

基於卷積神經網路之智慧型人物影像追蹤系統

林宥豪¹ 蔡玉娟²

¹ 國立屏東科技大學資訊管理系

² 國立屏東科技大學資訊管理系

摘要

監視器之儲存裝置雖然可記錄歷史影像，但從龐大的歷史影像資料庫檢索特定人物影像是非常耗時的工作。因此，本研究利用影像辨識及深度學習技術建置一套「基於卷積神經網路之智慧型人物影像追蹤系統」，主要包含三大功能：(1)人物影像偵測—使用灰階及高斯模糊模型過濾場景中之雜訊後，藉由高斯混合模型偵測前景，並使用型態學之侵蝕膨脹演算法取得動態物件之輪廓，再使用深度學習技術判斷影像是否為人物；(2) 人物影像識別—透過 RGB 與 HSV 色彩模型辨識人物色彩特徵，並標示其出現時間；(3) 智慧型人物影像追蹤—使用卷積神經網路學習及追蹤特定人物。實驗結果顯示本研究提出的系統具有高精確度的人物偵測及追蹤效能。

關鍵詞：影像辨識，人物影像識別，深度學習，人物影像追蹤。

一、緒論

1.1 研究背景與動機

智慧影像監控 (Intelligent Video Surveillance, IVS)，為近年視訊監控產業的發展重點，透過影像分析與智慧搜尋，提升視訊監控系統的安全性與便利性。智慧影像監控應用相當廣泛，從居家安全防竊賊，到觀光旅遊人群計數都可涵蓋，透過監視器也可記錄車流量與辨識車牌。另外，監視器畫面也可結合人臉辨識技術，來達到身分識別的功能。

隨著光學技術與硬體的進步，使得監視器成本越來越低，進而廣泛應用於環境安全監控中，但大量的監視器是必會產生許多資料量，傳統監視器最終還是要以人來進行調閱動作，但從龐大的歷史影像資料庫中調閱畫面卻是非常耗時的工作，調閱者必須長時間查看有無可疑份子，反而

造成負擔，這是需要面臨的問題，而近年來深度學習與影像辨識技術蓬勃發展，透過新技術解決傳統監視器造成的問題，輔助並降低人員負擔，讓調閱監視器越來越方便。

1.2 研究目的

本研究藉由影像辨識與深度學習技術，針對監視器影像智慧監控提出一套「基於卷積神經網路之智慧型人物影像追蹤系統」，在實驗環境門口裝設 IP Camera，並搭配合適的拍攝角度與高度，以提高辨識成功率。為了達成上述成果，本研究將達成以下目標。

(一) 行人資料集

經由監視器拍攝影像，使用灰階、高斯模糊等方法，過濾場景中之雜訊後，藉由高斯混合模型偵測前景，並使用型態學侵蝕膨脹處理，取得該動態物件之輪廓，

建立大量行人影像樣本，以利於後續使用卷積神經網路進行行人偵測。

(二) 人物影像偵測

使用卷積神經網路將行人資料集進行訓練，訓練完畢後，將影像輸入來偵測行人。

(三) 人物影像識別

透過 RGB 與 HSV 色彩模型進行行人識別，標示其行人出現時間、結束時間以及行人影像，將所有資訊存入到資料庫。

(四) 智慧型人物影像追蹤

將資料庫做智慧查詢，區分出不同行人，並計算每個行人出現次數，與總共出現之時間片段，使調閱監視器更為方便。

1.3 研究流程

本研究之研究流程圖 1，如圖所示。

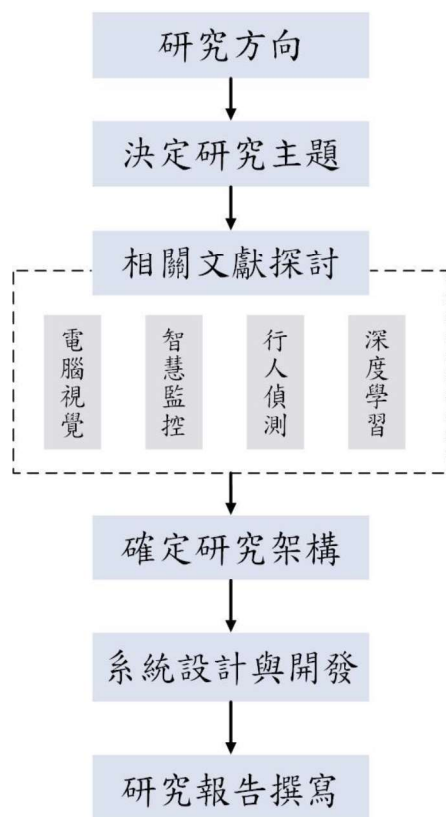


圖 1. 研究流程圖

1.4 論文架構

本論文之架構主要分為五個章節完成，

分別為緒論、文獻探討，研究方法、系統實作、結論與建議，其各簡述如下：

(一)緒論—主要論述本研究之研究背景、研究動機、研究目的、研究流程及論文架構。

(二)文獻探討—主要探討與本研究相關層面之文獻，包含智慧影像監控技術與應用、行人偵測與識別、影像辨識技術以及深度學習技術之探討與研究。

(三)研究方法—主要目的在於詳述本研究之系統架構，說明影像前處理方式以及後續透過深度學習偵測行人等過程，並介紹如何將影像辨識技術與演算法應用到本研究之中。

(四)系統設計與實作—主要描述系統設計與系統實作開發環境，並呈現系統影像辨識之分析情況與結果。

(五)結論與建議—主要呈現本研究預期之結果與預期之貢獻，以及提出研究過程中所遭受之限制並做總結。

二、文獻探討

2.1 智慧影像監控技術

2.1.1 移動物體偵測技術

動物體偵測技術是目前最常見的影像偵測技術，廣泛應用於監控系統，當監視器畫面有變化時，可記錄異動畫面，減少錄製不必要之影像，也可應用在警報系統，當有異動時發出警報具即時性的效益，其方法有：光流法、連續影像相減法、背景相減法等。

(一) 光流法(Optical Flow)

當物體移動時，影像中物體的亮度會產生變化，光流是指影像亮度變化的分佈情形，其計算成本較高，容易受到雜訊干擾。

Mansur et al.[1]提出基於光流法的視覺測量，並使用在四旋翼無人機導航研究上，測試的結果表明，該視覺定位算法可以確

定準確的位置和方向。

(二) 連續影像相減法(Temporal Differencing)

在連續的影像中，讓相隔幾張影像作相減，相減後若等於零表示沒有物體移動，若不為零，表示有移動物體。利用相減後所得的結果來辨別移動的區域。

Murali and Girisha [2]結合連續影像相減法與復相關，從連續監視影像中分割運動對象，該方法在有干擾或環境變化下，對移動物體的檢測都具有穩定性。

(三) 背景相減法(Background Subtraction)

背景相減法是種易於實現且廣受應用的方法，特別是在靜態背景的场景中，首先取得一張背景未有移動物體影像當作背景，把有物體進入的影像當作前景，前景與後景兩者相減後若等於零表示沒有物體移動，若不等於零表示有移動物體，其公式如(1)、(2)所示。

$$D_t(x, y) = |I_t(x, y) - I_b(x, y)| \quad (1)$$

$$B_t(x, y) = \begin{cases} 1, & D_t(x, y) > Th \\ 0, & otherwise \end{cases} \quad (2)$$

Amaresh et al.[3]提出了一種基於聲音定位和背景相減的即時入侵檢測系統。該系統由可旋轉監視器和麥克風組成，監視器通過聲音定位轉向聲音源，並使用背景相減法即時檢測入侵者。

Kumar and SenYadav [4]提出了一種基於背景相減法的影片對象提取與追蹤方法，並使用在複雜的環境中，該方法可支援靜態和動態背景下的物體追蹤。通過對背景資訊的初始化和更新，以提高準確度。

2.1.1 移動物體偵測技術

目標的歷史位置資訊去比較現在位置資訊進而達到追蹤的目的。用來追蹤在監視器連續畫面中的物件，通常用於追蹤特

定人或物。目前物體追蹤的方法可大致分成四類 Hu et al. [5]如表 1 所示。

表 1.物體追蹤方法[5]

