

Validation Reward

1.6  
1.4  
1.2  
1.0  
0.8

0

500

1000

1500

2000

2500

3000

Training Epoch

