

基於深度學習網路與 Arduino 開發版之動態手勢控制系統

A Dynamic Hand Gesture Control System based on Deep Learning Networks and Arduino UNO Board

李棟良

Donq-Liang Lee

劉政君

Jheng-Jyu Liou

蔡孟哲

Meng-Zhe Cai

張家寶

Chia-Pao Chang

葉珮玉

Pei-Yu Yeh

銘傳大學電腦與通訊工程學系

Department of Information and Telecommunications Engineering,
Ming Chuan University

摘要

手勢辨識近年來的研究相當受到歡迎，為了增加使用者的實用性，本研究將以不穿戴任何顏色手套的情況下，透過深度學習之技術，製作出一套動態手勢辨識系統，實作方法與操作步驟分為以下五部分進行：(1)將使用者的手部區域以 Kinect V2 攝影機裝置擷取深度影像，(2)利用深度影像之二值化來擷取手部區域，並去除多餘之雜訊，(3)以卷積神經網路(Convolutional Neural Network, 簡 CNN)來擷取手勢之特徵，再以支援向量器(Support Vector Machine, 簡 SVM)來辨識手勢影像 (4)以隱藏馬爾可夫模型(Hidden Markov model, 簡稱 HMM)來學習與辨識動態手勢，(5)以 Arduino UNO 開發版來實現動態手勢控制系統。

本研究需要克服的問題為：(1)動態手勢的即時處理與辨識 (2)手勢變化的變異性(例如，不同人比手勢會有不同的習慣，或手掌會有不同的大小與掌形) (3)穩定性(不會因為背景及光線等環境因素的影響，而造成辨識錯誤)。

關鍵字：Kinect V2 感測器、深度學習網路、Arduino 開發版、隱藏馬爾可夫模型、手勢辨識

Abstract

Research of hand gesture recognition has received much attention in recent years. In this research, a dynamic hand gesture recognition system based on the deep learning technology was built. Our method consists of the following five processing steps: (1) capture the depth image by

using the KinectV2 sensor, (2) obtain user's hand region through image thresholding, (3) learning and recognition of static hand gestures by using the Convolutional Neural Network (CNN), then use Support Vector Machine(SVM) to recognize hand gestures (4) learning and recognition of dynamic hand gestures by using the Hidden Markov Model (HMM) (5) Realization of a hand-driven control system through the use of the Arduino UNO board.

We have to overcome the following problems: (1) real-time processing and recognition of the dynamic hand gestures (2) Variants of hand gestures (e.g., hand gestures of different people may have different sizes and shapes ; different habits of making hand gesture, etc.) (3) Stability (recognition errors result from background clutter or illumination change).

Keywords: Kinect V2 sensor, deep learning networks, Arduino UNO board, Hidden Markov Model (HMM), hand gesture recognition

1. 簡介

過去人與電腦溝通的方式都是藉由鍵盤與滑鼠，近年來演變成用手直接觸碰螢幕，像是智慧型手機，就是利用觸控感應，取代過去的按鍵，未來若是可以由手和手指的姿態來取代電腦滑鼠的控制，就能讓使用者用更自然的方式與科技產品互動。

本研究目標是如何在不穿戴任何設備的情況下，能夠順利的辨識手識，並且不會因為複雜的光源及背景的影響而造成辨識困難。我們以 kinect v2 感測器來擷取手勢，進而利用捲積神經網路 Convolutional Neural Network(CNN)來做學習，讓手部辨識結合電腦，提供高效率且方便使用的人機介面服務。目前很多研究都透過 Kinect v2 攝影機以進行手勢影像的處理，像是劉禹庭[1]使用 Kinect 深度影像，將手掌位於深度影像裡的距離數值，將手掌轉為二值化，以此便能得到手掌之圖型。而鄧有光[2]將深度影像作平滑處理，影像經過侵蝕(Erosion)和擴張(Dilation)的處理之後就能夠去除其中的雜點，使影像變的較平滑，提高手掌的完整度。而在辨識系統方面，許明翔[3]提出 CNN(Convolutional Neural Network)與 HMM(Hidden Markov Model) 來建構手勢辨識系統的方法，其 CNN 之第一卷積層採邊緣強化與梯度偵測等固定之 filter patch 來加快學習過程，最後結合 HMM 對手勢進行辨識。林本煌[4] 提到辨識模的建構是採用分層隱藏馬可夫模型當作核心，進而結合手勢單位的分割及組合 來進行手識辨識，因傳統演算法的運算時間較為冗長且方法過於複雜，而提出一種改良型分層隱藏式馬可夫模型應用於手勢辨識。

