

機器人自我調整強化學習演算法

陳楚融

動機

- 目前越來越多機器人被用於遠端作業，當機體殘缺時，無法得到即時的維護，因此希望機器人能擁有適應機體損傷的能力。此研究參考動物的記憶能力，探討預先創造的經驗如何幫助適應傷害
- 若機器人在受損後仍可執行一定程度內的作業，如：行走、搬運等，或自行返回基地，可提升使用效率並降低回收成本



繼續作業

維修站

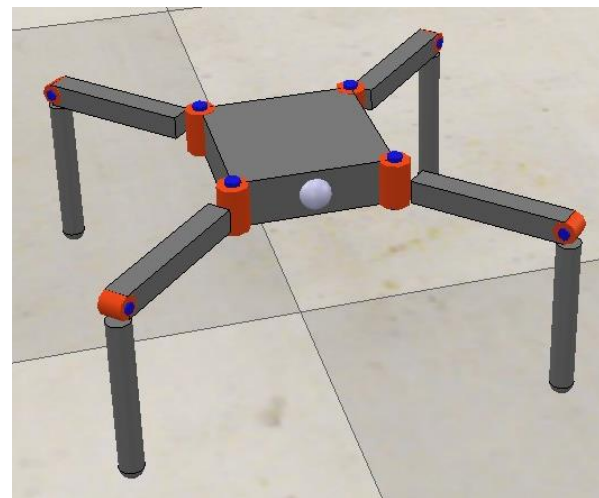
研究方法

- 實驗環境：

- 程式語言：Python3.7
- 機器學習框架：Tensorflow2
- 物理模擬器：CoppeliaSim

- 以模擬器建構四足機器人，
訓練其直線前進

- **Reward Function** 評估項包含：
速度、加速度及馬達輸出力矩大小
速度與加速度可衡量行進穩定度，
透過限制力矩可增加運作效率

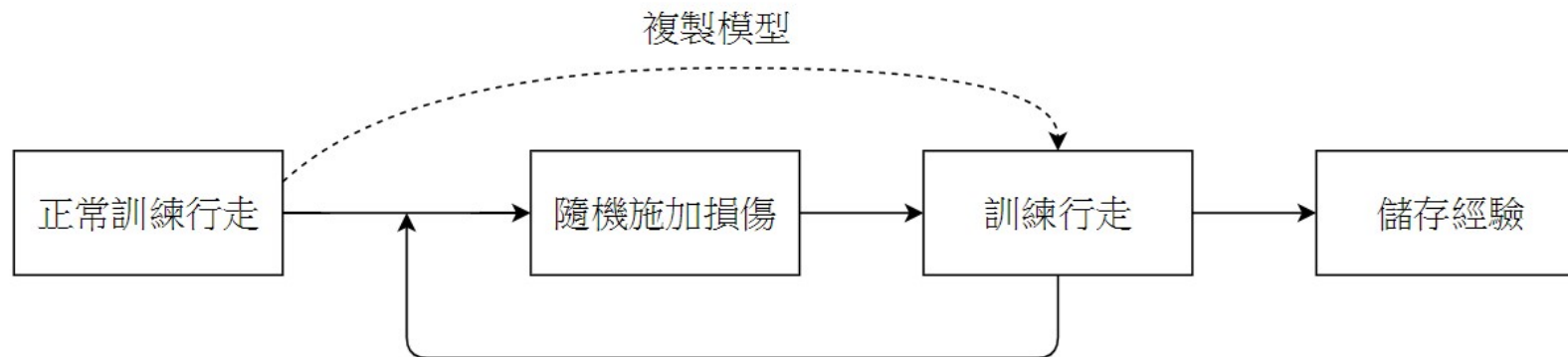


實驗用四足機器人



實驗場景

研究方法



• 實驗流程：

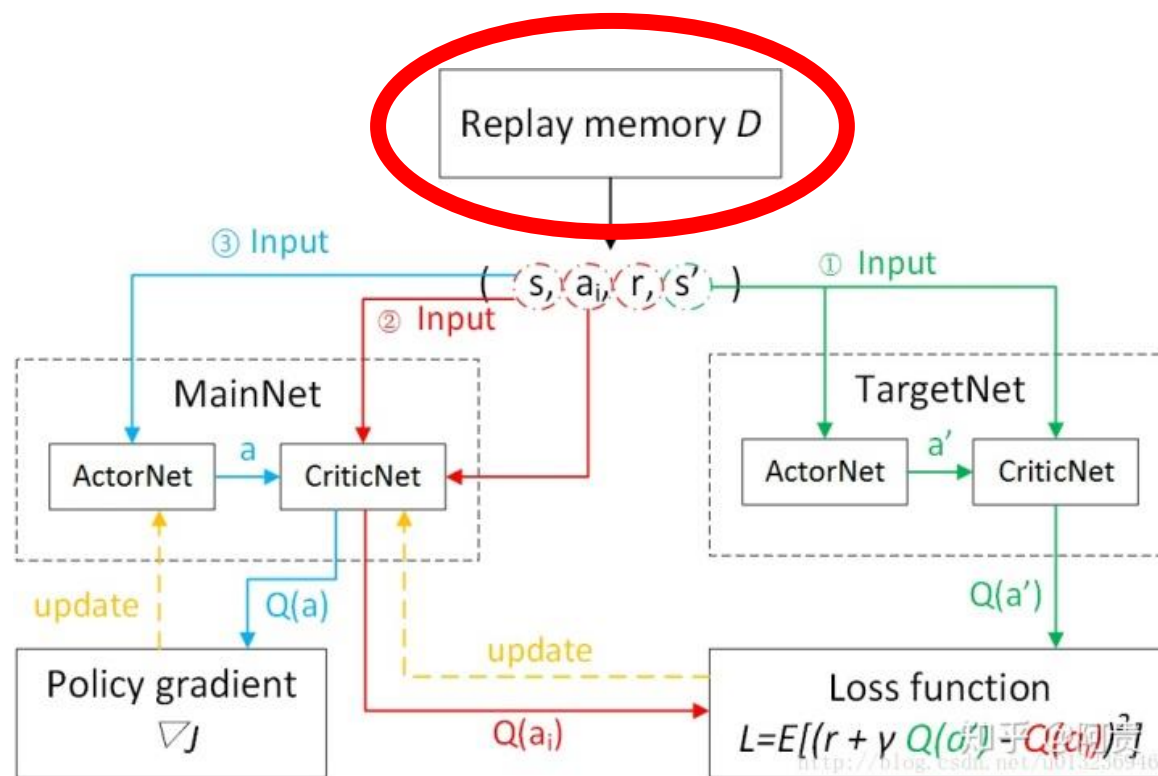
- 以 **DDPG (Deep Deterministic Policy Gradient)** 算法訓練模型在正常狀態下行走。
- 重複執行：
 - 隨機挑選損傷類型施予機體，如：鎖死馬達、肢幹斷裂
 - 訓練行走並蒐集經驗並儲存

	維度	資料
State	25	軀幹：座標 x, y, z 、速度 v_x, v_y, v_z 、傾角 ψ, θ, ϕ 馬達：角度 q_1, q_2, \dots, q_8 、角速度 $\omega_1, \omega_2, \dots, \omega_8$
Action	8	馬達：輸出力矩 $\tau_1, \tau_2, \dots, \tau_8$

研究方法

- **DDPG 算法簡述：**

- 採用兩組網路：**Actor**、**Critic**，類似生成對抗網路，由**Critic**評估**Actor**所採取之動作並更新
- 儲存狀態、動作、獎勵構成之樣本，提供訓練時抽樣。本研究即透過加入預先生成之經驗樣本，使實際受損時可利用相似經驗加速訓練

