1   A3C
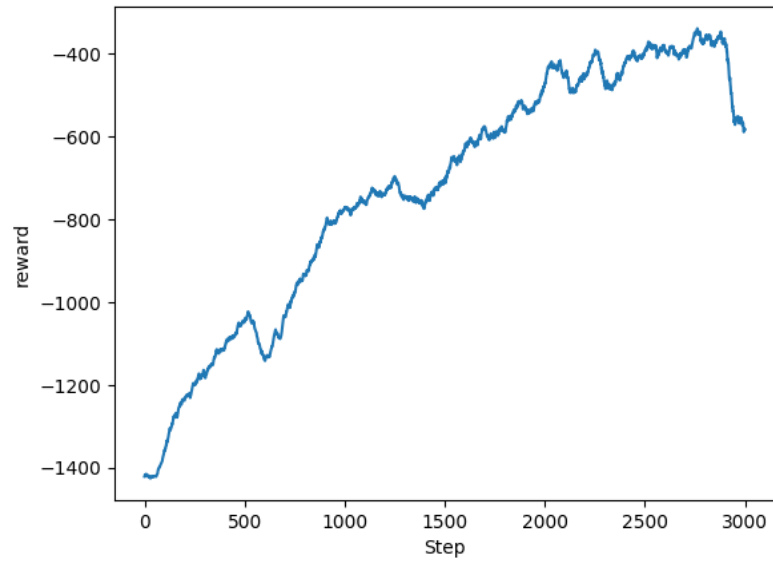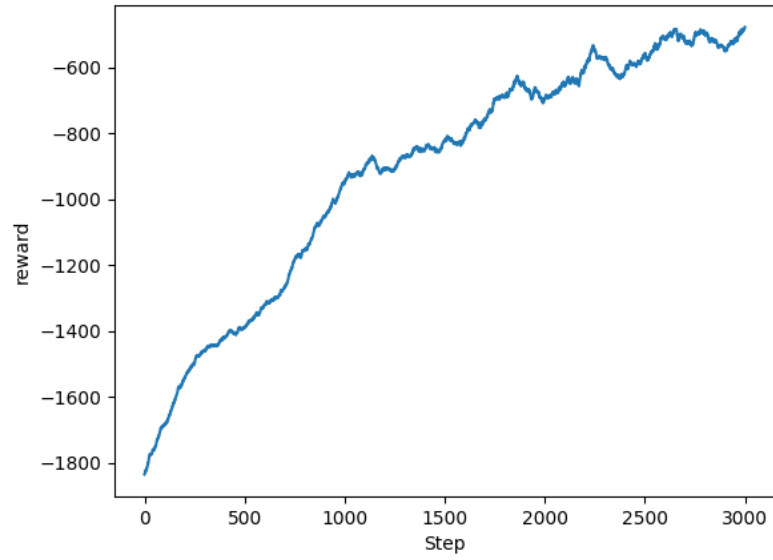
A3C算法的损失函数如下：

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

训练结果如下：

2   DDPG

DDPG算法的损失函数如下：

$$y = r + \gamma Q'(s', a'|\theta^{Q'})$$

$$L = \frac{1}{N} \sum (y - Q(s, a|\theta^Q))^2$$

$$J = -\frac{1}{N} \sum Q(s, \mu(s|\theta^\mu)|\theta^Q)$$

训练结果如下：