2019 年臺灣國際科學展覽會
優勝作品專輯

作品編號　100011

參展科別　工程學

作品名稱　運用 DDPG 建構氣動式肌肉上臂運動強化學習模型

得獎獎項　大會獎：三等獎

候補作品 2

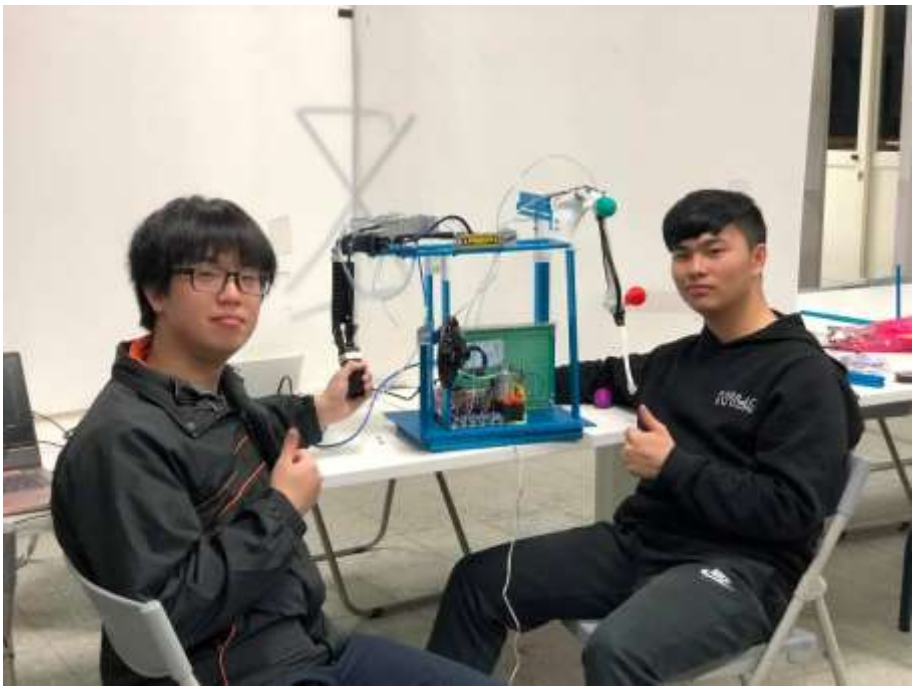
就讀學校　臺北市立內湖高級工業職業學校

指導教師　陳昭安、牛暄中

作者姓名　楊紹民、王典爵

關鍵詞　仿生手臂、強化學習、氣動肌肉

作者簡介



• 楊紹民

我叫楊紹民，我平時喜歡一些有關機器人方面的事務，我認為製作出一台屬於自己且與 別人不一樣的機器人是一件非常酷的事情！因此我也開始專攻研究於機器人這部分。希望能 夠在未來打造一台可以造福人類的機器人！在這科展中，我負責的是硬體規劃、控制電路的 部分，我總是會不斷的更新我們的機器手臂，讓我們的硬體能夠在每一次運行中都能快速又 準確的到達位置。 •

• 王典爵

我叫王典爵，我喜歡撰寫程式來改善自己的生活，以及製作遊戲增添生活樂趣。我認 為虛擬與現實的結合可以使人在忙碌的生活中以一定程度的放鬆與增加壓力，並能透過思想 上的發散與收斂來令自己有所成長。在這個科展中，我負責撰寫人工智慧與影像處理的程式 技術，透過人工智慧與仿生手臂的結合，我希望讓這類仿生機器人能夠更有智慧，並且更加 貼近我們生活。

摘要

研究探討透過強化學習讓機器學習各種人類上臂運動。延續「運用氣動肌肉缸模擬上臂肌肉控制之研究」，透過有限的動作組可以控制上臂肌肉，然而因應環境條件的多變，模擬人類透過學習產生多樣多變的反應，在仿生的領域中有其必要。比較強化學習中的 Actor-Critic 與 DDPG (Deep Deterministic Policy Gradient)兩種模式，我們透過 Gym 建構具動作與環境限制的簡易訓練環境。比較兩個模型的細節後，最後選用了 DDPG 為我們主要的強化學習方法。首先我們利用 Tensorflow 模擬學習模式並記錄模擬移動的學習過程。我們運用到仿生手臂的實體，藉由影像辨識取得手臂的狀態，回饋至學習模型。仿生手臂運用學習資料進行移動，接著我們觀測系統所學習的移動是否可完成指定動作或工作。在軟體模擬中，我們證實了藉由達成數次目標的學習後，DDPG 可完成較細緻的移動。而 DDPG 在仿生手臂上的實作，則需透過輸出動作給氣動仿生手臂系統，來控制仿生手臂移動至目標位置。在未來，機器人將不單單只是運用馬達來當作動力來源，也能運用氣動肌肉缸成為動力元件。並且，人形機器人將會做出更像真實人體的動作。

Abstract

This study explores the use of reinforcement learning to allow machines to learn a variety of human upper arm movements. Continuing the study of "Pneumatic robotic arm device simulating human upper arm muscle movement", the upper arm muscles can be controlled through a limited number of movements. However, within the changing environment, to simulate diverse and varied response of human beings after learning is necessary in the field of bionics. Comparing the two models of Actor-Critic and DDPG (Deep Deterministic Policy Gradient) in reinforcement learning, we construct a simple training environment with movements and environmental constraints through Gym. After comparing the details of two models, we finally chose DDPG as our main reinforcement learning method. First we use Tensorflow to simulate learning patterns and record the learning process of simulated movement. We apply the entity to the bionic arm to obtain the state of the arm by image recognition and feed back to the learning model. The bionic arm uses the learning material to move, and then we observe whether the movement learned by the system can complete the specified action or work. In the software simulation, we confirmed that DDPG can perform more detailed movements after learning through several goals. The implementation of DDPG on the bionic arm requires an output movement to the pneumatic bionic arm system to control the bionic arm moving to the target position. In the future, robots will not only use motors as a source of power, but also use air muscle as power components. The humanoid robots will make movements that are more like real humans.

壹、 前言

一、 研究動機

在生態界裡，動物後天的行為是為了能夠達成特定的目的，透過在環境中不斷的學習及嘗試而獲得的行為。在人工智慧的領域中，強化學習便是藉由不斷的訓練代理(Agent)來獲取獎勵(Reward)，得出能隨著環境改變而作出相對應動作的模型。延續「運用氣動式肌肉缸模擬上臂肌肉控制之研究」，目前運用氣動肌肉所製作出的機械手臂與利用馬達所組合成的機械手臂上，有個極大的差異，那就是準確度的問題，有些馬達能夠將自身的為了解決這項問題，我們希望能夠結合機械學習的部分，讓系統透過學習精準的到達目標位置，建構出一套具有學習能力的仿生機器人，並希望以這個做為基礎，建構出局部甚至是全身的仿生機器人。

圖 1 波士頓機器人

二、 研究目的

1. 探討氣動肌肉缸的基本物理性質
2. 探討人體骨骼及肌肉構造並分析
3. 探討骨骼模型控制手臂彎曲角度變化量
4. 分析並建構強化學習的學習環境

   在多種強化學習方法中分析並選擇一種作為本研究之主要研究對象，並依據仿生手臂設備建構出相應的虛擬學習環境。
5. 探討如何結合強化學習與仿生手臂控制

   探討如何以強化學習方法為基礎，將強化學習以及仿生手臂設備相結合，並加以控制，達到在實體空間學習與應用的成效。

貳、 研究方法或過程

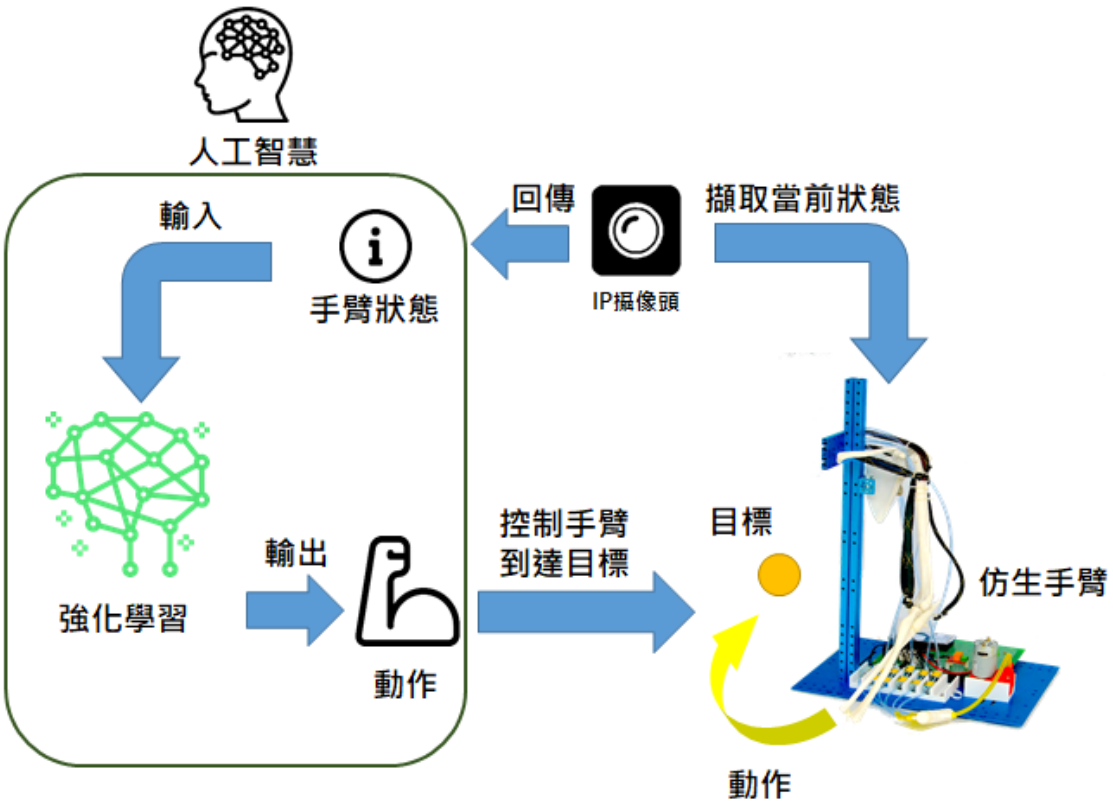


圖 2 研究流程圖

本研究流程為：架設仿生手臂設備，透過 IP 攝像頭擷取仿生手臂與目標的影像資訊，作為強化學習的狀態輸入，透過強化學習方法以使仿生手臂接觸目標為目的，輸出動作值並控制仿生手臂，達成不斷使仿生手臂接近目標的循環。

一、人體肱肘部骨骼及肌肉構造探討

為了模擬出更真實的人體動作，我們希望能夠建構出仿人體骨骼構造的仿生機器人，因此我們開始研究有關人體骨骼及肌肉構造的相關研究。

（一） 肘部

肘部重要的結構包含骨和關節、韌帶和肌腱、肌肉、神經、血管，肘部骨頭分為 肱骨、尺骨、橈骨，外(內)側有個骨性突起物，稱為外(內)上髁，大多數負責伸(屈)腕和伸(屈)指的肌肉都匯聚於此。

（二） 軟骨

橈骨與尺骨交界處為平滑，橈骨頭邊緣附蓋關節軟骨，關節軟骨是任何關節中附著在骨骼末端最重要的物質，呈白色、有光澤、如橡膠般有彈性並且很光滑，使骨頭磨擦時不會受到傷害並吸收震盪。

