

東海大學資訊工程研究所

碩士論文

指導教授：許瑞愷 博士

智能球探 - 球員數據自動產生系統

Intelligent Scout - A Player Tracking System for
Sports

研究生：林建良

中華民國 一〇八年 二月

東海大學碩士學位論文考試審定書

東海大學資訊工程學系 研究所

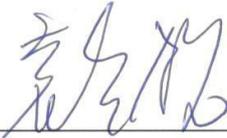
研究生 林 建 良 所提之論文

智能球探-球員數據自動產生系統

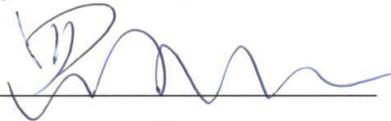
經本委員會審查，符合碩士學位論文標準。

學位考試委員會

召 集 人

 簽章

委 員



評 瑞 性

指 導 教 授

評 瑞 性 簽章

中華民國 108 年 1 月 18 日

摘要

基於電腦視覺的人工智能運動分析技術，不僅迅速有效的自動追蹤球員軌跡、收集賽事數據，更能解決傳統人為判斷與識別上過於主觀的問題。近年來，已經發展為主流運動賽事分析與應用的創新商業模式。然而，傳統的數據紀錄方式仰賴人工作業，這對於時間與成本上造成相當的負擔。

因此，本研究致力於研發數據產生系統，藉由電腦視覺在影像處理與分析上的優勢，於賽事影像上進行識別作業，紀錄球員在球場上的表現，通過仿射轉換的方法將影像資訊投影至平面圖，用來產生賽事中球員的絕對座標、奔跑速度、移動距離等資訊，該系統架構能夠有效地靈活化整體作業之流程，達成自動化數據紀錄之成效。

關鍵詞：自動化數據蒐集、仿射轉換、物件偵測與識別、電腦視覺

Abstract

Based on computer Vision, the artificial intelligent motion analysis technology not only quickly and effectively track the player's trajectory automatically, but also collect event data. It can solve the problem of tradition subjective judgment and recognition. In recent years, it has developed into an innovative business model for the analysis and application of mainstream sports events. However, the traditional method of data recording relies on manual work, which imposes a considerable burden on time and cost.

Therefore, this research is devoted to the development of dataset generation system. Through the advantages of computer vision and image processing in detection and recognition, Identifying the game on the broadcast sport video and recording the performance of the player of the court. We use the affine transformation method to project game image information to the floor plan to generate the absolute coordinates, running speed, moving distance and other information of the players in the events. This system architecture can effectively flex the overall operation process and achieve the results of automated data records.

Keywords: automated data collection, affine transformation, object detection and identification, computer vision

誌謝

回顧這兩年半的時光，研究所生活實在是過得很充實，特別要感謝許瑞愷老師對我的指導，在我遇到障礙時指引我方向，還要感謝羅文聰老師以及陳倫奇老師在碩士期間給我的許多協助，讓我們在討論與報告時能夠給予客觀的方案與建議，並且思考這些方法是否合宜，在無形中累積經驗與研究能量。在這邊也要特別感謝曾建浩學長，讓我在國家高速網路與計算中心學習，並且耐心的討論與指導我，雖然有幾次由於實力不夠而無法跟上預期進度，但要感謝學長不厭其煩的帶領我如何架構與實驗，然後在多次的會議中激盪想法。

謝謝實驗室的所有人，在艱辛的過程中互相加油。也要謝謝爸爸、媽媽、大姐及二姐，在我面臨延畢時能夠諒解我，你們的鼓勵讓我在挫折中更有力量去面對。感謝這所有的一切。

感謝人 建良

目 錄

摘 要.....	I
Abstract	II
誌 謝.....	III
第一章 緒論	1
1.1 研究動機與背景	1
1.2 研究動機與目的	4
1.3 章節概要	5
第二章 文獻探討	6
2.1 物件偵測(Object Detection).....	6
2.1.1 區域檢測進程介紹.....	6
2.1.2 滑動窗口檢測(Sliding window detector).....	7
2.1.3 錨點偵測(Anchor detector)	7
2.1.4 方向梯度直方圖(Anchor detector)	8
2.1.4 深度神經網路(Deep Neural Network).....	9
2.2 區域快速卷積神經網路(Faster Region-based Convolution Neural Network)	10
2.3 殘差網路(Residual Network).....	12
2.4 霍夫線偵測(Hough Line Detect)	13
2.5 仿射轉換(Affine Transform)	14
第三章 研究方法	15
3.1 方法流程圖	15
3.1.1 流程步驟.....	15
3.1.2 資料選擇.....	15
3.1.3 資料處理.....	16
3.1.4 資料轉換.....	17
3.1.5 電腦視覺及座標映射.....	18
3.2 模型訓練調配	18
3.2.1 Faster R-CNN 錨點參數調配	18
3.2.2 ResNet 識別網路架構.....	19

3.3 自動化數據紀錄	20
3.3.1 球場追蹤.....	20
3.3.2 座標映射.....	24
第四章 結果討論分析	28
4.1 NBA 球場廣播視頻.....	28
4.2 Faster R-CNN 分類結果	28
4.2 ResNet 識別結果.....	29
4.3 仿射轉換結果討論	30
4.4 實時運行時間	32
第五章 結論與未來展望.....	33
參考文獻.....	34



圖目錄

圖 2.1 區域偵測發展進程	7
圖 2.2 滑動窗口的比對執行	7
圖 2.3 錨點於原始影像上的偵測方式	8
圖 2.4 HOG 的圖像轉換過程	8
圖 2.5 深度神經網路的偵射方式	9
圖 2.6 FASTER R-CNN 模型架構圖	10
圖 2.7 正規化示意圖	11
圖 2.8 殘差網路中的區塊運算架構	12
圖 3.1 方法流程圖	15
圖 3.2 大眾媒體上播送的賽事影像	16
圖 3.3 LABELIMG 的標籤影像示意圖	16
圖 3.4 標籤資訊轉換成文件集格式	17
圖 3.5 圖像資料包裝成數據集格式	17
圖 3.6 (左)球賽廣播影像 (右)球場平面圖	20
圖 3.7 影像色彩遮罩	21
圖 3.8 通過型態學處理之影像	21
圖 3.9 通過邊緣偵測處理之影像	22
圖 3.10 通過霍夫線偵測之影像	22
圖 3.11 目標線偵測失敗之影像	23
圖 3.12 偵測失敗時保留影像資訊之處理	23
圖 3.13 (左)原始影像上的參考點位置、(中)平面圖上的參考點位置、(右)透過仿射轉換後的原始影像	24
圖 3.14 仿射轉換中參考點的匹配圖	25
圖 3.15 (左)偵測影像垂直線與水平線之間的交點、(右)交點坐標反饋回原始影像	25
圖 3.16 (左上)檢測器檢測之資訊、(左下)檢測資訊轉換為平面資訊、(右)以 CSV 方式針對能夠紀錄的資訊進行紀錄	26
圖 3.17 球員位置投影之系統執行圖	27
圖 3.18 賽事數據紀錄之系統執行圖	27

圖 4.1 (左)湖人訓練模型過程圖、(右)塞爾提克訓練模型過程圖29
圖 4.2 計分板遮蔽參考線之示意圖31

表 目 錄

表 3.1 IMAGENET 各年份之模型架構數據.....19
表 4.1 FASTER R-CNN 模型準確度統計29
表 4.2 (左)湖人球員身份之識別結果、(右)塞爾提克球員身份之識別結果30



第一章 緒論

1.1 研究動機與背景

隨著近年來資訊科技應用的發展，人們對於人工智慧與機器學習的使用模式也逐漸地被改變，包括交通運輸、居家生活、公共場所等都已習慣用於自動化場域記錄的應用系統來達到視訊檢索與調閱之作用，然而，在多繁複的遮蔽狀況與過於緊貼的人群環境下進行分析仍然是一項挑戰。在傳統球類數據蒐集中，僅能透過專業人員進行場域記錄，無法達到自動化數據認知與紀錄之能力，倘若需要降低時間與人力成本，傳統式的紀錄方式對於人力作業與事件時效性而言會造成相當的負擔。

而人工智慧帶來了最廣泛且顛覆性強的商業應用，該技術成長快速、變動性高，於質與量之間存在著很高的價值，再加上巨量資料所帶來的影響，使得電腦視覺、影像分析與機器學習的應用帶來相當大的衝擊。過往計算機於影像處理僅能夠對應單張或序列影像進行分析與圖像處理，然而，面對海量般的影像串流資料，傳統影像處理技術恐怕無法有效率的進行圖像分析與訊息萃取作業。因此，如何讓電腦視覺更有效率的處理智能辨識與資訊萃取，為自動化分析技術所導入的必要性。

電腦視覺在物件檢測任務中，是從數位影像中所提供龐大且複雜之影像數據集，提取有用且具有高度價值的資訊，然而，有效地抽取出可信度高的資訊，是一件相當困難的事情。對於近代而言，已知一些不錯的特徵擷取方法。例如：LBP(Local Binary Patterns)解決紋理圖像上的分類問題[4]，Haar-like feature 提出以數字圖像作為檢測圖像的表徵格式[5]，SIFT(Scale-invariant Feature Transform)則是基於圖像屬性不變之原理來偵測與描述影像中的局部性特徵[6]。方向與梯度直方圖(Histograms of Oriented Gradients，簡稱 HOG)[19]是將影像劃分成無數個影像細胞，提出基於影像細胞的方向性與強度值來表示影像的特徵，從上述多

