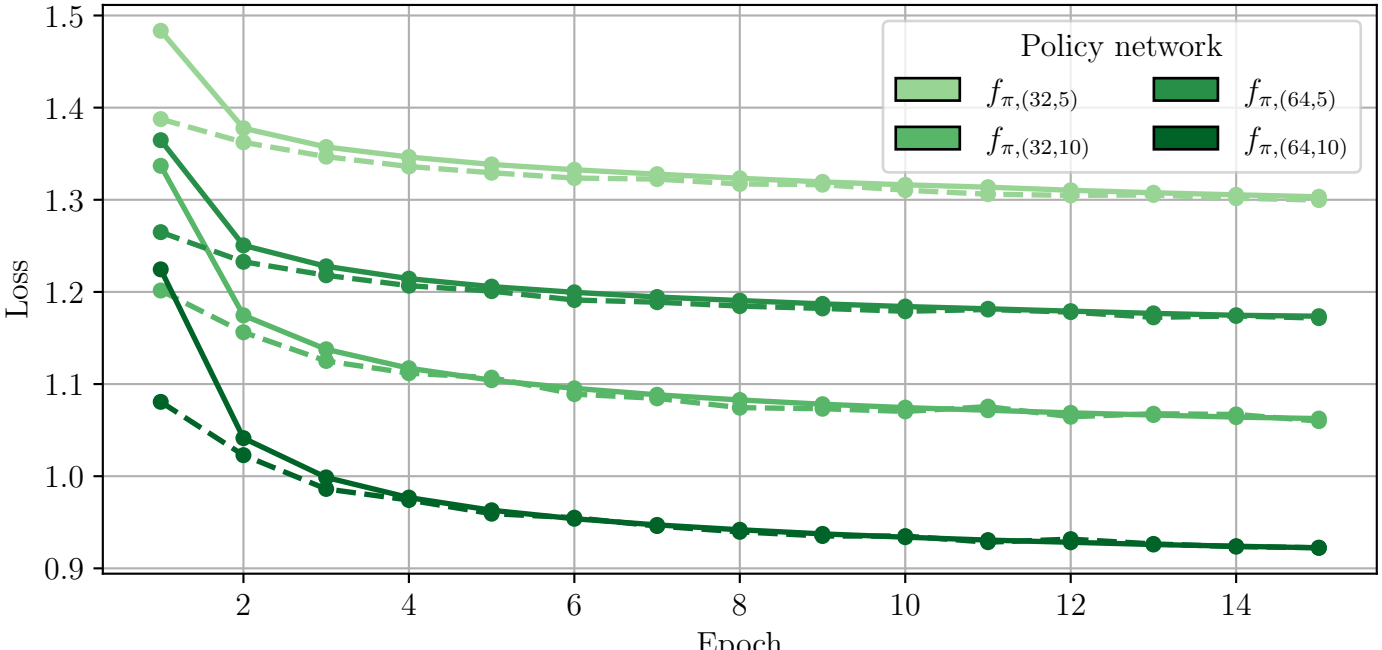
Training (full lines) and validation loss (stapled lines) for policy networks
Policy network
$f_{\pi,(32,5)}$
$f_{\pi,(64,5)}$
$f_{\pi,(32,10)}$
$f_{\pi,(64,10)}$
Loss
Epoch
1.5
1.4
1.3
1.2
1.1
1.0
0.9
2
4
6
8
10
12
14