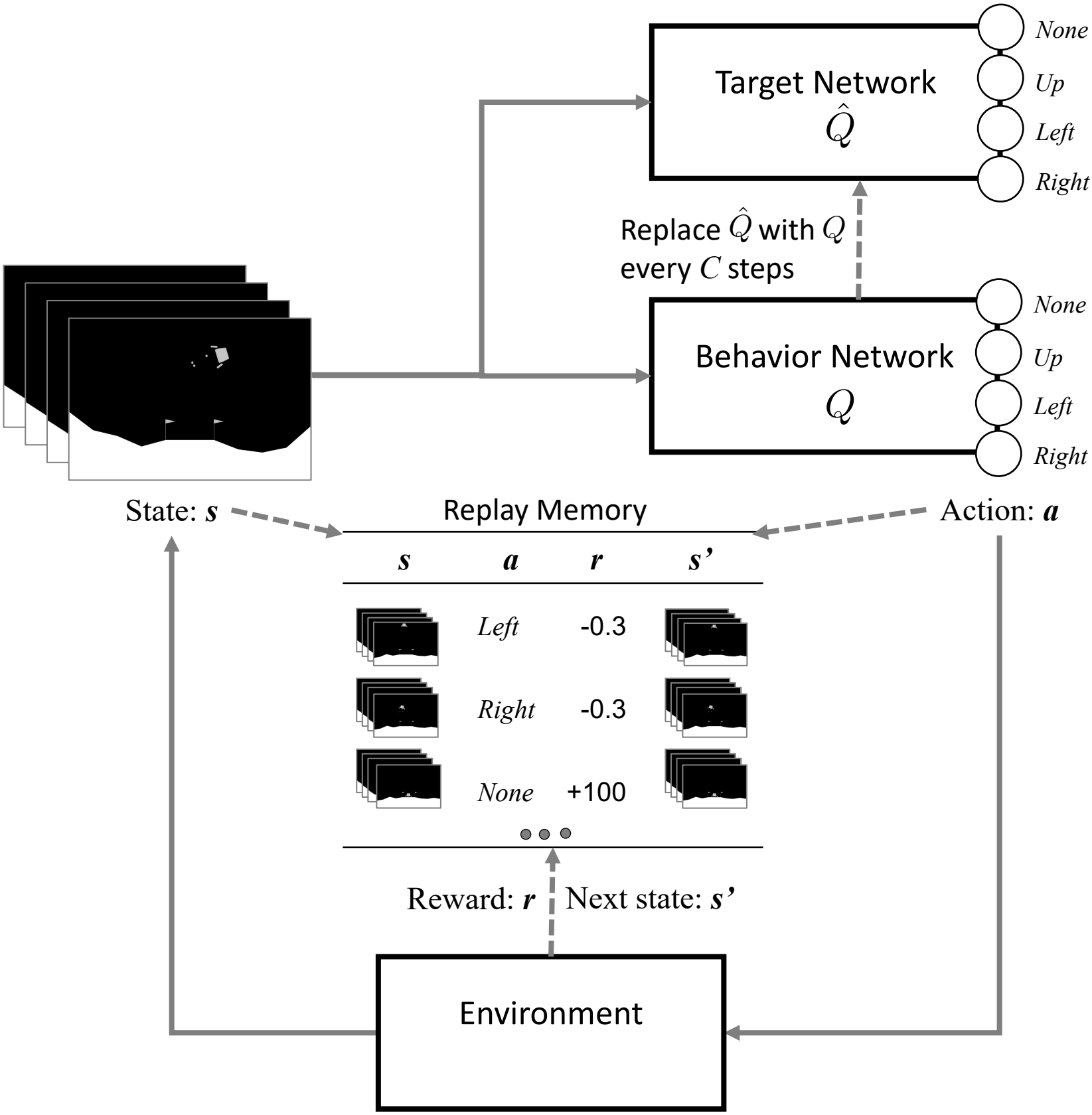
Target Network
$\hat{Q}$
None
Up
Left
Right
Replace $\hat{Q}$ with Q
every C steps
Behavior Network
Q
None
Up
Left
Right
State: s
Replay Memory
Action: a
s
a
r
s'
Left
-0.3
Right
-0.3
None
+100
Reward: r
Next state: s'
Environment