（三） 肌肉

肱二頭肌是使到手臂彎曲的肌肉，它們的肌腱止於撓骨粗隆和前臂筋處，基本功能是「彎舉」。它是由肘關節為轉動點組成「單關節」活動。從實踐證明，使肱二頭肌處於「頂峰收縮」狀態，它的最佳夾角為 50°~55°左右；肱三頭肌肌腱從上臂後側穿過，連接到尺骨上，使肘關節「伸直」運動。

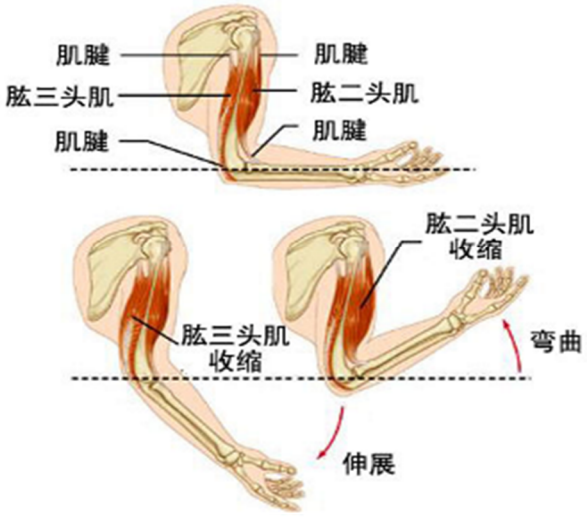


圖 3 肘部肌肉的運作說明圖

（四） 韌帶與肌腱

1. 韌帶

韌帶為可彎曲、纖維樣的彈性結締組織，連接骨與骨，附於骨的表面或與關節囊的外層融合，以加強關節的穩固性，以免損傷。內側副韌帶位於肘關節內側邊緣，外側副韌帶位於肘關節外側邊緣，兩條韌帶共同連接肱骨與尺骨，保持尺骨穩定於肱骨末端滑動。

2. 肌腱

肌腱是一堅韌的結締組織帶，通常將肌肉連接到骨骼，並可承受張力。腱類似韌帶和筋膜，都是由膠原蛋白組成。不過，韌帶連接骨骼，而筋膜則連接肌肉，當肌腱與肌肉一起作用則可產生動作。

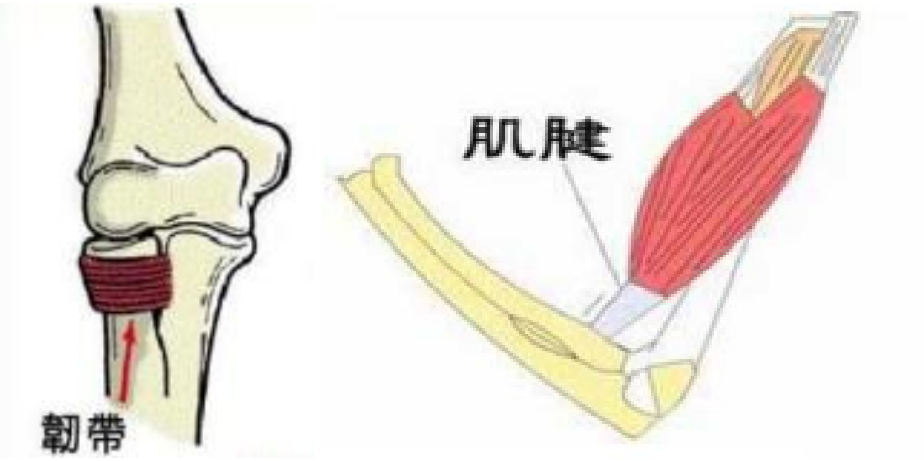
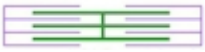
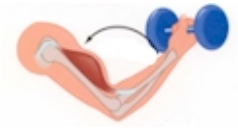
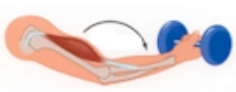
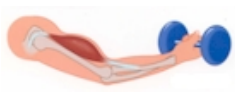


圖 4 韌帶與肌腱說明圖

（五） 肌肉收縮的類型

根據研究與文件探討有關肌肉的收縮情形，大致上得到了如下表格的結論：

表 1 肱二頭肌的收縮運動

肌肉收縮	等張收縮		等長收縮
	向心收縮	離心收縮	
長度	縮短 	伸長 	不變 
張力	轉變	轉變	轉變
速度	轉變	轉變	沒有動作
例子	肱二頭肌 收縮並縮短 （用力>重量） 	肱二頭肌 收縮但伸長 （用力<重量） 	肱二頭肌 收縮但長度不變 （用力=重量） 

二、　氣動肌肉缸研究

（一）　氣動肌肉缸設計

利用編織網當作外部材質，當氣球膨脹時，編織網隨之收縮，以利氣球表面平均承受內部氣體壓力，並且達到長度縮短，截面積變寬的現象，此現象也將會產生一股拉力。

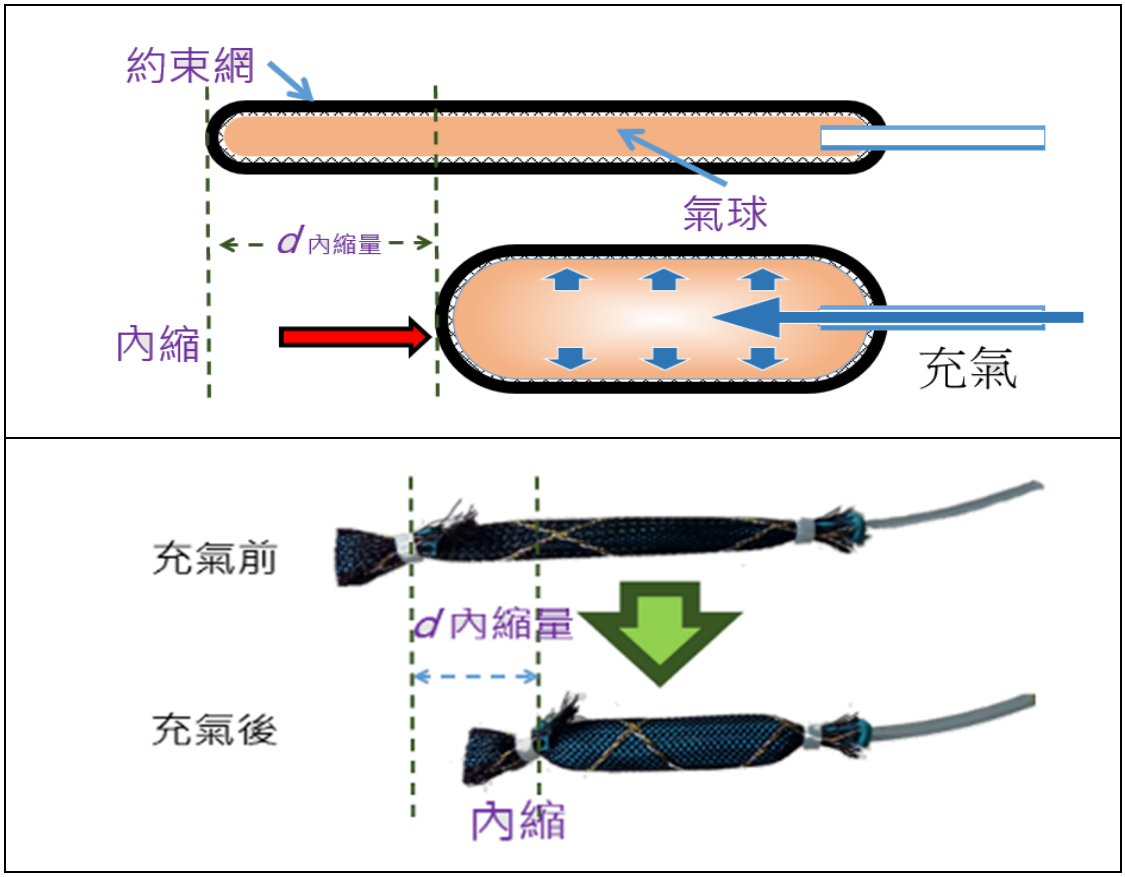


圖 5　氣動肌肉肛的運作原理

（二） 氣動肌肉缸製作方法及流程

藉由上述的設計構思與幾次的製作與改良，大致上得到了如下述的製作方法及步驟。

表 2 氣動肌肉缸製作流程

	
1. 量取欲做長度+5cm(預留範圍為綑綁用)的氣球並剪取。	2.將約束網套套至氣球。
	
3.尾端及開端綑上束帶防止氣體流失。	4.完成人工肌肉製作。

1. 改變內部結構材質

生膠管在於製作氣動肌肉缸時，是一款較優質的材料，其優點也更為接近肌肉構造。但同時也需要較大的氣量驅動 ，經實驗發現，空氣幫浦的氣量略為不足。因此，往後的實驗平台與骨骼模擬中，使用的是較方便操控的氣球。

表 3 氣球與生膠管之比較

	氣球	生膠管
圖片		
優點	容易取得、穩定優、方便操控	抗壓好、拉力強、耐用
缺點	容易爆裂、抗壓性不佳、拉力微弱	所需壓力大、不易取得、長度不易掌控、穩定性不佳

2. 輸氣系統

在模擬氣動肌肉缸的運作過程中，以打氣筒灌氣的氣體量有限，無法持續輸氣。所以將打氣筒改為空氣幫浦、水幫浦等不同送氣方式，再去探討不同送氣方式的差異。經實驗發現，空氣幫浦為最佳送氣來源，且能運用 Ardiono 進行控制，以利於往後的實驗操作。

