

東海大學資訊管理研究所

碩士學位論文

基於深度學習之網球軌跡影像分析

Image Analysis of Tennis Trajectory Based on Deep Learning

指導教授：林正偉 博士

研究生：鍾奉原 撰

中華民國 109 年 7 月

東海大學資訊管理學系碩士學位  
考試委員審定書

資訊管理學系研究所 鍾奉原 君所提之論文

基於深度學習之網球軌跡影像分析

經本考試委員會審查，符合碩士資格標準。

學位考試委員會 召集人： 高弘忠 (簽章)

委員： 沈家平  
林正清  
\_\_\_\_\_  
\_\_\_\_\_

中華民國 109 年 7 月 1 日

## 誌謝

在這東海資管讀碩士這兩年裡，過得相當得充實，學會了很多以前沒有掌握的技術能力，以及相關的技術理論。在這兩年在東海資管當了資料結構及程式設計的TA，學會了如何教導他人以及如何去幫助他人解決問題，也認識了不少大學部的學弟妹。

首先我要感謝我的指導教授林正偉老師，在我讀研究所這兩年來對我的指導，在研究的進行中給於實質的建議，使得我在研究中遇到瓶頸的時候，能夠找到一個方向進行突破，繼續向前邁進完成我的研究論文。感謝口試委員們的意見及建議，給於我修正論文使得論文可以更加的完善。

我要感謝我的同學們，在這兩年裡認識了一群同班同學，雖然大家的指導教授各不相同，有時候會討論到自己的研究做得怎麼樣，遇到了什麼問題等等，大家每學期都會聚餐 1-2 次，雖然每次聚餐的吃得金額是有點高，不過大家聚在一起吃飯、聊天是相當的開心。

最後我要感謝我的父母，當初決定要讀研究所，沒有跟家人進行討論過，就報名了研究所，等到拿到研究所的入取通知書，才告訴爸爸媽媽我要讀研究所，他們非常的尊重我的選擇，希望我好好地讀研究所，只求身體健康、平安順利就好，如今我完成了我最後的學業，不負他們的期待。

鍾奉原 謹誌

中華民國 109 年 7 月

論文名稱：基於深度學習之網球軌跡影像分析

校所名稱：東海大學資訊管理學系研究所

畢業時間：109年07月

研究生：鍾奉原

指導教授：林正偉

論文摘要：

體育選手在比賽時的表現，目前大多由教練或團隊成員以人工的方式進行紀錄及分析，比如透過錄影的方式進行賽後的影像播放及狀態分析。本研究提出了針對的網球軌跡進行擊球方向的分析，幫助選手及教練能夠更快、更方便地進行比賽資料的蒐集及分析。基於深度學習模型，我們使用三種方法進行網球比賽擊球方向分類，分別為 C3D、TrackNet+C3D 和 TrackNet+Trajectory Analysis，在 30 FPS 的比賽影片上使用連續 16 幀影像資料進行網球軌跡追蹤與擊球方向分類。C3D 與 TrackNet+C3D 可以分析近方選手回球的四個方向，TrackNet+Trajectory Analysis 可以同時分析近方與遠方選手的共八個回球方向。我們使用 Youtube 上面的網球單打比賽影片進行實驗，使用 C3D 方法分析擊球方向的分類精準度為 49%、TrackNet+C3D 的分類精準度為 79%，而 TrackNet+Trajectory Analysis 的分類精準度為 89%。目前這三種方法的處理時間仍相對較慢，未來將進行模型優化，協助網球教練或者分析人員用於實時處理，更快的了解選手比賽時的狀態，協助進行戰術分析。

關鍵詞：深度學習、網球、軌跡追蹤、軌跡分類。

Title of Thesis : Image Analysis of Tennis Trajectory Based on Deep Learning

Name of Institute: Tunghai University, Graduate Institute of Information Management

Graduation Time : 07/2020

Student Name : Fong-Yuan Zhong

Advisor Name : Jeng-Wei Lin

Abstract :

Currently, the performance of sports players during a game is mostly recorded and analyzed manually by coaches or team members, such as taking video recording and after the game doing video playback and performance analysis. This research puts forward the tracking of tennis ball and analysis of its direction, helping tennis players and coaches to collect and analyze tennis game data faster and more conveniently. Based on the deep learning models, we use three methods to track the tennis ball and classify its direction, namely C3D, TrackNet+C3D and TrackNet+Trajectory Analysis, using continuous 16 frames from tennis match video of 30 FPS (frame per second). C3D and TrackNet+C3D can analyze four directions of the ball hit by the player in the bottom side, and TrackNet+Trajectory Analysis can analyze eight directions of the ball hit by the players in the both sides at the same time. We use the match videos of tennis singles on Youtube to conduct the experiments. Using the C3D method to analyze the direction of the ball, the classification accuracy is 49%, and those of TrackNet+C3D and TrackNet+Trajectory Analysis are 79%, and 89%, respectively. At present, the processing time of these three methods is still slow. In the future, model optimization will be carried out for real-time processing, to help tennis coaches and players to understand the player's state during the game in real time.

Keywords : Deep learning, Tennis, Trajectory Tracking 、 Trajectory Classification

# 目次

第一章、緒論 .....	1
第一節、研究背景與動機 .....	1
第二節、研究目的 .....	1
第三節、研究流程 .....	2
第二章、文獻探討 .....	3
第一節、卷積神經網路 .....	3
第二節、軌跡追蹤 .....	5
第三節、TrackNet .....	8
第四節、C3D .....	9
第五節、模型評估 .....	10
第三章、資料集及標記工具 .....	12
第一節、資料集 .....	12
第二節、標記 .....	13
第三節、標記工具 .....	14
第四章、實驗 .....	19
第一節、C3D .....	19
第二節、TrankNet+C3D .....	20
第三節、TrackNet +Trajectory Analysis .....	22
第五章、實驗結果 .....	29
第一節、C3D 結果 .....	29
第二節、TrackNet+C3D 結果 .....	32
第三節、C3D VS TrackNet+C3D .....	34
第四節、TrackNet +Trajectory Analysis 結果 .....	36
第五節、時間成本 .....	39
第六章、結論 .....	40
參考文獻 .....	41

## 表次

表 2-1	網球追蹤比較 .....	7
表 3-1	軌跡資料集標記文件格式 .....	13
表 3-2	軌跡分類資料集標記文件格式 .....	13
表 3-3	擊球片段時間紀錄 .....	16
表 4-1	C3D 軌跡分類標籤示意表 .....	20
表 4-2	TrackNet+C3D 軌跡分類標籤示意表 .....	22
表 4-3	XY 方向組合 .....	24
表 4-4	軌跡組合表 .....	26
表 4-5	軌跡組合標籤介紹 .....	27
表 5-1	C3D 多類別混淆矩陣 .....	29
表 5-2	C3D 精密度 .....	30
表 5-3	C3D 召回率 .....	30
表 5-4	C3D F1 Score .....	31
表 5-5	TrackNet+C3D 多類別混淆矩陣 .....	32
表 5-6	TrackNet+C3D Precision .....	32
表 5-7	TrackNet+C3D Recall .....	33
表 5-8	TrackNet+C3D F1 Score .....	33
表 5-9	模型評估比較表 .....	34
表 5-10	TrackNet +Trajectory Analysis 計數統計 (單一片段) .....	37
表 5-11	TrackNet +Trajectory Analysis 計數統計 (多片段) .....	38
表 5-12	模型訓練時間比較表 .....	39
表 5-13	模型預測時間 .....	39

## 圖次

圖 1-1	研究流程圖 .....	2
圖 2-1	類神經網路概念圖 .....	3
圖 2-2	卷積運算示意圖 .....	4
圖 2-3	ReLU 函數示意圖 .....	4
圖 2-4	最大池化示意圖 .....	5
圖 2-5	TrackNet 模型架構圖 .....	8
圖 2-6	網球偵測熱圖 .....	8
圖 2-7	C3D 模型架構圖 .....	9
圖 2-8	混淆矩陣示意圖 .....	10
圖 2-9	ROC 曲線示意圖 .....	11
圖 3-1	前處理示意圖 .....	12
圖 3-2	LabelingTool 標記工具 .....	14
圖 3-3	LabelingTool 使用流程圖 .....	15
圖 3-4	video_mark 標記工具 .....	16
圖 3-5	擊球分類標記流程 .....	17
圖 4-1	C3D 軌跡分類架構圖 .....	19
圖 4-2	本實驗 TrackNet 模型圖 .....	21
圖 4-3	TrackNet+C3D 軌積分類架構圖 .....	21
圖 4-4	TrackNet+Trajectory Analysis 軌跡分類架構圖 .....	22
圖 4-5	Trajectory Analysis 步驟流程 .....	23
圖 4-6	Trackclass 步驟 1 概念圖 .....	24
圖 5-1	ROC 曲線圖 .....	35
圖 5-2	PR 曲線圖 .....	35
圖 5-3	TrackNet+Trajectory Analysis 分類計數 .....	36



# 第一章、緒論

## 第一節、研究背景與動機

現在深度學習的快速發展，傳統的影像辨識技術，可透過卷積神經網路進行代替，且傳統影像辨識技術，需要花費大量時間撰寫公式，去處理影像特徵，如今只需透過卷積神經網路，即可以取得影像特徵，進行影像辨識、分類、物體定位等等。

在體育項目中的球類運動，影像辨識的技術尤為重要，能為球員取得資料的分析，不管是棒球、網球等等球類運動，皆有這個需求，棒球項目球員的資料分析想當然就耳熟能詳，棒球教練要分析對手的球路，守備狀況，不過此工作需要教練、或者相關分析人員需要觀看眾多的球賽影片才能分析，只要能夠過影像辨識的技術進行分析，即可以減少教練需要花費大量時間觀看的成本。

在體育競賽中，比賽影片常用於賽後檢討以及戰術分析。以高爾夫球為例，攝影機可捕捉關鍵的揮桿動作，透過影片回放或測量影像中肢體的變化角度或者相對位置，教練即可指出揮桿動作的問題進行修正。在籃球比賽而言，影片則常用來於評估球員的反應狀態，透過大數據分析，即可預測採取攻擊或防禦戰術後，得失分變化期望值，在勢均力敵的比賽中，使用不同戰術所帶來的比分上的變化對於勝負為至關重要。在 NBA 當中，企業界使用高解析攝影結合影像處理技術來計算球的傳導及紀錄球員的跑位資料等。然而網球、羽球、棒球等運動，因球的體積小且移動速度快且模糊的問題，相對其研究項目較少。

在本研究中，針對日常的網球比賽影片，本研究規劃三種方法，用來偵測並定位一般攝影設備所錄製的網球比賽影片中的網球，進而進行網球選手擊球之分類，用於協助業餘及職業團隊進行資料收集及應用。可為比賽選手進行更有效的分析比賽狀況，減少人工記錄的時間及人力成本。

## 第二節、研究目的

為了瞭解網球選手的狀態，以往的選手狀態是由教練經由影片回放，手動紀錄的方式，來了解選手的狀態表現，本研究使用影像辨識技術，來記錄選手的回擊球狀況，可將教練以往手動的方式改為自動，以便更快速了解選手的狀況。

本研究目的，在於協助網球比賽中之賽事分析，可以分析選手之擊球路線。分析選手之比賽擊球狀況。用於協助網球教練用於賽事結束之後的賽後探討及資料蒐集與分析。

### 第三節、研究流程

本研究之研究流程分為 7 個步驟，首先訂定題目方向，然後進行文獻蒐集請參考第二章、確認研究方法請參考第四章、資料集蒐集請參考第三章、進行實驗實作、進行實驗結果分析請參考第五章，最後依照實驗結果進行討論定下結論請參考第六章。其研究流程如圖 1-1 所示。

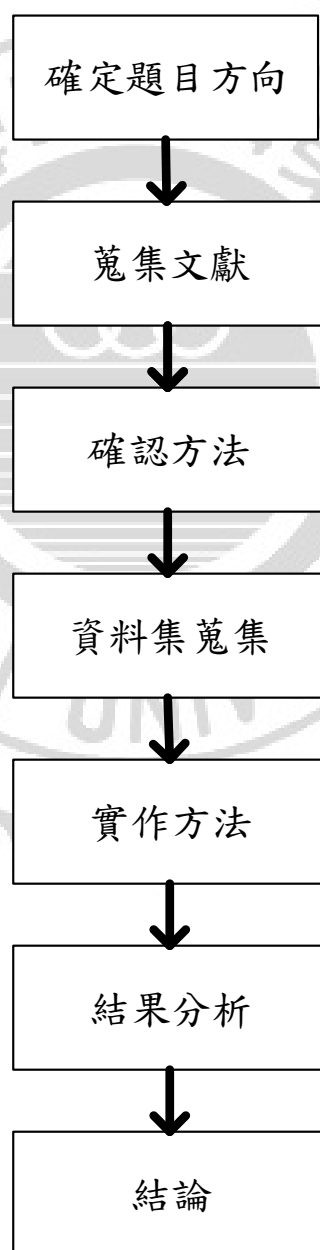


圖 1-1 研究流程圖

## 第二章、文獻探討

### 第一節、卷積神經網路

在最近這幾年深度學習日益盛行，深度學習在圖像分類問題上的成功案例越來越多(Krizhevsky et al.,2012; Simonyan & Zisserman, 2014)，如：物體偵測和截取(Redmon et al., 2016; Ronneberger et al., 2015)、網絡安全、電腦遊戲、活動識別(Jiang & Yin, 2015; Zhou, Xie, 2014)、文字/圖片語義理解、及智能商店等應用。

