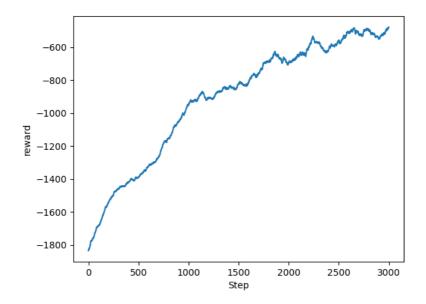
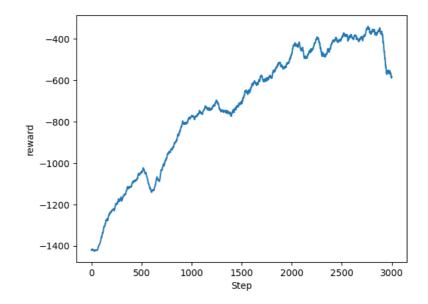
1 A3C

A3C算法的损失函数如下:

$$egin{aligned} L_{ ext{actor}} &= -\log(\pi(a_t|s_t; heta))A(s_t,a_t) \ &L_{ ext{critic}} &= \delta^2 \ &\delta &= R + \gamma V(s_{t+1}) - V(s_t) \ &L &= L_{ ext{actor}} + L_{ ext{critic}} - c \cdot H(\pi) \end{aligned}$$

训练结果如下:





2 DDPG

DDPG算法的损失函数如下:

$$egin{aligned} y &= r + \gamma Q'(s', a'| heta^{Q'}) \ L &= rac{1}{N} \sum (y - Q(s, a| heta^Q))^2 \ J &= -rac{1}{N} \sum Q(s, \mu(s| heta^\mu)| heta^Q) \end{aligned}$$

训练结果如下:

