

深度強化學習 於沉浸體驗遊戲之 開發

7112056073
7112056466
7112056016

蔡季憲
李展緯
吳昺儒



Table of contents



01.

Introduction

02.

Related work

03.

Design





01.

Introduction

Introduction :

Background:

當前遊戲開發中，創造一個能讓玩家深入沉浸的遊戲環境是**極具挑戰性的目標**。透過深度強化學習，我們不僅能開發出能進行**複雜決策的遊戲AI**，更能創造出與玩家進行真實互動的經驗。這種技術的應用，能夠使遊戲**AI**不僅反應更加自然，還能主動學習和適應玩家的行為，從而提供一種無法預測的遊戲體驗，尤其在**模擬現實互動的遊戲**中顯得尤為重要。



Introduction :

Objective: 「盲人找硬幣」遊戲設計

通過對這個Agent的持續訓練和優化，「盲人找硬幣」展示了深度強化學習在實際遊戲開發中的應用潛力，特別是在提升遊戲AI與人類玩家之間互動的自然性和真實性方面。這不僅對遊戲開發者在技術應用上提供了寶貴經驗，也為玩家創造了更加豐富和吸引人的遊戲世界。這樣的案例分析有助於理解深度強化學習如何推動遊戲設計邁向新的高度，特別是在創造沉浸式遊戲體驗的領域中

Introduction :

Motivation:

- (1)增加玩家的沉浸感 : 透過AI的自我學習和進化，創造一種玩家幾乎感覺不到界限的真實互動。使得每一局遊戲都獨一無二，增加遊戲的重玩價值。
- (2)賦予AI人性 : 持續的學習和環境互動，使AI可以更好地模擬人類對手或合作夥伴的行為，使遊戲體驗更加真實和多元
- (3)提升人工智慧的自主性 : 利用深度強化學習讓AI學習如何在不同情境下作出最佳決策，讓其能夠在不同的環境及前提條件都能嘗試模擬學習出最佳策略

Introduction :

Challenge

(1)高度複雜的環境模擬：

創建一個既真實又充滿未知的遊戲環境，要求我們在技術上做出創新。

(2) AI行為的人性化：

保證AI行為的自然和合理，使得遊戲的每一輪都充滿挑戰和新鮮感。

Introduction :

Solution

(1) 複雜環境設計 :

利用強大的遊戲引擎來構建多變化的遊戲環境，每一輪的遊戲都不相同，增加了遊戲的不可預測性和挑戰性。

(2) 持續的AI訓練 :

透過迭代訓練，持續優化AI的決策樹和行為模式，確保AI能在各種情境中表現得更加自然和智能



02.

Related Work

Related Work

Paper	Objective
OpenAI Five	用深度強化學習訓練AI在《Dota 2》等多人競技遊戲中與頂尖玩家競爭。此項目證實了AI在處理動態隊伍戰略和即時決策中的高效能
Human-level control through deep reinforcement learning:	介紹一種新型代理--DQN，此網路將強化學習與深度神經網路相結合，實現深度強化學習的人類水平控制。
A Study of Human-Like Deep Reinforcement Learning Agents	藉由動態調整獎勵函數維持良好的探索能力，並同時能在訓練後期更有效率地學習排除不像人的策略。最終得以同時像人並擁有良好的表現

Related Work

Paper	Objective
Cognitive Mapping and Planning for Visual Navigation	Cognitive Mapping and Planning (CMP)神經網路架構，用於在新環境中進行視覺導航。對三種不同智能體（反應式、基於LSTM和CMP）的表現進行的詳細比較，發現CMP在所有情況下都表現最佳
Visual Analysis for drone with Reinforcement Learning in Virtual Environment	基於虛擬環境的無人機過框虛擬競賽，使用深度強化學習的方法訓練成功過框完賽的模型，並結合現實中無人機時常運用的 ROS 系統作為指令傳遞的通訊架構縮小虛擬與現實的差異。

Related Work

Paper	Objective
Visualization Analysis for Deep Reinforcement Learning – A Case Study of Side-scrolling Video Game	針對橫向卷軸類型的遊戲環境，用A3C的深度強化學習架構，對環境和獎勵機制做調整，以增強AI進行遊戲的靈活度和適應性。接著，蒐集遊戲歷程和訓練資料。最後，制定視覺化分析。根據分析，可以提高研究人員對模型表現的解讀，降低改良深度學習模型的門檻。
Virtual Reality in Metaverse over Wireless Networks with User-centered Deep Reinforcement Learning	這篇論文介紹了一種基於深度強化學習的多用戶虛擬實境（VR）模型，用於無線網絡中的邊緣計算。主要貢獻包括：針對多用戶情境設計的計算卸載方案，一種新的混合獎勵PPO（HRPPO）算法，以及詳細的DRL場景設計。實驗結果顯示，此方法在不同條件和限制下達到了顯著的成效。



02.