2. 系統概述

2.1 系統架構

研究架構如圖 1，操作步驟分為以下五部分進行(1)將使用者的手部以 kinect v2 攝影機來擷取手勢影像，(2)將擷取之手勢深度影像做區域處理及影像處理(3)使用手勢影像提供給

卷積神經網路 Convolutional Neural Network(CNN) 與支援向量機 Support Vector Machine(SVM)做學習，(4)利用學習好的 CNN 對動態手勢產生 HMM 序列值來形成動態手勢之模組，(5)動態手勢控制系統之實現體並擷取特徵資料。



圖 1 研究架構圖

2.2 Kinect v2 擷取手勢深度影像

本研究主要使用 kinect v2 的深度影像鏡頭，圖 2 為本研究之操作環境，將 kinect v2 放置於平坦桌面上，手勢擺動範圍離 kinect v2 約 35-75 公分之間，因影像的擷取非常快速，每 1 秒約擷取 30 張，會導致辨識負載過量，影響辨識速率，因此本研究將每次擷取影像，設定為每 4 張取 1 張，總共取 9 張。本研究所擷取的影像規格為 512*424 的灰階圖片(如圖 3 之深度影像)，而我們所使用的神經網路判斷圖片的規格為 227*227 因此，我們將圖片在一開始擷取時，統一改為一致的大小格式，不僅加快訓練速度，也避免造成學習上錯誤。

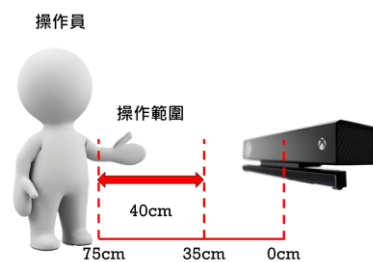


圖 2 操作環境示意圖



圖 3 深度影像

2.3 影像前置處理作業

2.3.1 深度影像轉二值化影像

由於本研究主要以辨識手勢為主，因此透過 Kinect v2 紅外線距離量測功能，並利用二質化函數可過濾掉背景影像或將背景塗黑，只留下使用者(User)的手部影像。經多次實驗結果，我們將最佳手勢偵測範圍設定為 35 公分~75 公分，其手勢活動空間為 40 公分範圍，其結果如圖 4。

2.3.2 形態學填補空洞與去除雜訊

背景去除完後，影像會有一些小空洞或是些微的雜訊(如圖 5)，此時我們選擇利用形態學來填補空洞。數學形態學 (mathematical morphology)，簡稱形態學，是影像處理的其中一環，專門用來處理、分析影像中的形狀。形態學有兩種基礎運算，分別為膨脹 (Dilation)

與侵蝕 (Erosion)，再延伸出斷開(open)與閉合 (closing)。斷開 (open) 為開啟擁有平滑影像，去除細小連接點以及可以去除細微突出的部分，閉合(closing)為擁有平滑影像的效果，與開啟不同的是會將細微斷裂或紋路的地方連接起來，並且去除細小的空洞。我們所採用的結構元素為 3*3 的菱形矩陣，進行先膨脹後侵蝕，重複此動作 3 次，便可得到如圖 6 之結果。



圖 4 透過二值化函數之結果

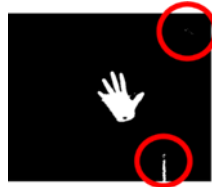


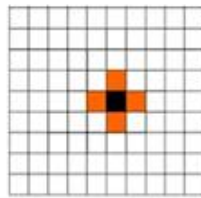
圖 5 未去除雜訊之影像



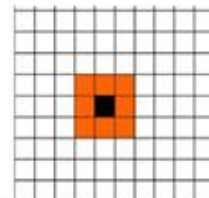
圖 6 去除雜訊後之影像

2.3.3 物件連通去除空間雜訊取重點影像

物件連通法主要是將輸入影像中所有物件,依照相互連結與否的屬性作分類[5]，可分為四連通 (4-Connected Component) (如圖 7a)與八連通 (8-Connected Component) (如圖 7b), 本研究採用的是八連通。物件連通是將所有的物件做標記,把 相鄰的點標記成同一標記,代表是同一個物件,以便做後續的處理。四連通是指對應像素位置的上、下、左、右,而八連通所對應的像素位置為上、下、左、右、左上、右上、左下、右下。標記後取最大面積之影像即為手部區域(如圖 8)



