NO

YES

Current\_Game < NoOfGames

NO

Done flag = True ?

1. Current\_game =0
2. NoOfGames = 100
3. Done flag = False
4. Reset env and observe current state (s)
5. LastAction =0
6. Score =0

NO

YES

Current timestep % target update step = 0

YES

NO

Sample MiniBatch

**CurrentMemory + BatchSize < Total Memory**

Update Policy Nw from Replay Buffer

Update target NW from Policy Nw

NO

YES

1. Store transition of MDP in Replay Buffer
2. S = S’

r = -100

Done and NoLife ?

1. Perform action (a) and observe reward (r) and next state(s’)
2. Score +=r

1 - E

E

NO

YES

NoOfFrames < 3 ?

Action = LastAction

Chose ArgMax action

Chose Random Action

E - greedy

Current\_memory < total ? ?memory

NO

YES

1. Perform action (a) and observe reward (r) and next state(s’)
2. State (S) = Next state (S’)

YES

NO

Main

Done flag = True ?

1. Reset env
2. Done flag = FALSE

Construct Policy Nw Q and Target Network Q

Make Env