區域式追蹤	Region-Based Tracking	根據與移動目標對應的區域變化來追蹤目標，將影像變動的區域視為有移動物體。
主動式輪廓追蹤	Active Contour-Based Tracking	將移動物體以邊界輪廓來表示，並賦與其空間特性，如影像邊緣與形狀等，最後在連續影像中動態更新這些輪廓來達到追蹤對象之目的。
特徵追蹤	Feature-Based Tracking	是基於物體特徵的追蹤方法，透過萃取物體特徵並比對連續影像間的特徵來達到追蹤的目的。
模型追蹤	Model-Based Tracking	利用已知的動態物體特徵建立模型，並利用該模型去預測往後連續影像的變化，最後透過搜尋和比對來達到追蹤的目的。

2.1.3 人臉偵測與識別技術

人臉偵測用於檢測影像中使否有人臉，再透過識別技術判斷該人臉是誰，該技術常用在監視保全系統與生物辨識。

Viola and Jones[6]提出了一個人臉檢測框架，能夠非常迅速地處理圖像，同時實現高檢測率。使用積分影像(Integral Image)特徵的 AdaBoost 方法，其 Cascade 分類器在人臉偵測方面速度快且準確度高。

Troitsky[7]提出了一種快速的兩層多

重人臉檢測算法，該算法由於不同的因素（視點依賴，非均勻性，光照等）而具有較少的約束條件。該算法由兩個獨立的部分組成：位置和驗證。對於位置部分來說，該方法基於一些臉部元素的多個檢測點，用於初步臉部定位；對於驗證部分，該方法基於相互區域來分析，並萃取線條特徵，估計相互亮度和亮度變化。

2.1.4 行人偵測技術

相較於人臉偵測，行人偵測焦點不在人臉上，而是去判別行人的姿態、衣服顏色及輪廓等，相對來說比較困難，常會受到光影、行人重疊及行人姿態改變等問題。

Farenzena et al.[8]提出了一種基於外觀的行人再識別(Person Re-Identification)方法，透過預處理，將行人影像劃分頭部、身軀、腿部、以及左右對稱軸，提取除了頭部以外的各區域多種特徵。

Dong Seon Cheng et al.[9]提出了一種基於圖形結構的行人再識別方法。當沒有人臉資訊，透過選擇性地在身體部位中尋找對應關係來識別個體。使用了一個身體圖形結構來表示行人，包括頭部、胸部、大腿和小腿，然後提取每個部分的顏色特徵進行比對

Bazzani et al.[10]提出了一種基於外觀的行人再識別方法，透過攝影機影格將一組個體的幀縮放成高度資訊，包含了行人外觀的全局特徵統計描述，並將行人身體分為上半身、下半身兩部分，提取各部位局部特徵，最後將全局與局部特徵做加權來辨識行人。

Kharjul et al.[11]提出了行人檢測系統的應用，使用 Adaboost 算法從圖像中分割行人，並使用支持向量機(SVM)進行分類。

2.2 智慧影像監控應用

智慧影像監控在功能性方面大致分為：

移動偵測、人流偵測、車流輛偵測、方向性偵測、警戒線/區、人臉偵測及煙霧火焰偵測等。

陳建霖[12]利用 IP Camera 拍攝畫面，使用三向相減法、光流法及高斯混合模型等演算提出了一種以電腦視覺為基礎的老人跌倒偵測演算法，針對老年人做控管，當有異常狀況發生時，可即時通知醫護人員。

陳信銘[13]提出一種電子圍籬技術應用在智慧型影像監控中，透過高斯混合模型、前景偵測技術及梯度直方圖特徵之行人偵測等電腦視覺技術，應用於校園監視器監控。

范家榕[14]提出觀光景點人潮及車流估計的智慧型監控系統，透過膚色偵測與背景相減，偵測人臉區塊，進而計算影像內人群數目；將停車格以 Canny 邊緣偵測方式，再藉由角點偵測以及車子輪廓與面積辨識車體特徵，進而計算影像內車流數目、停車格使用和剩餘數目。

邱柏訊[15]提出影像式智慧停車場安全監控技術，用以影像式火焰煙霧偵測與影像式停車位偵測，利用梯度計算、色彩分析、背景偵測及高頻衰減分析以判定火焰煙霧的發生；利用特徵抽取程序及特徵相似度比對程序來計算判定停車位使用狀況，並應用於停車場管理。

2.3 影像辨識技術

2.3.1 色彩空間

色彩空間是指使用數值來表示顏色的數學模型，目前常見的模型有 RGB、HSV 以及 YCbCr。最常使用的色彩模型為 RGB，在電腦視覺中，每張影像以像素(pixel)為單位，以整數 0 到 255 表示像素的顏色，一個像素有三組數值，RGB 代表紅光(Red)、綠光(Green)及藍光(Blue)的亮度，其人眼視覺與電腦視覺的圖片結構如圖 2 所示。

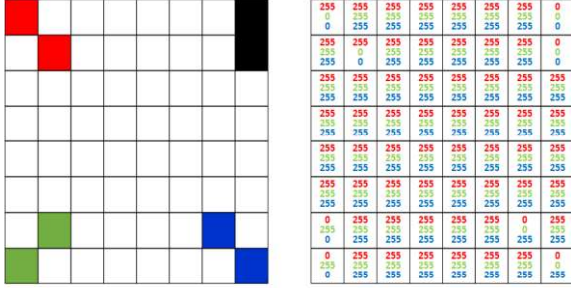


圖 2. RGB 圖片結構圖

另一常見的色彩模型為 HSV，分別代表 Hue, Saturation, Value 即色相、飽和度、明度，色相(H)是指色彩，如紅色、藍色、綠色等。飽和度(S)是指色彩的飽和度，取 0-100%的數值，數值越高顏色越純。明度(V)取 0-100%。轉換公式如 3 式，HSV 模型示意圖如圖 3 所示。

$$H = \begin{cases} 0^\circ, & \text{if } \max = \min \\ 60^\circ \times \frac{g-b}{\max-\min} + 0^\circ, & \text{if } \max = r \\ 60^\circ \times \frac{b-r}{\max-\min} + 120^\circ, & \text{if } \max = g \\ 60^\circ \times \frac{r-b}{\max-\min} + 240^\circ, & \text{if } \max = b \end{cases} \quad (3)$$

$$S = \begin{cases} 0, & \text{if } \max = 0 \\ \frac{\max-\min}{\max}, & \text{otherwise} \end{cases}$$

$$V = \max$$

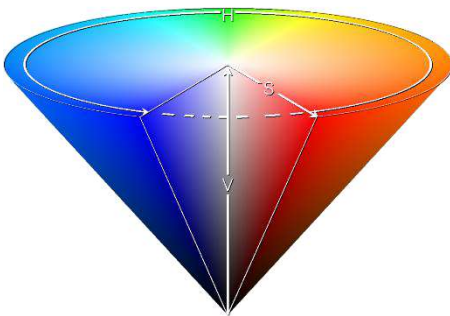


圖 3. HSV 色彩模型示意圖

另一常見的色彩模型為 YCbCr，常用在影像資料壓縮，其中 Y 是亮度 (Luminance)，Cb 是指藍色濃度偏移量成份，Cr 是指紅色濃度偏移量成份，轉換公

式如 4 式，YCbCr 模型示意圖如圖 4 示。

$$\begin{aligned} Y &= 0.299R + 0.578G + 0.114B \\ Cb &= 0.564(B - Y) \\ Cr &= 0.713(R - Y) \end{aligned} \quad (4)$$

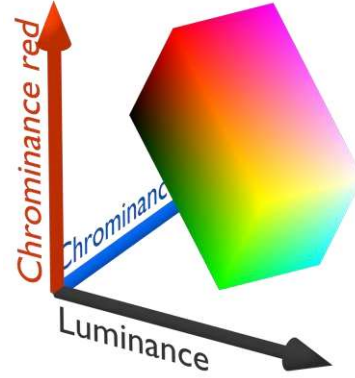


圖 4. YCbCr 色彩模型示意圖

Shaik et al.[16]提出了人體膚色檢測 HSV 和 YCbCr 色彩空間的比較研究，比較了各色彩空間的特性。實驗結果顯示，由於顏色（色度）和亮度（亮度）資訊及其非均勻特性的混合，RGB 色彩空間不適合用於顏色的檢測和顏色分析。使用 YCbCr 顏色空間用於膚色分割最有效率。