深度學習為類神經網路方法所構成的數學模型架構的總稱，而深度學習模型為使用大量資料來訓練學習出來的數學模型，基本的深度學習架構為一個輸入層、多個隱藏層和一個輸出層所構成，如圖 2-1 所示，圖中隱藏層每個節點都代表一個神經元，存在其函數模型之權重值。在影像辨識是領域中最常看見的深度學習網路架構為卷積神經網路，其龐大的卷積類神經網絡（CNN），隱藏層與隱藏層之間最常見的運算方法包括：卷積（convolution）、池化（pooling），最後再結合完全連接層或者 Softmax 層所構成，例如：卷積神經網路中被廣泛使用的 VGG-16 模型 (Simonyan, & Zisserman, 2014)，主要由 convolution 層、maximum pooling 層和 ReLU 層以及全連接層所組成的模型。

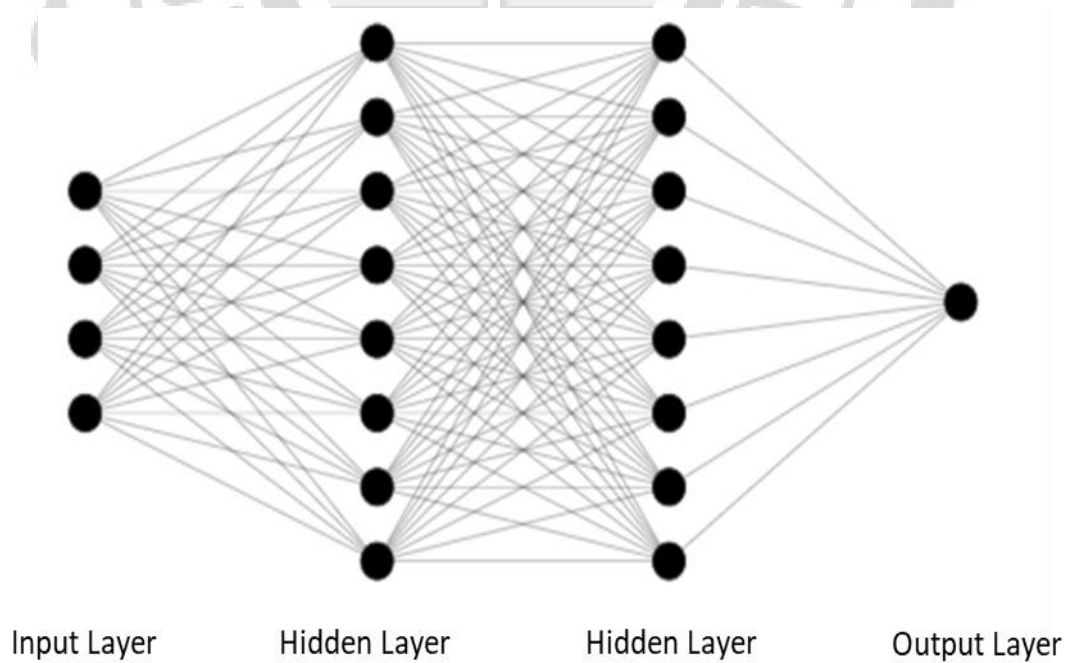


圖 2-1 類神經網路概念圖

卷積神經網路是由 LeCun 卷積之父 (LeCun,1989) 在 1989 年提出的一個神經網路架構，每層的資料都是由  $W \times H \times D$  構成的數據矩陣構成， $W$ 、 $H$ 、 $D$  分別為數據的寬度、高度、及深度。卷積的計算是由一個  $w \times h$  大小的卷積核在  $W \times H \times D$  的範圍內上移動計算的濾波器。步幅通常設定為 1 為卷積核移動的步數。卷積運算如圖 2-2 所示

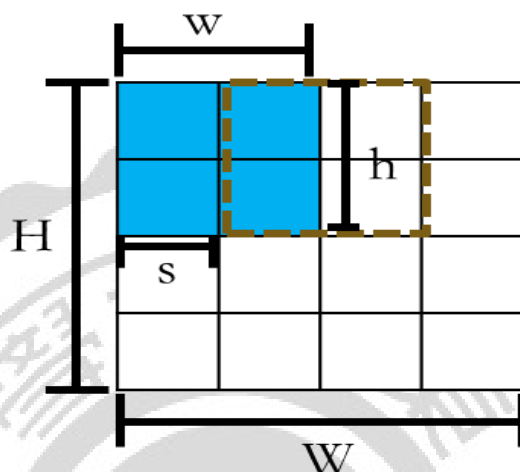


圖 2-2 卷積運算示意圖

在深度學習中，由於問題大多為非線性的問題，所以會使用激活函數產生非線性的結果，而卷積運算結果為線性的，無法解決非線性問題，因此會使用激活函數將原先線性結果轉化為非線性結果，ReLU 激活函數為目前深度學習中最常見的激活函數。ReLU 激活函數在當輸入值為負時，則該函數輸出 0，否則將輸入值直接輸出。ReLU 運算如圖 2-3 所示

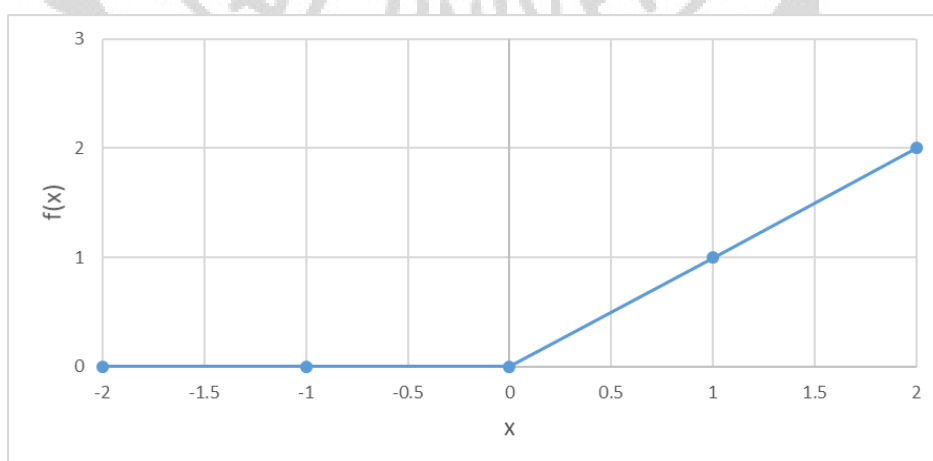


圖 2-3 ReLU 函數示意圖

在卷積網路中常見到一種方法為池化運算，池化運算具有將數據降維以及特徵聚合的特性。目前的卷積網路的池化運算大多使用最大池化（max pooling），在該運算中一個區塊中的數據會由最大值給取代。例如原數據為 4\*4 矩陣經由 2\*2 的池化算結果如圖 2-4 所示：

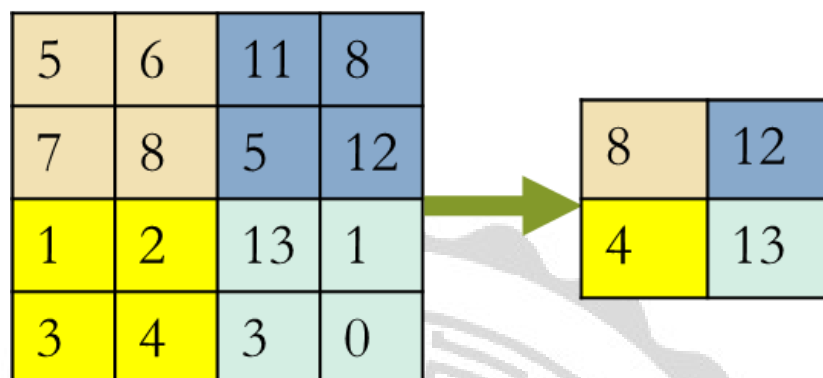


圖 2-4 最大池化示意圖

物件偵測是深度學習影像辨識中的問題之一，其中 VGG-16 為常用的神經網路模型。圖片中可能同時存在多個物件且物件大小不一，R-CNN(Girshick.R,2014; Girshick.R,2015; Ren et.al,2015)家族結構化地檢測畫面，從中先找出多個可能包含物件的區域，再對這些區域進行仔細的物件偵測與分類，但是由於計算速度較慢則無法達到即時應用的需求。所以為了加快處理速度，YOLO (Redmon, 2016)家族使用同時偵測及辨識方法，在有限的搜尋空間中偵測物體，提高了處理速度，精簡版的本 Tiny YOLO 甚至可以在 Raspberry Pi 等簡易型電腦上運行運算處理。完全卷積網路（FCN）(Long et.al., 2015)以像素為單位進行分類，為了還原特徵圖編碼過程所減損的尺寸使用上取樣及逆卷積網路常被用在生成對應原尺寸大小的資料陣列，如：熱度圖可表示目標物件的存在位置。

## 第二節、軌跡追蹤

近年來，球類運動依據球軌跡分析選手表現受到關注，目前已有多個追蹤球行進軌跡的演算法以及系統被提出來，如針對排球(Chen et al., 2012)、籃球(Fu et al., 2011)、桌球(Myint et al., 2015)和團隊球類運動(Wang, et al, 2014) 等軌跡研究。對於職業運動而言，自動化的影像分析工具，具有更高的價值。商業用途的運動影像分析工具使用高速攝影機進行比賽錄影，以得到較佳的影像，甚至在多個不同的位置使用多部錄影機來取得不同角度的影像，因此，軟硬體的成本相當高。例如，廣

泛運用於職業網球比賽的鷹眼系統(Hawk-Eye)，它可以計算網球行進的軌跡，並透過視覺化畫面協助裁判與球員解決有爭議的判決。然而，該系統需要在特定位置以及角度部署多部高速攝影機，並需要專人操作，成本相當的高。

在影片中嘗試定位網球的位置有不少的相關研究，但因網球體積小，在影像中面積占比小，容易與畫面中相似形狀混淆或顏色，則產生錯誤偵測，而且球移動速度快，會造成影像中形成殘影及模糊，則產生失敗偵測。網球的球軌跡相關研究(Yu et al., 2004)，如：使用多台攝影機結合 3D 技術進行網球位置偵測及追蹤(Renò et al., 2016)，使用粒子濾波器在低質量的影片中進行網球的位置追蹤(Yan et al., 2016)，以雙層資料關聯法從各幀影像處理含有失敗偵測和錯誤的結果中，算出最有可能的球軌跡位置(Zhou et al., 2014)。使用是四軸飛行器上的攝影機搭配隨機森林方法進行軌跡的預測(Qazi, 2015)。

使用多部攝影機進行定位網球軌跡運算的相關研究有不少，(Pingali, Opalach, & Jean, 2000)使用多部攝影機進行球與幀差和檢測。在攝影機中跟踪球的位置，獲得 3D 軌跡。(Renò et al., 2017)使用四個照相機在對重構移動實體所使用的在 3D 空間中，然後執行領域知識為基礎的推理找出球和重建自己的軌跡通過拆分和重新合併軌跡。(Yu et al., 2009)因攝影機拍攝內容會產生誤差，提出了應用的系統在需要的地方插入用於補充三維虛擬內容視頻，用於輔助球的定位及軌跡追蹤。

(Reno et al., 2018)提出一個卷跡神經網路(CNN)，針對網球球場上影像進行判斷畫面中，是否有網球的存在，將原始影像中的畫面進行裁切，總裁切數量為 116385 張圖片，類別為有球及無球，以 2 分法進行網球存在的一個模型。(D'Orazio et al., 2004)使用定向圓霍夫變換的修改版來檢測圖像中最適合容納球的區域，再將神經分類器應用於所選區域，以確認是否已正確檢測到球。

(Archana & Geetha, 2015) 先將各幀的圖像使用中值濾波器進行平滑處理去除細小的圖像雜訊，計算背景模型後，進行背景去除以取得前景。因網球的移動速度快，所以從前後兩幀的差異中找出可能為網球的部分，再使用邏輯中的 AND 運算兩者，得到在前景且移動相對較快的物件，再以形狀、大小和長寬判斷，找出可能為網球的物件，接下來應用擴張和侵蝕技術來找到網球影像。最後使用神經網路進行軌跡的預測。

