

## 人工智慧之校園安全影像辨識機器人

王士豪，東吳大學巨量資料管理學院

### 一、簡介

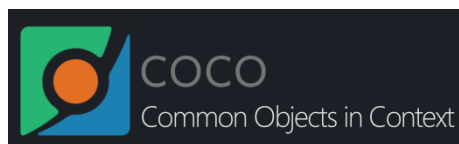
本計畫主要目的是想解決校園安全問題，校園安全關乎到所有在此活動的人的健康與安危，如果能在第一時間提醒教職員或學生所處的危險情境，勢必能夠避免人員發生傷害，進而提升校園的安全。

在校園中，透過安裝本計畫的人工智慧技術辨識機器人，試圖找出可能存在的危險動作，透過即時的文字與語音反饋，提醒現場特定人員相關警示與防護用語，而非標語式的宣導。

正因人工智慧技術在圖像辨識及自然語言生成有著重大的突破，因此，可降低校園災害的發生，一個能夠及時且直接的安全守護機器人，勢必能大幅降低人員處於危險之處。

### 二、使用工具介紹

#### (1) 資料來源



為了使本計畫所提出的機器人能夠學習相關的危險場景，本計畫採用 Microsoft 所提供的公開資料 COCO 2017 dataset 為主要的訓練資料集，其中包含 118K 張照片及 591K

句的照片場景描述內容。過濾出以上凡是標題或照片元素中擁有災害發生前的特徵資料，以利能夠訓練機器辨識跟安全有關的場景。

#### (2) CNN 模型

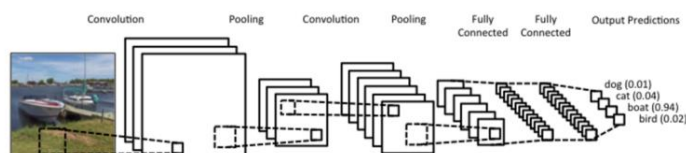


圖 1 CNN 流程圖

卷積神經網絡，原文名為 Convolutional Neural Network 簡稱 CNN，CNN 在影像辨識的技術層面上特別成熟，是目前在市面上最受歡迎的影像辨識架構之一，而 CNN 的架構主要分為卷積層、池化層及全連接層。

卷積層(Convolution Layer)就是將原本圖片與  $3 \times 3$  的特徵向量矩陣做相乘後再相加後，得到一個暫存的特徵向量矩陣(Feature Map)，其中將會使用 Relu 函數，將矩陣中的負值改為零，使物體的形狀，更加明顯。

池化層(Pooling Layer)，透過 Max Pooling 函式，將 Feature Map 劃分為若干個  $2 \times 2$  的矩形區域，並輸出每個區域

的最大值為新的特徵向量矩陣 (Pooled Feature Map)。

最後透過全連接層 (Fully Connected Layer)，將 Pooled Feature Map 轉為一維向量，屆時能夠連接到基本的神經網絡或不同的模型，做後續的演算開發及訓練，而本計畫就將 Fully Connected Layer 取代為 FCN 模型使用。

### (3)FCN 模型

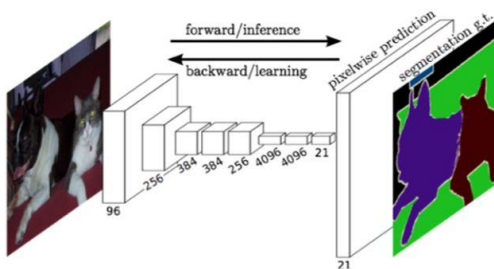


圖 2 全卷積網絡示意圖

全卷積網絡 (Fully Convolutional Networks)，簡稱 FCN，FCN 可以接受任意尺寸的輸入圖像，而將 Pooled Feature Map 恢復到輸入圖像的相同尺寸，從而可以對每個像素都產生了一個預測，同時保留了原始輸入圖像中的空間信息，最後再採樣的特徵圖上進行逐像素分類。

