

中華民國第 63 屆中小學科學展覽會  
作品說明書

科 別：電腦與資訊學科

組 別：高級中等學校組

作品名稱：基於深度學習之排球運動影片球體追蹤

關 鍵 詞：深度學習、排球、物件追蹤

編 號：

## 摘要

本研究旨在使用一般攝影機拍攝的排球比賽影片，利用深度學習模型追蹤排球軌跡，以減少多臺高速攝影機的需求，並節省人工觀看影片的時間成本。

本研究利用 TrackNetV2 這個基於卷積神經網路的深度學習框架，自行標記排球比賽資料集後進行訓練，訓練後的結果可以成功追蹤影片中的排球，常用的深度學習評價指標數值為 Accuracy（正確預測的樣本數占總樣本數的比例）=93.1%，Precision（預測為正樣本中真正為正樣本的比例）=96.5%，Recall（被正確識別為正樣本的佔所有正樣本的比例）=95.6%。本研究成功地使用一般攝影機追蹤排球軌跡，證明了深度學習模型在排球比賽分析中的應用價值。

# 壹、前言

## 一、研究動機

近年來，物件追蹤技術被廣泛應用於球類運動上，例如鷹眼系統（Hawk-Eye）已被應用於網球、羽球等運動，以追蹤和記錄球的路徑。然而由於其架設成本高，多數運動員無法負擔。為了解決此問題，國立交通大學網路最佳化實驗室開發了深度學習架構 TrackNet，可以由一般攝影機拍攝之網球比賽影片追蹤網球軌跡。（鍾奉原，2020）

排球在我校是非常興盛的一種運動，經常舉辦班際球賽等排球比賽，因此我們便想利用 TrackNetV2 追蹤排球比賽影片中的排球，以協助分析排球比賽，減少人工花費大量時間觀看的成本。

## 二、研究目的

應用 TrackNetV2 追蹤由一般攝影機（非高速攝影機）所錄製之排球比賽影片中的排球，藉由自動追蹤球的路徑協助排球比賽的賽事分析。

## 三、文獻回顧

傳統影像物件辨識是根據物件的外顯特徵和統計特徵進行偵測的，但排球比賽中的排球會因打擊的關係而導致有變形的情況出現，再加上移動速度過快，快門相對於球的速度來說較慢，因此容易有影像殘留及模糊的現象出現。

### （一）TrackNet

TrackNet（黃昱銓，2018）提出了一個以 CNN（Convolutional Neural Network，卷積神經網路）為基礎的深度學習架構（如圖 1）。但不同於其他深度學習網路，它容許一次輸入多張連續幀，可從中學習球的影像特徵及軌跡特性。然後仿效 FCN（Fully Convolutional Network，全卷積網路）的生成階段，生成用於偵測及定位球的熱度圖（如圖 2）。最後再依據熱度圖計算畫面中可能存在的網球。這個模型不僅能從模糊影像中定位球的存在，更可以進一步的判斷受遮擋的網球位置，可以很好的達到我們的需求。