		
打氣筒	空氣幫浦	水幫浦

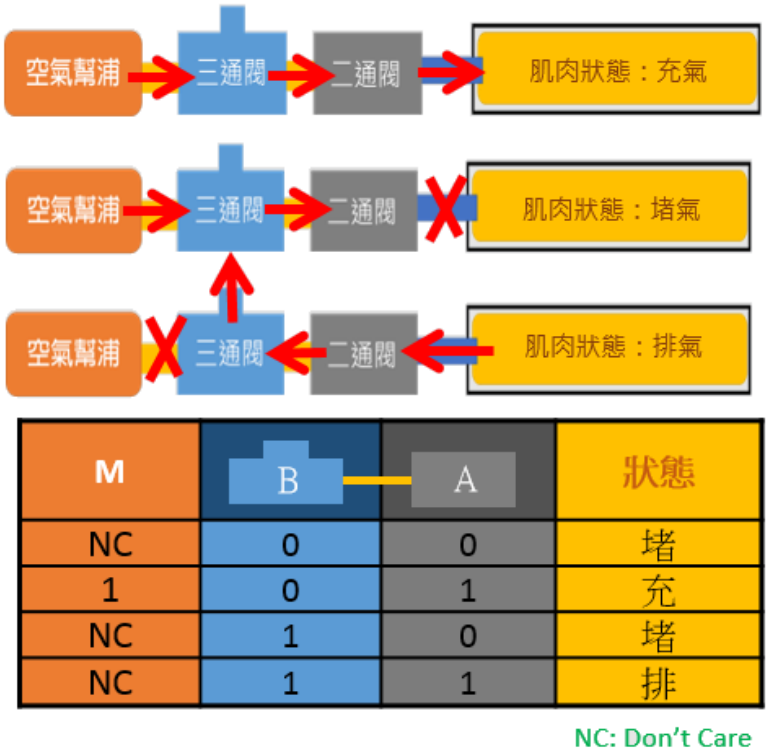


圖 6 肌肉控制流程圖

三、　人體骨骼構造製作

模擬肱二頭肌的實驗當中，若使用真實人體骨骼進行實驗，需要注入龐大的氣體量，為了利於觀察並設計，故採用小型的人體骨骼構造進行模擬。

（一）　3D 列印出肱部及肘部的骨骼模型。

圖 7　骨骼 3D 模組

（二）　運用氣球製作的人工肌肉，並兩端分別固定於骨頭兩側。

圖 8　氣動肌肉缸固定肌肉於骨骼的情況

（三）　將幫浦的出氣口與氣動肌肉缸緊密相接。

圖 9　氣動肌肉缸固定肌肉於幫浦的情況

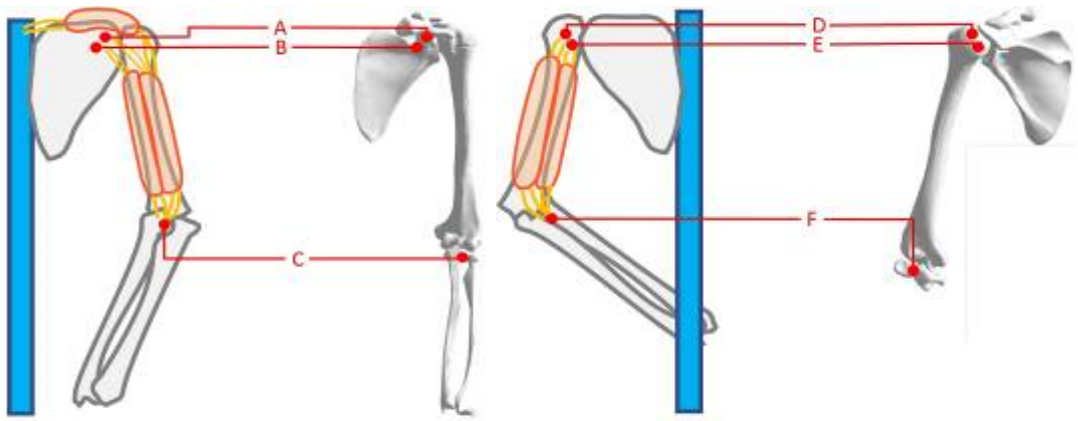


圖 10 骨骼背視圖

採用氣球製作的人工肌肉需要的氣壓比生膠管小，簡易的幫浦和電磁閥便可以控制內部壓力模擬肌肉的舒張、收縮，本模型藉由此方式，可輕鬆的模擬人體手臂構造，同時也較為輕便。

（四） 侵蝕與擴張

為形態學處理，運作後結果圖像形似收縮與擴大，故得其名，可經由運算去除影像雜訊以及達到連接破碎前景物的功能，其主要是用二值化影像的膨脹(Dilation)與收縮(Erosion)的組合達到消除雜訊的目的。

四、 影像辨識

為了更實際的模擬人體手臂運作，加入影像辨識技術，結合機器學習達成模擬人體實際手臂動作。

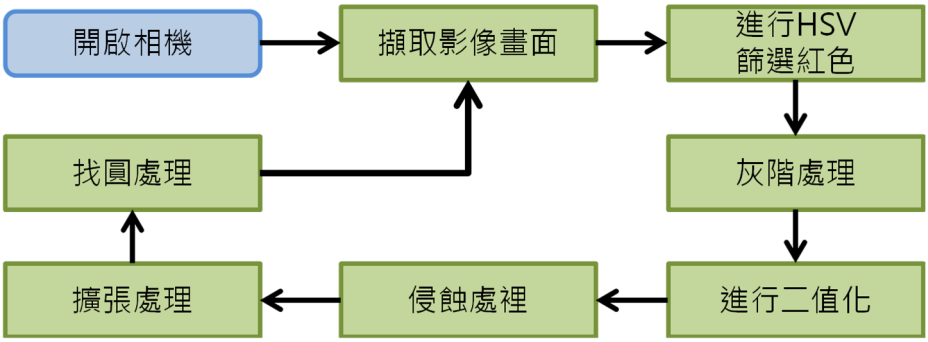


圖 11 影像辨識流程圖

（一） HSV 篩選紅色

是一種將 RGB 色彩模型中的點在圓柱坐標系中的表示法。

色相（H）： 色彩的基本屬性，就是平常說的顏色名稱，如黃色、藍色、紅色，取 0°-360°。

飽和度（S）： 是指色彩的彩度，數值愈高色彩愈鮮豔，愈低則愈淡，接近白色，取 0-100% 的數值。

明度（V）： 亮度，愈高顏色愈明亮，愈低愈靠近黑色，取 0-100%。

把顏色描述在圓柱坐標系內的點，這個圓柱的中心軸取值為自底部的黑色到頂部的白色而在它們中間的是灰色，繞這個軸的角度對應於「色相」，到這個軸的距離對應於「飽和度」，而沿著這個軸的高度對應於「亮度」，「色調」或「明度」。

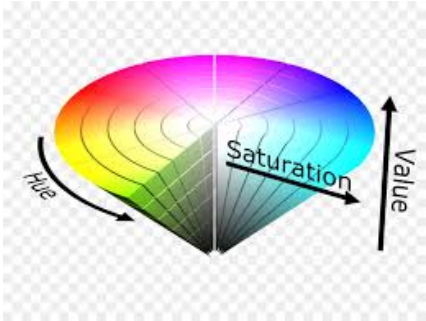


圖 12 HSV 3D 模型截面圖

（二） 灰階

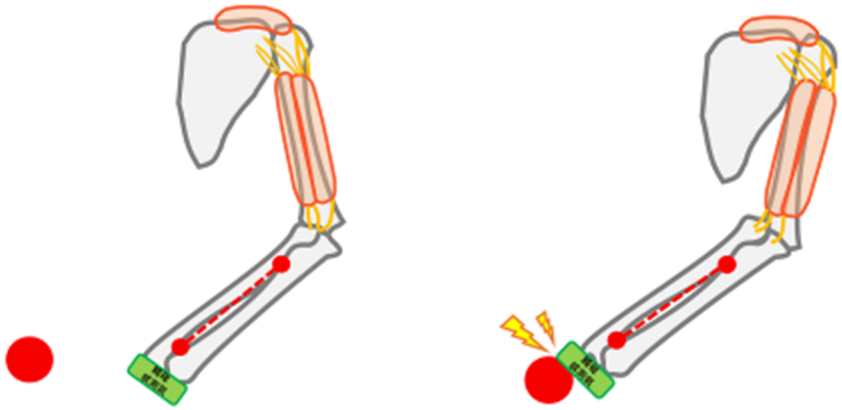
灰階影像除了純黑和純白之外，還可以包含深淺不同的灰色，即在黑色與白色之間加上不同的明暗度，每個像素佔有 8 bits 的空間，所以明暗度也就有 256 種變化，但也僅止於由黑到白的亮度變化而已，無法出現紅色、綠色等其他色彩，其利用的目的就是方便二質化。

（三） 二值化

是圖像分割的一種最簡單的方法，二值化可以把灰度圖像轉換成二值圖像。把大於某個臨界灰度值（閥值）的像素灰度設為灰度極大值（即白色），把小於閥值的像素灰度設為灰度極小值（即黑色），從而實現二值化。

五、　機器學習方法探討

人工肌肉與馬達最大的差異就是準確度以及回傳值的部分，人工肌肉需藉由其他感測器才可得知目前的位置，而馬達可直接將數值回傳。上圖(**錯誤！ 找不到參照來源。**)為本研究之系統架構，透過影像辨識，可以取得仿生手臂的狀態，成為強化學習模型的輸入，並透過強化學習模型輸出相應的動作，藉此接近目標。建構具有學習能力的仿生機器人。



(a)初始狀態　　(b)達成目的後狀態

圖 13 機器學習示意圖

在機器學習中，本作品預計使用強化學習的演算法，其方式為給定獎勵與處罰(Reward)，經由在環境中接收狀態，並產生動作，判斷產生的結果是否符合預期，若符合，則給予獎勵，並驅使結果朝著獎勵走，如不符合預期則反之，達到完成目的的學習效用。

如上圖(圖 13)之示意圖所示，觸碰到圖中大紅點為其學習目的，故於手臂前端加入觸碰感測器，起初經由給定亂數進行嘗試，直至達到目的為止，即為完成一次的學習，隨著學習成功次數逐漸增高，同時意味著學習的完整性。

我們利用以下分類標準來初步篩選出欲使用的強化學習演算法，並將對於本研究較有優勢者加上星號(*)標示。

對於本研究具有優勢的演算法須具備如以下優勢

- 因應不定的充放氣時間：輸出動作具連續性
- 能夠快速即時的訓練與反饋成效：及時修正動作

（一） 決策標準

在訓練過程中，選擇下一個動作時所做的決策標準。

1. Value based（基於價值）

在動作中直接選擇價值最高者，適合用於環境簡單且動作離散(discrete action)的代理。

2. Policy based（基於策略）

*依照概率隨機選擇動作，適合用於環境較複雜且隨機性且動作連續(continuous action)的代理。

3. Value & Policy

*結合價值與策略，利用 Value based 的算法對 Policy based 的算法進行調整，幫助 Policy based 算法做決策。

（二） 更新方法

於訓練時更新策略時所使用的方法，主要以更新頻率做為區別。

1. Monte Carlo update（蒙地卡羅方法）

每次訓練回合結束後才調整策略。

2. Temporal Difference update（時間差分方法）

*不必等到一個訓練回合結束才更新，每一步都會進行調整。

（三） 更新參考策略

於訓練時更新策略時所使用的策略，主要以更新的參考對象做為區別。

1. On Policy

更新策略時，直接依據更新訓練回合當下的策略。

2. Off Policy

*更新策略時，同時參考過去的策略與當前的策略。

下表(表 4)為我們參考資料後對於幾種強化學習方法的架構所進行的分類

表 4 強化學習方法比較圖

分類依據 / 學習方法	決策標準	更新方法	更新參考策略
Monte-Carlo Learning	Value based	Monte-Carlo	*Off Policy
Q Learning	Value based	*Temporal Difference	*Off Policy
Sarsa	Value based	*Temporal Difference	On Policy
DQN	Value base	*Temporal Difference	*Off Policy
Policy Gradients	*Policy based	Monte-Carlo	*Off Policy
Actor-Critic	*Value & Policy	*Temporal Difference	*Off Policy
DDPG	*Value & Policy	*Temporal Difference	*Off Policy
A3C	*Value & Policy	Monte-Carlo	*Off Policy
DDPO	*Value & Policy	*Temporal Difference	On Policy

由於 Actor-Critic 與 DDPG(Deep Deterministic Policy Gradient)在此分類下的優勢是相同的，因此我們進一步的分析與比較此兩種的特點。

（一） Actor-Critic (下稱 AC)

由兩個神經網路構成，每一步都由 Critic 對 Actor 進行調整，提高更新效率。

1. Actor (演員) 神經網路

一個決策標準為 Policy based 的神經網路，通常使用 Policy Gradients，其優點為能依照概率於連續動作空間中選出動作，並受 Critic 的影響調整動作的概率，Actor 神經網路為較積極探索的神經網路。

2. Critic (評論家) 神經網路

一個決策標準為 Value based 的神經網路，其優點為能在每一步就進行更新，並用於函數逼近。Critic 神經網路藉由學習環境與獎勵的關係，預測當前狀態的潛在獎勵，透過為 Actor 的行為評分來修正 Actor 的神經網路。

（二） Deep Deterministic Policy Gradient (下稱 DDPG)

基於 Actor-Critic 演算法的優化方法，Policy Gradient 的部分使用 Actor-Critic，加入了 DQN 演算法的精隨，增加神經網路，也利用 Deterministic 策略來改變隨機輸出動作值為直接輸出動作值。