Zhu et al.[17]使用 YCbCr 色彩模型來去除影像中的陰影，提出了對圖像去除陰影時可能發生的問題和挑戰，並提出陰影去除的概念，其中試圖克服一些問題，例如照明、顏色恆定性和反射等。實驗結果表明，該方法能有效檢測和消除陰影。

2.3.2 灰階

在電腦視覺中，常會將彩色影像轉換成灰階影像，以 0 到 255 來表示，0 為最暗(黑色)，255 為最亮(白色)，可透過公式將 RGB 進行轉換，而其轉換公式如 5 式所示。

$$\text{Gray} = R*0.299 + G*0.587 + B*0.114 \quad (5)$$

2.3.3 二值化

透過二值化設定門檻值(Threshold)將影像轉換成黑白影像，若像素質大於門檻值則視為背景，將像素質設成 255，否則設為 0。其轉換公式如 6 式所示，轉換結果如圖 5 所示

$$T(f) = \begin{cases} 255, & a \geq f \geq b \\ 0, & \text{otherwise} \end{cases} \quad (6)$$



圖 5.二值化轉換圖

2.3.4 前景偵測

透過前景偵測，可找出正在移動的物件。目前最常見的方法是建立高斯混合模型(Gaussian Mixture Model, GMM)來判斷前景，影像是由 RGB 三種數值所構成，使用高斯混合模型來觀察像素值在一段時間內的變化與機率分佈，進而去判斷背景，其高斯混合模型前景偵測如圖 6 所示。



圖 6.高斯混合模型前景偵測

Stauffer and Grimson[18]提出使用高斯混合模型(Gaussian Mixture Model, GMM)來判斷前景來達到位移偵測之目的，主要是用於視頻監控領域，實驗指出該方法能在戶外下雨下雪的環境，穩定執行 16

個月。

Kaewtrakulpong and Bowden[19] 將 GMM 的訓練過程做了改進，將訓練過程分為兩部分，第一部分使用 EM 演算法進行權值均值方差更新，第二部分的過程採用原先 Stauffer and Grimson[18]中的方法進行更新，取得了更好的檢測效果。

Zivkovi and DerHeijden[20]提出遞迴方程來用於不斷更新高斯混合模型的參數，取得了很好的效果和穩定性，來達到背景相減的目的。

2.3.5 膨脹與侵蝕

透過型態學的膨脹可以將影像特徵更加突顯，再透過侵蝕可以取得輪廓並過濾雜訊，萃取出更好的影像，其膨脹與侵蝕效果圖如圖 7 所示。

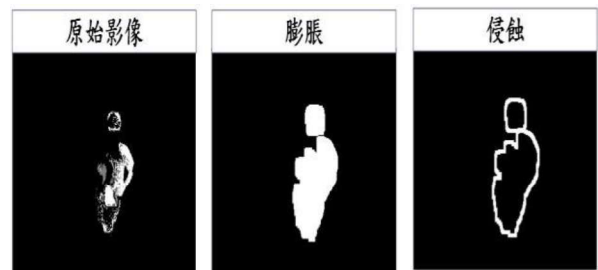


圖 7.膨脹與侵蝕

2.3.6 特徵擷取

Dalal and Triggs[21]提出利用方向梯度直方圖(Histogram of oriented gradient, HOG)演算法來進行行人特徵擷取，HOG 主要用於計算局部圖像梯度的方向信息的統計值。首先將圖像分成小的連通區域，稱作細胞單元 (cell)。接著計算細胞單元中各像素點的梯度方向直方圖，最後把直方圖組合構成特徵描述器。

Uijlings et al.[22]提出了選擇性搜索(Selective Search)在物件偵測上，結合了窮舉搜索法和分割的優點，透過圖片結構萃取出可能包含物件的候選區域。

Liu et al.[23]提出了一種高階特徵表

示方法，將行人影像依照身體比例進行分割並萃取特徵，使用結構金字塔圖層進行學習與特徵共享，最後使用自適應權重進行特徵融合以適應不同場景下的行人。

2.4 深度學習技術

2.4.1 類神經網路

類神經網路主要是使用數學模型來模擬生物神經網路的資訊處理系統。每個階層中包含許多人工神經元(Artificial neuron)，每個神經元會跟下一層的神經元連接，並將上層的神經元輸入加總並進行活化函數(Activation function)的轉換，當成神經元的輸出，輸出值經過權重計算(weight)後傳遞給下一層的神經元。這些神經元通常是以平行且分散的方式作運算，用於同時處理大量的資料，如語音辨認、影像識別等。

(一) 感知器演算法(Perceptron algorithm)
Rosenblatt[24]提出了感知器演算法，感知器是指收到多個數入訊號之後，再當作一個訊號輸出，是最早的神經網路概念，其公式如 7 式所示，概念圖如圖 8 所示。

$$y = \begin{cases} 0 & (b + w_1x_1 + w_2x_2 \leq 1) \\ 1 & (b + w_1x_1 + w_2x_2 > 0) \end{cases} \quad (7)$$

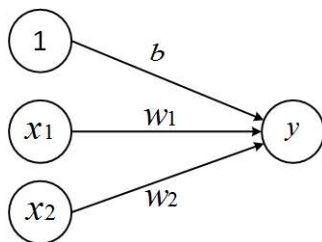


圖 8. 感知器概念圖

其中 x_1 、 x_2 為輸入訊號， y 為輸出訊號， w_1 、 w_2 代表權重(weight)， b 為偏權值，當輸入訊號傳送到神經元時，分別乘上權重，並計算加總，如大於 0 輸出為 1。

(二) 活化函數(Activation function)

由於感知器必須人工設定適當的權重，且感知器無法解決互斥或閘(XOR Gate)等問題，進而沒落。而神經網路透過活化函數可進一步從資料當中，自動學習適當的權重參數，常見的函數為 sigmoid 函數，其公式如 8 式所示。

$$h(x) = \frac{1}{1 + e^{-x}} \quad (8)$$

其中的 e 為納皮爾常數(Napier's constant)，透過 sigmoid 函數可讓輸出數值介於 0 到 1 之間，相較於感知器只能傳遞 0 或 1 兩個數值，神經網路傳遞的是連續性的訊號，進而搭配學習率進行數值修正，其 sigmoid 函數圖如圖 9 所示。

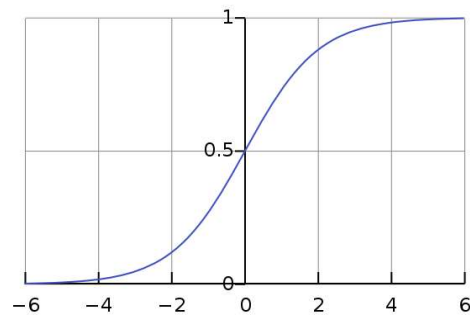


圖 9. sigmoid 函數圖

近年在深度學習中常用的函數為線性整流函數(Rectified Linear Unit, ReLU)，Krizhevsky et al.[25]發現使用 ReLU 函數收斂速度相較於 sigmoid 快上許多，由於 sigmoid 函數需要指數運算，而 ReLU 函數只需要設置閾值，進而加快了運算速度，其公式如其公式如 9 式所示。

$$h(x) = \begin{cases} x & (x > 0) \\ 0 & (x \leq 0) \end{cases} \quad (9)$$

ReLU 將輸出超過 0 的直接輸出，0 以

下的輸出為 0，其 ReLU 函數圖如圖 10 所示。

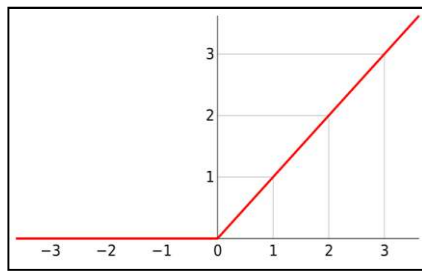


圖 10. ReLU 函數圖

2.4.2 卷積神經網路

卷積神經網路相較於一般神經網路多出了卷積層與池化層，透過過濾器(filter)進行卷積運算，降低模型複雜度，再透過池化來降維，縮小垂直、水平空間的運算。

(一) 卷積層(Convolutional Layer)