項特徵提取的方法來看，機器學習於特徵提取上已有相當可觀的成就。其中，深度卷積神經網路(Deep Convolution Neural Network，簡稱 DCNN)在基於圖像分析上有著相當卓越的研究。Yann Lecun 等人[8]認為影像中，像素點的強度與方向是具有意義的，提出透過卷積核計算局部特徵萃取作業，降低維度使大部分不相關的像素點進行濾除，考慮到主要特徵有被過濾的可能性，使用激活函數將主要的像素進行增強，透過池化取樣方式提取與保留主要有用的像素，用來簡化像素之間複雜且瑣碎的演算，最後以全連接運算進行決策輸出，此為卷積神經網路(Convolutional Neural Network，簡稱 CNN)之技術架構。之後又以 Alex Krizhevsky 等人[9]透過訓練 CNN 網路模型，提出將神經元中的激活函數從 Tanh 函數換成修正線性單元(ReLU)，使得 CNN 網路於訓練階段更快速達到收斂。因此，運用深度學習架構於電腦視覺技術之開發，已嚴然成為智慧影像分析的熱門研究議題。

基於電腦視覺技術，我們從球類廣播影像中的全景影像作為依據，已知傳統的偵測方法採用可變形部位模型(deformable Parts model，簡稱 DPM)[11]來偵測球員，但該方法於影像鏡頭聚焦時容易產生誤差，球員與球員之間的緊貼與碰撞事件上仍然無法準確地分析出來，且滑動窗口方式對大尺度影像的偵測上過於耗時[13]，因此球員的進攻與防守影像上改用更快速之區域卷積神經網路(Faster Region-based Convolution Neural Network，簡稱 Faster R-CNN)[23]進行球員區域偵測，該演算方法雖然能對區域的偵測作業上有著卓越的準確率，但在遠景物件的辨識作業中較不穩定，考慮到偵測作業上存在部分錯誤，我們將錯誤的區域進行紀錄，並且與辨識模型的樣本混合進行訓練，提升影像在偵測球員作業中的準確率。再來，我們更改傳統的 Faster R-CNN 分類類別方式，以球隊類別來進行初步的分類的依據，降低後續球員於影像辨識的複雜性，最後透過 DCNN 中的殘差網路(Residual Network，簡稱 ResNet)[24]架構進行辨識球員，提高在球員辨識上的準確性。

本研究將探討在球場廣播影像上偵測、辨識、投影等應用，藉此執行自動化球場記錄作業，已利於後續的語意轉換、戰術分析等相關研究[22]。本文架構主要分為三個部分，第一個部分為圖像標記作業，蒐集大量球場廣播影像之片段，進行切片抽樣與標籤作為訓練所需之樣本數據，接著使用 Faster R-CNN 中的錨點(Anchors)偵測方式進行潛在區域的預選區域(Pre-Region of Interest)作業，經過運算歸納出影像中前景的物件區域，且針對球員的明顯特徵進行辨識其球隊類別。第二部分為物件學習與認知，將使用基於 DCNN 的 ResNet 網路架構，透過計算殘差(Residual)的方式，將每位球員細微的特徵差異進行萃取，並以分類器訓練辨識所需的模型，幫助我們執行定義與標籤類別。第三部分為座標投影與產生數據，運用霍夫線轉換的演算方程，將球場區域進行球場線偵測與追蹤，由於參考點在動態影像上會造成失真問題，因此球場線的偏移量計算是必要的。為了取得球員在平面戰術板上的絕對位置，透過校準後的參考點進行仿射轉換作業，最後將轉換的結果與相關分析資訊紀錄起來。

從近年來智慧運動的發達，多家企業紛紛投入執行數據化的訓練模式，進行球員的訓練與數據考察，讓球隊在選擇時能夠以更客觀的資訊理解球員的能力，或是從資訊中發現潛在受傷的狀況，也有多媒體影音公司透過該技術來建立影像特效，帶給各位不一樣的影音體驗，對於上述的需求，智慧化的數據分析在未來趨勢上勢必會造成改革，而有效率取得數據的方法，乃為該篇研究的重點，對於這些未來的趨勢與應用，其方法與問題值得我們去深入探討。

1.2 研究動機與目的

電腦視覺(Computer Vision)這方面的技術，是在對於大量數據化後的影像透過分類器的計算給於這些影像區域相對應類別的方法，視覺模型的想法來自於 1981 年，美國神經生物學家 David Hubel 和 Torsten Wiesel 對於動物視覺系統的處理信息方式有了進一步的發現，後人將神經傳遞的處理方式數學化，認為人在決策上是由數以億萬計的神經元細胞彼此之間互相影響與分析決策，透過這種認知模式我們了解到，將圖片切割成無數個圖片細胞，並且對圖像進行細微的色度與空間轉換來修正認知，透過近似於生物般的決策模式，使用計算機來分析與模擬資訊，這些模型對於圖像中的物件進行機率般的分析與預測，我們稱這項研究技術為電腦視覺(Computer Vision)。

電腦視覺的相關研究日甚一日，其中物件之間的偵測(Detection)、辨識(Identification)、追蹤(Tracking)被廣泛運用在生活中，像是利用 Google TensorFlow、Microsoft Coffe 及 IBM Keras 等深度學習分析套件，幫助工程師與研究員進行人工智慧領域上的各種問題。

資訊紀錄的方式就傳統而言，需要透過人工作業以紙筆來進行，該方法簡單且易於作業，但對於日積月累下的紙本數量而言，龐大的資訊量使得作業人員與分析人員難以閱覽，縱使以計算機進行收錄也很難執行搜尋檢索作業，因此，激發了智慧型手機 APP 與運動數據平台的產生，讓數據記錄作業時能夠更加輕鬆方便，但仍需仰賴人工作業進行，已知許多先進國家，像美國 STATS 以及 SportlogiQ 採用電腦視覺與感測器技術來追蹤籃球與冰球的紀錄作業，歐洲的 OPTA 也透過各項先進的設備與技術來蒐集足球的資訊，上述之數據紀錄公司與數據分析公司 Second Spectrum 及球類賽事播報平台 Whoscored、Squawka 進行合作，創作出不一樣的分析方式與播送體驗，因此，本文希望藉由計算機來處理人為視覺作業上的分析，將分析後得到的結果進行紀錄與保存，開發出自動化數

據產生之系統。

由於該系統為監督式分析方式，我們必須給予經過人工分類的數據集來替代傳統抽樣不佳的圖像特徵及過於主觀性的特徵提取方式，訓練計算機學會偵測區域與識別物件，並且從錯誤中不斷的修正。然而，對於圖像處理而言，可以選擇多種形式來描述表示，其中，圖像最基本的單位像素，可以看作是一種表示強度的向量值，或者具有幾何形態意義的邊、線等特定形狀區域，通過抽象般的特徵表示則更能夠容易達到學習之成效。

本研究目的在紀錄球場環境中，每個人移動的位置與軌跡資訊，進行數據分析並且學習專業戰術與知識，幫助業餘球員執行相同品質的訓練與戰術教學。這項技術還能夠幫助教練團分析自己與其他對手之間的優勢與劣勢，也能夠幫助電視廣播公司，創建星攝影像(star-camera video)，串流分析與特效顯示球員進攻的精彩片刻。在許多職業運動球員的選擇作業上，能夠以過去的數據來預測未來成長的可能性，而本文智能球探的應用系統將提供球團可行性的洞察，將每段賽事的時間點中，球員的落點位置進行記錄與保存，幫助分析人員與球團人員在戰術分析與人才選擇的領域上能夠以更客觀又全面的建議來達到目的。

1.3 章節概要

本研究將探討在球場廣播影像上偵測、辨識、投影等應用，藉此執行自動化球場記錄作業，已利於後續的語意轉換、戰術分析等相關研究。以下介紹各個章節，第二章為文獻探討，我們將現有的偵測方式與投影紀錄的相關文獻進行概述，第三章為研究方法，我們將詳述整個用到的技術架構以及紀錄的方法流程，第四章為輸出結果與其他現有的方法流程比較並進行討論與建議，最後第五章為結論及未來工作。

第二章 文獻探討

本章將探討電腦視覺中，區域偵測與物件辨識的相關進程，然後再介紹本文所使用到的兩模型，Faster R-CNN(Faster Region-based Convolution Neural Network)與 CNN ResNet 101(Convolution Neural Network with Residual Network 101)，最後是霍夫線偵測(Hough Line Detection)的介紹。

2.1 物件偵測(Object Detection)

物件偵測在視頻影像中是一項重要的課題，例如車輛與行人的偵測等，該技術對於物件辨識與區域分析上有著可觀的助益。然而，在區域偵測的領域中有個最經典的模型架構，我們稱呼該架構為基於區域的快速卷積神經網路(Faster Region-based Convolution Neural Network，簡稱 Faster R-CNN)，我們將在下列介紹其發展進程。