1. Deterministic

在 Actor-Critic 中，在連續動作中做出決策是基於概率，認為在同一狀態中的連續動作空間為常態分佈。而 DDPG 透過 Deterministic 的決策系統，做出連續動作的選擇更加的果斷，直接選定動作，提升輸出動作時的穩定性。

2. Deep Q Learning

Deep Q Learning 演算法的特色在於，將神經網路再細分為現實與估計網路，故在 DDPG 算法中，Actor 與 Critic 將會以延伸出現實網路與估計網路(Target u)。

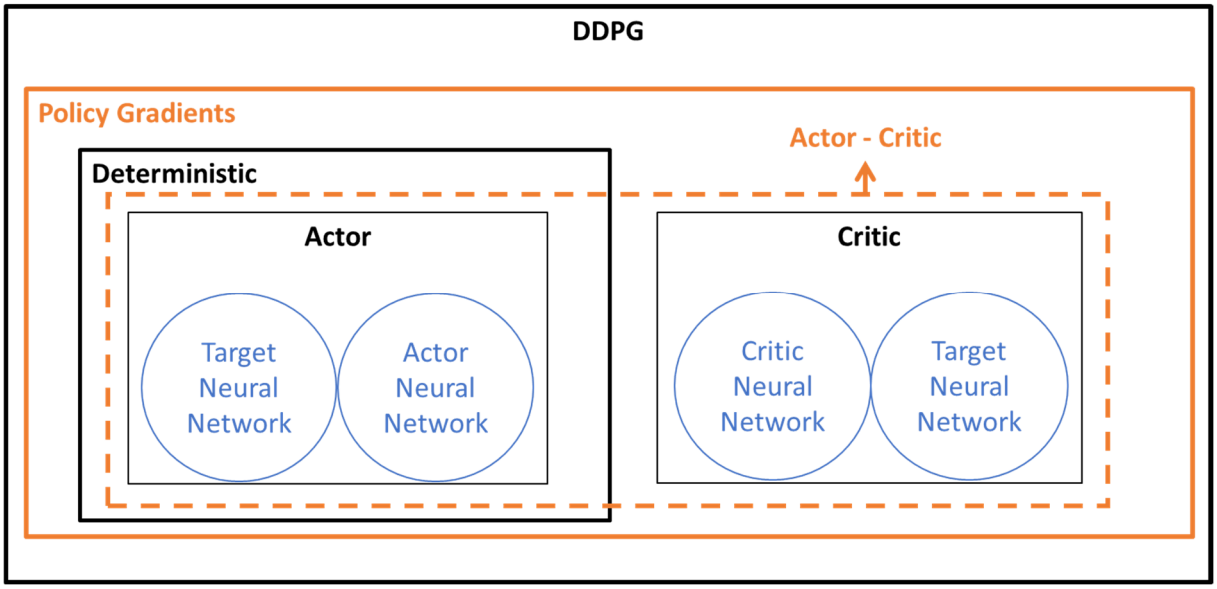


圖 14 DDPG 演算法架構

（三） 比較 AC 與 DDPG 在實作上呈現的差異

本研究選擇 OpenAI 所提供的 gym 作為環境搭建的依據，並選用 Pendulum-v0 鐘擺環境，在這個鐘擺的環境中，鐘擺會以隨機的位置開始，目標是將其向上擺動，並使其保持直立。簡單的目標任務以及連續的輸出動作，較接近本研究的需求，故我們將此環境作為兩種演算法比較的基礎。

下圖(

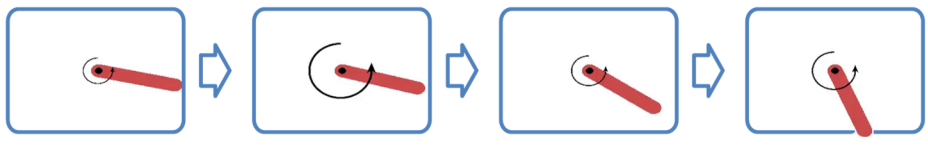


圖 15)為連續動作的表示，紅色棒狀物為鐘擺，固定於畫面中央，受物理模擬的引力影響，目標為使其直立。圍繞著鐘擺的順、逆時針箭頭為一時間單位所輸出的力，其半徑愈大表示輸出的力值愈大。

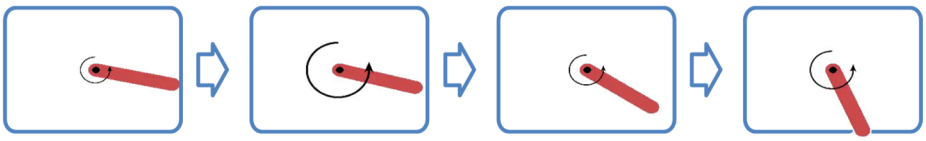


圖 15 Pendulum-v0 環境連續動作圖

六、 建構強化學習環境

本實驗目的為：藉由強化學習方法在仿生手臂的種種限制下控制仿生手臂使手臂前端接近目標物。

為了捕捉仿生手臂的動作，利用不同顏色的球狀標示物裝置在仿生手臂的肩部、肘部與手臂前端、以及目標物，透過影像辨識找圓的方法，定位出手臂各端點與目標物的位置，進而取得仿生手臂在環境中的狀態。

由於在實體空間中訓練機器學習通常需要耗費大量的時間，並容易受環境中極大的隨機性影響，故本實驗透過建構虛擬環境來模擬真實環境。

以下為觀察與分析實體結構。

表 5 分析仿生手臂結構表

狀態	限制
肩部座標	肩部座標固定與角度限制
肘部座標	前臂一端必須連接上臂前端，使肘部不斷接
前臂前端座標	前臂前端座標不得與肩部座標重合
前臂與上臂夾角	前臂與上臂的最小夾角需介於 60°-173° （above 探討骨骼模型控制手臂彎曲角度變化量）

獎勵設計：前臂前端對於目標物在平面上的直線距離為一負值獎勵，距離愈遠則獎勵愈低；前臂前端直接接觸到目標物就得到正值獎勵，接觸時間愈長則獎勵愈高。

我們決定建構一個虛擬環境令強化學習模型能在其中訓練，是由於虛擬環境具備快速方便且安全的優勢，能夠大幅提升學習模型的速度。我們所利用的環境使用了兩個相連結的支架與隨機可動的目標點，來對應現實的骨骼與實際目標。訓練過程則是以支架的端點越靠近目標點而得到更多的獎勵作為訓練目標，並因應不同環境來學習不同的肢體動作。

圖 16 學習環境示意圖

參、 研究結果

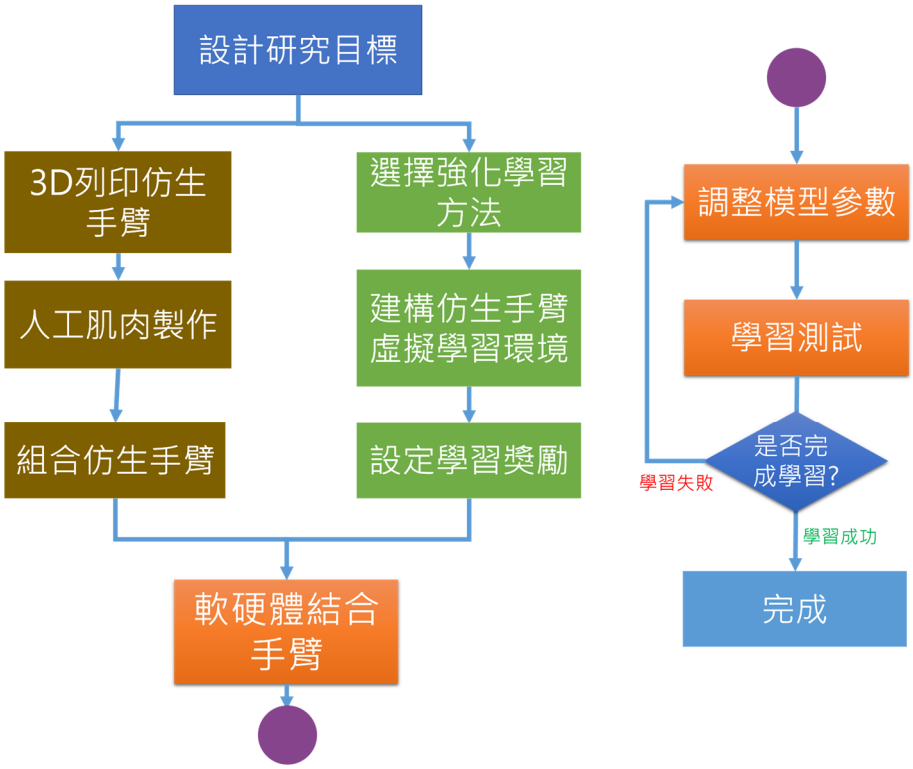


圖 17 研究架構

一、氣動肌肉仿生上臂

（一） 空氣幫浦送氣量測試

在取得空氣幫浦作為動力來源時，首先最重視的不外乎就是幫浦的送氣效能，為了有利於往後的實驗數據測量，故撰寫了控制幫浦送氣時間程式並進行細微的控制，且將幫浦的送氣量與時間關係進行對照並記錄如下列圖表。

表 6 幫浦送氣量(ml)與時間對照表

次數 / 秒數(Sec)	第一次測量	第二次測量	第三次測量	平均值 (Avg)
0.1	8.1	8.1	8.1	8.10
0.2	16.1	16.1	16.1	16.10
0.3	24.0	24.0	23.9	23.96
0.4	31.8	31.9	31.9	31.83
0.5	39.9	39.9	40.0	39.93
0.6	48.1	48.0	48.0	48.03
0.7	56.0	55.9	56.0	55.96
0.8（推算）				63.89
0.9（推算）				71.82
1.0（推算）				79.75

由於在本測試實驗中所使用的為打氣筒，故只能測量至 60ml 的測量刻度，但經由分析後發現，所繪製成的圖形近似於線性關係，經由這個關係我們可以推估往後進氣量的值，將其結果擷取部分繪製成圖表，如下圖所示。

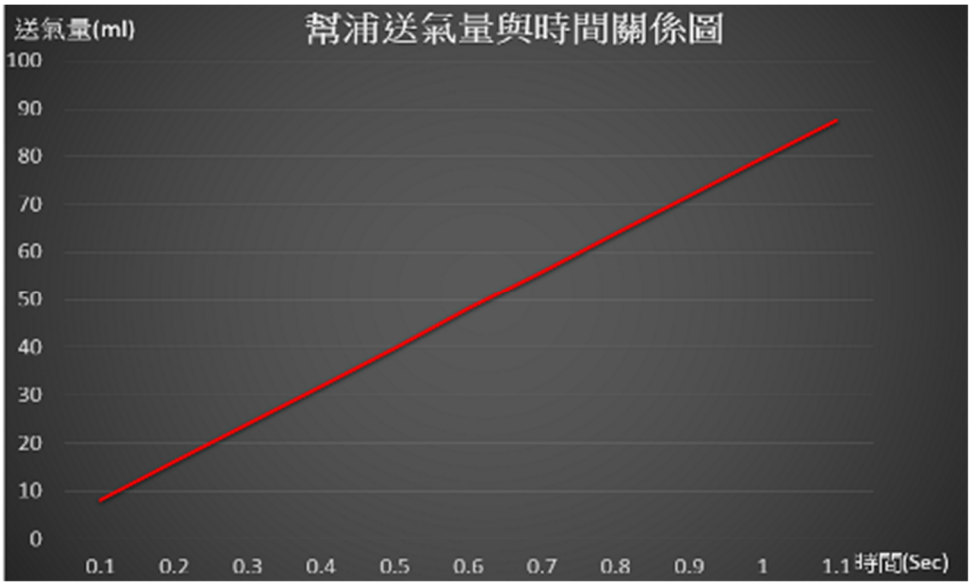


圖 18 幫浦送氣量與時間關係圖

（二） 探討氣動肌肉缸充飽後的長度變化量與進氣關係

當測試完幫浦的進氣量與完成氣動肌肉缸(以下簡稱肌肉)的製作後，可加以測量各不同長度肌肉的收縮效能，並依照所需的進氣時間，對照幫浦的送氣量，取得估算的肌肉進氣量，有利於往後對於肌肉的進氣量與角度計算等更多作用。