以上所探討關於網球軌跡追蹤研究或系統之比較如表 2-1 所示，進行探討技術、技術類型與精準度進行比較。

表 2-1 網球追蹤比較

系統或者研究	技術	技術類型	精準度
Hawk-Eye	光纖影像技術		誤差 2.5mm
Reno et al, 2016	影像減法+ 3D 定位	傳統	誤差 10cm
Yan et al, 2016	粒子濾波器	傳統	-
Zhou et al, 2014	雙層資料關聯法	傳統	65.6%
Qazi, 2015	影像減法+ 隨機森林	傳統+Machine Learning	94%
Pingali, 2000	影像減法+ 3D 定位	傳統	-
Renò et al, 2017	影像減法+ 3D 定位	傳統	83%
Yu et al, 2009	相機校正法+ 3D 定位	傳統	誤差 14cm
D'Orazio et al, 2004	定向圓霍夫+ 神經網路	傳統+Machine Learning	93.78%-
Archana, 2015	濾波器+影像減 法+神經網路	傳統+Machine Learning	90%

從表 2-1 可看出，在領域中學者大多使用傳統影像處理方式進行，僅少部分學者或系統使用機器學習(Machine Learning)的方法進行軌跡追蹤，其精準度為'-'，代表該研究並無特別提到精準度，從表中可看出使用了機器學習之方法其精準度是相當優異。Hawk-Eye 系統是目前大型球類比賽常使用來進行軌跡追蹤之系統，其成本花費是相當的高。

### 第三節、TrackNet

TrackNet 為(Huang et al., 2019)提出的卷積神經網路模型，其用途為在網球廣播影片中找出網球所在，因此可構成影片中的網球之移動軌跡。其模型使用 VGG-16 模型取得其網球特徵加上 DeconvNet(Noh et al., 2015)模型進行圖像的恢復。該研究將模型分為一幀影像及三幀影像輸入，在該研究發現三幀輸入能夠更好的學習到球的特徵，可改善在影像中網球因障礙物所阻擋而看不到的軌跡。該研究使用單幀或三幀影像進行輸入，最後生成一張檢測網球的熱度圖。該熱點圖與原始影像中的解析度為相同。該熱點圖中的白點為網球在影像中所在位置。TrackNet 模型架構如圖 2-5 所示。TrackNet 模型輸出的熱圖如圖 2-6 所示。

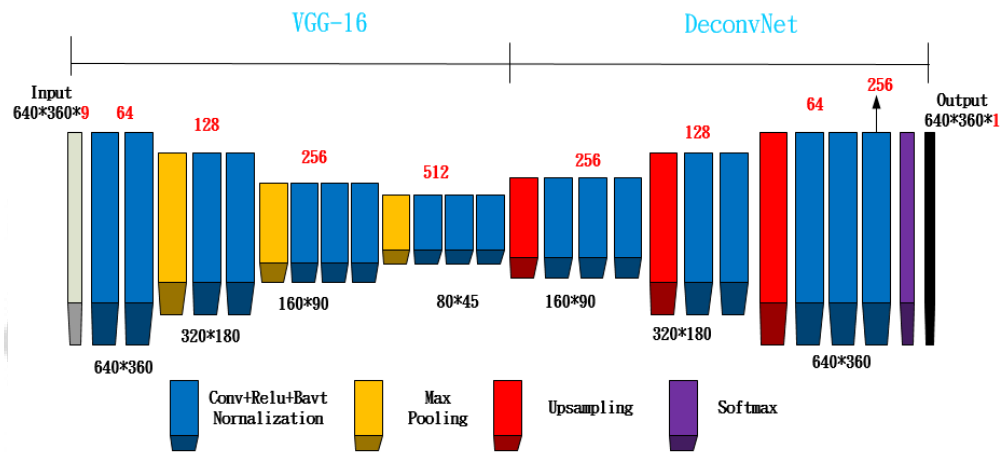


圖 2-5 TrackNet 模型架構圖

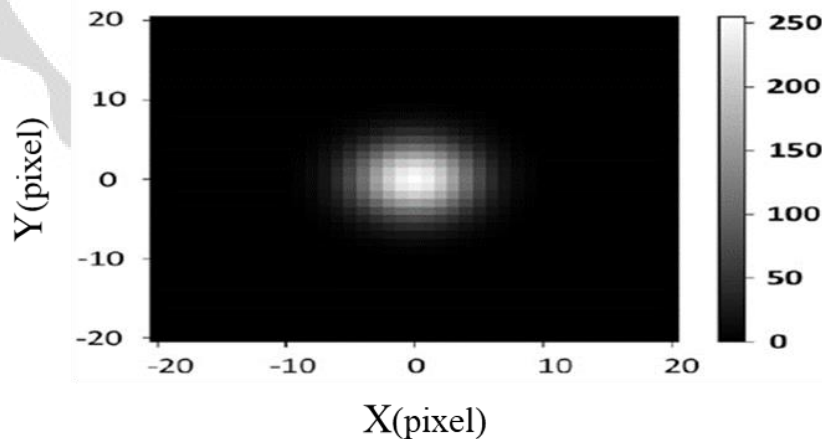


圖 2-6 網球偵測熱圖

該研究使用連續三幀作為輸入的 TrackNet 模型達到 99.8%的 Precision 和 96.1%的 Recall。能夠看出有相當高的準確度。本研究使用 TrackNet 模型進行網球軌跡的座標行為預測。



## 第四節、C3D

C3D(Tran et al., 2015)是 2015 年提出，C3D 提出了時間概念的 CNN 網路，平常的 CNN 網路是 2D 的將單張圖像的資料放入 CNN 網路中進行識別，而 C3D 網路因為有提出將影片中的 16 幀影像進行 CNN，所以可以達到動作的識別，因為動作是由連續的影像所構成。其精準度在 UCF-101 資料集中可達到 78.3%，在 C3D 當中提出 2 個特點 1.與二維 ConvNets 相比，3D ConvNets 更適合於時空特徵學習，2.卷積核的選擇  $3 \times 3 \times 3$  卷積核心的均勻架構是 3D ConvNets 中最好的架構。

C3D 提出有時間長度的卷積核為  $d \times k \times k$ ，其中  $k \times k$  為我們熟知的卷積核， $k$  為空間大小，而不同的  $d$  為時間深度。其資料集中的視頻轉換成  $c \times l \times h \times w$  的格式， $c$  為通道數、 $l$  為時間長度、 $h$  為影像高度、 $w$  為影像寬度。

C3D 所提出的網路架構為 5 個卷積層、5 個池化層、2 個全連接層和 Softmax 層。C3D 模型架構如圖 2-7 所示。

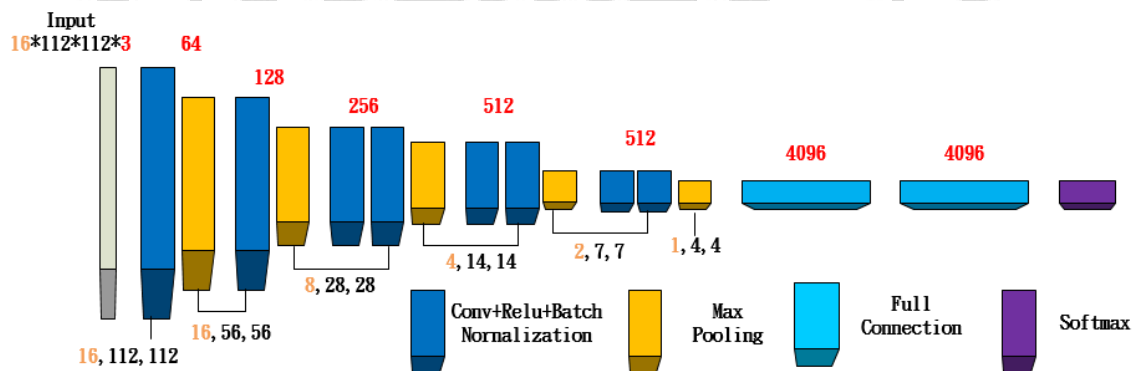


圖 2-7 C3D 模型架構圖

本研究將使用 C3D 模型特性進行實驗，C3D 模型帶有的時間概念的卷積運算對於在進行軌跡分類上可以達到幫助，因在擊球軌跡上需要多張連續影像才可構成，所以本研究將採用 C3D 模型進行實驗。

## 第五節、模型評估

在本研究針對模型的評估使用了多個評估方式，進行模型評估包含了混淆矩陣、準確度、精密度、召回率、F1 Score、Micro F1、ROC 和 AUC。

混淆矩陣：

混淆矩陣為在機器學習領域中是很容易用到的一個結果分析表，在混淆矩陣中每一列為預測值，每一行為真實值。混淆矩陣中名詞介紹：混淆矩陣如圖 2-8 所示。

TP：實際為真，預測為真。

FN：實際為真，預測為假。

FP：實際為假，預測為真。

TN：實際為假，預測為假。

Actual	Predicted		TP	FN
1	1	→	2	1
1	2			
2	1		FP	TN
2	2		1	-
1	1			

圖 2-8 混淆矩陣示意圖

準確度(Accuracy)：代表了在總樣本中預測成功的比率，等於正確預測的數量（包括正確預測為正或負），除以總預測數（包括正確預測為正或負，以及錯誤預測為正或負）。公式如下：

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FN} + \text{FP} + \text{TN})$$

精密度(Precision)：為預測某樣本中的該樣本的比率

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

召回率(Recall)：為真實值樣本預測中預測到該樣本中比率

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

F1 Score：Sasaki, Y. (2007)為該樣本對於整體預測分數

$$\text{F1-score} = 2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$$

## ROC 曲線

ROC 曲線：ROC 曲線全名為 Receiver Operating Characteristic curve(Fawcett, Tom 2006)首先在由二次世界大戰中的一名工程師發明的，用於敵軍載具上的信號檢測，在近期 ROC 曲線多用於無線電、醫學、生物學、犯罪心理學領域等，並且也使用在機器學習的應用。ROC 曲線將定義 X 軸為偽陽性率 (FPR)，定義 Y 軸為真陽性率 (TPR)。ROC 曲線請參考圖 2-9。

TPR：在所有實際樣本為正樣本中，被正確地判斷為正之機率。

$$TPR = TP / (TP + FN)$$

FPR：在所有實際樣本為負樣本中，被錯誤地判斷為正之機率。

$$FPR = FP / (FP + TN)$$

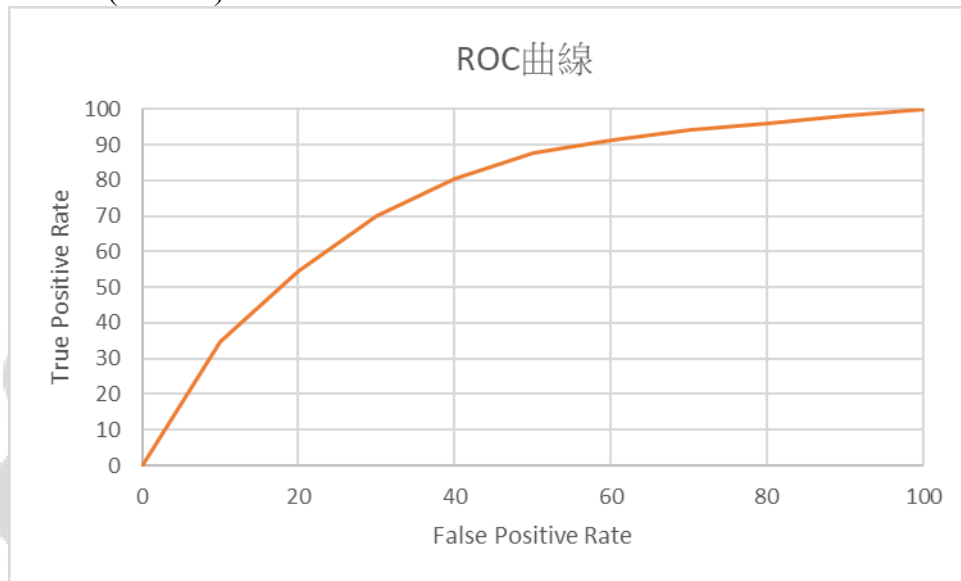


圖 2-9 ROC 曲線示意圖

在 ROC 曲線中有個名詞為 AUC 全名為 Area under the Curve，AUC 為曲線下面積，代表為 ROC 曲線下的面積。可用來比較多個模型中曲線下面積做為模型優劣的指標，AUC 數值越高其代表模型預測能力越強。AUC 曲線越接近直角，代表模型越優良。當  $AUC = 1$ ，為完美分類器。當  $0.5 < AUC < 1$ ，代表有預測價值。AUC = 0.5，代表為隨機猜測，模型沒有預測價值的能力。AUC < 0.5，預測能力比隨機猜測還要差。