(a)



(b)

圖 7 物件連通法(a)四連通(b)八連通

2.3.4 透過定界框框選完整手勢

根據上述步驟所處理的二質化影像，除了主體(手)的部分其餘全是黑的，多餘的黑色部分，在深度學習裡是不必要的，所以我們使用定界框 bounding box (如圖 8a)框選完整手勢，可大量刪減多餘部分，且提升辨識準確率，結果如圖 8b。

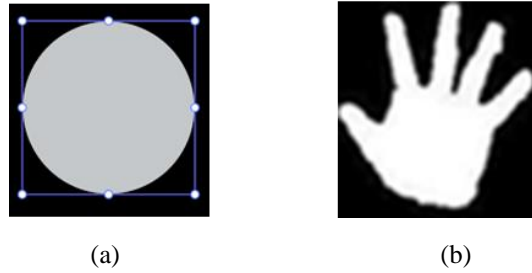


圖 8 bounding box(a)示意圖(b)實作圖

2.3.5 線性方程式的轉換

根據 Kinect v2 所擷取到的深度影像，是以距離數據所組成的灰階圖(如圖 10a)，手勢的細節非常地不分明，我們採用線性方程式(如圖 9)來做調整，線性方程式公式如下。

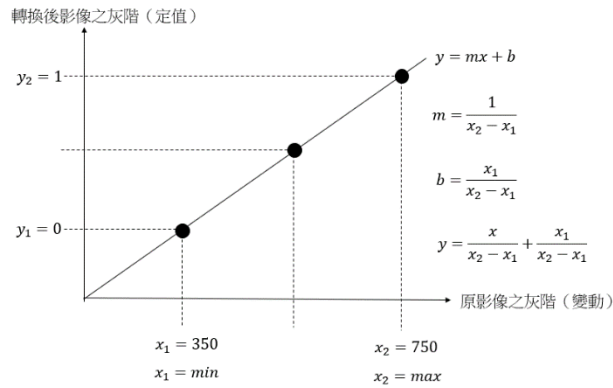


圖 9 線性方程圖

由線性方程式做計算，並將手勢圖裡的每個像素點帶入函式裡，即可得出較多細節的手勢灰階圖(如圖 10b)。



圖 10: (a) 距離灰階圖(b)線性標準結果圖

2.4 利用 CNN 與 SVM 進行靜態手勢辨識

本研究採用 matlab 官方[8]之範例一樣的做法，將處理過後的手勢影像，經 CNN 擷取特徵，再透過 SVM 來學習與分類。

2.4.1 卷積神經網路(CNN)之架構

Convolutional Neural Network(CNN) [6]、[7]主要由兩種網路架構所組成，一種是 convolution layer，另一種是 subsampling layer，最後一層就是把 subsampling layer 攤開，稱為 full connected layer。卷積神經網路中的每一個卷積層都緊跟著一個用來求局部平均與二次提取的計算層，這種特有的兩次特徵提取結構減小了特徵解析度。在實際應用中，往往使用多層卷積，然後再使用全連接層進行訓練，多層卷積的目的是一層卷積學到的特徵往往是局部的，層數越多，學到的特徵越複雜，參考(圖 11)。

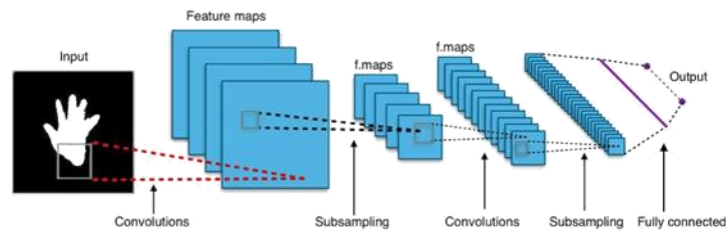


圖 11 Convolutional Neural Network 結構圖

2.4.2 卷積神經網路 Convolutional Neural Network(CNN)之學習

本研究先以 Convolutional Neural Network(CNN)做靜態手勢訓練。利用 kinect v2 擷取影像，總共採集 5 個人的手勢，每個人每種手勢比 100 張，一種手勢總共採取 500 張畫面作為訓練影像，總共採集 8 種不同的手勢，分別為(1)left、(2)right、(3)good、(4)bad、(5)five、(6)up、(7)down、(8) zero (9)未偵測到的，統一將數據轉成相對應的格式(如圖 12)。輸入解析度為 227x227 的手勢圖片將手勢圖片隨機分成百分之九十作為學習資料和百分之十作為樣本測試樣層並降低維度,一連串的卷積層>採樣層>卷積層>採樣層>卷積層...連續重複，圖片會變得越來越小且越來越細，直到獲得整張圖片的特徵向量(如圖 13)，包含以上步驟總計二十三層(包括輸入和輸出)。