表 7 氣動肌肉缸充飽後長度變化與進氣關係表

肌肉長度（*cm*）		變化量 *d*（*cm*）	充氣時間（*Sec*）	效能（*d/l*）	進氣量（*ml*）
原始長度（*l*）	充飽長度				
8.0	5.9	2.1	0.5	26.25%	39.93
10.5	7.8	2.7	0.6	25.71%	48.03
12.0	8.6	3.4	0.8	28.33%	63.89
15.5	11.8	3.7	1.0	23.87%	79.75
19.0	14.6	4.4	1.0	23.15%	79.75

（三） 指針式實驗平台測試實驗

在真實的人體肌肉構造中，肌肉與肌肉間的拉扯會互相協調、干擾，導致無法達到仿生效果，故設計本實驗，用以模擬以兩支氣動肌肉缸調節角度的變化量，如下圖所示。

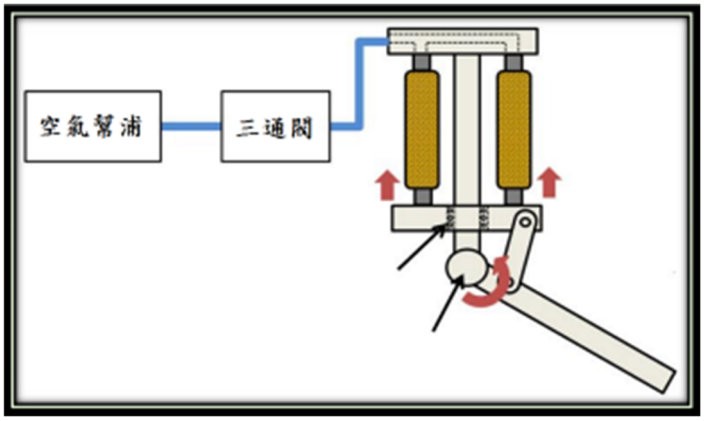


圖 19 平台示意圖

下圖為模擬上圖之指針式實驗平台構造，此平台運用 Makeblock 結合 3D 列印，其中包含了 Arduino Uno 板、變壓器、空氣幫浦、電磁閥、空壓管與氣動肌肉缸等元件，並可藉由此平台搭配 Arduino 與 C# 程式控制幫浦、三通電磁閥進行角度的變化量測試。

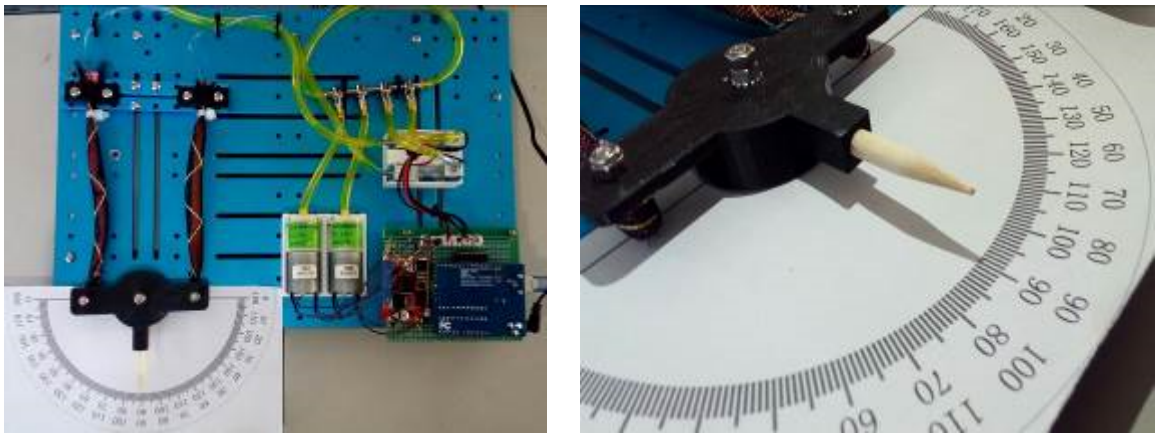


圖 20 觀察角度變化

為了方便操控幫浦與三通電磁閥，使角度實驗測試的流程更為順暢，故利用肌肉控制流程所得到的結果，加以分析並設計了一套控制指令碼，灌入 Arudino 開發板並利用其進行控制，其指令碼大致如下表所示。

指令	操作	指令	操作
左側肌肉充氣	68	左側肌肉堵氣	13
左側肌肉放氣	24	右側肌肉堵氣	57
右側肌肉充氣	67	充氣（秒數 0.05s）	d
右側肌肉放氣	23	充氣（秒數 0.1s）	D
回歸原始狀態	Q	放氣	P

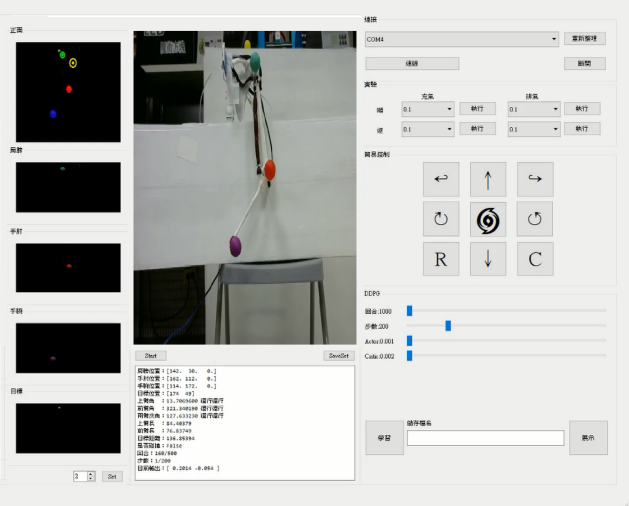


圖 21 程式控制面板

此外，並設計使用者介面與 Arduino 的結合，如左圖所示，可利用此面板直接轉成 Arduino 指令碼並進行指針式實驗平台的操控，利用不同的指令碼組合，可改變運作方式，進而調節角度。經由逐次實驗，可得到各角度變化量所需的充放氣時間。

以下為充放的轉換與時間不同之角度變化之測試表

表 8 左肌肉充氣右肌肉放氣對應時間與角度變化量表

左充(Sec) / 右放(Sec)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.05	2°	2°	2°	2°	2°	2°	2°	2°	2°	2°
0.10	4°	4°	4°	5°	5°	5°	5°	5°	5°	5°
0.15	7°	7°	7°	7°	7°	7°	7°	7°	7°°	7°
0.20	7°	8.5°	8.5°	8.5°	8.5°	8.5°	8.5°	8.5°	8.5°	8.5°
0.25	8°	8°	8°	10°	10°	10°	10°	10°	10.5°	10.5°

表 9 右肌肉充氣左肌肉放氣對應時間與角度變化量表

右充(Sec) / 左放(Sec)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.05	2°	2°	2°	2°	2°	2°	2°	2°	2°	2°
0.10	5°	5°	5°	5°	5°	5°	5°	6°	6°	6.5°
0.15	8°	8°	8°	9°	10°	10°	10°	10°	10°	10°
0.20	11°	11°	11°	11°	11°	11°	11°	11°	11°	11°
0.25	11°	12°	12°	14°	14°	14°	14°	14°	15°	15°
0.30	11°	12°	13°	14°	14°	14°	14°	15°	15°	15°
0.35	11°	12°	14°	15°	15°	15°	15°	15°	15°	15°
0.40	11°	13°	14°	15°	15°	15°	16°	16°	16°	16°

由上述資料可得出以下結論：

1. 開始對兩個氣缸充氣至 90 度的角度且不能飽和。
2. 兩肌肉充放搭配大約有 1 ° 的解析度。
3. 排氣時間越小，搭配充氣肌肉的充氣，可解析到更細的刻度。
4. 排氣時間越久，搭配充氣肌肉的充氣，可更快趨近於飽和。
5. 雖然製作的規格及方法一致，但左右的肌肉規格很難相同。
6. 若要調節出更大的角度，兩端施力點要往中心靠近。

以下為假設其中一支肌肉之長度並推算角度，求 θ^t ：

設 S 為點到點之最小長度 ≒ 14.253 cm (經測量)

設 O 為點到點之原始長度 ≒ 14.7 cm (經測量)

設 2B 為 8 cm (經測量) 則 B 為 4 cm

設 a 長度為 O － S = 0.447 cm 設 c = 4 cm

設 b 長度 (非定值)= $c^2 - a^2$ ≒ 3.976 cm

利用餘弦定理求得 θ :

$\cos\theta = \frac{b}{c} = 0.994$

Arc cos 0.994 ≒ 6.279 °

$\theta + \theta' = 90°$ ，因為 $\theta' + \theta'' = 90$

所以 $\theta = \theta''$。 $\theta'' + \theta''' = 90°$，又因 $\theta''' + \theta^t = 90$ ，故 $\theta'' = \theta = \theta^t$ ≒ 6.279 °。

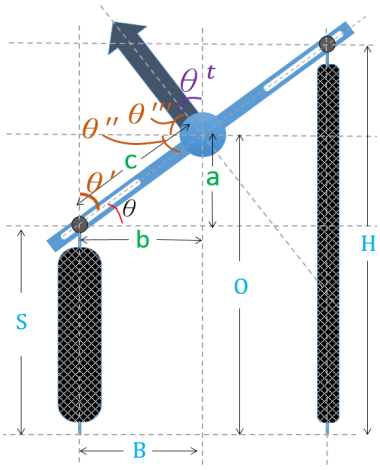


圖 22 推算示意圖

設 S 為點到點之最小長度 ≒ 14.217 cm (經測量)

設 O 為點到點之原始長度 ≒ 14.7 cm (經測量)

設 2B 為 8 cm (經測量) 則 B 為 4 cm

設 a 長度為 O － S = 0.483 cm ，設 c = 4 cm，設 b 長度 (非定值)= $c^2 - a^2$ ≒ 3.939 cm。

利用餘弦定理求得 θ :

$\cos\theta = \frac{b}{c} = 0.984$

Arc cos 0.994 ≒ 10.26 °

$\theta + \theta' = 90°$ ，因為 $\theta' + \theta'' = 90°$，所以 $\theta = \theta''$。

$\theta'' + \theta''' = 90°$，又因 $\theta''' + \theta^t = 90°$，所以 $\theta'' = \theta = \theta^t$ ≒ 10.266 °。

設 S 為點到點之最小長度 ≒ 13.628 cm (經測量)

設 O 為點到點之原始長度 ≒ 14.7 cm (經測量)