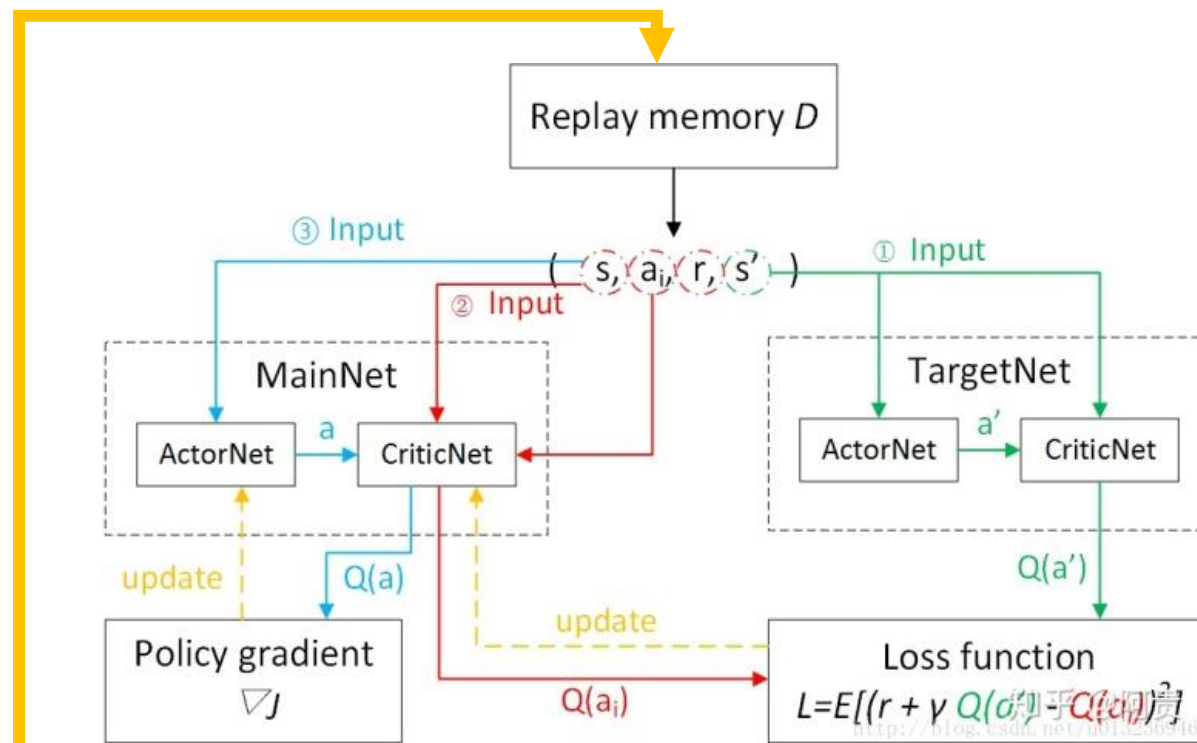
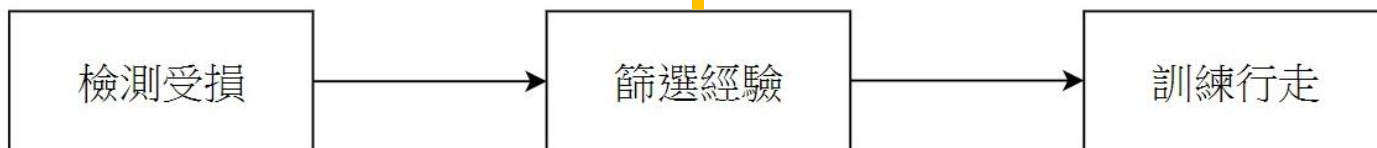


DDPG 流程

研究方法

- 適應流程：

- 檢測表現低於門檻，確認是否受損
- 以行為-表現之關係檢測受損型態
- 篩選相似經驗輔助訓練模型



DDPG 流程

優勢與劣勢

- 優勢：

- 目前已有研究利用動作-表現之映射表進行運算，然而機器複雜化時，此映射可能受維度災難(空間指數性增長)影響。強化學習之網路模型可避免此情況。

- 劣勢：

- 強化學習在複雜環境的控制較難達成穩定。
- 需要事先進行大量訓練

目標

- 能成功增加適應損傷之效率並提升表現
- 拓展至更複雜、不易控制的機體如六足、二足
- 在真實機器人上達到預期功效