2.1.1 區域檢測進程介紹

前面提到 Faster R-CNN 為神經網路上的區域檢測演算法，然而在進程的演化中，我們可以將區域檢測的看成幾個階段，早期以支持向量機來訓練方向梯度直方圖 (Histogram of oriented gradient，簡稱 HOG) 的表徵圖像，或是訓練可變型元件模型(Deformable Part-based Model，簡稱 DPM)，用來比對當前帶檢測影像上的行人與車輛。亦也有透過影像處理方式，以紋理與顏色等元素進行區分，但在選擇物件上需要仰賴分類器進行，而良好的特徵才會有好的分類結果，在偵測時間的效率上由於傳統方式難以達到實時應用，因此我們以深度學習的架構模式，採用錨點檢測法進行，減少了滑動窗口中比對待檢測區域的數量以降低運算所消耗的時間，藉由多種尺度與比例的邊界框(Bounding box)進行可疑區域的選取作業，主要區分前景與背景之區域，再以迴歸方式對邊界框進行分析類別，我們稱呼這種偵測的方式為區域快速卷積神經網路(Faster R-CNN)。

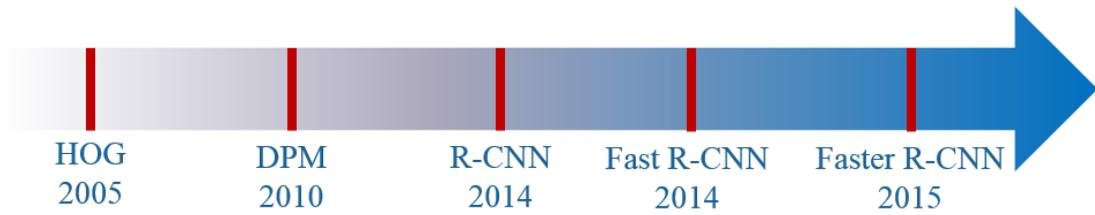


圖 2.1 區域偵測發展進程

2.1.2 滑動窗口檢測(Sliding window detector)

基於固定窗口的大小，將滑動窗口與影像上的待檢測窗口進行比對，由於影像上物件具有尺度問題，因此必須不斷更新檢測影像的尺度大小，若窗口已檢測完，縮放其窗口尺度已用來對應不同尺度下的目標物件。滑動窗口內的數據對象對應一個固定窗口的序列 ($FW_i, FW_{i+1}, \dots, FW_{i+n}$)，如下圖 2.2 所示

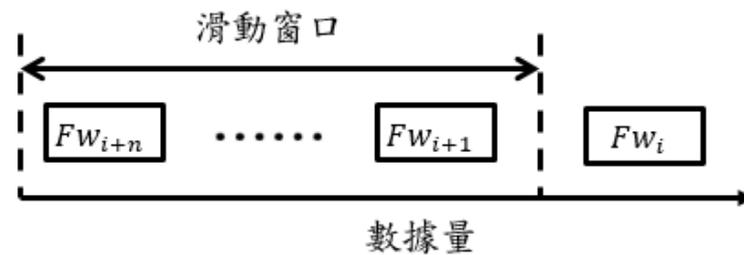


圖 2.2 滑動窗口的比對執行

2.1.3 錨點偵測(Anchor detector)

錨點偵測是一項基於滑動窗口的偵測方式，該方式是由 m 個比例與 n 個尺度的邊緣框組成，用來解決物件於影像上尺度轉換上的問題，每一個錨點是其影像上的像素點，根據錨點的步幅來滑動檢測區域。如圖 2.3 中顯示了偵測物件的過程，其中，錨點框位是以 $m \times n$ 個各自進行檢測，每個錨點框位都會透過線性迴歸的計算方式得到物件的置信分數，若超過門檻值則歸類為物件。

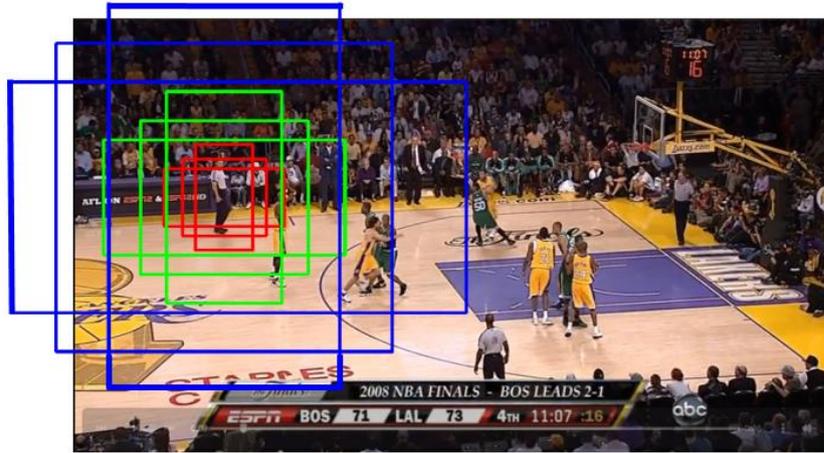


圖 2.3 錨點於原始影像上的偵測方式

2.1.4 方向梯度直方圖(Anchor detector)

早期以方向梯度直方圖作為一種影像特徵轉換後的表徵方式，該方法能夠基於幾何方式，對物件的局部特徵進行抽取，計算影像物件梯度大小與方向變量，係構成特徵向量並描述局部區域之邊緣與形狀，因此，對於待檢測圖像上之目標形狀變量和光照變化等外部干擾原因，係可擷取局部特徵區域以降低其干擾因素，對於物件幾何形態與光照變化都有良好之不變性。圖 2.4 所示，影像中物件明顯且局部特徵(軀幹、頭部、身體)皆有明顯的表徵格式，且對背景雜訊有良好的濾除方式來提高匹配結果。

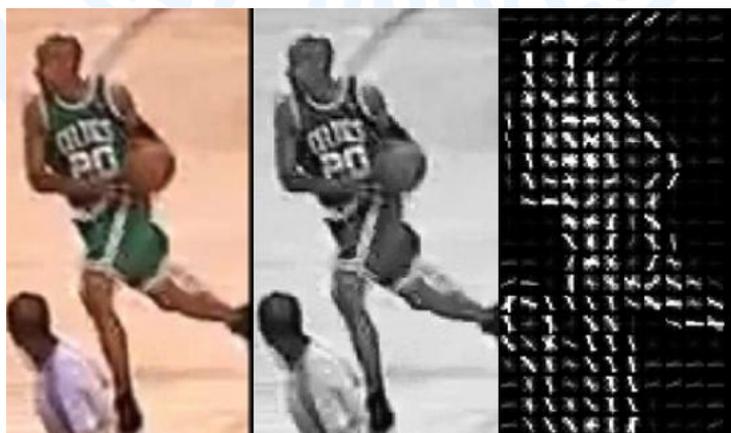


圖 2.4 HOG 的圖像轉換過程

2.1.4 深度神經網路(Deep Neural Network)

深度神經網路模型於影像偵測的應用中，是在 2013 年於 Google 研究所提出的一種偵測方法，該方法是基於像素點將影像區域進行劃分，並且透過每個像素之間的遮罩(Mask)區域進行訓練與偵測，由於目標物件或影像有著尺度轉換的區域問題，所以縮放比例的目的是為了更確定物件在影像上的位置。圖 2.5 中說明，將影像劃分區塊後進行檢測遮罩區域，但過小的物件有著不確定因素，所以需要縮放調整來校準位置。

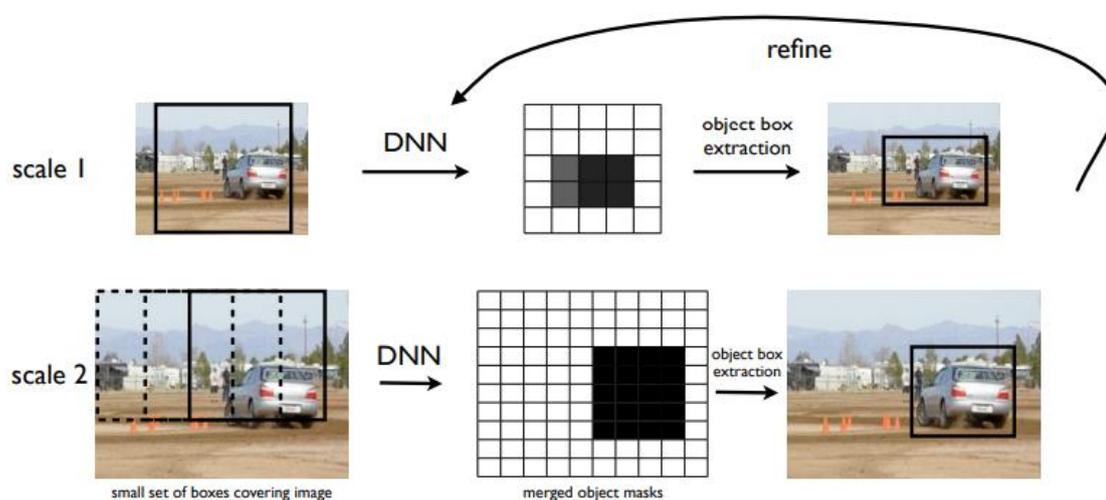


圖 2.5 深度神經網路的偵測方式

2.2 區域快速卷積神經網路(Faster Region-based Convolution Neural Network)

