| | |
|---|---|
| Number of actors | 256 |
| Actor parameter update interval | 400 environment steps |
| Sequence length $m$ | 80 (+ prefix of $l = 40$ in burn-in experiments) |
| Replay buffer size | $4 \times 10^6$ observations ($10^5$ part-overlapping sequences) |
| Priority exponent | 0.9 |
| Importance sampling exponent | 0.6 |
| Discount $\gamma$ | 0.997 |
| Minibatch size | 64 (32 for R2D2+ on DMLab) |
| Optimizer | Adam (Kingma & Ba, 2014) |
| Optimizer settings | learning rate $= 10^{-4}$, $\varepsilon = 10^{-3}$ |
| Target network update interval | 2500 updates |
| Value function rescaling | $h(x) = \text{sign}(x)(\sqrt{\lvert x\rvert + 1} - 1) + \epsilon x,\ \epsilon = 10^{-3}$ |

Table 2: Hyper-parameters values used in R2D2. All missing parameters follow the ones in Ape-X (Horgan et al., 2018).