### 第三章、資料集及標記工具

#### 第一節、資料集

在影音平台 Youtube 2017 年夏季世大運網球男子單打決賽，解析度為 1280×720，總時長 75 分鐘，幀速率為 30 fps。

前處理：

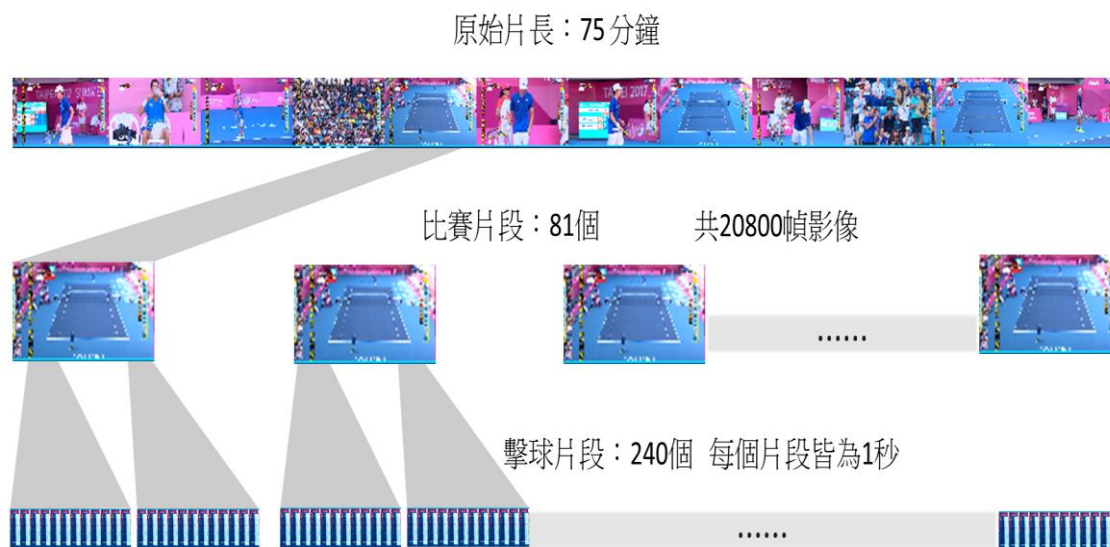


圖 3-1 前處理示意圖

圖 3-1 為本研究資料集前處理示意圖，其流程為使用影片剪輯工具去除掉非比賽的部分，分割出了 81 個片段，為發球後到死球的過程，為比賽片段，其 81 個比賽片段使用 FFmpeg 工具將片段中的各幀影像儲存下來，儲存出 20580 幀影像資料，用於 TrackNet 模型進行學習網球定位追蹤。接下來將原 81 個片段影像會依照各片段影片中的擊球軌跡，再進行一次分割片段動作，總共切出 240 個擊球片段資料，擊球片段時長皆為 1 秒，用於 C3D 模型進行擊球軌跡分類學習。前處理結束將進行資料標記動作，請參考本章第二節、標記進行標記部分之敘述。

## 第二節、標記

前處理的 81 個片段的各幀圖片球之座標都將進行標記，作為軌跡追蹤資料集，以供進行軌跡追蹤網路學習。

軌跡追蹤網路資料集標記文件中記錄方式如表 3-1 所示，標記文件中記錄各個圖片檔，路徑以及球坐標 X 及坐標 Y 的記錄，共標記 20508 張影像中的網球位置座標。

表 3-1 軌跡資料集標記文件格式

Path,x,y
aa/clip1/1.jpg, 10,20
aa/clip1/2.jpg, 200,500
aa/clip1/3.jpg, 50,600
.....
.....
.....

前處理的另外 240 個片段資料依照標籤，進行分開存放，做為軌跡分類資料集，以供進行軌跡分類網路學習。訓練集為 180 個樣本，測試集為 60 個樣本。

軌跡追蹤網路資料集標記文件中記錄方式如表 3-2 所示，標籤文建會記錄各個影片片段之路徑資料夾，以及分類標記代號。分類標記代號分別有 0、1、2、3 四種標記代號。0 代表已方選手從左半場打到左半場、1 代表已方選手從左半場打到右半場、2 代表已方選手從右半場打到左半場、3 代表已方選手從右半場打到右半場。

表 3-2 軌跡分類資料集標記文件格式

Path,Label
aa/clip1_1, 0
aa/clip1_2, 1
aa/clip2/,2
.....
.....
.....

軌跡資料集標籤文件以及軌跡分類資料集標記文件，此 2 個標記文件依照資料數量以 75%與 25%分為訓練集與測試集，分為軌跡資料訓練集標記文件、軌跡資料測試集標記文件、軌跡分類訓練集標記文件和軌跡分類測試集標記文件 4 份文件。在後續的實驗方法中提出了三種方法進行實驗。軌跡資料訓練集標記文件、軌跡資料測試集標記文件，用於進行網球追蹤模型之訓練以及網球追蹤模型之測試。軌跡分類訓練集標記文件和軌跡分類測試集標記文件，用於進行網球擊球軌跡分類模型之訓練以及網球擊球軌跡分類模型之測試。

### 第三節、標記工具

本研究使用了兩套工具進行資料標記，分別使用 LabelingTool 及 video\_mark 工具。LabelingTool 為 TrackNet 提供之標記工具，用於標記球之座標及相關屬性；video\_mark 標記工具，為本研究自行開發之標記工具，用於紀錄擊球片段時間點之標記工具。

網球座標軌跡標記：

首先介紹 LabelingTool 標記工具的操作方式，LabelingTool 標記工具介面如圖 3-2 所示。此標記工具，須將影片轉為各幀影像，紀錄各幀影像中網球位置以及網球屬性，本研究只需使用 LabelingTool 標記工具中的網球座標紀錄方式，其餘網球屬性雖有進行標記，不過並無使用到其標記內容。

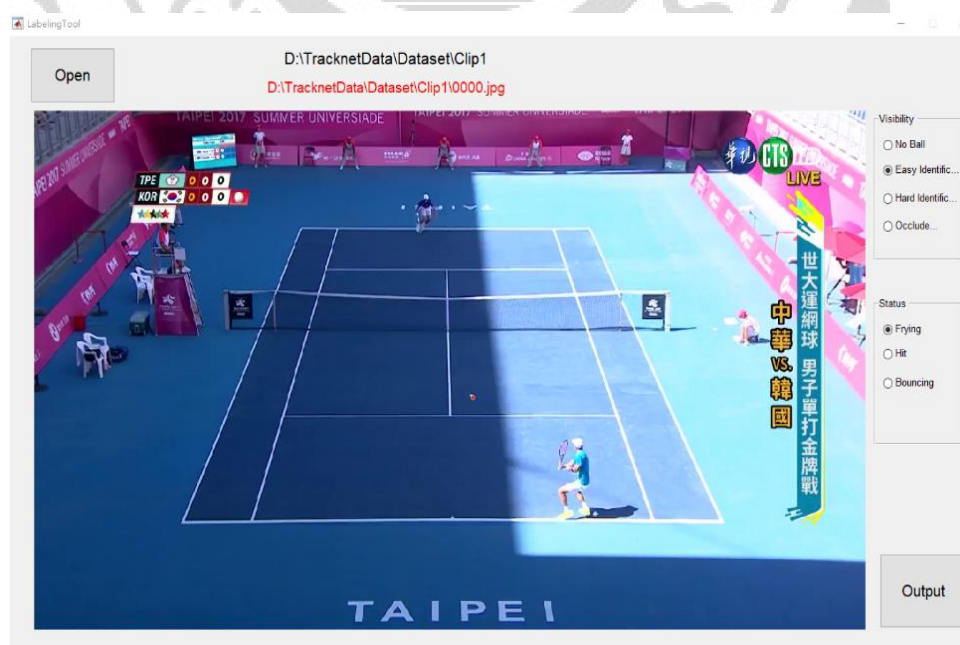


圖 3-2 LabelingTool 標記工具

LabelingTool 標記工具使用說明：

LabelingTool 標記工具使用步驟，如圖 3-3 所示。更詳細內容說明請參考下文說明。

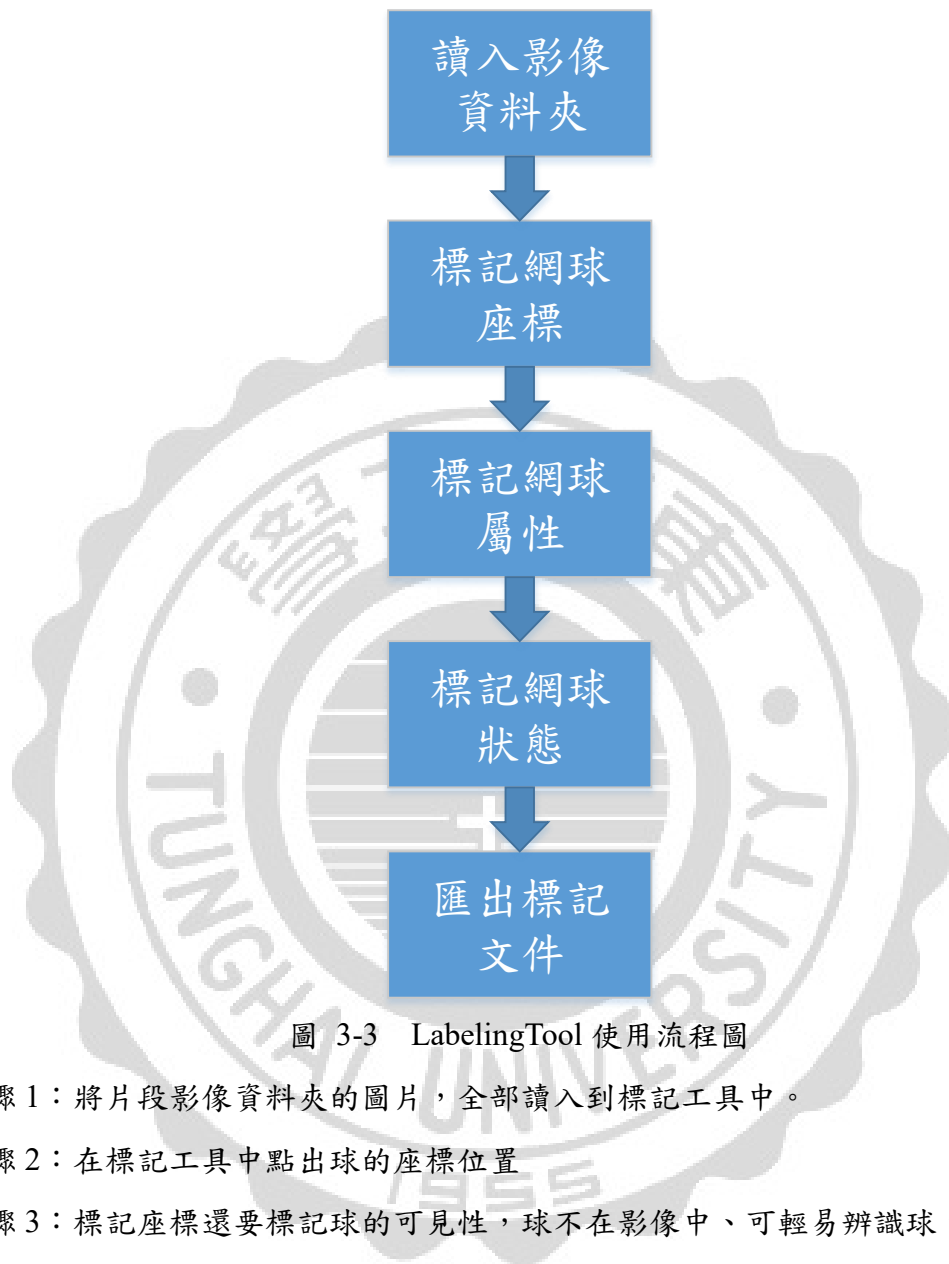


圖 3-3 LabelingTool 使用流程圖

步驟 1：將片段影像資料夾的圖片，全部讀入到標記工具中。

步驟 2：在標記工具中點出球的座標位置

步驟 3：標記座標還要標記球的可見性，球不在影像中、可輕易辨識球、無法輕易識別、球被障礙物阻擋。

步驟 4：標記球的狀態，球在飛行中、球在彈跳、球被擊中。

步驟 5：依序標記到結束

步驟 6：匯出標記文件



擊球片段處理：

擊球片段處理的方式主要是由 video\_mark 標記工具的部分進行完成，video\_mark 標記工具為本研究開發用於紀錄擊球片段時間點，將完整影片擊球時間點進行紀錄，video\_mark 標記工具介面如圖 3-4 所示，其使用說明請參考下文所敘述，



圖 3-4 video\_mark 標記工具

表 3-3 為 video\_mark 標記工具匯出之擊球時間點之文件，其擊球時間點將以每行為一個擊球時間點，其格式為 HH:MM:SS:MS，分別代表小時、分鐘、秒鐘、毫秒。