輸入資料通過卷積層做卷積運算將權重做內積。透過過濾器(Filter)來掃描圖像中的某些特徵，像是直線、橫線等。與輸入的圖像做卷積運算，可增強圖像特徵並降低雜訊，運算後會產生多個向量稱為特徵映射(Feature Maps)。

在影像中，相同的特徵可能會出現於其他位置，所以可用相同過濾器來萃取特徵，降低網絡複雜性。由於權值共享的特性，使得一個過濾器只能提取一種特徵，當過濾器越多，可以提取圖像不同的特徵也越多，對圖像裡的內容就越了解。

(二) 池化層(Pooling Layer)

池化層使用一個固定尺寸的矩陣在輸入圖像上移動，挑出該範圍內最大(Max Pooling)的像素值當作輸出，接著移動步幅，再繼續挑選區域內的最大像素值，最後輸出的圖像矩陣就是由各區域的最大像素值所組成。池化層利用影像局部相關性對影像進行降維採樣(Down sampling)減少空間維度。池化層的做法分成最大池化(Max Pooling)、重疊池化(Overlapping Pooling)

以及平均池化(Mean Pooling)。

Krizhevsky et al.[25]採用重疊池化(Overlapping Pooling)減少圖像資訊的損失，並觀察到在訓練期間具有重疊採樣的模型較不會有過度擬合的現象。

(三) 全連接層(Fully-Connected Layer)

卷積神經網路模型在最後幾層會連接至全連接層，全連接層與傳統神經網路一樣，每個神經元都與前一層所有神經元連接。

Hubel and Wiesel[26]通過對貓視覺皮層細胞的研究，提出了接受域(receptive field)的概念，可以視為第一代卷積神經網絡的概念。

電腦視覺深度學習在 2012 年的 ILSVRC 大型視覺辨識競賽(ImageNet Large Scale Visual Recognition Challenge)受到矚目，在 2012 年 AlexNet 獲得了壓倒性的勝利，顛覆了過去影像辨識的手法後，陸續有學者使用深度學習來處理影像。

Krizhevsky et al.[25]訓練了一個龐大的深層卷積神經網絡 AlexNet，將 ILSVRC-2010 比賽中的 120 萬張高分辨率圖像分為 1000 個不同的類別，其網路包含 6000 萬個參數，65 萬個神經元，5 個卷積層和 3 個全連接層連結 1000 類的 softmax 分類器，使用 GPU 非飽和神經元和卷積運算，提升訓練的速度。其中為了減少全連接層中的過適問題(overfitting)，提出了 dropout 正規化的手法，其網路架構圖如圖 11 所示。

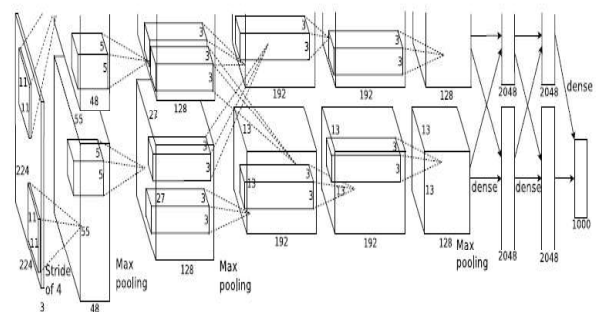


圖 11. AlexNet 架構圖[25]

2.4.3 深度學習應用

Girshick et al.[27]結合了 Uijlings et al.[22]的選擇性搜索方法與 Krizhevsky et al.[25]的 AlexNet 深度網路模型，提出了

R-CNN(Regions with CNN features)用於物件偵測，該方法比起過去傳統的方法效能更高，達到了 53.3% mAP，其流程圖如圖 12 所示。

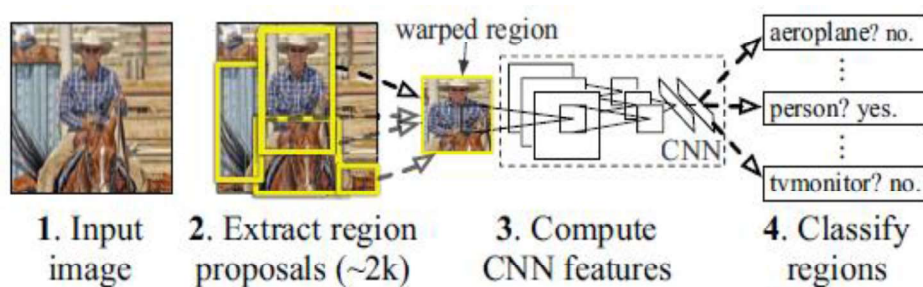


圖 12. R-CNN 流程圖[27]

R-CNN 使用了 Selective Search 方法從一張影像產生約 2000 個候選區域，再對每個候選區域，使用 AlexNet 深度網路萃取特徵，最後將特徵送入每一類的 SVM 分類器，判別是否屬於該類。

Long et al.[28]利用深度學習應用在圖像分割領域，提出了全卷積神經網路(Fully Convolutional Network, FCN)，對全部像素進行類別分類，來達到圖像分割的目的。FCN 將一般 CNN 全連接層使用卷積層取代，在輸出處理時可保持相同空間大小。Kang and Wang[29]提出一個快速的全卷積神經網路 (Fully Convolutional Neural Network, FCNN)的人群分割方法，透過 1×1 卷積核取代 CNN 中的全連接層，並應用在人群分割。

Sun et al.[30]使用卷積神經網路學習臉部特徵，並使用聯合貝式模型作為分類器，選取七組特徵分別訓練分類器，並將結果使用支持向量機(SVM)進行組合，得到的最終模型在 LFW(Labeled Face in the Wild)資料集上的人臉驗證任務中，達到了 99.15%的準確率。

Wu et al.[31]提出了一種輕型卷積神經網路(Light CNN)框架，與一種新的激勵函數 Max-Feature-Map，可以學習更適用的臉部表情訊息。