Proposed Design

Proposed Design

遊戲引擎: Unreal Engine(UE)

- (1) 高品質的視覺效果 : 先進的渲染功能，包括光線追蹤和高級的光影效果，這使得遊戲和應用可以達到接近電影級別的視覺效果，更能貼近現實世界，達到深度沉浸的境界
- (2) 完整的開發工具套件 : 豐富的開發工具，包括物理模擬、動畫、音頻處理等，同時還有一個完整的可視化腳本系統
- (3) 躍的社群和支援 : Unreal Engine 擁有一個龐大的開發者社群，提供大量的教程、論壇和第三方資源

Proposed Design

方法設計

- (1) 架構 : 使用DQN演算法，利用了混合型神經網絡架構，結合卷積神經網路(CNN)來分析視覺信息，這能夠讓AI更加精確地理解遊戲環境的動態。
- (2) 獎勵 : 獎勵機制根據遊戲的具體情況動態調整，鼓勵AI在面對短期挑戰和長期目標時做出最適決策
- (3) 適應性訓練 : 實施基於遊戲發展階段和AI學習進度的適應性訓練策略，進行階段性的難度調整和策略優化

Proposed Design

遊戲設計

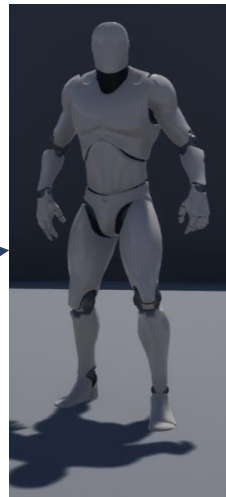
1.遊戲環境：四方型房間且中間有凸起障礙物



2.硬幣(尋找目標)：在模擬的現實環境中隱藏，且可取得硬幣範圍隨著訓練動態縮小



3.Agent 盲人：負責尋找硬幣

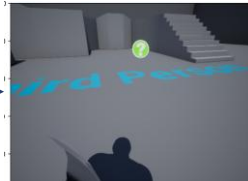


4.reward:

- a.取得目標硬幣 +100
- b.移動步數達最大步數，仍未取得硬幣 -100
- c.結束時，移動步數大於平均步數，每多一步-1，相反則+1

5.Action:人物的前、後、左、右移動，共有四個

6.State: 盲人身上的攝影機的第一人稱視角畫面



Proposed Design

訓練流程 - : 尋找效率訓練

初始化 : 定義目標硬幣位置, 初始化相機參數, 初始化Agent 盲人起始位置

任務 : Agent 盲人需在最短時間內找到硬幣

獎勵機制 : 基於找到硬幣所需的時間來計算獎勵, 時間越短, 獎勵越高

目標 : 訓練Agent 盲人提高尋找效率, 逐漸學會預測和解读硬幣可能的藏匿地點。

Proposed Design

評估與優化

在多次迭代後，統計並分析Agent的表現數據，觀察學習進展和行為演變。使用這些數據來調整訓練參數，如獎勵大小、隱藏點的選擇範圍等，以優化整體的AI表現

重點程式碼片段

```
action = agent.act(state)
x, y, z, pitch, yaw, roll = get_camera_image()

if action == 0: # 向前移動
    client.request('vset /action/keyboard w 1')

elif action == 1: # 向後移動
    client.request('vset /action/keyboard s 1')
    # client.request('vset /camera/0/rotation 0 180 0')

elif action == 2: # 向左移動
    client.request('vset /action/keyboard a 1')
    # client.request('vset /camera/0/rotation 0 -90 0')

elif action == 3: # 向右移動
    client.request('vset /action/keyboard d 1')
    # client.request('vset /camera/0/rotation 0 90 0')

# 觀察新的狀態
res = client.request('vget /camera/0/lin npy')
next_state = read_npy(res)

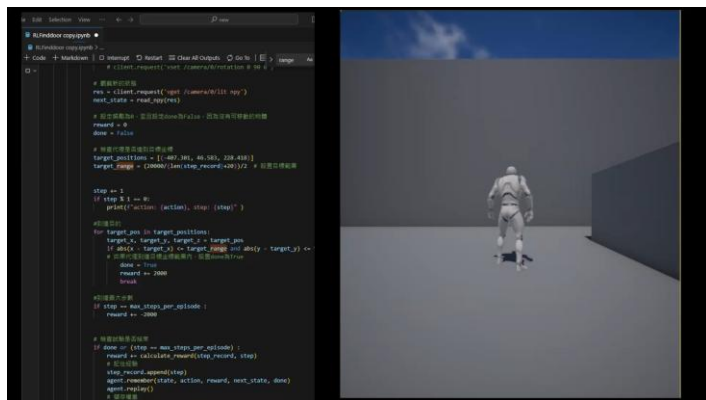
# 設定獎勵為0，並且設定done為False，因為沒有可移動的物體
reward = 0
done = False
```

```
# 檢查代理是否達到目標坐標
target_positions = [(-407.301, 46.583, 228.418)]
target_range = (20000/(len(step_record)+20))/2 # 設置目標範圍
```

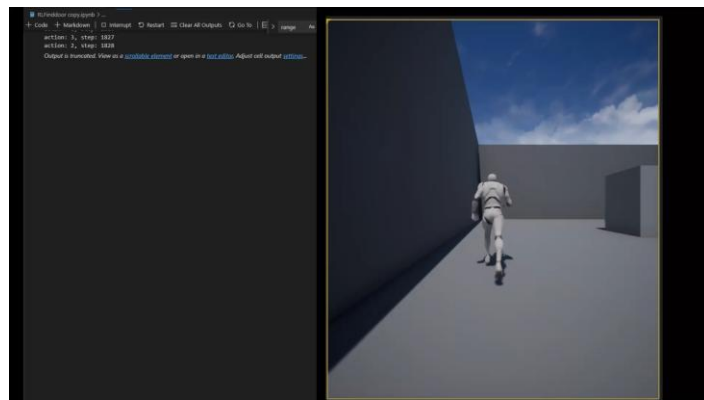
四個給agent盲人執行的action

定義硬幣在3D空間中的座標和隨著訓練去縮小硬幣的判定範圍

Demo 影片



epoch1



epoch100

貢獻

針對3D遊戲中的npc設計一套自主學習模型，讓他可以在遊戲中探索透過玩家的遊戲的指令或操作，自己找到遊戲物件。



Thanks!