Faster Region-based Convolution Neural Network(Faster R-CNN)，是 Ross Girshick 於 2015 年所提出的一種基於特徵影像上的偵測方法，也是區域目標網路(Region-based Network)的第三篇研究論文，目的是為了訓練計算機能夠認知前景與背景物件的差異，從中學習並且執行區域選取，Faster R-CNN 模型架構中最重要三個核心概念為，抽取影像特徵圖的前置網路(last convolution layer)、取得可疑目標區域目標的提案網路(region propose network)、以及提取相同大小的感興趣池化(range of interest pooling)與回歸決策，等三個部分所建構而成。在圖 2.6 中顯示了 Faster R-CNN 的模型架構，我們在這邊簡單概述每個核心部分的重點。

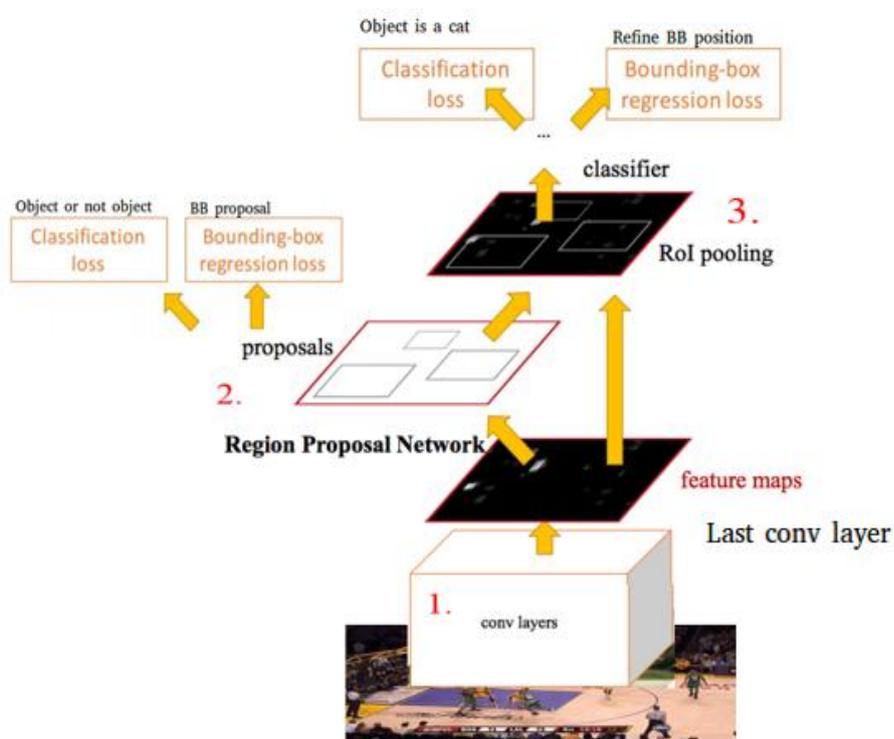


圖 2.6 Faster R-CNN 模型架構圖

前置網路的目的在取得影像上的所有特徵，架構上具有多個卷積層以及池化層等相關提取特徵的演算法所構成，用來處理後續在區域偵測上的作業。而提案網路是採用錨點偵測法在卷積後的特徵影像上進行滑動檢測，目的是為了評估每個錨點框在該特徵點上是否有前景物件存在，若物件存在，則會對該錨點的邊框進行調整。最後，感興趣池化的目的是由於每個提案的區域大小與比例不相等，若要以全連接層執行的決策迴歸，在影像上中每個物件大小是需要相等的，圖 2.7 中顯示了訓練物件時，使用的兩個正規化方法，因此，為了將不同大小與尺度的提案區域達到固定長度輸出(Fixed Length output)，感興趣池化的正規化方式是必要的。而正規化後的區域通過全連接層(fully connect layer)以及決策輸出層(Softmax layer)，計算類別的權重，同時利用在邊緣框迴歸(bounding box regression)獲得每個提取區域之真實偏移量，用於更加精確的選取目標區域。

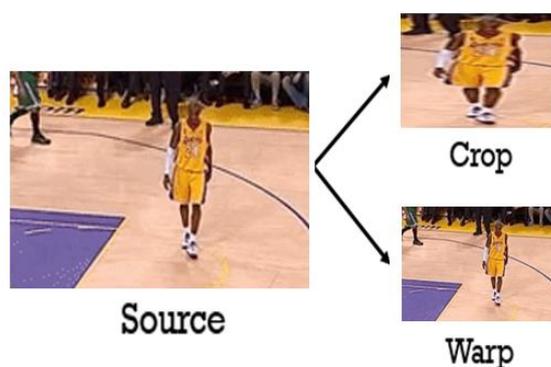


圖 2.7 正規化示意圖

2.3 殘差網路(Residual Network)

傳統的網路以多層卷積與池化方式進行提取特徵作業，但在非常深的網路架構中添加更多層的時候會表現得更差，其主要是由於"梯度消失"所造成特徵不明顯的問題，當不存在過度凝合(over fitting)狀況時，我們知道深層的網路架構不會比淺層的網路架構還要來的差，簡單來說，我們建構了 N 層時實現了一定的準度，那麼在邏輯上，我們加上 1 層變成 $N+1$ 時他至少也能實現相同的準度，我們稱該做法叫"恆等映射"，以此類推，但是該方法在實際的情況下卻表現得更差。

藉此，我們嘗試假設恆等映射的方式是難以學習的，那麼我們就不再重複學習本來有的影像，而是學習這兩者影像之間的差異，也就是所謂的"殘差(residual)"，在圖 2.8 中我們展示了殘差網路的區塊運算過程，通常恆等映射會計算輸入權重 x 到輸出權重 $H(x)$ 之間的關係，但是透過殘差方式，我們計算出結果

$$F(x) = H(x) - x \quad (1)$$

所以

$$H(x) = F(x) + x \quad (2)$$

其中我們就是學習了原始輸入權重 $F(x)$ 與輸出權重 $H(x)$ 之間的殘差 x 。

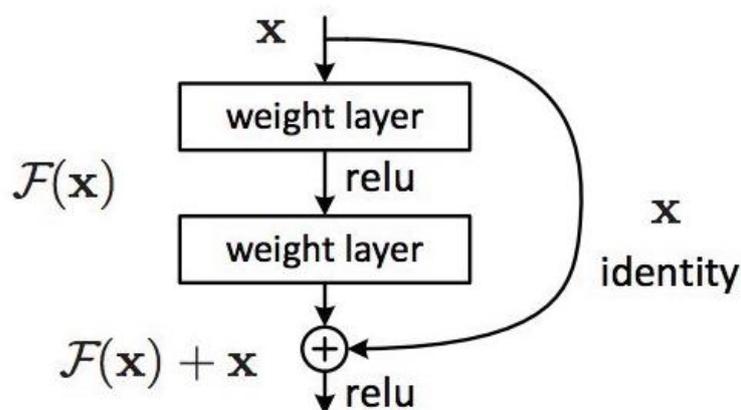


圖 2.8 殘差網路中的區塊運算架構

在訓練細微的特徵時，ResNet 能夠透過一系列層的"模塊(block)"與"捷徑(shortcut)"連接進行非常深的特徵萃取，這個"捷徑"將該"模塊"的輸入與輸出連接再一起執行"加法(add)"運算，透過"補零"或"投射"來得到匹配大小。這樣的網路架構模式在 2015 年時贏得了 ImageNet 的圖像識別冠軍。

2.4 霍夫線偵測(Hough Line Detect)

球場線的提取是追蹤與轉換比賽廣播視頻的關鍵步驟之一，在球場線的提取作業上運用了色彩空間的遮罩轉換以及型態學濾波之外，最重要的線條偵測採用霍夫線偵測法。該方法是 Hough 於 1962 年所提出，如下方公式我們得知一個點是否落在線上

$$y = mx + c \quad (3)$$

當直線呈垂直時，斜率會是無限大，這種時候是無法偵測出直線的，因此，霍夫線偵測法是基於假設參考點位置，如下

$$x \cos(\theta) + y \sin(\theta) = r \quad (4)$$

透過起始點與目標點之間得到角度 theta 以及距離 rho，並且將座標值帶入後需要透過累加器來統計直線上的落點數量做為識別直線的一種手段。

2.5 仿射轉換(Affine Transform)

仿射轉換的技術是在平面向量中進行線性變換的方法，目的是為了將座標從空間中映射到另一個空間的一種方法，該方法主要針對旋轉與平移兩個部分進行，從公式

$$A = \begin{bmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{bmatrix} \quad (5)$$

公式 A 中的矩陣表示了旋轉所需的參數

$$B = \begin{bmatrix} b_{00} \\ b_{10} \end{bmatrix} \quad (6)$$

公式 B 中的矩陣表示了位移所需的參數

$$M = [A \ B] = \begin{bmatrix} a_{00} & a_{01} & b_{00} \\ a_{10} & a_{11} & b_{10} \end{bmatrix} \quad (7)$$