Gao et al.[32]提出一種結合 Adaboost 算法和卷積神經網路的頭部檢測人群計數方法。透過此方法可大幅減少分類時間。

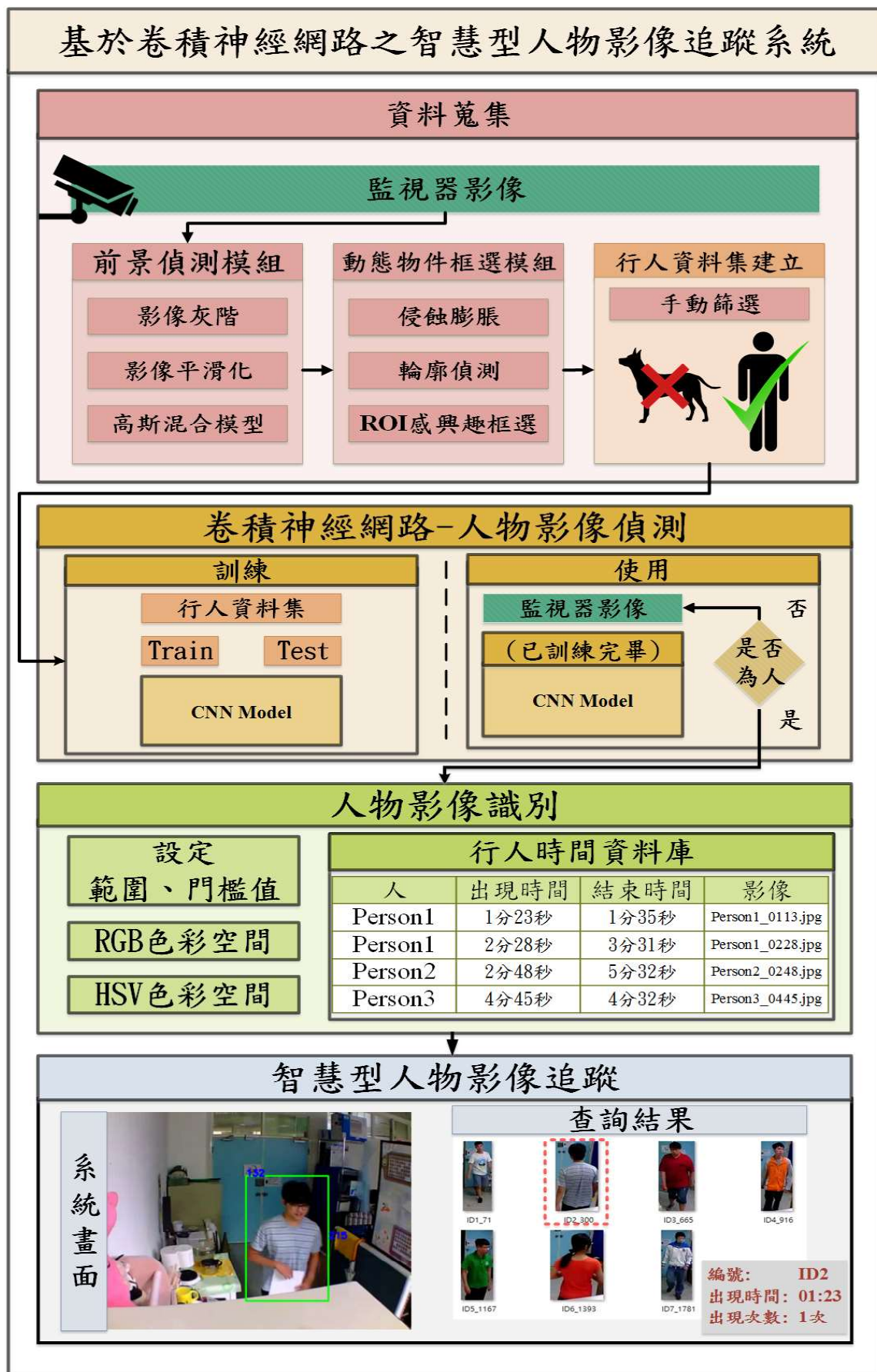


圖 13.本研究之系統架構圖

三、研究方法

本研究主要是基於影像辨識與深度學習技術提出一套「基於卷積神經網路之智慧型人物影像追蹤系統」其主要目的是設計出一套可以智慧搜索監視器影像中的所有人物的系統，以大幅減少調閱監視器影像的時間與麻煩。基於第二章的文獻探討智慧影像監控技術、智慧影像監控應用、以及影像辨識技術與深度學習技術理論與基礎探討後，本章節將分為三個主要機制與一個行人時間資料庫做細部介紹。

3.1 系統架構

本系統「基於卷積神經網路之智慧型人物影像追蹤系統」，系統架構圖如圖 13 所示。主要功能有：(1) 資料收集—透過前景偵測萃取出移動影像，再透過手動篩選

出行人影像；(2) 卷積神經網路人物影像偵測—透過資料集訓練出判斷行人之神經網路，過濾非人影像；(3) 人物影像識別—透過 RGB、HSV 兩大色彩空間，判斷行人顏色與亮度，進而識別行人。(4) 智慧型人物影像追蹤—查詢行人標記、出現時間點及結束時間點計算該行人在影片中出現之次數與每次的時間。

3.2 資料蒐集

3.2.1 前景偵測模組

透過預處理將圖片進行高斯模糊將影像平滑化過濾雜訊，再使用高斯混合模型偵測前景，當畫面有物件移動時，就會被偵測出為前景像素，其流程如圖 14 所示。

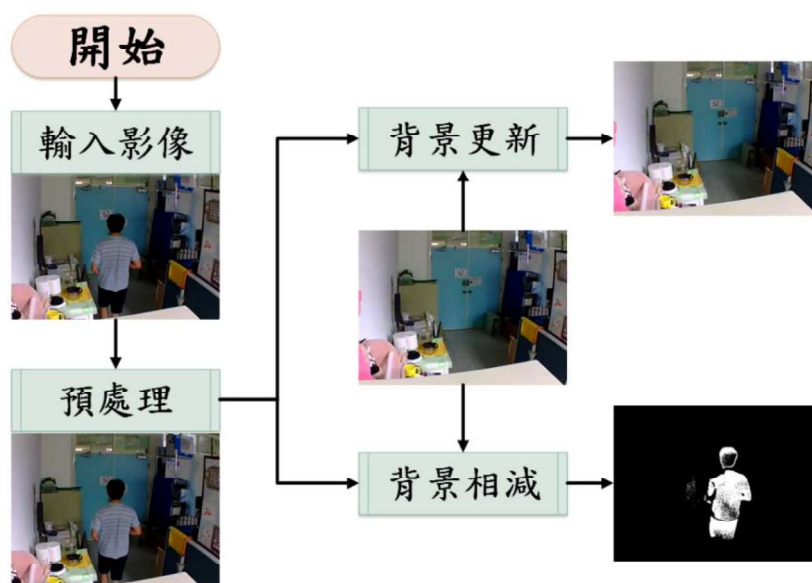


圖 14.位移偵測流程圖

3.2.2 動態物件框選模組

物件有時會受到光影影響，缺少一些資訊，必須將物件進行侵蝕與膨脹後再偵

測該物件輪廓，最後使用感興趣框選出物件，其流程如圖 15 所示。

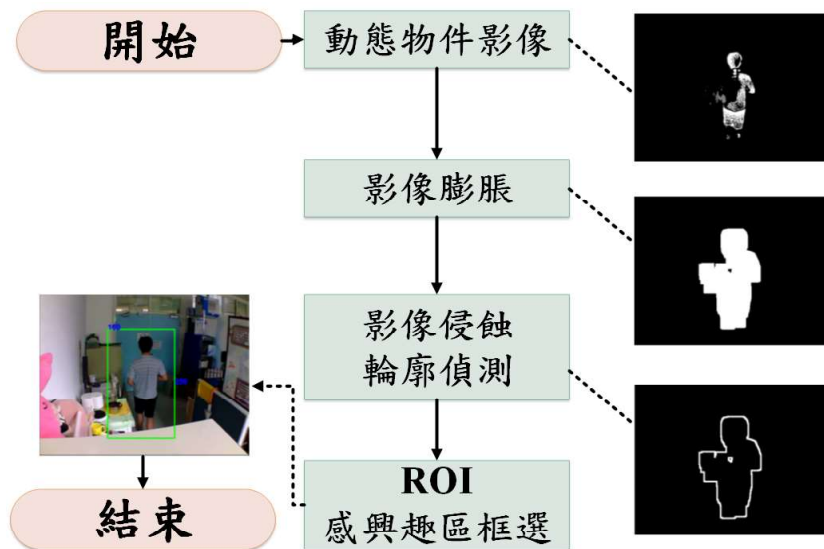


圖 15.動態物件框選流程圖

3.2.3 行人資料集

前兩者模組將動態物件進行影像擷取，產生許多影像，但在這些影像當中有些因為光影的關係誤框，或是框選的影像不是行人，必須先手動篩選掉一些不良影像，以利卷積神經網路學習。

3.3 卷積神經網路人物影像偵測

本模組分為兩階段，第一階段為訓練階段，使用行人資料集篩選過影像，將行人影像導入卷積神經網路進行訓練，其訓練流程如圖 16 所示。

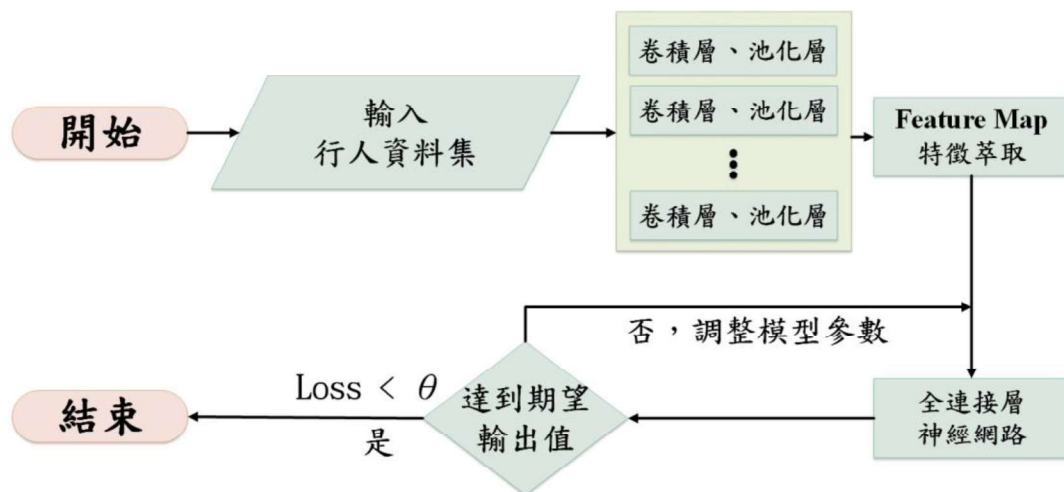


圖 16.深度學習模組流程圖(訓練階段)