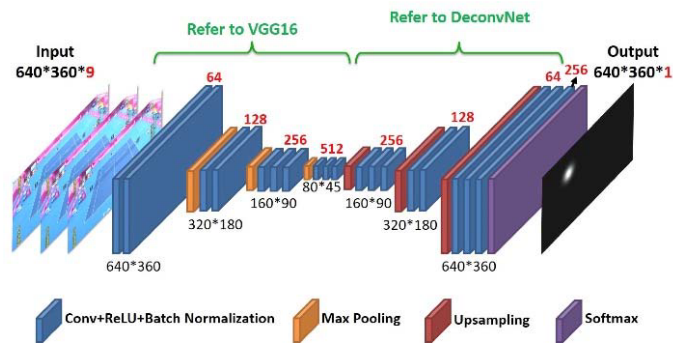


圖 1 、TrackNet 架構圖（黃昱銓，2018）

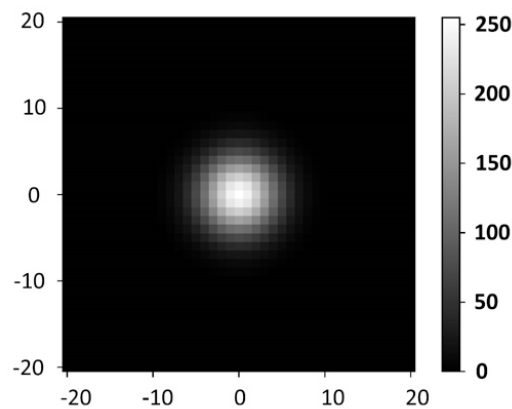


圖 2 、熱度圖（黃昱銓，2018）

## （二）TrackNetV2

TrackNetV2 (Nian-En Sun et al., 2020) 在 TrackNet 的基礎上改進了它的處理速度、準確度、記憶體用量等。他們將 TrackNet 的模型重新設計，讓它從一個 MISO (Multiple-In Single-Out) 的模型變成一個 MIMO (Multiple-In Multiple-Out) 的模型，處理速度從原本的 2.6 FPS (Frame per second) 上升到 31.8 FPS。他們也參考了 U-Net 的 skip connection (殘差連接)，將卷積層所得的特徵圖 (feature map) 串聯到相應的上採樣層 (upsampling layer)，以此方式將低層次與高層次的特徵結合起來以提高準確度。他們所使用的資料集為來自 18 場羽球比賽共 55563 幀的羽球比賽影片。

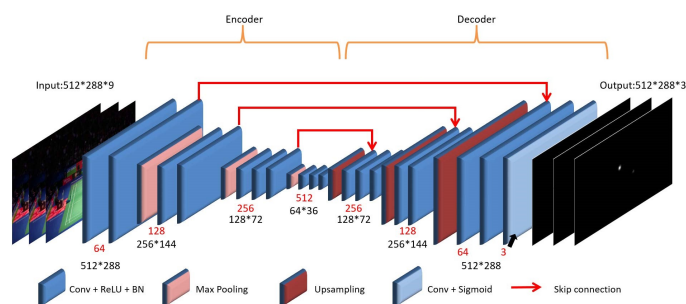


圖 3 、TrackNetV2 架構圖 (Nian-En Sun et al., 2020)

### (三) 卷積神經網路

卷積神經網路 (Convolutional Neural Network, CNN) 是一種深度學習的神經網路模型，常被應用於影像識別方面。CNN 的結構包含卷積層 (Convolution Layer)、池化層 (Pooling Layer)、全連接層 (Fully Connected Layer)。

#### 1. 概念

##### (1) 權值共享

在一般的圖像內有許多的特徵是相同的，如特定的輪廓或線條，那我們就可以讓相同幾個神經元組成的卷積核去學習這個特徵，透過滑動窗口對整張圖片進行卷積，進而達到節省參數的效果。(Cinnamon AI Taiwan, 2019)

##### (2) 保留位置資訊

圖片中的像素 (Pixels) 與其鄰近的像素會有一定的關聯度，若使用 FC (Fully Connected, 全連接網路結構) 的結構來訓練圖像資訊的話，要先通過一個展開 (Flatten) 的步驟，把高維的資訊拉成一條直線，如此一來就會大量失去特徵之間的空間資訊。(Cinnamon AI Taiwan, 2019)

#### 2. 結構

##### (1) 卷積層 (Convolution Layer)

卷積層負責提取圖像中的局部特徵，其原理是透過許多的卷積核 (kernel) 在圖片上進行滑動擷取特徵。將圖片與特定的卷積核進行卷積運算。

0	0	9	0	0
0	1	11	1	0
8	9	12	9	8
0	1	11	1	0
0	0	9	0	0

 $\ast$ 

1	0	-1
1	0	-1
1	0	-1

 $=$ 

-24	0	24
-26	0	26
-24	0	24

$$\begin{aligned}
&12 \times 1 + 9 \times 0 + 8 \times (-1) + \\
&11 \times 1 + 1 \times 0 + 0 \times (-1) + \\
&9 \times 1 + 0 \times 0 + 0 \times (-1) = 24
\end{aligned}$$