公式 M 中的矩陣表示了某個平面轉換至另一個平面所需要的參數，因此，假設將平面 M 上的坐標 A(x, y) 轉移到另一個平面 T，當我們想知道落在平面 T 上的絕對位置 B(x', y') 時，我們可以透過

$$x' = a_{00}x + a_{01}y + b_{00} \quad (8)$$

$$y' = a_{10}x + a_{11}y + b_{10} \quad (9)$$

進行坐標之間的投影轉換，簡單來說，在座標 A 轉換至座標 B 的過程中，矩陣 A 掌握了旋轉的變量，而矩陣 B 則是掌握了移動的變量。最後，仿射轉換的矩陣參數是能夠透過初期設定在原始影像上的三個點坐標，以及平面影像上的三個點坐標，透過聯立方程的求解方式取得，從這邊我們可以了解到映射投影上的核心概念。

第三章研究方法

3.1 方法流程圖

此章節將闡述本論文的方法及流程。由圖 3.1 可以觀覽本文方法全貌的運作狀況，下面的小節也會以此圖進行補助說明。

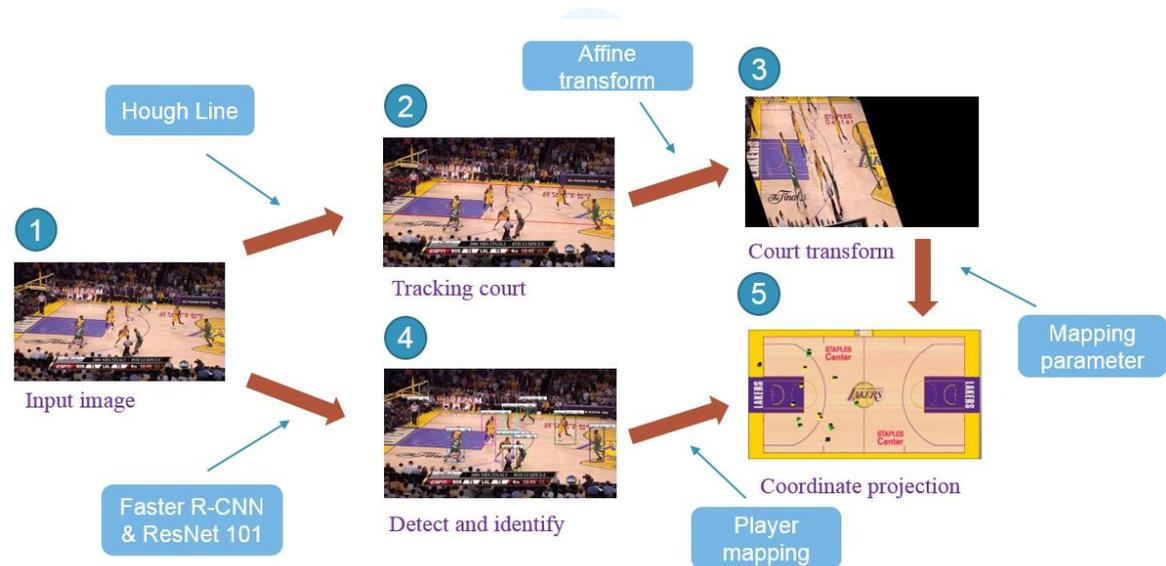


圖 3.1 方法流程圖

3.1.1 流程步驟

根據圖 3.1 所示，這是一個標準資料處理與轉換的流程圖，透過偵測球場物件、識別球員身分、追蹤球場界線、計算仿射參數及球員座標映射等基本步驟達到我們要的自動化紀錄應用。

3.1.2 資料選擇

本論文使用的 Data set 為球類賽事廣播影像，我們選取隨機一段場次與球場，該 Data set 為大眾影音媒體上隨處可見的資料，我們事先將影像進行錄製並且透過 OpenCV 影像處理套件幫我們執行影像切片作業，方便我們對每一偵進行處理與分析。



圖 3.2 大眾媒體上播送的賽事影像

3.1.3 資料處理

透過多媒體網站抓取下來的影片，我們將進行影像前處理動作，而我們將針對兩種狀況進行前處理：

- (一) 由於賽事影片由多張圖片組成，我們必須對每張圖片進行分析，所以我們需要透過影像處理套件 OpenCV 對影像進行切片(convert)作業，幫助我們將影片以每秒幾幀的圖像做為輸出。
- (二) 我們使用 LabelImg 這項工具將我們所有圖片中的物件進行標籤動作，標籤物件後的結果我們將會劃分目標物件區域以及背景區域於我們的 xml 標籤文檔中，因此兩種類別在後續的區域偵測模型訓練上有最佳的代表性。

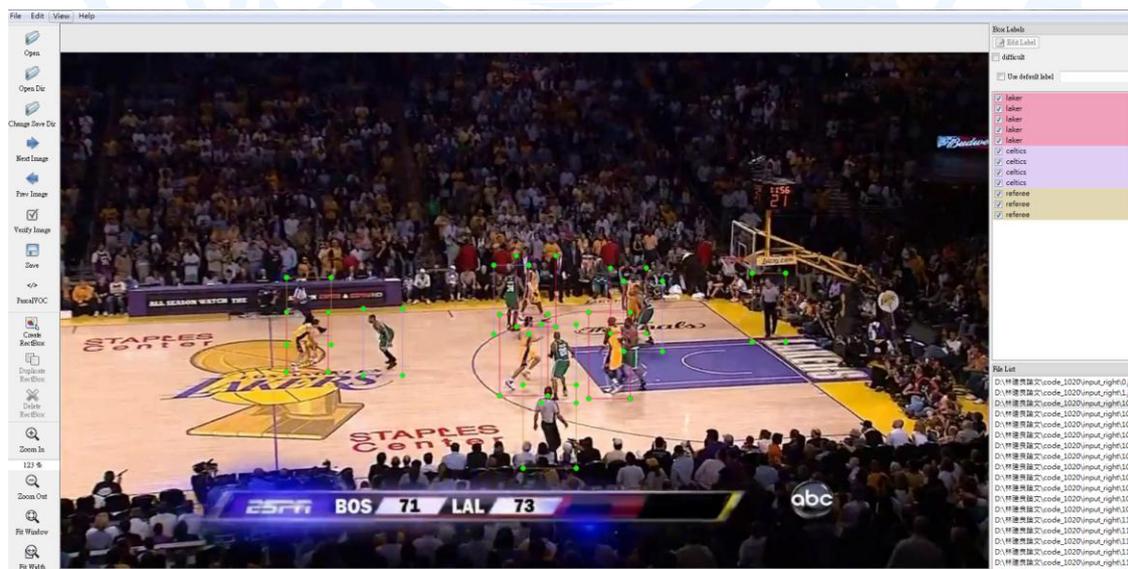


圖 3.3 labelImg 的標籤影像示意圖

3.1.5 電腦視覺及座標映射

最後的兩個步驟為電腦視覺及座標映射，自動化數據紀錄所使用的方法為電腦視覺中的偵測演算法 Faster R-CNN，Faster R-CNN 透過特徵影像(Feature map)與迴歸分析(regression analysis)的關係學習偵測以及檢測之模型，架構簡單的 Tree 分類方式，藉由球衣及其他明顯的相關特徵能夠在 Faster R-CNN 做初步的辨識，以及利用 ResNet 對每個球隊的球員進行識別身份，而 Faster R-CNN 與 ResNet 的運行有參數調配與細部說明的必要，我們留於下一節進行討論，整個流程最後的座標映射結果採用線性領域的仿射轉換方式，也將於下個部分進行詳述。

3.2 模型訓練調配

在前一節方法架構的部分，我們有提到 Faster R-CNN 與 ResNet 的參數調配，其中 Faster R-CNN 有兩個最主要的可變參數所控制，第一個為錨點比例數量 m ，第二個為錨點尺度數量 n ，透過兩個主要參數能夠得到每個錨點下具有 $m \times n$ 個框位種類。另一方面，就現今已經存在多種網路架構模型，我們將在下一個小節中說明我們為何選擇 ResNet 網路架構進行識別模型的訓練。

3.2.1 Faster R-CNN 錨點參數調配

自動化影像分析的過程中，首要面對的問題即是如何有效的將球員所在之影像區域與背景區域偵測出來。為了有效提前景物件之候選區域，本文使用 Faster R-CNN，該演算透過人工方式事先給予訓練所需之樣本與標籤資料做為訓練前景物件之區域演算，其中，基於球場動態廣播視頻所拍攝出的畫面進行偵測，由於遠景物件上的目標尺度變化量不大，因此，本文認為可以減少調配錨點框位的比例數量與大小數量，再來我們統計動態視頻縮放時球員的區域大小，來調配框位比例與尺度的參數，用來降低訓練時間上校準所造成負擔。

3.2.2 ResNet 識別網路架構

在第 2.3 章節中，本文說明了傳統的恆等映射網路進行太細微的特徵訓練時，會產生梯度消失的過度學習狀況，梯度消失的問題比起模型訓練上的過度凝合 (over fitting) 問題還要來的嚴重，主要是由於在訓練時難以察覺，即準確率會先上升然後達到收斂，再持續增加深度導致準確率下降，不光是在測試樣本上誤差增大，訓練集本身誤差也會增大，因此，基於這些考量，如果我們取消訓練影像中的相同特徵，而是訓練特徵彼此之間的差異的話，就能夠在某種程度上解決了這些問題。