表 3-3 擊球片段時間紀錄

時間紀錄
00:00:01.33
00:00:02.57
00:00:03.60
00:00:06.67
00:00:34.83
00:00:52.87
.....



video\_mark 工具使用說明：

擊球分類標記需使用 video\_mark 工具及 FFmpeg 工具進行處理，其步驟流程圖 3-4 所示，更詳細內容說明請參考下文說明。

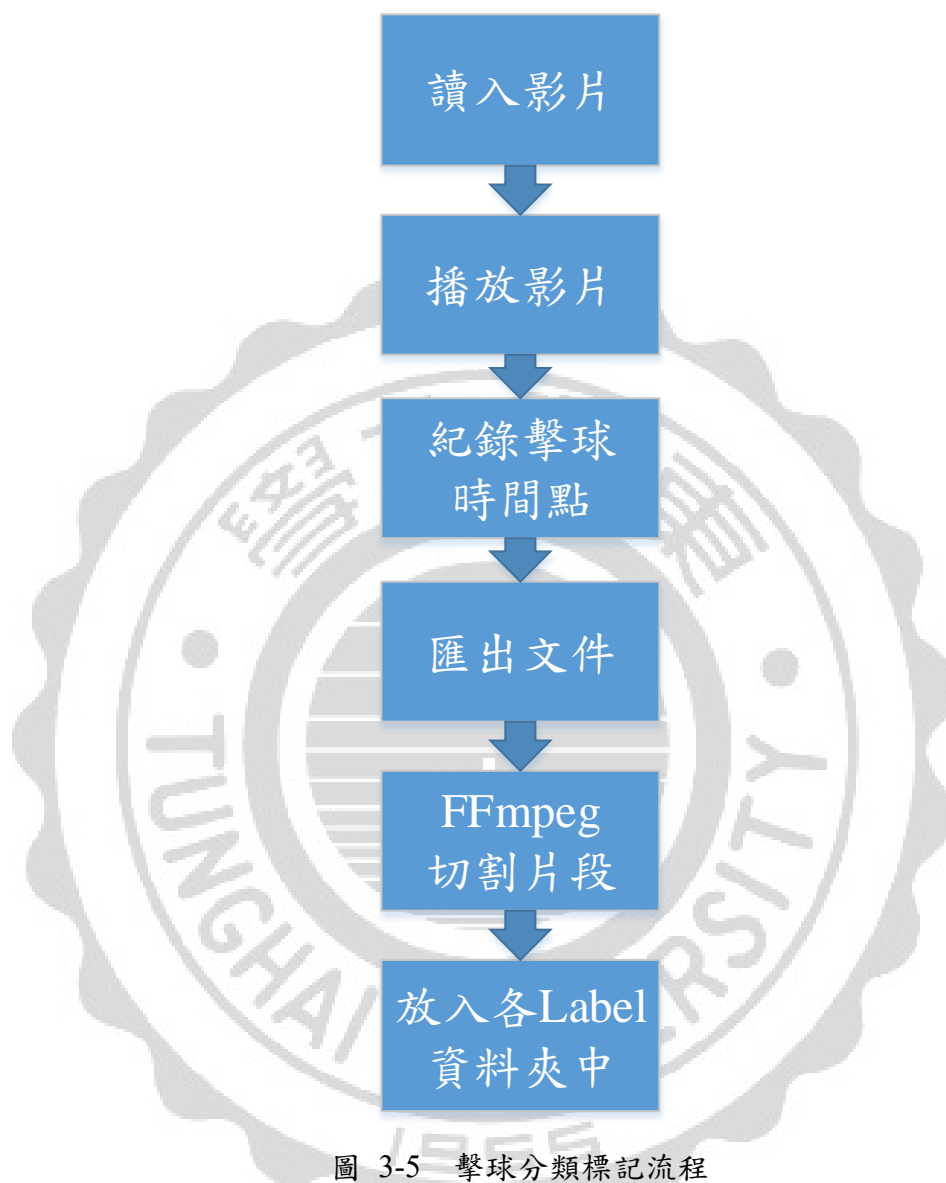


圖 3-5 擊球分類標記流程

步驟 1：將影片讀入到 video\_mark 中。

步驟 2：在擊球時，點及右側的紀錄分割點。

步驟 3：最後會產生一個 txt 檔

步驟 4：使用 FFmpeg 工具進行切割影片段。

步驟 5：將切割完的擊球片段進行人工標記分類處理，分別存放於各個 Label 資料夾底下。

FFmpeg 工具：

FFmpeg 全名為 Fast Forward MPEG (Moving Picture Experts Group) 是一個開源多媒體處理框架軟體，可執行視訊和音訊多種格式的影音解碼、編碼、編碼轉換、混合、抽取、串流、過濾等功能，且使用 CLI(Command-Line Interface)操作方式容易上手，只需使用指令參數進行操作可進行快速轉檔等動作，無需使用 GUI(Graphical User Interface)介面進行操作使用。

使用方式：

FFmpeg 工具主要是使用 CLI 介面進行操作，所以必須使用指令化方式進行操作，在本研究使用的指令為切割影片以及取得影片各幀圖片的指令，這 2 個指令如以下敘述：

1. `ffmpeg -i "$file" -vf fps=$fps "%05d.jpg"`  
\$file 為影片原檔，\$fps 為 fps，%05d.jpg 為輸出圖片的命名規則。
2. `ffmpeg -i $file -ss $time -t 00:00:01.00 xxx..mp4`  
\$file 為影片原檔，\$time 為標記工具紀錄的欲切割影片的時間，而後方的數字為欲切割影片的時長，而最後的為輸出影片檔名。

本研究使用上述 FFmpeg 工具之指令進行軌跡資料集以及軌跡分類資料集的製作，資料集的部分處理需要使用到上述指令進行處理，方可完成影片轉各幀圖片、以及影片分割的功能。

## 第四章、實驗

為了進行預測網球軌跡之分類，本研究使用三種方法進行了三種實驗。此三種方法分別為 C3D、TrackNet+C3D、TrackNet+Trajectory Analysis。前兩種方法的分類方法有使用的深度學習模型進行擊球軌跡分類，而在第三個實驗方法中的分類方法並無使用到深度學習模型進行分類處理。

為了前兩個實驗的公平性，在軌跡分類資料集的部分為訓練資料集、測試資料集，分別的樣本數為 180 筆和 60 筆樣本供前 2 種實驗的模型學習以及測試。

### 第一節、C3D

在此實驗中本研究僅使用 C3D 架構，將連續 16 幀擊球片段影像輸入經過 C3D 模型後進行分類。此實驗使用 C3D 模型之架構進行分類，僅變更改輸出分類的數量大小，其模型輸入層以及其他卷積層、池化層、激活函數等等參數設定並無做其他更改。此實驗之架構圖如圖 4-1 所示：

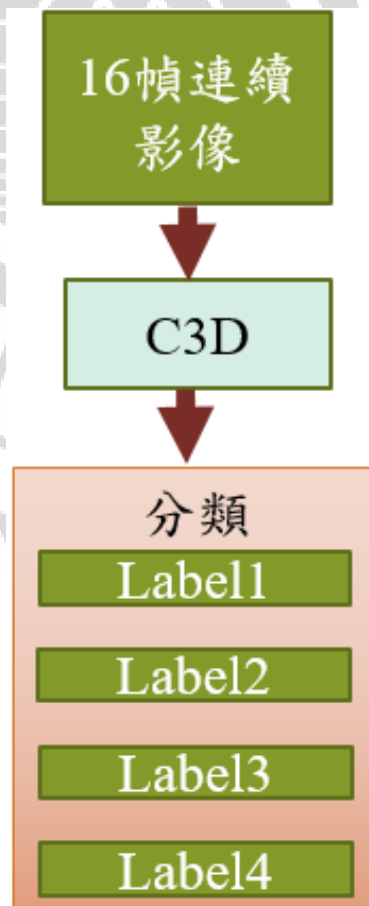


圖 4-1 C3D 軌跡分類架構圖

表 4-1 C3D 軌跡分類標籤示意表





Label	示意圖	
	左半場	右半場
Label1		
Label2		
Label3		
Label4		

表 4-1 為此方法資料標籤的的示意圖，資料標籤有 4 種，這四種標籤為用左右半場切割出各 2 種標籤，依照己方選手擊球方向分為上述四種依照擊球方向分為上述四種。

## 第二節、TrankNet+C3D

在此實驗中本研究使用 TrackNet+C3D 架構，首先使用 TrackNet 模型將連續三幀影像中進行輸入，輸出網球的位置產生網球位置熱圖影像，接下來累計 16 張網球位置熱圖影像輸入至 C3D 模型到當中進行擊球軌跡的分類預測。

此實驗使用 TrackNet+C3D 架構，其原因為改善僅使用 C3D 模型進行擊球軌跡預測之雜訊資料眾多，以便用於優化 C3D 模型資料輸入，用於減少雜訊資料，以得到更好的擊球軌跡分類預測模型。

本研究因運算設備顯示卡記憶體之限制，原始 TrackNet 模型原始影像輸入尺寸無法進行使用，所以本研究將 TrackNet 原始模型中的輸入尺寸進行了變更，原始 TrackNet 模型為輸入尺寸為 360\*640\*9 的圖片大小更改為了 288\*512\*9 的輸入尺寸的影像大小進行輸入，其餘 TrackNet 參數皆無更改，圖 4-2 為此實驗前半段針對網球追蹤的所使用之 TrackNet 模型架構圖。本研究所使用 TrackNet 模型架構圖如圖 4-2 所示。

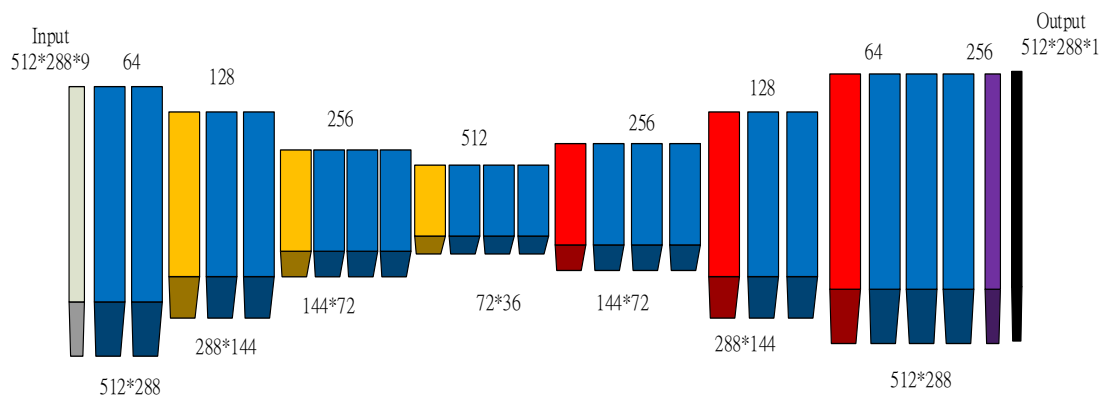


圖 4-2 本實驗 TrackNet 模型圖

針對軌跡分類之模型中的 C3D 模型與上個實驗的 C3D 模型設定參數為一模一樣。本實驗擊球軌跡分類架構圖如圖 4-3 所示，前半部為使用 TrackNet 模型利用該模型進行網球位置的預測產生其網球位置之熱圖，後半段加上 C3D 模型進行擊球軌跡之分類預測。

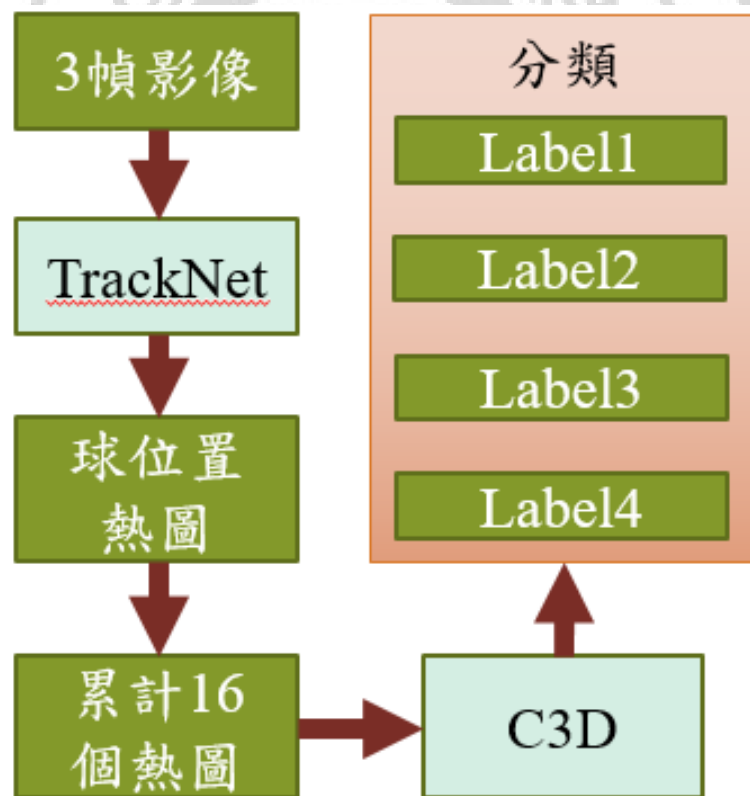


圖 4-3 TrackNet+C3D 軌積分類架構圖

表 4-2 TrackNet+C3D 軌跡分類標籤示意表





Label	示意圖	
	左半場	右半場
Label1		
Label2		
Label3		
Label4		

表 4-2 為 TrackNet+C3D 資料標籤的的示意圖，資料標籤有四種，這四種標籤為用左右半場切割出各兩種標籤，依照已方選手擊球方向分為上述四種。

### 第三節、TrackNet+Trajectory Analysis

使用 TrackNet 將影像中的球的位置產生網球位置熱圖，再從熱圖中透過 OpenCV 找出網球位置座標，再使用 Trajectory Analysis 進行網球軌跡分類。Trajectory Analysis 為本研究所自行規劃的軌跡分類演算法。本實驗之軌跡分類架構圖如圖 4-4 所示。