圖 4 、卷積層 (Brandon Da Silva, 2018)

## (2) 池化層 (Pooling Layer)

池化層是用來大幅降低參數量級 (降維)、降低過擬合問題、緩解卷積層對位置的敏感度。池化層的計算與卷積層一樣，都是透過滑動視窗框選的局部數值進行數值運算，最常用的計算方式為 Max pooling，在框選的局部數值中挑出最大值。(李馨伊, 2020)

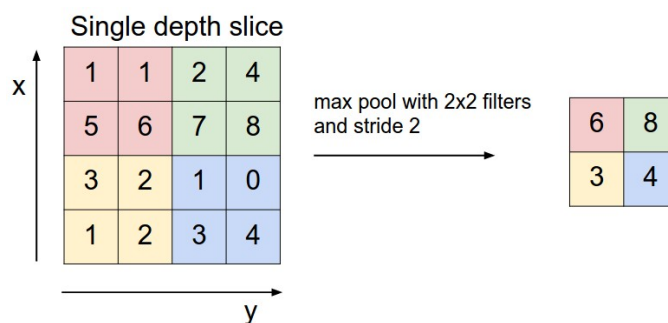


圖 5 、池化層 (Max pooling) (cs231n, 2021)

## (3) 全連接層 (Fully Connected Layer)

在經過攤平 (Flatten) 變為一維數據後，全連接層 (FC) 主要在做最後的特徵提取，並且利用最後一層 FC 當作分類器。

# 貳、研究設備與器材

## 一、硬體設備

- (一) 中央處理器 (CPU) AMD Ryzen 9 5950X 16-Core Processor
- (二) 記憶體：32GiB
- (三) 圖形處理器 (GPU)：NVIDIA GeForce RTX 3090 (GPU 記憶體 24 GiB)
- (四) 作業系統：Linux (Ubuntu 20.04.01)

## 二、軟體及工具環境

- (一) Python 3.7.13：具有高效能的高階資料結構，直譯式且物件導向的程式語言。
- (二) moviepy.editor 1.0.3：是一個 Python 用來編輯影片的函式庫。
- (三) PyQt5 5.15.7：是一個 Python 用來創建圖形使用介面 (Graphical User Interface, GUI) 應用程式的函式庫。
- (四) pandas 1.3.5：是一個 Python 用來操作及分析資料的函式庫。
- (五) PyMySQL 1.0.2：是 Python 用來連接到 MySQL 資料庫伺服器的一個介面。
- (六) opencv-python 4.6.0.66：是一個用於解決電腦視覺相關問題的 Python 函式庫。
- (七) imutils 0.5.4：整合了 OpenCV、numpy 和 matplotlib 的相關操作的套件，主要用於進行圖像處理。
- (八) Pillow 9.2.0：是一個 Python 中用於影像辨識的套件庫，可以用來旋轉照片、嵌字、合成照片、變更圖片解析度等。
- (九) piexif 1.1.3：可以讀取與修改圖片 Exif (可交換圖檔格式) 的套件，通常與 Pillow 搭配使用。
- (十) scikit-learn 1.0.2：是一個開源資料分析函式庫，具有各種分類、回歸和集群分析。
- (十一) keras 2.9.0：是一款用 Python 編寫而成的開源神經網路函式庫，能搭配 TensorFlow、Theano 等運作。
- (十二) TensorFlow 2.9.2：是現在重要的深度學習框架之一，支援各式不同的深度學習演算法。
- (十三) CUDA 10.1.243：是用於圖形處理單元 (graphical processing units, GPU) 的平行運算平台和程式設計模型。