對於本文而言，在遠景物件的影像上進行物件識別是一項挑戰，影像上存在著物件過小和模糊不清等因素，因此，再挑選模型架構上必須謹慎，就目前而言我們透過 ImageNet 的物件識別結果做為依據，如下表 3.1 中所示，在 2015 年贏得了物件識別的模型架構 ResNet 在 1000 類的識別任務中達到 3.57% 的 Top 5 準確率，與其他網路架構相比有著較大的差距，同時參數卻比其他網路來的低，再時間與準確上都非常突出，因此我們選擇以 ResNet 網路架構來執行球員之間的身份識別任務。

模型	AlexNet	ZF Net	GoogleNet	ResNet
時間	2012	2013	2014	2015
層數	8	8	22	152
Top 5 錯誤率	15.4%	11.2%	6.7%	3.57%

表 3.1 ImageNet 各年份之模型架構數據

3.3 自動化數據紀錄

本文目的是將球員數據進行紀錄，而單純的將廣播視頻上將球員區域偵測出來是不夠的，我們需要的是透過絕對位置上的資訊來分析，下圖 3.6 示例了球賽廣播影像以及球場平面圖中，如何將視頻內的球員座標映射至平面上是一項挑戰，在動態影像的相機校準(Camera Calibration)方式領域中，有基於球場線模型的校準方式[15]，也有透過廣播影像辨識出球場矩形是否成立後的校準方式[14]，但在本研究中，我們希望將每一次進攻的數據進行紀錄，而且是基於實時的應用上，因此在訓練與比對上我們希望可以將過程簡單化。



圖 3.6 (左)球賽廣播影像 (右)球場平面圖

3.3.1 球場追蹤

由於動態影像必須追隨球員的進攻方向來不斷位移與縮放攝影鏡頭，這樣的拍攝作業造成影像環境的不穩定，提高了廣播影像與平面圖的匹配難度，本文透過影像色彩遮罩的方式，基於球場具有一致性的顏色調配，能夠利用影像處理技術得到球場於影像上所涵蓋的面積區域，如下圖 3.7 所示。白色區域是通過球場色彩區間之區域。



圖 3.7 影像色彩遮罩

在遮罩影像上雖然能夠明顯得到球場區域，但不管是球場區域內或區域外都伴隨著許多的雜訊，這樣會產生偵測邊緣線的不穩定性，因此我們使用形態學的侵蝕(Erosion)與擴張(Dilation)方式，針對球場內外進行濾除雜訊之作業，下圖 3.8 示例了執行形態學處理後的球場廣播影像，與執行前相比，球場區域外的雜訊能夠有效的濾除，而球場內則是將不必要的線條濾除，藉此提高球場邊緣線偵測時的穩定性。



圖 3.8 通過型態學處理之影像

接下來，我們需要找尋影像上的邊緣線，透過黑色與白色之間的關係，使用坎尼邊緣偵測(Canny Edge Detect)方式得到影像上的邊緣線，下圖 3.9 示例了通過邊緣偵測後的影像，該影像對於球場的邊緣線具有良好的描述性。



圖 3.9 通過邊緣偵測處理之影像

最後，我們使用 2.4 節中所介紹的霍夫線偵測方法，將球場的直線檢測出來，由於該影像上會產生出許多球場邊緣線之外的線條，該線條對於我們的追蹤方式是沒有意義的，所以我們透過參考點位置以及產生直線的角度作為選擇直線的依據，考慮到影像在縮放與位移時造成參考點偏移以及角度變化等問題，我們對每一偵影像所偵測出的直線資訊更新，該資訊則會做為比對下一偵直線所需的依據。下圖 3.10 示例了透過邊緣偵測後的邊緣線來偵測直線，而透過我們初始輸入的參考點資訊與角度資訊來獲得我們在後續作業上的球場線。



圖 3.10 通過霍夫線偵測之影像

由於偵測方式是有可能在某些偵數點下產生誤差，或是球員軀體遮蔽球場線等狀況，造成目標線消失的問題，如下圖 3.11 所示，禁區之參考線由於球員遮蔽或是影像縮放的尺度過大，造成目標線無法被霍夫線偵測法所偵測出來，因此我們需要考慮當該狀況發生時要如何處理。



圖 3.11 目標線偵測失敗之影像

將每一偵的目標線偵測出來是幾乎不可能的，但偵測的結果能夠以訊號學的方式做處理，換句話說就是，即使該偵頻下目標線沒有被偵測出來，仍然與上一偵之間存在著關係，我們透過每一偵的更新作業，讓沒有偵測出來的影像目標線依舊存在。下圖 3.12 示例了當圖 3.11 的狀況發生時，會透過最後更新的資訊將目標線之資訊仍然保留在影像上，縱使該線條配對的方法是會造成誤差的，但這些微的部分我們認為是可以接受的。



圖 3.12 偵測失敗時保留影像資訊之處理

3.3.2 座標映射

本文透過座標映射原理取得絕對位置之資訊，將球場影像與球場平面圖之相同參考點進行線性轉換，使得球場區域能夠與平面圖相似，根據 2.5 節中我們提到，需要由兩個平面空間，每個平面空間個別提供三個參考座標做為旋轉與位移的參數，追蹤六個參考座標可以透過 3.3.1 章節中的直線偵測來追蹤球場，藉此得到相關變化量之參數。下圖 3.13 中展示了影像球場座標轉換的過程，藉著左圖與中圖之參考點做為參數，得到與平面圖相似的視頻影像。



圖 3.13 (左)原始影像上的參考點位置、(中)平面圖上的參考點位置、(右)透過仿射轉換後的原始影像

在此部分，我們必須強調追蹤參考點上所遇到的困難進行探討，在上個章節中，雖然能夠追蹤球場之參考線，但影像伸縮造成參考點的伸縮量無法準確計算，僅能夠針對位移的偏移量進行校準，再加上伸縮導致參考點消失的問題，使仿射轉換上所需的參考點資訊無法取得，因此，我們得知單純的透過每偵之間線的偏移進行參考點的更新是不可行的，所以我們必須透過其他方式反饋推論出消失的參考點。如下圖 3.14 中顯示了仿射轉換時每個參考點之間的匹配關係，其中 A、B、C 為固定位置且各自匹配 A'、B'、C' 三個點，在示意圖中，當畫面拉進時，我們無法透過球場線之偏移關係取得 A' 在廣播影像上的座標。

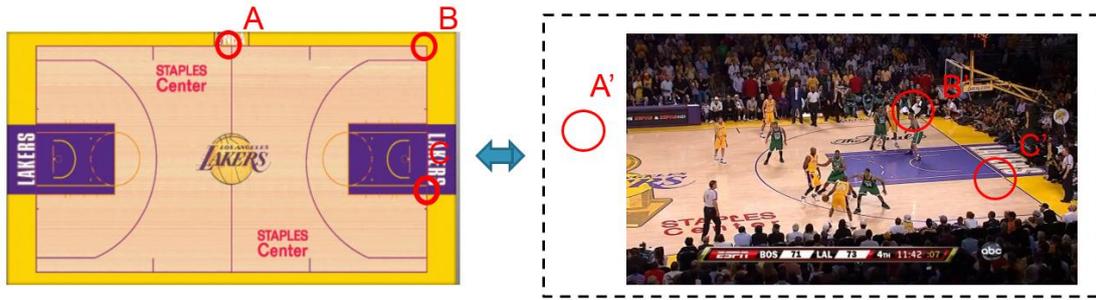


圖 3.14 仿射轉換中參考點的匹配圖

從上述得知參考座標在仿射轉換的過程中是必要的，本文藉由轉換後的圖片來找尋影像上的垂直線，得到初始水平線的線性方程，計算水平線與垂直線之間的兩個交點做為實際座標，並且計算交點與交點之間的距離，基於鏡頭之縮放比與實際距離比相等的關係上，用來推算消失於影像上的參考點坐標，解決鏡頭縮放時所造成的參考點校準問題，如下圖 3.15 展示了於仿射影像上找尋的交點過程，找尋後的交點再次仿射回原始影像上。



圖 3.15 (左)偵測影像垂直線與水平線之間的交點、(右)交點坐標反饋回原始影像

根據圖 3.14 能夠得到中線點與角頂點之間的距離，以及角頂點與禁區點之間的距離，透過等比規則就能夠統計出實際影像上，消失的中線點坐標位置，藉由此方法對每一偵進行統計達到球員座標的座標映射參數，最後，將區域偵測出來的球員位置及殘差網路辨識出的球員身分進行映射，如下圖 3.16 所示，得到了在廣播影像上球場資訊投影至球場平面圖的資訊，並且將每一偵的實時資訊，如速度、位置、奔跑距離、裁判位置等，紀錄成文件標籤資訊檔(CSV File)，幫助未來納入至資料庫或大數據檢索系統之應用。