當訓練完畢後，進行第二階段—使用階段，當有新的影像輸入時，透過已訓練之卷積神經網路判斷該影像是否為人，若判斷為

人，就進入行人識別，否則過濾該影像，其流程如圖 17 所示。

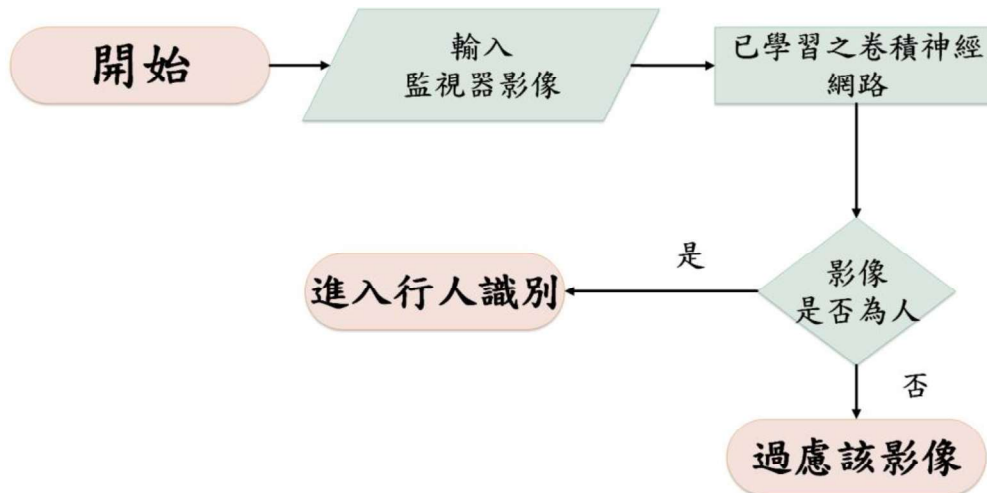


圖 17.深度學習模組流程圖(使用階段)

3.4 人物影像識別

由於人物遠近會產生行人亮度與顏色的偏差，先框選一個合適的範圍，當行人影像進入該範圍時，再進行 HSV、RGB 直方圖轉換，萃取出行人身體上的顏色特徵，

進而進行資料比對，如果該行人未出現過在資料庫中則判斷為新行人，其識別流程如圖 18 所示。

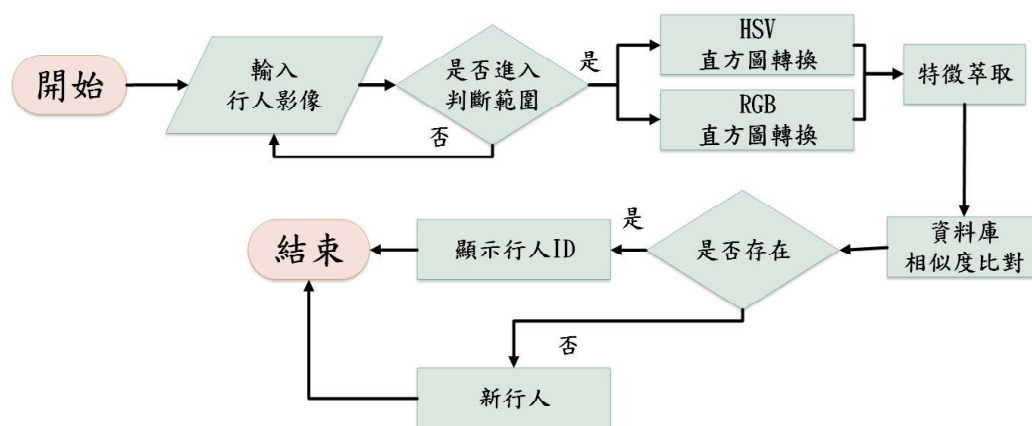


圖 18.行人識別-人臉辨識流程圖

3.5 智慧型人物影像追蹤

將行人識別判斷為新行人的影像，存進資料庫中，並記錄時間標記，透過系統

查詢可查詢該行人出現時間、結束時間以及出現次數，其流程如圖 19 所示。

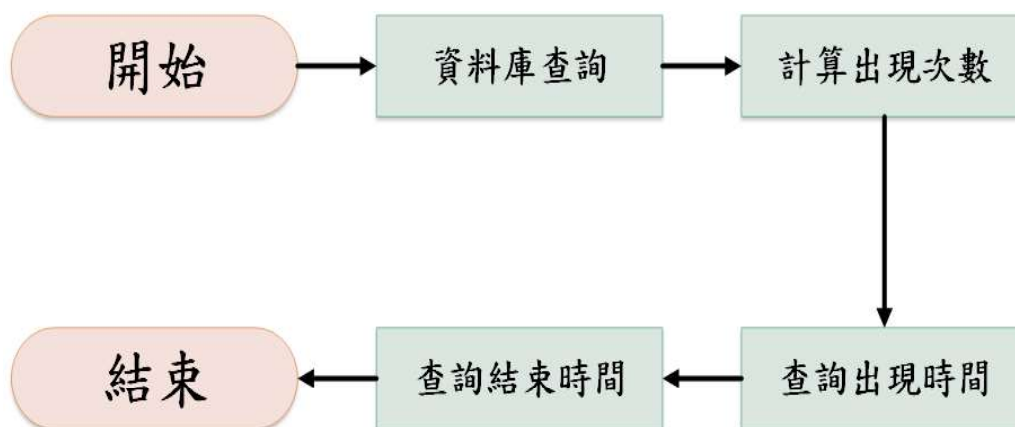


圖 19.智慧搜尋人物流程圖

四、實驗結果與分析討論

4.1 實驗環境

本研究利用卷積神經網路與影像辨識技術追蹤監視器行人，實驗場域為屏東科技大學管理學院實驗室內門口，使用監視器為 Dlink DCS-5020L，其系統環境之作業平台為 Window10 作業系統，使用處理器為 Intel(R) Xeon(R) CPU E5-2630 v4 2.20 GHz(2 個處理器)，隨機存取記憶體(RAM)為 32GB，顯示卡(GPU)分別為 Nvidia GTX 1080ti(11GB)與 Nvidia GTX 950(2GB)交叉使用，環境為 Cuda9.0 與 cuDNN7.0.5。程式使用 Anaconda 5.1(Python 3.6)，並使用 opencv-python 3.4.0.12 與 tensorflow-gpu 1.6.0 套件進行開發。

4.2 系統展示

本系統設計共四大功能，包含資料蒐集、卷積神經網路人物影像偵測、人物影像識別、智慧型人物影像追蹤。前兩者系統功能主要將監視器影像中所有行人畫面進行擷取動作，配合手動篩選出合適影像後，進行卷積神經網路的訓練；後兩者系統功能使用影像辨識技術將行人影像區分，進而給予每個行人特有 ID，使調閱監視器更加方便，其系統畫面如圖 20 所示。



圖 20.主系統畫面

4.2.1 資料蒐集

將背景相減動態物件影像資料集手動篩選後，其影像分別為 Train:252、Test:368 張影像，資料集如圖 21 所示。

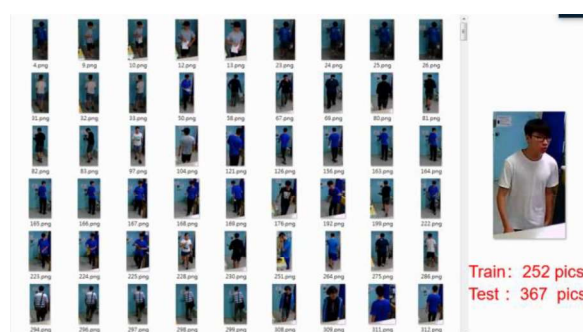


圖 21.行人資料集

4.2.2 卷積神經網路人物影像偵測(訓練)