## 參、研究過程或方法

### 一、研究構思與架構

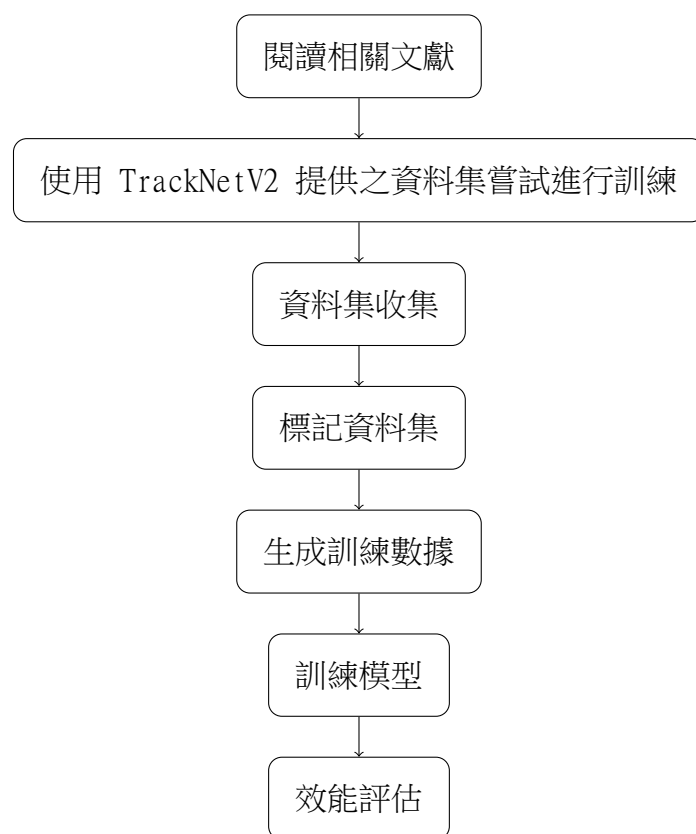


圖 6 、研究流程圖

圖 6 為我們的研究流程圖。首先我們閱讀 TrackNet、TrackNetV2 及其他物件追蹤相關論文，以了解相關技術，並嘗試使用 TrackNetV2 所提供的羽球資料集進行訓練，熟悉訓練流程並修改部分因版本更新無法使用的程式碼。

接下來我們由網路下載排球比賽影片，切割成「發球～死球」的數個影片，利用 Labeling Tool 標記排球中心在每幀的所在位置，然後生成熱度圖等訓練數據後進行訓練，最後由其他比賽擷取測試資料進行測試並評估性能。

表 1 、訓練集與測試集整理

集別	比賽名稱	擷取長度	功能
訓練集	108UVL 大專排球聯賽女一級 季軍賽	1240 秒	人工標記排球位置後 作為訓練資料訓練模型
測試集	108UVL 大專排球聯賽女一級 冠軍賽	76 秒	模型訓練完成後 使用此測試資料測試模型性能



## 二、研究過程

### (一) 資料集收集

我們由 108UVL 大專排球聯賽女一級季軍賽擷取比賽片段，為去除與「排球追蹤無關部分，每段子影片擷取發球到死球的段落，影片名稱為「局數 \_A 方比數 \_B 方比數」，如 1\_00\_00.mp4。影片幀數為每秒 30 幀，解析度為  $1920 \times 1080$ 。

原影片長度 2 小時 14 分 35 秒，擷取後資料集全長共 1240 秒，37200 幀。



圖 7、發球



圖 8、死球

我們先記錄每個子影片的名稱、開始與結束時間在一個 csv 檔內，再使用 moviepy.editor 切割影片。

記錄格式：局數 \_A 方比數 \_B 方比數，開始時，開始分，開始秒，結束時，結束分，結束秒（如表 2 所示）。

表 2、紀錄格式示意圖

1_00_00	0	5	35	0	5	46
1_00_02	0	6	18	0	6	22
1_03_03	0	7	39	0	7	43

## 肆、研究結果

## 伍、討論

## 陸、結論

## 柒、參考文獻資料

- 一、 10 程式中 (2021 年 10 月 6 日)。[Day 24] 機器學習- 不能忽視的過擬合與欠擬合。  
檢自：<https://ithelp.ithome.com.tw/articles/10278254>