圖 3.16 (左上)檢測器檢測之資訊、(左下)檢測資訊轉換為平面資訊、(右)以 CSV 方式針對能夠紀錄的資訊進行紀錄

最後，在圖 3.17 與圖 3.18 中各別顯示了球員座標投影與賽事數據紀錄之系統執行圖，本文將數據結構化並轉換成文件檔保存外，同時也進行了數據可視化的影像呈現方式進行保存，用意在於補助分析數據與實際影像上的比對作業，讓分析師能夠更加簡便的使用基礎數據來進行後續工程作業。

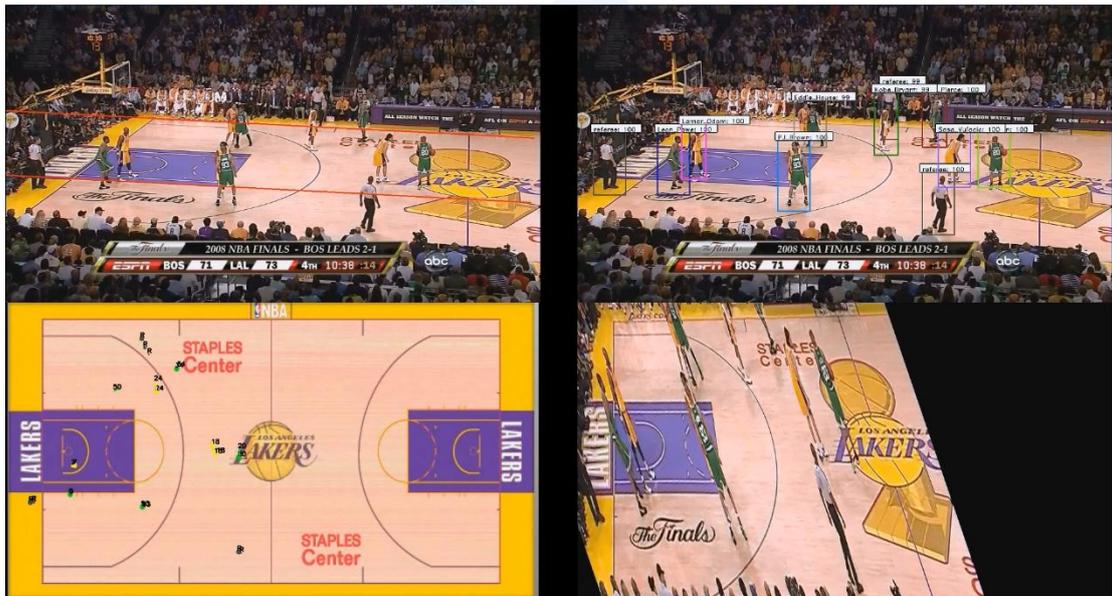


圖 3.17 球員位置投影之系統執行圖

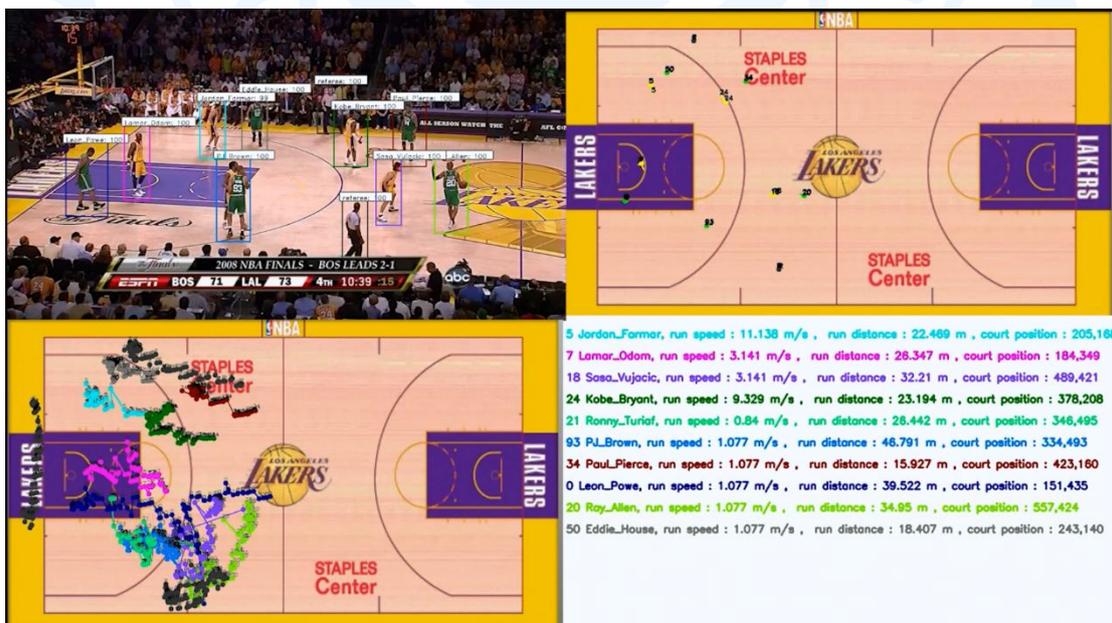


圖 3.18 賽事數據紀錄之系統執行圖

第四章 結果討論分析

這個章節將展示本篇論文之實作成果，本文採用 NBA 冠軍戰的賽事廣播視頻進行測試，透過本論文提出的結果進行討論與建議。

4.1 NBA 球場廣播視頻

以美國職業籃球冠軍賽做為模型測試以及影像處理之環境，我們可以將賽事廣播影像的目標分類看成是一個兩層的 Tree，其中最上層總共有裁判、湖人、賽爾提克三大類，接著第二層由湖人與塞爾提克這兩類做為細分其場上球員各 5 名球員。

藉由 Tree 的分析架構上，我們以深度學習方式對每個 sub Tree 下之分層步驟進行分析及模型訓練，透過偵測模型來初步進行三大類之球隊分類，再來以兩個識別模型針對所屬的兩邊球隊之球員身份進行識別。

4.2 Faster R-CNN 分類結果

執行 Faster R-CNN 輸出結果會將區域裁剪並保存起來，我們透過剪裁後的區域進行人工方式來統計模型的準確率。

下表為透過 NBA 2007-08 總冠軍中的第四節片段進行，該片段共 120 秒，以每秒 15 幀的方式進行實驗數據的採集，共 1800 多張影像做為偵測模型的測試環境，其中，模型共擷取了 13717 張區域影像，我們以人工檢測方式來驗證模型檢測的正確率，結果如下表 4.1 中所示，在 13717 張區域影像中，共有 13676 張有效區域影像，而無效或判錯的區域影像共 41 張，模型之準確率高達 99.7%。

	總區域數量	有校區域數	無效區域數	準確度
湖人	6736	6721	15	99.78
賽爾提克	6981	6955	26	99.63
總計	13717	13676	41	99.70

表 4.1 Faster R-CNN 模型準確度統計

4.2 ResNet 識別結果

識別模型的作業上，本文採用 ResNet 101 架構執行模型訓練，各別訓練了兩個類別，分別是兩邊所屬之球隊，且各別訓練兩隊於球場上之球員，每位球員的樣本數量落在 250 至 350 之間，考慮到訓練時樣本數量不足導致模型訓練不穩定的因素，我們不希望再添加更繁複瑣碎的人工作業，改以擴增生成方式，針對每張樣本進行旋轉、位移、縮放、色彩等微量變化，產生出新的訓練樣本，讓模型在訓練上能夠更加穩健且有效提升模型的準確性。

訓練過程如下圖 4.1 中所示，本文使用了將近 12500 個樣本數量來進行識別模型的訓練，其中，湖人之模型在 150 圈的訓練環境下最終得到了 99.81% 的訓練準確率，以及 96.26% 的觀測準確率，而塞爾提克在最終訓練之準確率達到 99.79%，以及 96.69% 的觀測準確率，我們知道訓練與觀測之間的距離表示過度學習的嚴重程度，然而兩者之間的誤差落在 4% 之內，對我們而言，太過模糊不清且訊息微弱之影像訓練上，這樣的誤差值是可以被接受的。

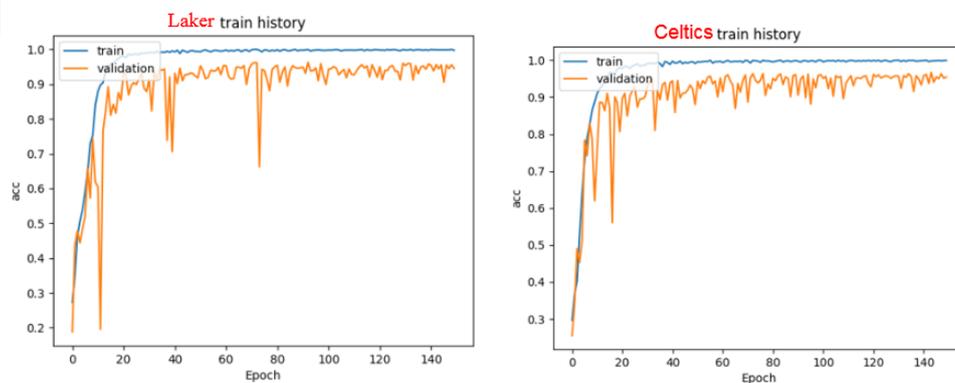


圖 4.1 (左)湖人訓練模型過程圖、(右)塞爾提克訓練模型過程圖

最後我們以不同場次之廣播影像片段做為實驗環境，該片段一共擷取了 13647 張區域影像，其中，湖人擁有 6715 張而塞爾提克擁有 6932 張，我們透過上述訓練之識別模型進行實驗，識別結果如下表 4.2 中所示，針對類別下之所有球員之身分進行識別，在湖人的模型識別中，我們得到了平均 98.81% 的識別準確率，而塞爾提克的模型識別中則是得到了平均 98.43% 的識別準確率。

