Algorithm Algorithm Title

1: Initialize each agent's Q network; (Q_1, Q_2)
2: **while** Train Episode **do**
3:      Take action and queue the sample to each replay buffer
4:      Obtain random sample batch (s, a, r, s')
5:      Calculate target Q
6:        $y = r + Q(s', a')$ where $a' = \arg\max(Q(s'))$
7:      Update each Q according to basic DQN update rule
8:        $Q \leftarrow Q + \alpha(\gamma y - Q)$
9: **end while**