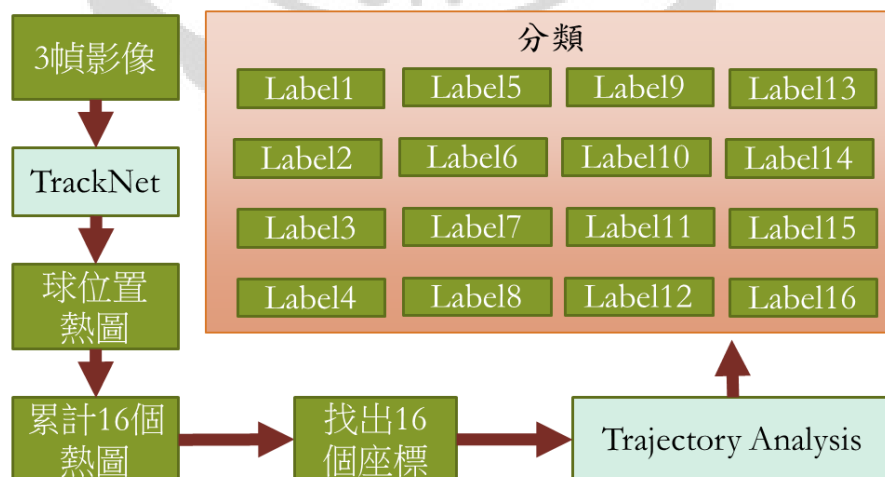


圖 4-4 TrackNet+Trajectory Analysis 軌跡分類架構圖

Trajectory Analysis 演算法步驟：

Trajectory Analysis 演算法其步驟如圖 4-5 所示，首先進行方向判斷、座標軌跡組合，最後分類軌跡組合，更詳細的介紹內容介紹，請參考下文的文字介紹。

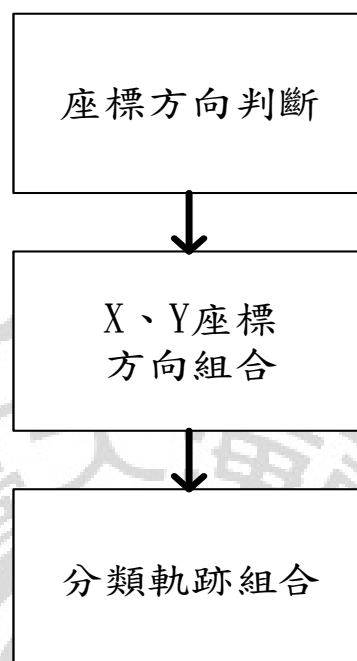


圖 4-5 Trajectory Analysis 步驟流程

步驟 1：

首先會輸入一個 Center\_X，Center\_Y 值為球場中心點座標，進行校正動作，預防影片中球場位置並無置中，則會影像到後面的運算流程。影像中的座標以左上角為原點(0,0)，座標值並無負數存在，X 座標以左上角為原點向右延伸，Y 座標以左上角為原點向下延伸。

接下來將各張影像中球位置座標 X、Y 部分進行 1、0 的標記稱為 NewX、NewY。X 座標：大於 CenterX 值的一半的為 1 則代表為右半場，反之為 0 則代表為左半場；Y 座標：兩兩座標相減， $>0$  為 1 則為向下方移動， $0 <$  為 0 則為向上方移動，相減等於 0 為上一個的 1 或 0。步驟 1 概念如圖 4-6 所示。

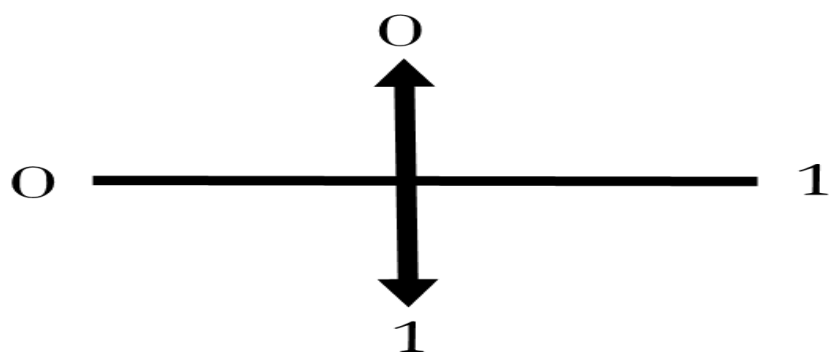


圖 4-6 Trackclass 步驟 1 概念圖

#### 步驟 1 範例

首先使用校正用座標 Center\_X=640, Center\_Y=320, 作為球場中心座標, 進行校正動作。

$X = \{780, 90, 120, 640, 80, 60, 1024, 320, 240, 40\}$   
 $Y = \{600, 200, 100, 522, 50, 60, 58, 80, 90, 80\}$   
 $NewX = \{1, 0, 0, 1, 0, 0, 1, 0, 0, 0\}$   
 $NewY = \{0, 0, 0, 1, 0, 1, 0, 1, 1, 0\}$

#### 步驟 2:

將 NewX 座標的 1、0 做組合, 右半場至右半場為 1、右半場至左半場為 2、左半場至左半場為 3、左半場至右半場為 4; 將 NewY 座標的 1、0 做組合, 下半場至下半場為 1、下半場至上半場為 2、上半場至上半場為 3、上半場至下半場為 4, 取 16 個連續 (1 或 0), 先判斷第一個值為 (1 或 0), 進行 CountIf 判斷在連續 16 個當中是否有不同的值, Countif 僅會對連續出現的值進行加總, Countif 僅會出現 8 或 16 代表是否有反曲點產生, 可找出是否方向改變之存在, 其步驟 2 之方向組合如表 4-3 所示。

表 4-3 XY 方向組合

X 座標方向組合		Y 座標方向組合	
右右	1	下下	1
右左	2	下上	2
左左	3	上上	3
左右	4	上下	4



步驟 2 範例：

X 座標：

範例(1)={1,1,1,1,1,1,1,1,1,1,1,1,1,1,1} ➡ 範例(1)=1

範例(2)={1,1,1,1,1,1,1,1,1,0,0,0,0,0,0} ➡ 範例(2)=2

範例(3)={0,0,0,0,0,0,0,0,0,0,0,0,0,0,0} ➡ 範例(3)=3

範例(4)={0,0,0,0,0,0,0,0,0,1,1,1,1,1,1} ➡ 範例(4)=4

Y 座標：

範例(5)={1,1,1,1,1,1,1,1,1,1,1,1,1,1,1} ➡ 範例(5)=1

範例(6)={1,1,1,1,1,1,1,1,1,0,0,0,0,0,0} ➡ 範例(6)=2

範例(7)={0,0,0,0,0,0,0,0,0,0,0,0,0,0,0} ➡ 範例(7)=3

範例(8)={0,0,0,0,0,0,0,0,0,1,1,1,1,1,1} ➡ 範例(8)=4

以上 8 個範例為步驟 2 的範例，由上面 8 個範例可以看出 X 座標及 Y 座標分別各自產生 4 種座標方向組合，而 X 座標方向組合與 Y 座標方向座標組合，進行兩兩組合共可組合出做出 16 種軌跡組合，詳細的軌跡組合內容請參考步驟 3 所介紹之內容。

步驟 3：











將 X 座標方向組合 1、2、3、4 與 Y 座標方向的組合 1、2、3、4 進行分類軌跡組合。可做出 16 種的軌跡組合，如表 4-4 所示：

表 4-4 軌跡組合表

X 組合	Y 組合	軌跡
1	1	1
1	2	2
1	3	3
1	4	4
2	1	5
2	2	6
2	3	7
2	4	8
3	1	9
3	2	10
3	3	11
3	4	12
4	1	13
4	2	14
4	3	15
4	4	16

表 4-4 為上述步驟 X 座標方向及 Y 座標方向組合之後的軌跡組合表，由 X 座標方向組合為核心與 Y 座標方向組合進行配對共可配對出 16 個軌跡組合，如表 4-4 所示可看出是如何進行軌跡組合，而軌跡組合之示意圖以及詳細的內容會由表 4-5 所示。

表 4-5 軌跡組合標籤介紹

標籤	圖示(圓點代表擊球後)	
	左半場	右半場
Label1		
Label2		
Label3		
Label4		
Label5		
Label6		
Label7		
Label8		
Label9		
Label10		

Label11		
Label12		
Label13		
Label14		
Label15		
Label16		

軌跡分類共為 16 種軌跡，這 16 種軌跡當中 8 種為擊球軌跡、8 種為飛行軌跡，其中 8 種擊球軌跡可分為己方選手及對方選手的擊球軌跡，己方選手及對方選手分別各有 4 種擊球軌跡，由表 4-5 當中綠色底色 Label 的擊球軌跡為己方選手的擊球軌跡，而黃色底色 Label 的擊球軌跡為對方選手的擊球軌跡。這 8 種擊球軌跡會在實驗結果中，進行出現次數進行統計，在影片中標示出各種擊球軌跡的出現次數，以協助網球教練或相關分析人員進行統計工作。

## 第五章、實驗結果

在前兩個實驗中測試資料集都是一樣的狀況下進行實驗結果之分析，測試資料集為 60 個樣本，而在 Label1、Label2、Label3 和 Label4 樣本數分布，樣本分別為 13、17、15 和 15。

本研究為探討前兩個實驗的成果之比較，進行了混淆矩陣、準確度、精密度、召回率、F1 Score 以及 Micro F1 及 AUC 和 ROC 統計方式進行驗證作為依據評估模型的優劣，以進行本研究實驗結果之分析。

第三個實驗，完成本研究目標的擊球軌跡分類結果之計數，因分類器部分未使用深度學習方法進行分類，所以並沒有使用混淆矩陣、準確度、精密度、召回率、F1 Score 以及 Micro F1 以及 AUC 和 ROC 統計方式進行驗證，僅使用真實計數與預測計數進行比較。

### 第一節、C3D 結果

混淆矩陣 (Confusion Matrix)：混淆矩陣中可以看出預測結果與實際樣本的結果分析。

C3D 多類別混淆矩陣：

表 5-1 C3D 多類別混淆矩陣

		predicted			
		Label1	Label2	Label3	Label4
Actual	Label1	8	3	1	1
	Label2	4	11	1	1
	Label3	1	0	2	12
	Label4	2	2	3	8

從表 5-1 中可以得知所有實際樣本與預測結果各標籤的分布狀況，由上表中可以更了解所有測試資料中的結果分布，可以看出在 Label1、Label2 的預測結果相對比較好，而在 Label3、Label4 的預測結果下是相對不優異的。

準確度：準確度為在此預測結果中成功預測成功的比例。

使用 C3D 進行軌跡分類準確度為 49.2%，在測試樣本總樣本 60 個樣本中預測

成功的樣本數為 29 個。

精密度 (Precision)：預測為該 Label 樣本中的準確度

表 5-2 C3D 精密度

Precision	百分比
Precision(Label1)	53.33%
Precision(Label2)	68.75%
Precision(Label3)	28.57%
Precision(Label4)	36.36%

由表 5-2 中可以看出在進行精密度的分析當中，可以看出預測為該 Label 當中，真的預測成功的百分比。在表 5-2 當中在預測 Label1、Label2 真實為 Label2 或 Label2 的比例是相當高的，而另外 2 個 Label 就相對準確度都相當的低。

召回率 (Recall)：該 Label 樣本中預測正確的準確度

表 5-3 C3D 召回率

Recall	百分比
Recall(Label1)	61.53%
Recall (Label2)	64.71%
Recall (Label3)	13.33%
Recall (Label4)	53.33%