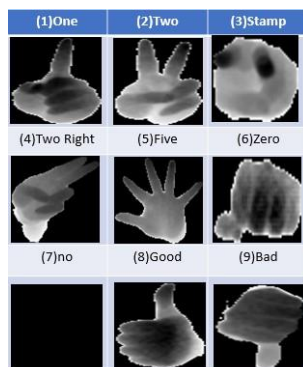


圖 12 9 種類別的訓練樣本

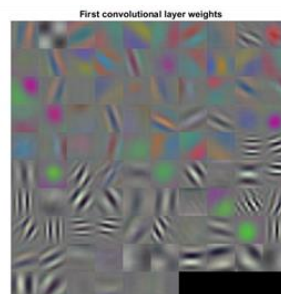


圖 13 第一層特徵向量

2.4.3 支援向量機 Support Vector Machine(SVM)之學習

經由上述步驟後，我們所採用的卷積神經網路，最後結合了支援向量機 Support Vector Machine(SVM) [9] (參考圖 14))通常用在機器學習 (Machine learning)。是一種監督式學習 (Supervised Learning)的方法，主要用在分類(Classification)和回歸 (Regression)上，在本研究中使用 CNN 來擷取出影像的高階特徵，再利用 SVM 來學習與分辨不同的手勢類別。

2.5 隱藏馬可夫模型與動態手勢辨識

以 CNN 與 SVM 架構所辨識出的靜態手勢之後，便可組成動態手勢序列，然後再以 HMM 架構來推測出相應的動態手勢。

2.5.1 隱藏馬可夫模型架構

Hidden Markov Model (HMM) [10]、[11]是統計模型，它用來描述一個含有隱含未知參數的馬可夫過程。其難點是從可觀察的參數中確定該過程的隱含參數。然後利用這些參數來作進一步的分析。簡單來說 Hidden Markov Model (HMM)提供了一套數學的理論以及工具，讓我們可以利用「看得到的」連續現象去探究、預測另一個「看不到的」連續現象。(參考圖 15)

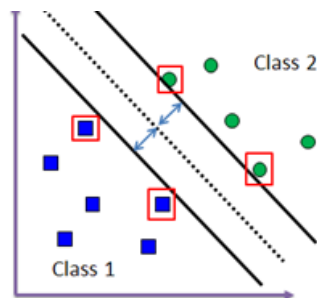


圖 14 svm 示意圖[9]

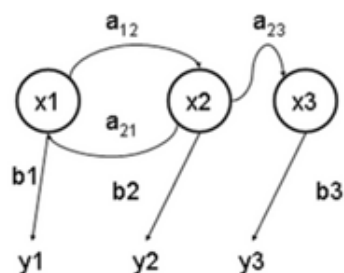


圖 15 Hidden Markov Model 參考圖(來源:維基百科) (1) x：隱含狀態 (2) y：可觀察的輸出 (3) a：轉換機率 (transition probabilities) (4) b：輸出機率 (output probabilities)

2.5.2 隱藏馬可夫模型之建立

動態手勢總共有 5 種，我們直接將手勢偵測系統所輸出之固定張數的取樣動態手勢畫面(手勢的連續動作)送至 CNN，去辨識並對應出每張手勢之 CNN 預設代碼(代碼為使用者自訂)，並將此代碼以一串總數為 9 張照片的序列來儲存，因 HMM 需要大量的數據來運算出各種動態手勢所對應的機率，所以我們的每一種動態手勢有 80~90 串的序列，最後再將動態手勢一一代入 HMM，取得對應的觀測值(如圖 16~圖 22)。