Laker	張數	正確	準確率	Celtics	張數	正確	準確率
Jordan_Farmar	1381	1366	98.91	Eddie_House	1515	1505	99.34
Kobe_Bryant	1454	1437	98.83	Leon_Powe	1535	1524	99.28
Lamar_Odom	1476	1470	99.59	Paul_Pierce	1162	1141	98.19
Ronny_Turiah	1239	1224	98.79	PJ_Brown	1444	1417	98.13
Sasa_Vujacic	1165	1138	97.68	Ray_Allen	1276	1236	96.87
average	6715	6635	98.81	average	6932	6823	98.43

表 4.2 (左)湖人球員身份之識別結果、(右)塞爾提克球員身份之識別結果

4.3 仿射轉換結果討論

在本章節中，我們將討論仿射轉換所面臨到的問題，首先是球場追蹤的偵測準確率，由於球場的影像投影方式，參考點的取得與追蹤是必須確保的，因此我們有必要對該偵測法做個實驗，統計出該方法的可靠信。再來是當影像面臨到沒有偵測到參考線的狀況下，透過保留上一偵的偵測結果來彌補所產生的誤差進行討論。

首先是針對球場線的偵測方法，我們統計了一段進攻的廣播影像片段，該片段一共擁有 561 張影像，其中球場線被偵測到的張數共 520 張，而沒有被偵測到的張數共 41 張，偵測的準確率高達 92.7%，在實驗的過程中，我們探究了部分廣播影像上的問題，如下圖 4.2 中，當計分板遮蔽球場線時，該偵測方式會造成極大的誤差存在，該狀況必須等待鏡頭還原或計分板關閉等狀況，讓偵測的資訊做更新來校準影像資訊。若我們排除計分板遮蔽的狀況，則能夠提高偵測之準確率。



圖 4.2 計分板遮蔽參考線之示意圖

彌補資訊的做法是用來解決參考線沒有被偵測的情況，然而影像仍然是處於移動狀態下，而資訊卻是取上一偵的參考線資訊，因此，兩者之間是存在著相當的誤差，本文認為每秒 20 偵的偵測方式所造成的誤差是非常小的，再加上我們有進行線條位移校準的門檻設定，用來測量每次校準時的兩條線距離，其中以平均 6 點像素點之水平距離誤差存在，而該誤差只存在於短暫幾毫秒之中，對於本實驗之仿射轉換上是不會構成巨大的影響。

4.4 實時運行時間

為了達成實時廣播影像於自動化紀錄之應用，我們紀錄了本研究系統之硬體規格，並統計了每個部分的階段所花費的計算時間。

本系統主要基於 Nvidia 所開發的 CUDA GPU 加速引擎，並且採用 Titan V 顯示卡來執行偵測與辨識的效能運算，CUDA 能夠對簡單且龐大的運算量進行加速，並且支援深度學習套件及矩陣運算套件，是近年來用於 GPU 運算上不可或缺的加速運算工具。另外 CPU 採用 INTEL I7-8700K 以及 16G 的暫存記憶體容量作為系統執行的硬體環境。

從偵測模型的訓練中，我們花費了 40 多個小時進行 Faster R-CNN 的區域模型訓練，而實際偵測上，每張圖片花費了 0.3 秒來偵測球員區域。識別模型的訓練作業上則是花費了 15 分鐘，在實際辨識上，每區域以 0.05 秒來辨識球員身份。最後，仿射轉換的座標投影處理花費了 0.05 秒，我們將所有區塊結合後，在每張影像上以 0.4 秒達成偵測、辨識、轉換及紀錄，而一次的進攻時間片段大約是 20 秒，約 400 張影像，以本文系統進行時間統計得到 170 秒的運算處理時間。

第五章 結論與未來展望

本研究基於深度學習與電腦視覺方式架構出一套自動化紀錄系統，基於前景與背景的關係，教導電腦從影像中辨識球場區域下的球員身份，並同時紀錄球員於平面圖上的絕對座標資訊。最後透過數據記錄系統，將影像轉換為結構化數據進行保存。

我們所提出的論文架構可以達到 98% 的偵測準確率，以及 96% 以上的辨識精確度，在實時的偵測效率上則是每張 0.4 秒的辨識速度，且達成自動化實時球類數據紀錄的應用，方便後續人員進行分析與研究。

未來工作中，我們可以將該架構方法套用再其他類別的體育賽事上，像排球、足球、羽球...等，並且改良該場域環境下的影像處理設定來達到相似的效果，使得研究者在致力於其他研究時能更佳方便使用。

參考文獻

- [1] S. Dev, B. Wen, Y. H. Lee, and S. Winkler, “Ground-based image analysis: A tutorial on machine-learning techniques and applications,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 79-93, 2016.
- [2] J. Ding, X. H. Hu, and V. Gudivada, “A Machine Learning Based Framework for Verification and Validation of Massive Scale Image Data,” *IEEE Transactions on Big Data*, vol. PP, no. 9, pp. 1-1, 2017.
- [3] S. B. Jemaa, M. Hammami, and H. Ben-Abdallah, “Data-mining process: application for hand detection in contact free settings,” *IET Image Processing*, vol. 7, no. 8, pp. 742-750, 2013.
- [4] S. Jia, J. Hu, J. Zhu, X. Jia, and Q. Li, “Three-Dimensional Local Binary Patterns for Hyperspectral Imagery Classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 4, pp. 2399–2413, April 2017.
- [5] L. Essannouni, and D. Aboutajdine, “Correlation of robust Haar-like feature,” *Electronics Letters*, vol. 47, no. 17, pp. 961–962, August 2011.
- [6] Y. Wang, L. Du, and H. Dai, “Unsupervised SAR Image Change Detection Based on SIFT Keypoints and Region Information,” *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 7, pp. 931–935, July 2016.
- [7] S. Tuermer, F. Kurz, P. Reinartz, U. Stilla, “Airborne Vehicle Detection in Dense Urban Areas Using HOG Features and Disparity Maps,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 6, pp. 2327–2337, December 2013.
- [8] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324,

1998.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Advances in Neural Information Processing Systems*, vol. 25, no. 2, pp. 1097-1105, 2012.

[10] Y.-S. Zhao, J. Wu, Y.-F. Zhu, H.-X. Yu, and R. Xiong, “A Learning Framework Towards Real-time Detection and Localization of a Ball for Robotic Table Tennis System,” *IEEE International Conference on Real-time Computing and Robotics*, pp. 97–102, July 2017.

[11] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *TPAMI*, 2010

[12] W.-L. Lu, J.-A. Ting, J. J. Little, and K. P. Murphy, “Learning to track and identify players from broadcast sports videos,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1704–1716, 2013.

[13] W.-L. Lu, J.-A. Ting, K. P. Murphy, and J. J. Little, “Identifying Players in Broadcast Sports Videos using Conditional Random Fields,” in *CVPR*, 2011.

[14] P.-C. Wen, W.-C. Cheng, Y.-S. Wang, H.-K. Chu, N. C. Tang, H.-Y. M. Liao, “Court Reconstruction for Camera Calibration in Broadcast Basketball Video”, *IEEE Transactions on Visualization and Computer Graphics*, Vol: 22, May 2016

[15] K. Okuma, A. Taleghani, N. de Freitas, J. J. Little, and D. G. Lowe, “A Boosted Particle Filter: Multitarget Detection and Tracking,” in *ECCV*, 2004.

[16] J. Han, D. Farin, and P. H. N. de With, “A real-time augmented-reality system for sports broadcast video enhancement,” in *International Conference on Multimedia*, 2007, pp. 337–340.

[17] K. Kumar, P. Parisot, and C. De Vleeschouwer, “Demo: Spatio-temporal template matching for ball detection,” in *ACM/IEEE International Conference on*

Distributed Smart Cameras, 2011, pp. 1–2.

[18] C. Verleysen and C. D. Vleeschouwer, “Recognition of sport players’ numbers using fast-color segmentation,” in IS&T/SPIE | Electronic Imaging, 2012.

[19] N. Dalal and B. Triggs. “Histograms of Oriented Gradients for Human Detection,” in CVPR, 2005.

[20] C. L. Zitnick and P. Dollar. “Edge Boxes: Locating Object Proposals from Edges.” In ECCV 2014.

[21] C. P. Papageorgiou, M. Oren, and T. Poggio. “A general framework for object detection.” In ICCV 1998.

[22] D. Cervone , A. D’Amour , L. Bornn , and K. Goldsberry, “POINTWISE: Predicting Points and Valuing Decisions in Real Time with NBA Optical Tracking Data”, in MIT SLOAN 2014.

[23] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks” in Neural Information Processing Systems (NIPS) 2015.

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun “Deep Residual Learning for Image Recognition” in CVPR 2015.