由表 5-3 中可以看出在進行召回率的分析當中，可以看出在真實 Label 中預測為該 Label 當中，真的預測成功的百分比。在表 5-3 當中在預測 Label1、Label2 真實為 Label2 或 Label2 的比例是相當高的，而另外 2 個 Label 就相對準確度都相當的低。

F1 Score：將該 Label 的 Precision 和 Recall 進行綜合考量

表 5-4 C3D F1 Score

F1 Score	百分比
F1 Score (Label1)	51.17%
F1 Score (Label2)	66.67%
F1 Score (Label3)	17.99%
F1 Score (Label4)	42.34%

由表 5-4 中可以看出各個 Label 在 F1-Score 上的百分比，F1-Score 由精密度及召回率進行評估，可以視為多方面考量的綜合評分。從表 5-4 中看出在僅有 Label2 在 F1-Score 得分上拿到了 60 以上的分數，也就是說在這個模型中僅有 Label2 是有預測效果的，而其他的 Label 是相對沒什麼預測能力。

Micro F1：為整體樣本進行的綜合考量

Micro F1 = 48.33%，Micro F1 為無需考慮類別直接對總樣本數進行 F1-Score 分析，所得到分數為 48.33% 是低於 50 所以說對於網球擊球分類軌跡的預測能力是不到 5 成的準確率。

從上述不同的統計方式中，可以看出 C3D 模型對於網球擊球軌跡分類的結果，預測能力並非相當優異，其問題點出在於資料集的不足，以及資料集中影像雜訊相當的多，因為球的影像資料相對在影像中其他影像中物體比例占比是相當的小。所以進行模型訓練過程中，可能相當大機率學習到雜訊影像資料，所以容易抓取不到特徵資料，解決方案為需增加大量數據，可以使 C3D 模型進行運行運算時，找出更加優異以及取得更多的資料特徵，進行 C3D 模型運算的優化，進而分類結果的準確度可得到提升。在第二個方法使用 TrackNet+C3D 方法，可改善僅使用 C3D 模型進行擊球軌跡分類的預測能力，解決影像資料中其他資料雜訊，TrackNet+C3D 的結果，請參考本章第二節、TrackNet+C3D 結果所敘述。

## 第二節、TrackNet+C3D 結果

混淆矩陣 (Confusion Matrix)：

表 5-5 TrackNet+C3D 多類別混淆矩陣

		predicted			
		Label1	Label2	Label3	Label4
Actual	Label1	11	1	0	1
	Label2	0	16	0	1
	Label3	0	1	9	5
	Label4	1	0	3	11

從表 5-5 中可以得知所有實際樣本與預測結果各標籤的分布狀況，上表可以更了解所有測試資料中的結果分布，可以看出在所有的 Label 的預測結果都還不錯，比只有使用 C3D 方法的效果好要好。

準確度：

使用 TrackNet+C3D 進行軌跡分類準確度為 78.7%，在測試樣本總樣本 60 個樣本中預測成功的樣本數為 47 個。

精密度 (Precision)：

表 5-6 TrackNet+C3D Precision

Precision	百分比
Precision(Label1)	91.67%
Precision(Label2)	88.89%
Precision(Label3)	75.00%
Precision(Label4)	61.11%

由表 5-6 中可以看出在進行精密度的分析當中，可以看出預測為該 Label 當中，真的預測成功的百分比。在表 5-6 當中在預測 Label1、Label2 真實為 Label2 或 Label2 的比例是相當高的，而另外 2 個 Label 就相對準確度都相當的低。



召回率 (Recall) :

表 5-7 TrackNet+C3D Recall

Recall	百分比
Recall(Label1)	84.62%
Recall(Label2)	94.12%
Recall(Label3)	60.00%
Recall(Label4)	73.33%

由表 5-7 中可以看出在進行召回率的分析當中，可以看出在真實 Label 中預測為該 Label 當中，真的預測成功的百分比。在表 5-7 當中在預測 Label1、Label2 真實為 Label2 或 Label2 的比例是相當高的，而另外 2 個 Label 就相對準確度都相當的低。

F1 Score :

表 5-8 TrackNet+C3D F1 Score

F1 Score	百分比
F1 Score(Label1)	88.00%
F1 Score(Label2)	91.43%
F1 Score(Label3)	66.67%
F1 Score(Label4)	66.63%

由表 5-8 中可以看出各個 Label 在 F1-Score 上的百分比，從表 5-8 可以看出 Label、Label2、Label3、Label4 的 F1-Score 得分都算是相當高的都超過 60 百分比，其中在 Label2 的部分 F1-Score 得分得到有 91 百分比是相當高的，代表對於 Label2 的預測，幾乎不會產生錯誤預測。

Micro F1 :

Micro F1=78.33%，使用 TrackNet+C3D 方法對於網球擊球軌跡預測的能力，算是能夠預測，至少有接近 8 成的預測能力。

從上述不同的統計方式中，可以看出 TrackNet+C3D 模型對於網球擊球軌跡分類的結果，預測能力相當優異，對比第一個實驗僅使用了 C3D 模型進行的擊球軌跡分類是優異很多。其問題為因在使用 C3D 模型進行分類前使用 TrackNet 模型將

影像畫面中其他影像雜訊去除，只保留網球位置的影像資料，對於 C3D 模型中的輸入資料減少大量雜訊影像，對於分類結果產生優良的差異結果。C3D 與 TrackNet+C3D 結果比較，請參考本章第三節、C3D VS TrackNet+C3D，所敘述此兩個實驗結果之比較。

### 第三節、C3D VS TrackNet+C3D

評估比較：

表 5-9 模型評估比較表

	C3D	TrackNet+C3D
ACC	49.2%	78.7%
Precision(Label1)	53.33%	91.67%
Precision(Label2)	68.75%	88.89%
Precision(Label3)	28.57%	75%
Precision(Label4)	36.36%	61.11%
Recall(Label1)	61.53%	78.57%
Recall(Label2)	64.71%	94.12%
Recall(Label3)	13.33%	60%
Recall(Label4)	53.33%	73.33%
F1 Score(Label1)	57.17%	84.62%
F1 Score(Label2)	66.67%	91.43%
F1 Score(Label3)	17.99%	66.67%
F1 Score(Label4)	43.24%	66.67%
Micro F1	48.33%	78.33%

在表 5-9 上可以看出 C3D 及 TrackNet+C3D 這個模型在不同的評估方式中在 TrackNet+C3D 的方法下都相對使用 C3D 的方法還要優異，對於網球擊球軌跡分類 TrackNet+C3D 明顯對於 C3D。

ROC 曲線：

C3D 模型所做的 ROC 曲線和 AUC 如圖 5-1 所示，從圖中可看出 TrackNet+C3D 模型比起 C3D 模型更靠近於左上角，其代表 TrackNet+C3D 模型優異於 C3D，而 C3D 模型在本實驗當中 AUC 取得 0.742 的分數，TrackNet+C3D 模

型在 AUC 取得 0.935 的分數，代表在本實驗 2 個分類器都有預測價值，而在 TrackNet+C3D 比 C3D 的 AUC 分數高也代表說 TrackNet+C3D 比 C3D 優異。

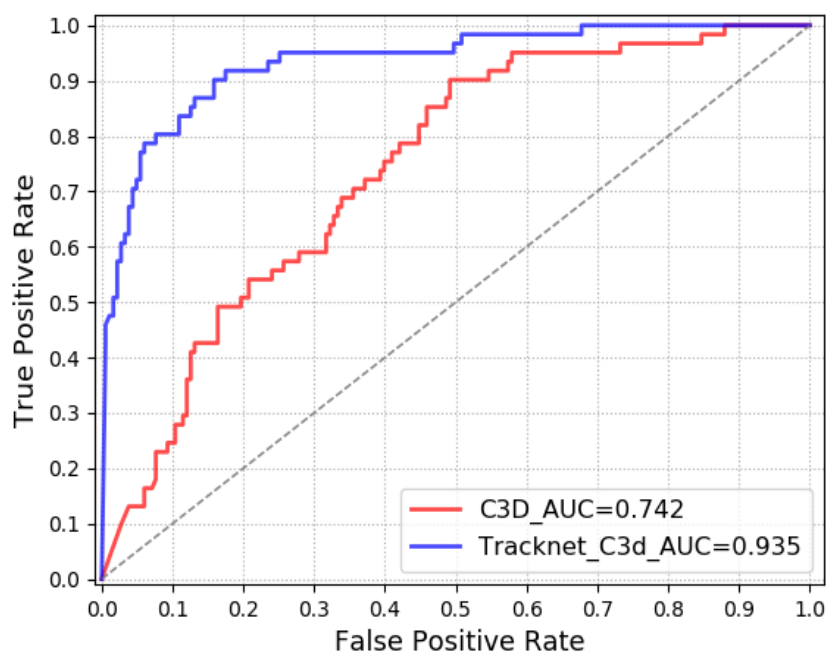


圖 5-1 ROC 曲線圖

PR 曲線：

PR 曲線可以看出類別不平衡的資料集的模型優劣，PR 曲線越接近右上角越好，代表模型的預測能力越優良，C3D 模型及 TrackNet+C3D 模型的 PR 曲線如圖 5-2 所示，從圖中可以看出 TrackNet+C3D 方法的 PR 曲線靠近右上角比起 C3D 方法相對優異很多。

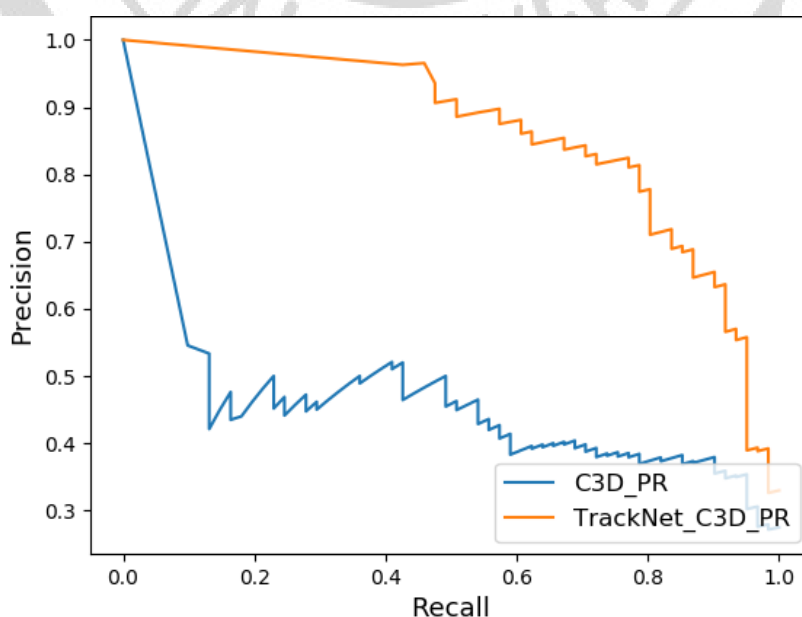


圖 5-2 PR 曲線圖

在本節中，依照上述比較可以得出 C3D 模型及 TrackNet+C3D 模型，在不同的統計方式下，明顯比較出對於擊球軌跡分類預測的能力上，TrackNet+C3D 模型是優於 C3D 模型。這 2 個模型方法未能進行 Real Time 預測，進而使用了 TrackNet+Trajectory Analysis 方法，進行擊球軌跡分類，其分析結果，請參考本章第四節、TrackNet+Trajectory Analysis 結果所述。

## 第四節、TrackNet+Trajectory Analysis 結果

TrackNet+Trajectory Analysis 為進行驗證 Real Time 方式進行軌跡分類，上述 C3D 與 TrackNet+C3D 之實驗的方法已無法進行 Real Time 方式進行軌跡分類，因此兩方法所使用資料集已無擊球以外片段，所以在學習擊球分類上僅有擊球，無法學習到擊球以外的片段，故而無法進行 Real Time 處理。