使用 Tensorflow 進行訓練，其網路架構由 6 層卷積層建構，訓練次數為 200000 次，Batch_Size 為 24，訓練時畫面如圖 22 所示。

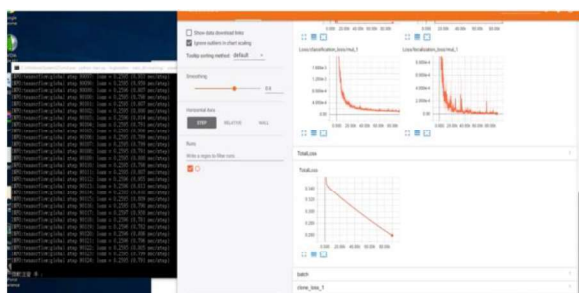


圖 22. Tensorflow 訓練畫面

4.2.3 卷積神經網路人物影像偵測(使用)

輸入監視器影像，使用卷積神經網路框選出行人部分，如圖 23 所示。

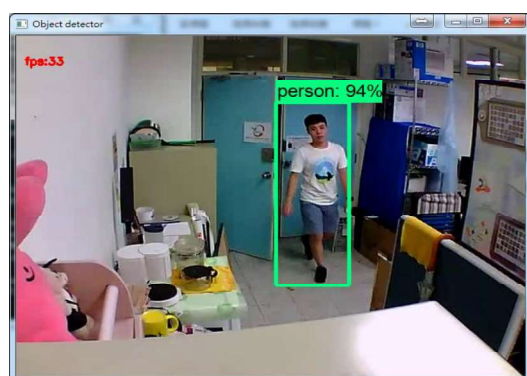


圖 23.卷積神經網路框選行人

4.2.4 人物影像識別

由於人物遠近會影響顏色與亮度的判斷，必須先框選一個固定範圍，當行人進入時，再進行行人識別動作，系統畫面如圖 24、圖 25。

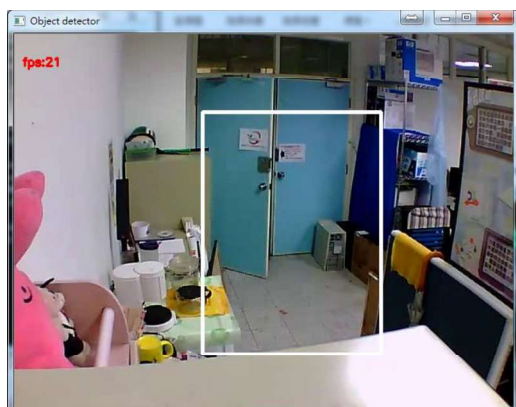


圖 24.框選範圍

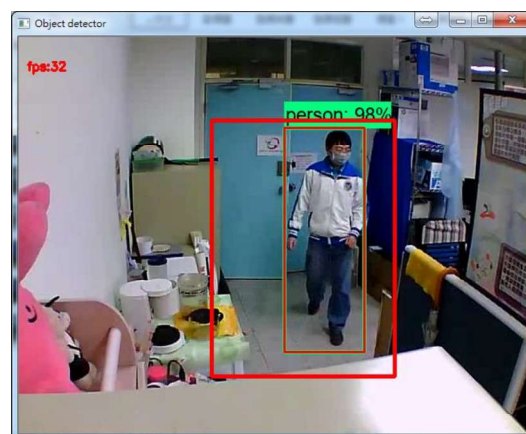


圖 25.行人進入範圍內開始識別

當行人進入範圍後，將進行 HSV 與 RGB 直方圖轉換，並開始比較資料庫內有無相似行人，其轉換圖如圖 26 所示。

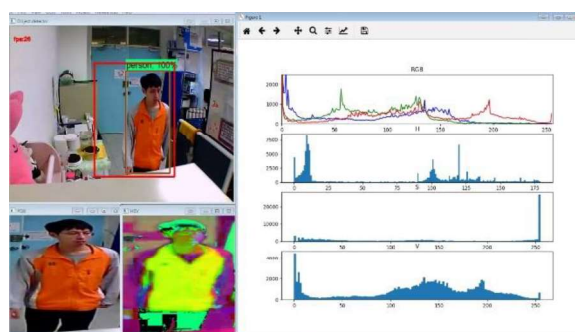


圖 26.直方圖轉換

當 RGB 與 HSV 相似度皆低於 0.75 時，判斷為新行人，增加行人 ID 與該行人出現時間並存入資料庫中，如圖 27 所示。

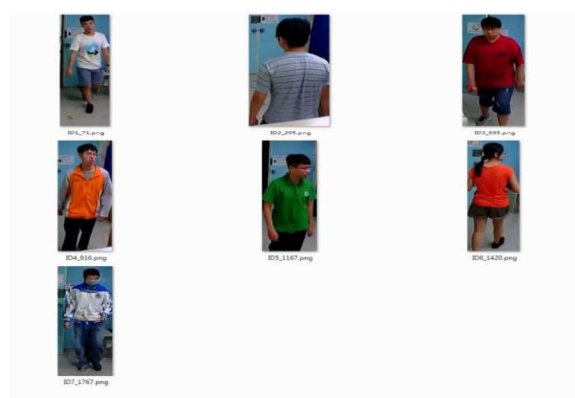


圖 27.行人 ID 資料庫

4.2.5 智慧型人物影像追蹤

系統對應行人 ID 與出現時間，系統右邊會出現所有行人按鈕，當點選按鈕時影片會跳轉至該行人出現片段，使調閱監視器者更加方便，如圖 28 所示。



圖 28.智慧搜尋人物系統

4.3 實際結果分析

依據本研究實驗設計與步驟流程，使用監視器拍攝影像導入本研究所開發之系統，下面將逐一分析，卷積神經網路行人框選結果與行人識別結果。

4.3.1 卷積神經網路人物影像偵測結果分析

導入本系統之測試影片總張數為 1912，實際行人出現總張數為 644，系統框選行人總張數為 613，其正確率約為 95%。

研究發現，系統未正確框出行人，是因為行人剛出現時，完整資訊並未完整呈現出來，導致系統沒有框出行人，如圖 29 所示，當行人資訊完整呈現後，系統即可框選出行人，如圖 30 所示。

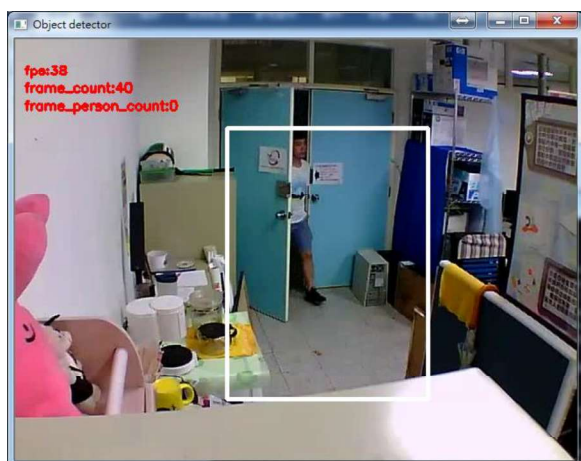


圖 29.系統未框選行人

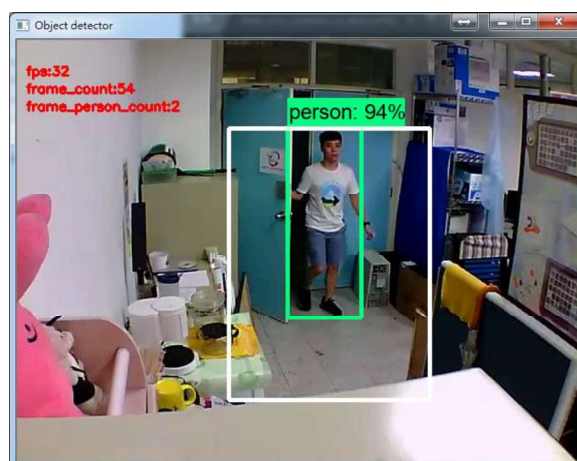


圖 30.系統框選行人

4.3.2 人物影像識別結果分析

本研究使用 RGB、HSV 兩個色彩空間進行行人識別，當相似度門檻值小於 0.75 時判斷為新行人，由於行人遠近會讓辨識產生偏差，所以增加一個固定範圍，當行人進入範圍內才進行辨識。