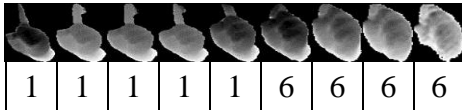


圖 16 動態手勢序列 1: sound up

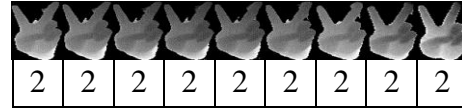


圖 17 動態手勢序列 2: sound down

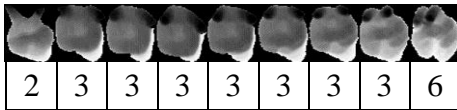


圖 18 動態手勢序列 3: open-close



圖 19 動態手勢序列 4: change

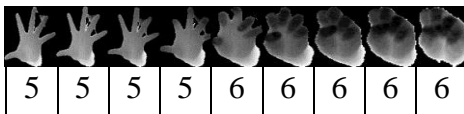


圖 20 動態手勢序列 5: play-stop

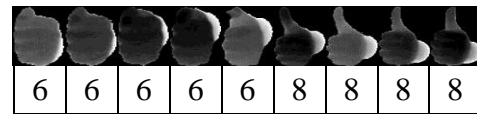


圖 21 動態手勢序列 6: turn up

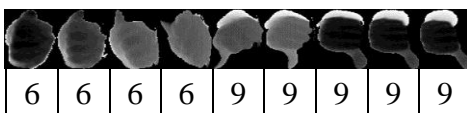


圖 22 動態手勢序列 7: turn down

2.5.2 隱藏馬可夫模型之機率推測

正常的馬可夫模型中，狀態對於觀察者來說是直接可見的，而隱藏馬可夫模型(Hidden Markov Model)的特色為只能看到觀測序列，但看不到狀態。我們將 CNN 靜態手勢模型辨識好的手勢序列，給隱藏馬可夫模型運算，即可得知其機率。

2.6 動態手勢控制系統

推測出相應的動態手勢後，可以利用 matlab 端發送相對字串至 Arduino 端，以啟動 Arduino 發送紅外線訊號，來控制家電。

2.6.1 Arduino

Arduino 是個 open-source electronics prototyping platform，所謂 open-source 開放源碼，意思是說 Arduino 的硬體電路設計圖是開放的，軟體原始碼也都是開放的；electronics 代表說創作物是要吃電的；prototyping platform 表示運用這套平台可以快速地建構出原型，打造產品。本研究使用 Arduino UNO 開發版。

2.6.2 接收與錄製紅外線遙控器之訊號

本研究控制家電，採用紅外線遙控[12]之家電，將紅外線接收器接上 Arduino UNO 開發版上，透過 Ken Shirriff 開發撰寫的 IRRemote 這套 Arduino 程式庫，來偵測該家電遙控器之紅外線訊號，該程式庫包含了該遙控器所使用的協定。

2.6.3 結合 Matlab 與 Arduino 發射紅外線遙控器之訊號

經由上述步驟，把所接收到的協定與紅外線訊號存下，並將辨識動態手勢之結果，搭配 Arduino UNO 的紅外線發射器，來完成靜態與動態手勢之操控家電。

3. 實驗結果

3.1 深度神經網路之訓練與測試

透過 Kinect v2 得到研究樣本資訊後，進行深度神經網路訓練，訓練約 500 張，測試約 50 張，現階段我們能夠順利辨識出手勢結果，其成功辨識率可達到 92% 以上(如表 1)，未來會進一步作改善，讓訓練樣本更具有多樣變化，以便 CNN 學習，提高所有手勢辨識之成功率。

	One	Two	Stamp	Two Right	Five	Zero	No	Good	Bad
One	0.98	0.02	0	0	0	0	0	0	0
Two	0.04	0.96	0	0	0	0	0	0	0
Stamp	0	0	0.95	0	0	0.05	0	0	0
Two Right	0	0.01	0	0.99	0	0	0	0	0
Five	0.03	0.04	0	0	0.93	0	0	0	0
Zero	0	0	0.05	0	0	0.95	0	0	0
No	0	0	0	0	0	0	1	0	0
Good	0	0	0	0	0	0	0	0.92	0.08
Bad	0	0	0	0	0	0	0	0.08	0.92

表 1 每種手勢取 50 張之測試結果

表 2 所顯示的資料為動態手勢的辨識結果，總共有 5 位測試人員，每個人皆測試 20 組動

態手勢，每種手勢皆有 100 筆測試資料，進而算出其成功率。

實際 測試	Sound up	Sound Down	Open-Close	Change	Play-Stop	Turn up	Turn down
Sound up	0.91	0.03	0	0.06	0	0	0
Sound Down	0	0.96	0	0.04	0	0	0
Open-Close	0	0.02	0.9	0.08	0	0	0
Change	0	0.07	0.08	0.95	0	0	0
Play-Stop	0	0	0	0	0.92	0.08	0
Turn up	0	0	0	0	0	0.86	0.14
Turn Down	0	0	0	0	0	0.16	0.84