設 2B 為 8 cm (經測量) 則 B 為 4 cm

設 a 長度為 O － S = 1.072cm

設 c = 4 cm

設 b 長度 (非定值)= $c^2 - a^2$ ≒ 3.863 cm。

利用餘弦定理求得 θ :

$$\cos\theta = \frac{b}{c} = 0.9657$$

Arc cos 0.9657 ≒ 15.0499 °

$\theta + \theta' = 90°$ ，因為 $\theta' + \theta'' = 90°$，所以 $\theta = \theta''$。

$\theta'' + \theta''' = 90°$，又因 $\theta''' + \theta^t = 90°$，所以 $\theta'' = \theta = \theta^t$ ≒ 15.0499 °。

設 S 為點到點之最小長度 ≒ 13.5 cm (經測量)

設 O 為點到點之原始長度 ≒ 14.7 cm (經測量)

設 2B 為 8 cm (經測量) 則 B 為 4 cm

右圖為 肌肉最長與肌肉最短之狀態

設 a 長度為 O － S = 1.2 cm

設 c = 4 cm

設 b 長度 (非定值)= $c^2 - a^2$ ≒ 3.81 cm。

利用餘弦定理求得 θ :

$$\cos\theta = \frac{b}{c} = 0.9525$$

Arc cos 0.9525 ≒ 17.73 °

$\theta + \theta' = 90°$ ，因為 $\theta' + \theta'' = 90°$，所以 $\theta = \theta''$。

$\theta'' + \theta''' = 90°$，又因 $\theta''' + \theta^t = 90°$，所以 $\theta'' = \theta = \theta^t$ ≒ 17.73 °。

由上述資料可得出以下結論：

(1) 推算中我們的肌肉最大可至 17.73°，但因誤差值等因素，導致角度變化量在實際

的實驗數據中最高只能至 16°。

(2) 由推算的方式可將角度求得更精密但較難顧慮誤差值。

（四） 探討骨頭模型手臂的角度變化量

在完成平台的肌肉模擬後，我們將氣動肌肉缸架在 3D 列印出的仿生骨頭上，藉由內外肌肉進氣量的控制，改變骨頭模型手臂的角度變化量並將其記錄，之後藉由角度變化量以利於控制手臂的抬舉，進行觀察並記錄。

表 10 內肌肉間隔秒數(次數)與角度變化量關係圖(初始角度 140°)

間隔秒數(s) / 次數	0.25	0.5	1
1	140°	140°	140°
2	138°	130°	125°
3	135°	105°	80°
4	127°	85°	60°
5	116°	70°	60°(飽和)
6	105°	60°	60°(飽和)
7	96°	60°(飽和)	60°(飽和)
8	86°	60°(飽和)	60°(飽和)
9	80°	60°(飽和)	60°(飽和)
10	70°	60°(飽和)	60°(飽和)
11	65°	60°(飽和)	60°(飽和)
12	60°	60°(飽和)	60°(飽和)

表 11 外肌肉間隔秒數(次數)與角度變化量關係圖(初始角度 140°)

間隔秒數(s) / 次數	0.5	1.0
1	140°	140°
2	140°	140°
3	140°	157°
4	145°	170°
5	155°	173°
6	168°	173°(飽和)
7	173°	173°(飽和)
8	173°(飽和)	173°(飽和)

表 12 內肌肉充滿時外肌肉逐漸充氣角度變化量關係圖(初始角度 60˚)

次數 \ 間隔秒數(s)	1
1	60˚
2	70˚
3	85˚
4	95˚
5	105˚
6	110˚
7	115˚
8	120˚
9	120˚(飽和)

由上述資料可得出以下結論：

1. 第一次進氣時，不論秒數，會因為氣動肌肉缸內部壓力不夠大，使角度變化量不明顯。
2. 在多次進氣後，角度變化量會隨著內部壓力逐漸提高，而愈趨明顯。
3. 當內肌肉充滿而外肌肉逐漸充氣時，間隔秒數過短，變化量太過不明顯，故使用 1 秒為間隔開始測試。
4. 內肌肉骨頭模型將於 60˚ 時氣動肌肉缸達到飽和狀態。
5. 外肌肉骨頭模型將於 173˚ 時氣動肌肉缸達到飽和狀態。
6. 因兩支氣動肌肉缸力量拉扯，內外肌肉骨頭模型將於 120˚時氣動肌肉缸達到飽和狀態。

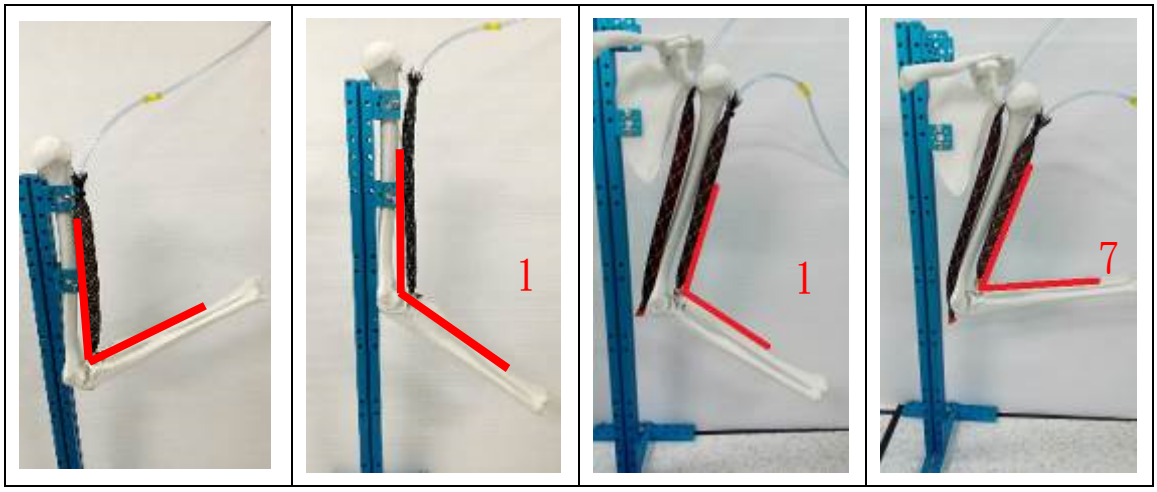


圖 23 骨骼不同角度情形

（五） 以骨頭模型為基礎追加三角肌模型

在前述實驗中，已可進行抬舉與伸直的作用。但在真實的人體手臂中，並不單只有抬舉與伸直的動作，應還有上舉的動作，為模仿真實人體手臂，手臂中追加了三角肌，以完成上舉的動作，藉由三角肌的控制，已可達到側舉功能。

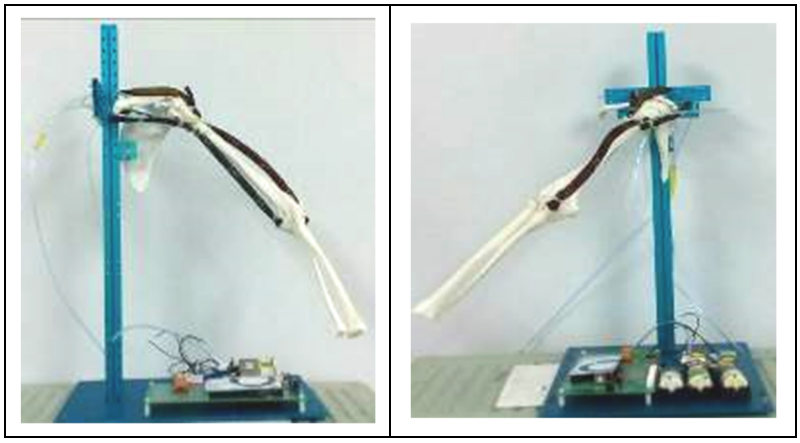


圖 24 控制手臂的側舉的示意圖

（六） 骨頭模型加入影像辨識

為了以最真實的方式模仿人體手臂構造，需要加入更多的氣動肌肉缸，而肌肉與肌肉間的結抗作用，會互相影響其進氣量與伸縮量，在操作的指令上會顯得非常繁瑣。為達到簡化操作指令的目的，加入影像辨識技術。由於本作品欲降低因效能而使畫面產生延遲的情形，故使用 OpenCv 作為影像辨識元件。為了判斷出手臂的位置，所採用的方式是將手臂貼上球狀標示物作為判斷的基準，藉由這樣的方式，能夠找到手臂所在的位置，也能夠透過相同方式辨別需要查找的目標物。

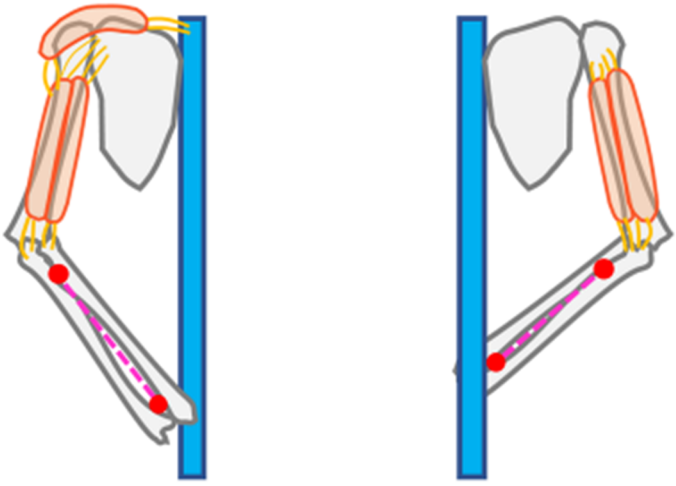
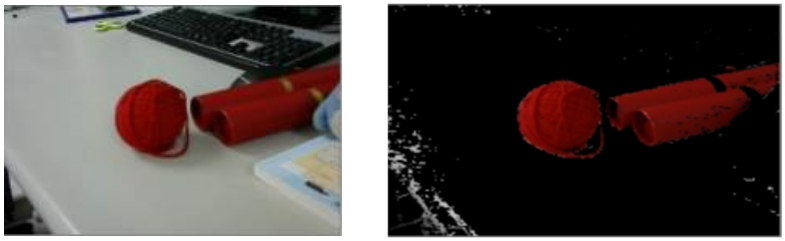


圖 25 判斷手臂位置示意圖

7. 利用 HSV 篩選紅色物體

由於在學習的過程中，需要一個明確的目標物，計畫以驅近飽和的紅色球體作為基準，利用 HSV 篩選畫面中紅色的物體。HSV 分別代表色相、飽和度與明度，將 H 值調到 360 度左右，可篩選出紅色。