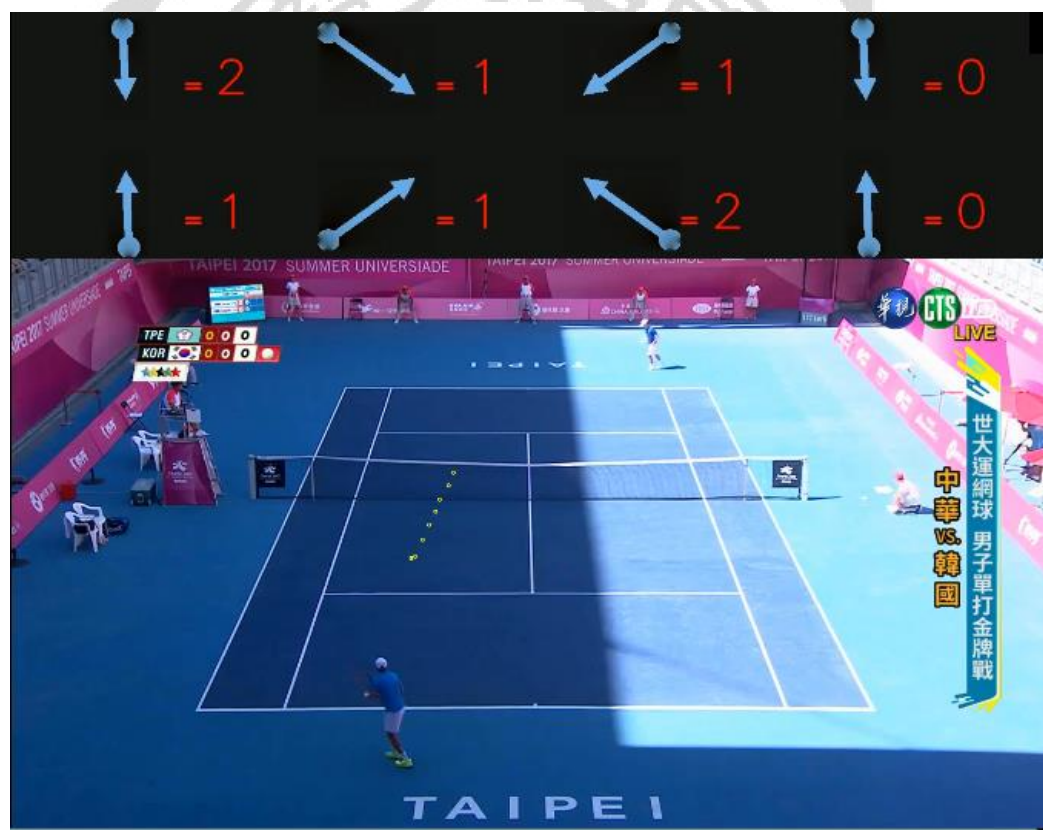










圖 5-3 TrackNet+Trajectory Analysis 分類計數

圖 5-3 為 TrackNet+Trajectory Analysis 在一段網球比賽影片中所進行的計數統計，可以方便去統計各個擊球軌跡的出現次數，在影片中的上方會即時出現 4 種擊球軌跡的出現次數。針對擊球分類計數統計的結果分析，會在表 5-10 進行詳細描述。

表 5-10 TrackNet +Trajectory Analysis 計數統計（單一片段）

Label		真實計數	預測計數(1)	預測計數(2)
		0	0	0
		0	0	0
		2	2	2
		1	1	1
		1	2	1
		2	3	2
		1	0	1
		1	0	1

在這實驗中因為使用自行演算之演算法的方法可以依據即時的網球位置座標進行運算推算出網球軌跡的分類，本研究目的達到的紀錄選手的擊球軌跡分類的計數統計如表 5-10 所示，表 5-10 為前處理比賽片段中 81 個片段中其中一個片段，所發生的擊球計數統計。在表 5-10 可看出真實影片軌跡分佈及本實驗預測結果比較。在上表中可以看出分類軌跡的計數統計準確度算是，預測計數(1)、預測計數(2)代表 Trajectory Analysis 的處理方式，預測計數(1)為原 Trajectory Analysis 的預測計數，預測計數(2)為原預測結果的軌跡做進行外插座標至邊緣的延伸軌跡處理，原先 Trajectory Analysis 的計數方式無法做到更精準的方式預測，因 16 幀影像中的座標位置不足，會產生其誤差。

表 5-11 TrackNet+Trajectory Analysis 計數統計（多片段）









Label		真實計數	預測計數
		13	11
		12	12
		22	24
		20	20
		14	12
		26	23
		14	16
		13	18

表 5-11 為本研究使用使用前處理中比賽片段 81 個片段當中的 20 個片段進行擊球統計進行比較，共有 134 個擊球軌跡進行分類，用於判斷 TrackNet+Trajectory Analysis 方法是否優異。此表中預測計數為進行為外插座標的處理方式進行計數統計，從此表中可看出存在部分擊球分類上發生錯誤統計，此分類方法其準確度可達到 89%，TrackNet+Trajectory Analysis 與 C3DTrackNet+C3D 進行擊球分類軌跡比較精準度是最高的。

## 第五節、時間成本

在本研究中之實驗中，訓練集由數據集中的 75% 所組成，剩下 25% 則用於測試集。本實驗中所使用的深度學習機器，配備為 Intel Xeon E5-1650 CPU 和 64G RAM 以及配備了一個 Nvidia Geforce Gtx 1070 Ti GPU，作業系統是 Ubuntu 16.04 LTS，深度學習使用的 API 是 Keras 及 Tensorflow，使用 Python 語言進行撰寫。

表 5-12 模型訓練時間比較表

Method	Training Time
C3D	約 32 分鐘
TrackNet+C3D	約 68 小時 15 分鐘+約 28 分鐘
TrackNet+Trajectory Analysis	約 68 小時 15 分鐘

表 5-12 可以看出訓練階段中，在本研究所使用模型之訓練時間，C3D 約 32 分鐘，TrackNet+C3D 約 68 小時 15 分鐘+約 28 分鐘，TrackNet+Trajectory Analysis 約 68 小時 15 分鐘。

表 5-13 模型預測時間

Method	16 幀處理時間
C3D	0.9592 秒
TrackNet+C3D	5.2838 秒
TrackNet+Trajectory Analysis	4.3205 秒

表 5-13 中可以看出在本研究當中所使用方法，運算時間因本研究使用 16 幀影像資料進行分類，所以表 5-13 中的時間為 16 幀影像處理時間，表中可以看出在 C3D 模型進行預測分類是最為快速，最慢的為 TrackNet+C3D 的方法，而 TrackNet+Trajectory Analysis，處理時間雖比 TrackNet+C3D 約快 1 秒，平均一幀處理時間為 0.27 秒，無法達到 Real-time 的處理速度。

## 第六章、結論

本研究使用 30FPS 之網球比賽影片，提出三個實驗方法 C3D、TrackNet+C3D、TrackNet+Trajectory Analysis，進行網球擊球軌跡分類。其中前 2 個實驗結果 C3D 與 TrackNet+C3D 中，可以發現在 TrackNet+C3D 的方法中的 Precision、Recall 以及在 F1 Score 的部分會比只單獨使用 C3D 方法進行網球擊球軌跡分類的部分還要好。由此可知進行擊球的軌跡分類，在畫面中進行處理去除球影像以外雜訊，在進行分類會相對優異很多，不過經由 TrackNet 模型此方法會浪費了很多的運算資源用於去除背景等雜訊資料。

在第三個實驗中使用 TrackNet+Trajectory Analysis 的方法中，在分類的部分因無使用深度學習方法進行分類，所以在運算資源上會相對優異很多，不過此方法在進行擊球分類上存在缺點，由於這方法在分類上需要做自己推算左右半場的分隔線，如攝影機跟隨選手而移動拍攝角度，則分類結果會發生錯誤判斷，而分類後計數統計發生錯誤統計。

從時間成本來看使用 C3D 時間成本最低，不過在分類擊球軌跡中是最差的，TrackNet+C3D 是時間成本最高的，預測能力相對次之。TrackNet+Trajectory Analysis 時間是次之，而預測能力是三個方法中最佳的。

本研究結果於其他相關研究之差異，本研究無需特別進行影像中的特徵萃取動作，可省下大量時間進行規劃演算法等流程，且該相關研究並無進行軌跡分類。目前研究之結果可用於協助網球教練或者相關分析人員進行網球比賽資料分析，可進行自動紀錄比賽選手擊球之結果。

未來展望：

本研究希望未來能對於 Real-Time 網球比賽影片進行擊球軌跡分類，目前三種方法的運算時間還無法達到 Real-Time 處理，本研究接下來會進行更進一步優化本研究的運算模型，減輕運算資源，以助於未來如若網球教練或者相關分析人員，使用本研究之結果，可以不用花費大量的運算成本即可使用，且可達到 Real-Time 處理，使得分析人員可以在比賽中進行紀錄分析。



## 參考文獻

1. 2017 年夏季世大運網球男子單打決賽， <https://reurl.cc/Nj8e55>
2. Archana, M., & Geetha, M. K. (2015). Object detection and tracking based on trajectory in broadcast tennis video. *Procedia Computer Science*, 58, 225-232.
3. Chen, H. T., Tsai, W. J., Lee, S. Y., & Yu, J. Y. (2012). Ball tracking and 3D trajectory approximation with applications to tactics analysis from single-camera volleyball sequences. *Multimedia Tools and Applications*, 60(3), 641-667.
4. Chen, Y., & Xue, Y. (2015, October). A deep learning approach to human activity recognition based on single accelerometer. In *2015 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 1488-1492). IEEE.
5. D'Orazio, T., Guaragnella, C., Leo, M., & Distanti, A. (2004). A new algorithm for ball recognition using circle Hough transform and neural classifier. *Pattern recognition*, 37(3), 393-408
6. Fawcett, Tom (2006), An introduction to ROC analysis, *Pattern Recognition Letters*, 27, 861-874.
7. FFmpeg, <https://reurl.cc/vDNxNe>
8. Fu, T. S., Chen, H. T., Chou, C. L., Tsai, W. J., & Lee, S. Y. (2011, November). Screen-strategy analysis in broadcast basketball video using player tracking. In *2011 Visual Communications and Image Processing (VCIP)* (pp. 1-4). IEEE.
9. Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448).
10. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587).
11. Hawk-eye, <https://reurl.cc/31dKyL>
12. Huang, Y. C., Liao, I. N., Chen, C. H., Ik, T. U., & Peng, W. C. (2019, September). TrackNet: A Deep Learning Network for Tracking High-speed and Tiny Objects in Sports Applications. In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (pp. 1-8). IEEE.
13. Jiang, W., & Yin, Z. (2015, October). Human activity recognition using wearable sensors by deep convolutional neural networks. In *Proceedings of the 23rd ACM international conference on Multimedia* (pp. 1307-1310).

14. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).
15. LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4), 541-551.
16. Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).
17. Myint, H., Wong, P., Dooley, L., & Hopgood, A. (2015, May). Tracking a table tennis ball for umpiring purposes. In *2015 14th IAPR International Conference on Machine Vision Applications (MVA)* (pp. 170-173). IEEE.
18. Noh, H., Hong, S., & Han, B. (2015). Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision* (pp. 1520-1528).
19. Pingali, G., Opalach, A., & Jean, Y. (2000, September). Ball tracking and virtual replays for innovative tennis broadcasts. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000 (Vol. 4, pp. 152-156)*. IEEE.
20. Qazi, T., Mukherjee, P., Srivastava, S., Lall, B., & Chauhan, N. R. (2015, December). Automated ball tracking in tennis videos. In *2015 Third International Conference on Image Information Processing (ICIIP)* (pp. 236-240). IEEE.
21. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).
22. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91-99).
23. Reno, V., Mosca, N., Marani, R., Nitti, M., D'Orazio, T., & Stella, E. (2018). Convolutional neural networks based ball detection in tennis games. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 1758-1764).
24. Renò, V., Mosca, N., Nitti, M., D'Orazio, T., Guaragnella, C., Campagnoli, D., ... & Stella, E. (2017). A technology platform for automatic high-level tennis game analysis. *Computer Vision and Image Understanding*, 159, 164-175.

25. Renò, V., Mosca, N., Nitti, M., Guaragnella, C., D'Orazio, T., & Stella, E. (2016, December). Real-time tracking of a tennis ball by combining 3d data and domain knowledge. In 2016 1st International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW) (pp. 1-7). IEEE.
26. Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
27. Sasaki, Y. (2007), "The truth of the F-measure"
28. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
29. Tran, D., Bourdev, L., Fergus, R., Torresani, L., & Paluri, M. (2015). Learning spatiotemporal features with 3d convolutional networks. In Proceedings of the IEEE international conference on computer vision (pp. 4489-4497).
30. Wang, X., Ablavsky, V., Shitrit, H. B., & Fua, P. (2014). Take your eyes off the ball: Improving ball-tracking by focusing on team play. *Computer Vision and Image Understanding*, 119, 102-115.
31. Yan, F., Christmas, W., & Kittler, J. (2005). A tennis ball tracking algorithm for automatic annotation of tennis match. In *British machine vision conference* (Vol. 2, pp. 619-628).
32. Yu, X., Jiang, N., Cheong, L. F., Leong, H. W., & Yan, X. (2009). Automatic camera calibration of broadcast tennis video with applications to 3D virtual content insertion and ball detection and tracking. *Computer Vision and Image Understanding*, 113(5), 643-652.
33. Yu, X., Sim, C. H., Wang, J. R., & Cheong, L. F. (2004, October). A trajectory-based ball detection and tracking algorithm in broadcast tennis video. In *2004 International Conference on Image Processing, 2004. ICIP'04.* (Vol. 2, pp. 1049-1052). IEEE.
34. Zhou, X., Xie, L., Huang, Q., Cox, S. J., & Zhang, Y. (2014). Tennis ball tracking using a two-layered data association approach. *IEEE Transactions on Multimedia*, 17(2), 145-156.