表 2：動態手勢辨識率

3.2 隱藏馬可夫模型之訓練與測試

本研究設定的動態手勢總共有 7 種，每種手勢都有獨立的 Hidden Markov Model。運算後可得知其學習機率值及狀態轉移之機率，可以參考表 3、表 4。每種動態手勢的狀態個數由使用者自行設定，我們依照動態手勢的複雜程度來設定狀態個數，如表 3 sound up 手勢狀態大多為手勢 1 及手勢 5 所組成，當 HMM 所測出的機率值符合我們所設的手勢及狀態時，就將此狀態個數列為最佳設定。

靜態手勢 狀態	手勢 1	手勢 2	手勢 3	手勢 4	手勢 5	手勢 6	手勢 7	手勢 8	手勢 9
狀態 1	0	0	0.1638	0.00428	0.0428	0	0.5088	0	0.2846
狀態 2	0.9446	0	0	0.0554	0.0554	0	0	0	0

表 3 動態手勢 1: sound up 之學習機率值

	狀態一	狀態二
狀態一	0.08281	0.1719
狀態二	0.0236	0.9764

表 4 動態手勢 1: sound up 狀態轉移之機率值

4. 結論

本研究應用深度學習之卷積神經網路與支援向量機整合手勢辨識，再透過隱藏馬可夫模型整合動態手勢序列來控制紅外線遙控家電，最後我們成功的達到使用者以手勢控制家電並提升便利性之目的。

而透過實驗的結果發現，動態手勢的暫態部分、相似較高之手勢以及電腦 GPU，影響了辨識的準確率與速率，希望未來能透過加強深度學習內部架構、改善內部程式碼以及更換高規格 GPU 來達到更完美的動態手勢辨識系統。而深度學習部分，因本研究所使用的圖像為灰階影像，細節度與資料量都遠低於彩色影像，並沒有完整的利用到深度學習，未來希望能透過影像擷取裝置擷取彩色手勢影像，以達到完整利用深度學習。而家電部分，目前只能使用紅外線家電，不夠廣泛，希望以後能結合更多元的設備使操控者有更多元的選擇。

5. 參考文獻

- [1] 劉禹廷、黃冠瑜、蔡人儀，”以視覺為基礎的見報系”中華大學資訊工程學系 100 學年度專題製作期末報告，指導教授：黃雅軒教授
- [2] 鄧有光、陳永輝、劉毅人”人機互動介面之即時手勢識別技術”第九屆離島資訊技術與應用研討會論文集
- [3] 許明翔，”具有深度學習精神之人類手勢影像辨識系統”國立台北科技大學自動化科技研究所碩士論文，民國一百零四年指導，指導教授：林顯易博士
- [4] 林本煌，”改良型分層隱藏式馬可夫模型於手勢識別之研究”國立台北科技大學電機工程研究所碩士學位論文，民國一百年七月指導，指導教授：練光祐教授 p1~p25
- [5] 吳怡明，”手勢辨識應用於遙控音樂撥放系統”國立台北科技大學電機工程系碩士學位論文，民國九十八年指導，指導教授：蔡超人博士 p25~P28
- [6] Convolutional Neural Net 筆記
<http://darren1231.pixnet.net/blog/post/336760136-convolutional-neural-net-%E7%AD%86%E8%A8%98>
- [7] 一文讀懂卷積神經網絡 CNN <https://read01.com/7Rx00O.html>
- [8] MATLAB 開發平台官方網站
<https://www.mathworks.com/help/nnet/convolutional-neural-networks.html>
- [9] 支持向量機器 (Support Vector Machine)
<https://cg2010studio.com/2012/05/20/%E6%94%AF%E6%8C%81%E5%90%91%E9%87%8F%E6%A9%9F%E5%99%A8-support-vector-machine/>
- [10] 維基百科-隱馬爾可夫模型
<https://zh.wikipedia.org/wiki/%E9%9A%90%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B>
- [11] 隱馬可夫模型：探索看不到的世界的數學工具 <http://pansci.asia/archives/43586>
- [12] 葉難 <http://yehnan.blogspot.tw/2012/02/arduino.html>