(a)原圖　　(b)HSV 擷取紅色處理

圖 26 HSV 處理

8. 將影像灰階與二值化處理以利後續分析處理

在完成 HSV 進行截取紅色後，由於色彩的比對不夠鮮明，而二值化為影像切割最簡易也最常用的方式，而將影像轉為二值圖像前，需要將影像轉成灰度圖像。

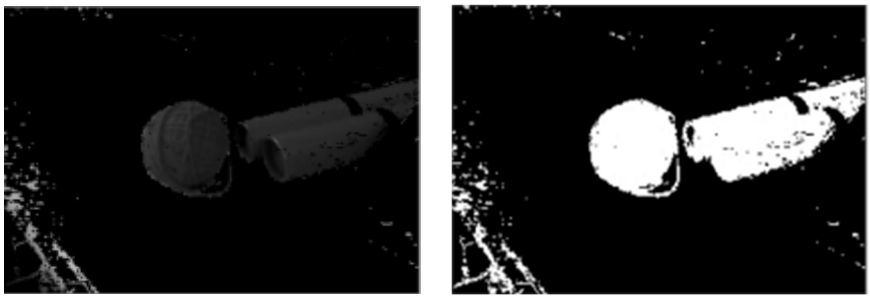


圖 27 二值化處理

9. 運用侵蝕與膨脹進行減少雜點

經由二值化後圖像仍有些許非紅色物體存在，而先侵蝕後膨脹是一種能夠消除雜點的影像處理技巧，為求得影像辨識的準確性，於是將二值圖像進行在處理。

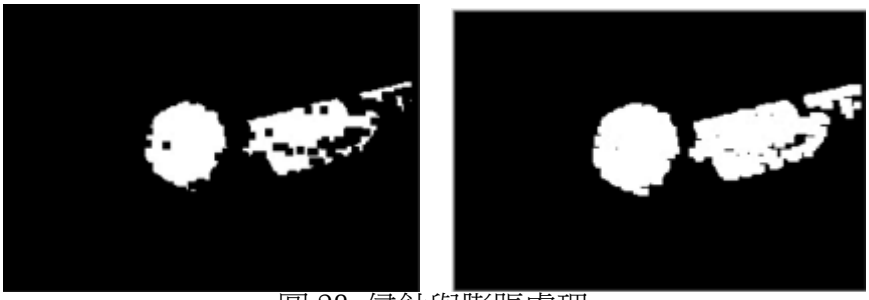


圖 28 侵蝕與膨脹處理

10. 利用找圓處理篩選出畫面中圓形的物體

由於我們欲查找的目標為手臂上的紅原點，故需要達到取圓的功用。通常使用的霍夫找圓過於耗費效能，於是利用侵蝕與擴張處理過後之圖像，判別邊界、半徑、圓心是否足以形成一個圓，再將所取得的圓形資料回傳至原圖作畫。

(a)原圖　　(b)找圓後圖片

圖 29 找圓處理

二、　機器學習

（七）　強化學習訓練狀況

藉由以上方式，我們已可辨識手臂所要學習的目標物及手臂的位置，利用影像辨識的方式，可以取得各個圓心在影像上的座標，並能夠依此計算手臂夾角。

將 DDPG 與 AC 的超參數(附錄 2)設為如下表(表 13)，並開始進行訓練。

表 13 DDPG 與 AC 訓練的參數

超參數	數值
Actor 學習率	0.001
Critic 學習率	0.001
Γ 折扣因子	0.9
最大訓練回合數	1000
回合最大步數	200

獎勵(Reward)機制設置為非直立時得到負值獎勵，因此獎勵值愈接近 0 時，表示代理(Agent)愈善於避免鐘擺下墜。

下圖(圖 30)為 AC(下圖左)與 DDPG(下圖右)分別以鐘擺環境訓練了 1000 回合後的數據簡表，X 軸為回合數，Y 軸為回合總獎勵值。DDPG 與 AC 皆在 200 回合時獲取較高的獎勵值，但觀察數值分布後能發現，DDPG 相對於 AC 有較快速的收斂(下圖綠色區塊)，以及在獎勵獲取(避免負值獎勵)上有更穩定的表現(下圖紅色區塊)，因此最終選擇了 DDPG 應用於本研究。

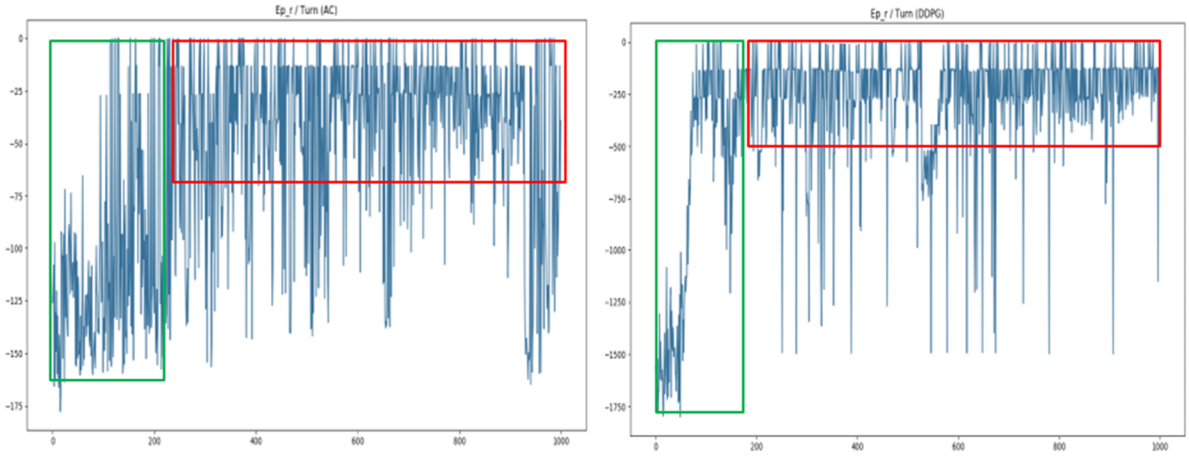


圖 30 AC 與 DDPG 之學習過程 （X 軸：回合數，Y 軸：回合總獎勵值）

結論

經過兩者比較後，我們選擇了 DDPG，因為 DDPG 能有效進行連續動作的選擇與訓練，並改善了 Actor-Critic 的 Critic 神經網路收斂不佳的問題，較為適合本研究的模型訓練。

（八） DDPG 於虛擬環境中的訓練實驗

我們在此研究中所使用的超參數參考如下表：

表 14 DDPG 超參數設定

超參數類別	步數	回合數	Actor 學習率	Critic 學習率	τ 值	γ 值
超參數數值	250	10000	0.0001	0.0001	0.001	0.9

其中類神經網路的隱藏層層數為 1，有 300 個神經元，使用 ReLu 線性整流函數做為激勵函數，並使重放緩衝區為 50000 單位。

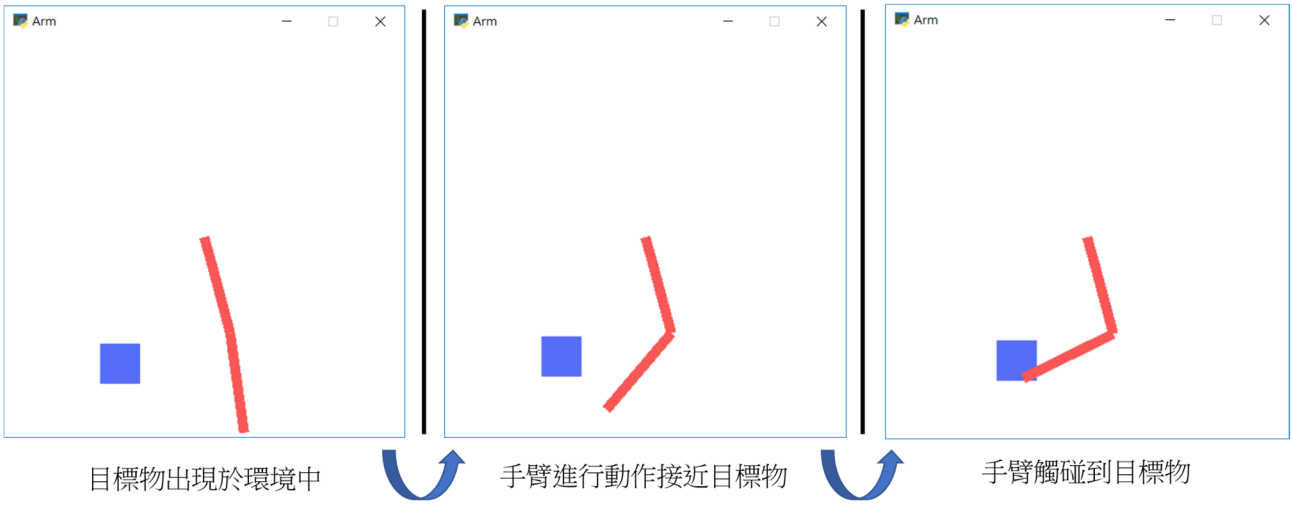


圖 31 簡易訓練過程

在訓練完成後我們可以在有目標位置與當前虛擬手臂夾角、長度的資料下，將虛擬手臂的端點盡可能的接近目標，若達到目標，便結束訓練回合，開始新的訓練回合，並記錄加總每回合中每一步對於目標的距離作為負值獎勵，以及若接觸到目標則加上正值獎勵(圖 32 左)，以及紀錄每回合的步數(圖 32 右)。

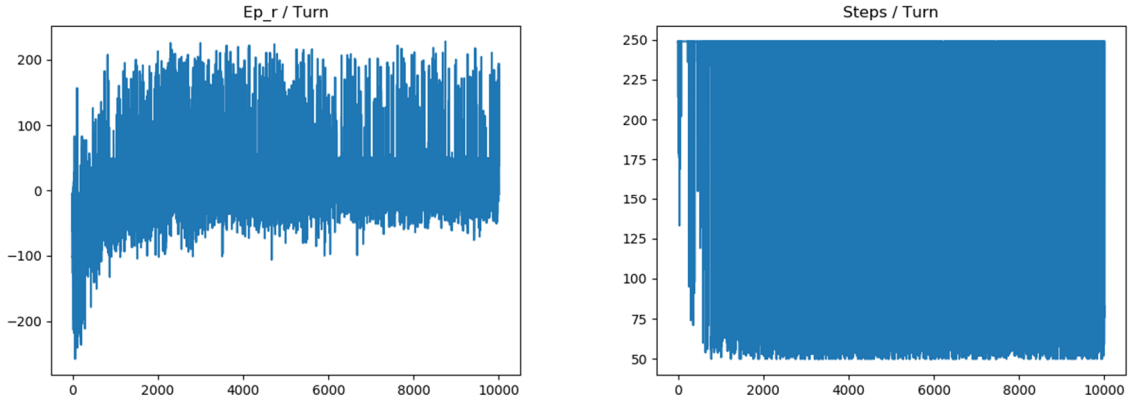


圖 32 DDPG 學習過程 （X 軸：回合數，Y 軸：回合總獎勵值）

由記錄所示，我們能觀察出以下資訊：

1. 每回合的總獎勵值在約 2000 回合後就無明顯提高。
2. 在約 6000 回合後能觀察到，總獎勵值趨近於 0 的情況增加。
3. 每回合步數在約 1000 回合後，步數降低的頻率也提高不少，最低則到 50 步。
4. 即使在約 1000 回合後，已能時常以約 50 步結束回合，還是有步數動盪，不穩定的於極值 250 步與 50 步之間游離。

圖 33 虛擬手臂與目標示意圖

而透過

（九） 強化學習模型與仿生手臂的結合

利用訓練完成的強化學習模型，與實體結合，利用影像辨識去取得角度，位置等參數。由於充氣的時間為連續性的，為避免影像辨識的誤差以及時間差造成手臂過度充氣的後果，我們以 Python 端與 Arduino 端進行一來一往的溝通與控制，並在每一次的動作間檢查狀態。

透過影像辨識得知目標物與初始狀態後，改變學習模型的狀態，經動作輸出後，再透過影像辨識確認狀態，達成仿生手臂接觸目標物的循環。

超參數類別	步數	回合數	Actor學習率	Critic學習率	τ值	γ值	Buffer size	Batch size
超參數數值	200	1000	0.001	0.002	0.001	0.9	5000	32