當系統沒有增加固定範圍時，行人總張數為 609 張，正確辨識行人 ID 總張數為 495 張，正確率為 81.2%。

當系統增加固定範圍後，行人總張數為 295 張，正確辨識行人 ID 總張數為 281 張，正確率為 95.5%。

五、結論與未來研究方向

5.1 結論

本研究基於卷積神經網路之智慧型人物影像追蹤系統，調閱監視器者使用本研究開發之系統，可智慧搜索監視器影片中所有人物與出現之時間點，減少調閱時重複觀看影片尋找可疑人物之麻煩，透過智慧搜尋人物搜尋同一個人物出現過幾次，與每次出現之時間點。

5.2 未來研究方向

深度學習與電腦視覺結合的應用越來越多，本研究以色彩空間為基礎來識別行人，未來期望能做到在行人識別領域透過

深度學習模型判斷是否為新的行人，而不是用相似度門檻值來進行判斷，或是加入人臉資訊作雙重判斷。

5.3 研究範圍與限制

本研究所設計之基於卷積神經網路之智慧型人物影像追蹤系統，主要有一個限制：行人識別以 RGB 與 HSV 色彩空間為基礎，雖然已設定範圍來減少誤判率，但當有無可避免之因素時，會造成誤判，如：實驗室光線昏暗(關燈狀況時)，因此不列入研究範圍。

參考文獻：

- [1] S.Mansur, M.Habib, G. N. P.Pratama, A. I.Cahyadi, andI.Ardiyanto, “Real time monocular visual odometry using Optical Flow: Study on navigation of quadrotors UAV,” 2017 3rd Int. Conf. Sci. Technol. - Comput., pp. 122–126, 2017.
- [2] S.Murali and R.Girisha, “Segmentation of motion objects from surveillance video sequences using temporal differencing combined with multiple correlation,” 6th IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2009, pp. 472–477, 2009.
- [3] H. S.Amaresh, Y. G. A.Rao, andR. S.Hallikar, “Real-Time Intruder Detection System Using Sound Localization and Background Subtraction,” 2014 Texas Instruments India Educ. Conf., pp. 131–137, 2014.
- [4] S.Kumar and J.SenYadav, “Video object extraction and its tracking using background subtraction in complex environments,” Perspect. Sci., vol. 8, pp. 317–322, 2016.
- [5] S. J.Mckenna, “Tracking Groups of people,” pp. 1–15.
- [6] P.Viola andM. J.Jones, “Robust real-time face detection,” Int. J. Comput. Vis., vol. 57, no. 2, pp. 137–154, 2004.
- [7] A. K.Troitsky, “Two-level multiple face detection algorithm based on local feature search and structure recognition methods,” Int. J. Appl. Eng. Res., vol. 11, no. 6, pp. 4640–4647, 2016.
- [8] M.Farenzena, L.Bazzani, A.Perina, V.Murino, andM.Cristani, “Person re-identification by symmetry-driven accumulation of local features,” Comput. Vis. Pattern Recognit. (CVPR), 2010 IEEE Conf., pp. 2360–2367, 2010.
- [9] L. B. and V. M.Dong Seon Cheng, Marco Cristani, Michele Stoppa, “Custom Pictorial Structures for Re-identification,” Br. Mach. Vis. Conf., p. 68.1--68.11, 2011.
- [10] L.Bazzani, M.Cristani, A.Perina, andV.Murino, “Multiple-shot person re-identification by chromatic and epitomic analyses,” Pattern Recognit. Lett., vol. 33, no. 7, pp. 898–903, 2012.
- [11] R. A.Kharjul, V. K.Tungar, Y. P.Kulkarni, S. K.Upadhyay, andR.Shirsath, “Real-Time pedestrian detection using SVM and AdaBoost,” Int. Conf. Energy Syst. Appl. ICESA 2015, no. Icesa, pp. 740–743, 2016.
- [12] 陳建霖(2013)，智慧型影像監控老人跌倒偵測，國立臺北大學資通科技產業碩士專班，碩士論文
- [13] 陳信銘、陳佳珮、邱柏訊、林多常、簡大為(2016)，中華電信研究院企客方案研究所，電工通訊季刊 2016 第 4 季 (2016/12)，14-23
- [14] 范家榕(2015)，觀光景點人潮及車流估

- 計之智慧型監控系統，國立屏東科技大學資訊管理系所，碩士論文
- [15] 邱柏訊、郭萱聖、陳信銘、林多常、簡大為、張慶年(2016)，中華電信研究院企客方案研究所，電工通訊季刊 2016 第 4 季 (2016/12)，33-45
- [16] K. B.Shaik, P.Ganesan, V.Kalist, B. S.Sathish, and J. M. M.Jenitha, "Comparative Study of Skin Color Detection and Segmentation in HSV and YCbCr Color Space," *Procedia Comput. Sci.*, vol. 57, pp. 41–48, 2015.
- [17] X.Zhu, R.Chen, H.Xia, and P.Zhang, "Shadow removal based on YCbCr color space," *Neurocomputing*, vol. 151, no. P1, pp. 252–258, 2015.
- [18] C.Stauffer and W. E. L.Gimson, "Adaptive background mixture models for real-time tracking," *Proc. 1999 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Cat No PR00149*, vol. 2, no. c, pp. 246–252, 1999.
- [19] P.Kaewtrakulpong and R.Bowden, "An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection," *Adv. Video Based Surveill. Syst.*, pp. 1–5, 2001.
- [20] Z.Zivkovic and F.Van DerHeijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognit. Lett.*, vol. 27, no. 7, pp. 773–780, 2006.
- [21] N.Dalal and W.Triggs, "Histograms of Oriented Gradients for Human Detection," 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. CVPR05, vol. 1, no. 3, pp. 886–893, 2004.
- [22] J. R. R.Uijlings, K. E. A.Van DeSande, T.Gevers, and A. W. M.Smeulders, "Selective Search for Object Recognition," 2012.
- [23] H.Liu, L.Ma, and C.Wang, "Body-structure based feature representation for person re-identification," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2015–Augus, no. 2006, pp. 1389–1393, 2015.
- [24] F.Rosenblatt, "The Perceptron - A Perceiving and Recognizing Automaton," Report 85, Cornell Aeronautical Laboratory. pp. 460–1, 1957.
- [25] A.Krizhevsky, I.Sutskever, and G. E.Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2012.
- [26] D. H.Hubel and T. N.Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, p. 106–154.2, 1962.
- [27] R.Girshick, J.Donahue, T.Darrell, and J.Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 580–587, 2014.
- [28] J.Long, E.Shelhamer, and T.Darrell, "Fully convolutional networks for semantic segmentation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 7-12-NaN-2015, pp. 3431–3440, 2015.
- [29] K.Kang and X.Wang, "Fully Convolutional Neural Networks for

- Crowd Segmentation,” 2014.
- [30] Y.Sun, X.Wang, andX.Tang, “Deep Learning Face Representation by Joint Identification-Verification,” pp. 1–9, 2014.
- [31] X.Wu, R.He, Z.Sun, andT.Tan, “A Light CNN for Deep Face Representation with Noisy Labels,” 2015.
- [32] C.Gao, P.Li, Y.Zhang, J.Liu, andL.Wang, “People counting based on head detection combining Adaboost and CNN in crowded surveillance environment,” *Neurocomputing*, vol. 208, pp. 108–116, 2

An Intelligent Facial Image Tracking System Based on Convolutional Neural Network

Yu-Hao Lin ¹, Yuh-Jiuan Tsay ²

**¹ Department of Management Information Systems, National Pingtung University of
Science and Technology, Taiwan**

**² Department of Management Information Systems, National Pingtung University of
Science and Technology, Taiwan**

Abstract

Although the storage device of the monitor can record historical images, but it is a very time-consuming task for retrieving specific facial image from historical image database. This study will thus be design to set up a system for "An Intelligent Facial Image Tracking System Based on Convolutional Neural Network ", which includes: (1) detect of facial image, using Grayscale and Gaussian Blur model to filter noise, Gaussian Mixture model to detect foreground, Erosion and Dilation algorithms to detect dynamic object contours, and deep learning techniques to detect facial image; (2) recognition of facial image, using RGB and HSV color models to recognize face position, and record the time of appearance; (3) track of facial image, using Convolutional Neural Network to learn and track specific facial image. The experiment results show that the proposed system can detect and track facial image with high accuracy.

Key words: Image recognition, Pedestrian recognition , Deep learning , Pedestrian tracking