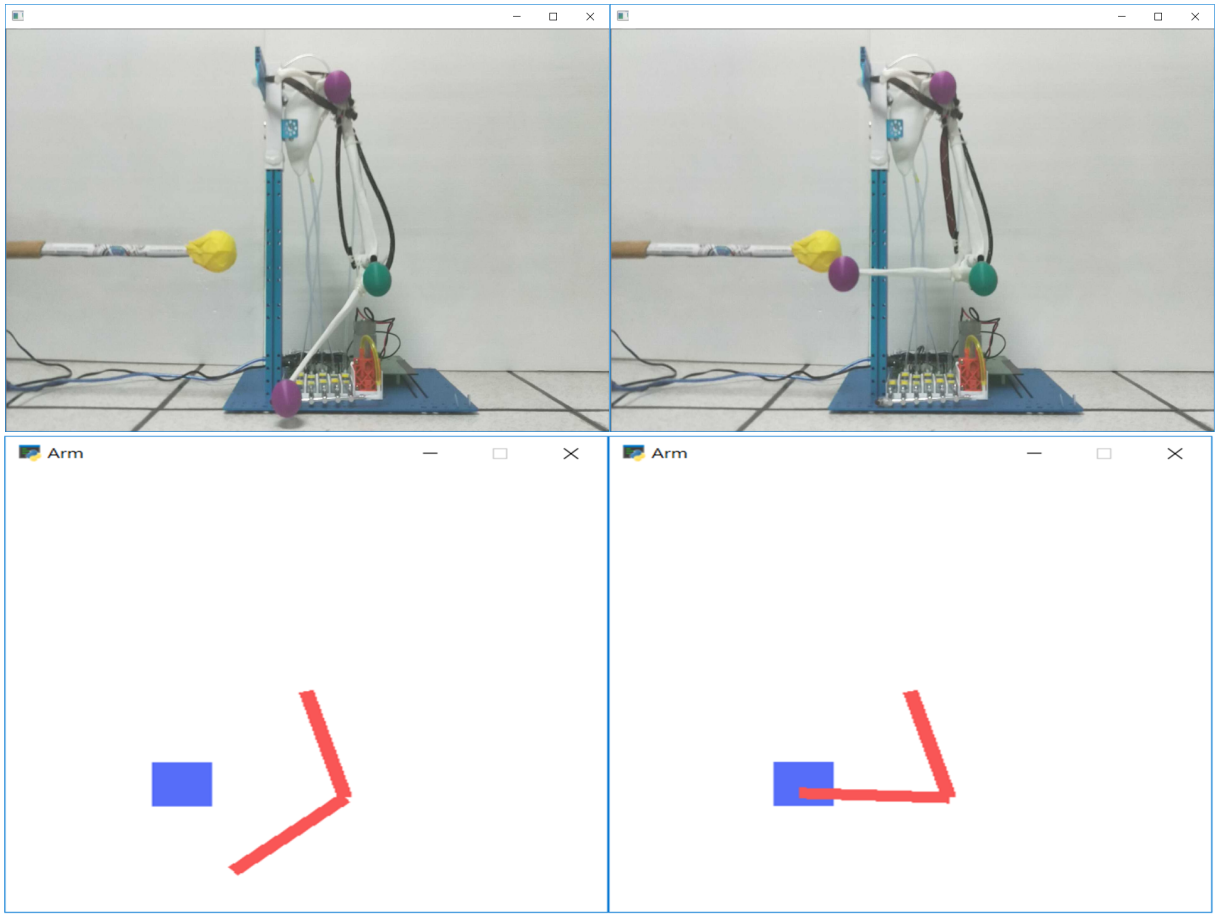


圖 34 仿生手臂動作示意圖

肆、 討論

一、肌肉個體差異

為確實模仿真實人體肌肉的特性，並不會每條肌肉的長度、韌性等都一樣，故所製作之氣動肌肉缸並不會每支都特性相同。然而，這樣的結果所導致的會是肌肉與肌肉間產生個體的差異，在實驗與紀錄上同時略顯複雜。

二、 充放氣誤差值

當我們需要操控精密的角度（例如 1 度）時，勢必需要極短（或趨近於 0）的充氣時間或放氣時間，然而當時間過短時，容易使元件反應不及，產生無法確切進氣的情況，使變化量不明顯等問題，最終使充放氣與呈現角度成為非線性關係，因此我們提出一種改善方式，利用深度學習與影像辨識，學習如何因應肌肉個體差異以及充放氣誤差值，在不同的肌肉狀態下，以特定的充氣時間得到所需的單位角度，優化 DDPG 於本研究的應用。

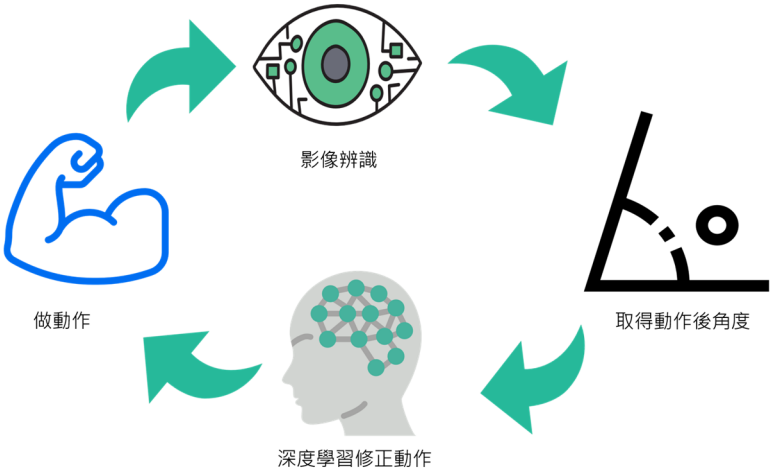


圖 35 利用深度學習修正動作示意圖

三、　動力來源的比較

一般市面上通常使用伺服馬達作為動力來源，而本作品所使用的是幫浦，因此我們整理了表格做說明，如下表所示。

表 15 動力來源比較圖

	伺服馬達	幫浦結合氣動肌肉缸
成本	1.幾個旋轉角就需要幾顆伺服馬達 2.價格昂貴	1.一組空氣幫浦 2.價格便宜
重量與體積	1.伺服馬達重 2.馬達本身重量會成為另一個關節的負擔 3.需要有容納伺服馬達的空間，若為多個伺服馬達，則體積龐大	1.幫浦重，但不會放在肢體上，造成關節的負擔 2.肌肉缸可以繞折彎曲，比較沒有空間限制。
效能與耗能	1.可作 0-360 的旋轉角度 2.可以精確控制角度 3.反應速度快 4.扭力 20kg 以上，擴充不容易，需購置更大扭力的馬達來解決 5.電力消耗大	1.旋轉角度與收縮的長度有關，旋轉角度不大 2.角度控制不易 3.反應速度跟充氣量有關 4.拉力大約 3kg 需搭配多條提高負重能力 5.效能相較伺服馬達耗能低
控制	PWM	充氣、放氣、堵氣
仿生	無法完全模擬生物行為	可以接近生物的動作行為

伍、 結論與應用

一、結論

1. 由於所使用的材質上無法達到人體組織的精細程度，肌腱、韌帶的位置及肌肉的大小，有部分的位置權宜的變動，以符合順利運作的目標。
2. 透過仿照人體肌肉組織，建構肌肉系統對於了解人體架構有很大的助益。
3. 本研究建構的上臂模型採用實際的骨頭 3D 模型，確實可以運作出人類部分的肢體動作。
4. 本研究所訓練的學習模型，即使有部分動作陷入局部最優的情況，還是能夠如期的達到接近目標的目的。
5. 氣動肌肉的充放氣時間與上下臂的夾角關係成非線性關係，可以透過深度學習來優化此問題。

二、 未來展望

預計建構出局部甚至是全身的仿生機器人，目前機械學習所學的維度只侷限於二維空間，希望未來能夠透過影像辨識，並建構出三維空間，在三維的空間下學習，讓仿生手臂做出更像是人類的動作。

陸、 參考資料及其他

一、附錄

（一） DDPG 虛擬碼(取自參考資料 6)

Algorithm 1 DDPG algorithm

Randomly initialize critic network $Q(s, a|\theta^Q)$ and actor $\mu(s|\theta^\mu)$ with weights θ^Q and θ^μ.
Initialize target network Q' and μ' with weights $\theta^{Q'} \leftarrow \theta^Q$, $\theta^{\mu'} \leftarrow \theta^\mu$
Initialize replay buffer R
for episode = 1, M **do**
  Initialize a random process $\mathcal{N}$ for action exploration
  Receive initial observation state s_1
  for t = 1, T **do**
    Select action $a_t = \mu(s_t|\theta^\mu) + \mathcal{N}_t$ according to the current policy and exploration noise
    Execute action a_t and observe reward r_t and observe new state s_{t+1}
    Store transition (s_t, a_t, r_t, s_{t+1}) in R
    Sample a random minibatch of N transitions (s_i, a_i, r_i, s_{i+1}) from R
    Set $y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$
    Update critic by minimizing the loss: $L = \frac{1}{N}\sum_i (y_i - Q(s_i, a_i|\theta^Q))^2$
    Update the actor policy using the sampled policy gradient:

$$\nabla_{\theta^\mu} J \approx \frac{1}{N} \sum_i \nabla_a Q(s, a|\theta^Q)|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu} \mu(s|\theta^\mu)|_{s_i}$$

    Update the target networks:

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$$
$$\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$$

  end for
end for

（二） 超參數

超參數為機器學習在開始訓練過程之前所設置的參數。

超參數名稱	敘述
回合數	代表 Agent 所要訓練的回合數
步數	每回合最高可改變狀態的次數
學習率	此值為每次迭代函數所更新的程度，過低可能造成動作學習不佳的狀況；過高則可能導致動作學習過於猛烈，無法準確達成任務。
τ 值	限制目標值的變化，控制學習穩定性。
γ 值	折扣因子，對於未來動作期望的折扣率

二、 參考資料

1.吳孟哲(2009)，穿戴式下肢輔具之氣壓式人工肌肉鑑別與控制，國立清華大學機械工程學系，碩士論文。

2.金冠霖(2013)，別以嵌入式系統實現人工肌肉氣壓缸之系統識別，國立臺灣師範大學機電科技學系，碩士論文。

4.吳政學，肌肉系統，崇仁護專自編教材。

5.FESTO，Fluidic Muscle DMSP/MAS FESTO

https://www.festo.com/cat/zh-tw_tw/data/doc_engb/PDF/EN/DMSP-MAS_EN.PDF

6. CONTINUOUS CONTROL WITH DEEP REINFORCEMENT LEARNING(2016)|Timothy P. Lillicrap, Jonathan J. Hunt

7. 機器學習實踐|莫煩 Python

https://morvanzhou.github.io/tutorials/machine-learning/ML-practice/

8. Policy Gradient Methods forReinforcement Learning with Function Approximation (2000)| Richard S. Sutton, David McAllester, Satinder Singh

9. CONTINUOUS CONTROL WITH DEEP REINFORCEMENT LEARNING(2015) |Volodymyr Mnih1

10. Asynchronous Methods for Deep Reinforcement Learning(2016)| Adrià Puigdomènech BadiaMehdi Mirza, Alex Grave,Tim Harley

11. Actor-Critic Algorithms(2000) | Vijay R. Konda John N. Tsitsiklis

12. Emergence of Locomotion Behaviours in Rich Environments(2017) | Nicolas Heess, Dhruva TB, Srinivasan Sriram, Jay Lemmon, Josh Merel

13. Proximal Policy Optimization Algorithms(2017) | John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, Oleg Klimov

14. A Survey on Policy Search for Robotics(2013) | Marc Peter Deisenroth,Gerhard Neumann and Jan Peters

【評語】100011

1. 參展者以簡易的氣球加網，透過簡易的氣壓管路以及 pump 來模擬人工肌肉及韌帶的作用。

2. 透過學習系統，讓人工肌肉系統可以在短時間內定位，或是到達較接近標的物的位置，效率很高。

3. 未來如能加上關節自由度的模仿，會讓整個系統更為接近仿生手臂的目標。