### (4)LSTMA 模型

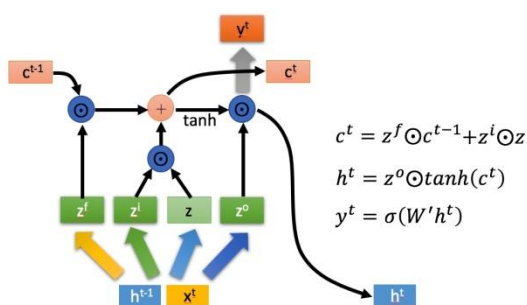


圖 3 LSTM 主要形式

長短期記憶網絡 (Long Short Term Memory Network with attention mechanism)，簡稱 LSTMA，LSTMA 模型在設計上避免了長期依賴的問題，為一種特殊的循環神經網絡 (RNN)，能夠學習到長期依賴關係。

### (5)Pytorch 套件



Pytorch 為一個開源的 Python 機器學習框架，應用於人工智慧領域，最初由 Facebook 的人工智慧研究團隊所開發。

### (6)CKIP 斷詞



由於本機器人會及時提醒現場特定人員相關警示與防護用語，而我們將使用 CKIP 的斷詞系統，達到自然語言生成之目的。

CKIP 為台灣中研院開發的中文斷詞系統，具有未知詞辨識能力及附加詞性標註功能。

### 三、研究方法

本計畫是為了使用人工智慧與資料科學技術來實現校園安全防護模型，主要模型訓練及運作流程為辨識圖像中的場景及生成場景相對應的安全防護相關用語。

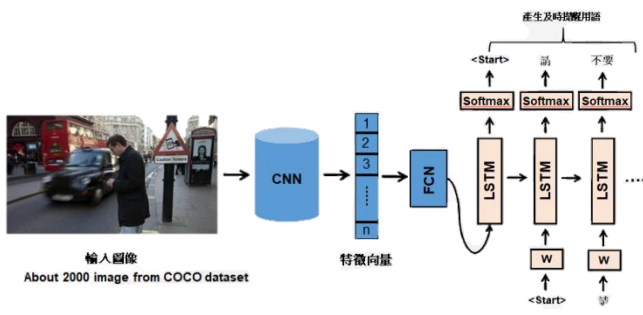


圖 4 機器人運作流程

利用人工智慧的深度學習技術，其中包含影像辨識模型及自然語言生成模型，主要是識別教職員生在校園中是否有出現危險徵兆。其中包含 CNN 模型辨識圖像及 LSTMA 模型生成自然語言，而使用這兩個模型可以大幅解決本計畫所也解決的校園安全防護任務。在工具上以套件 Pytorch 深度學習套件作為主要的人工智慧技術，以實現上述兩個主要模型。

由於本機器人涉及到圖像與文字資料，意味著需要大量的圖像與場景描述文字來訓練機器，本計畫將採用上述所提及的 COCO dataset 作為主要訓練資料。利用人工智慧技術學習圖片與場景描述詞間的關係，大量的圖片與句子可以讓機器人從資料中理解到何謂危險場景，可創造出一個校園安全守護機器人。

本計畫使用 CNN 模型來提取特徵向量，此特徵向量能辨識給定的圖像中的場景描述，然而此特徵向量將會掌握此場景的重要情景，舉凡下樓梯正在滑手機若有偵測到類似此一場景的圖像，將會接續使用 LSTMA 模型針對此場景特徵向量產生場景相對應的安全防護相關用語，如「**那位同學，下樓梯請勿滑手機，否則跌倒時，我會不捨！**」

最終選擇參數 Epoch 設為 40 訓練出的模型，此模型的 Loss function 為 0.0241 及 Perplexity 為 1.0244，做為本次機器人的最初樣品。

### 四、結果

透過上述的模型來做第一次的測試，於是先從訓練資料集隨機挑一張看似具有危險的照片做測試圖片，當使用者輸入圖 5-1，此時模型會產出「**同學，不要在這裡玩滑板，非常危險！**」。

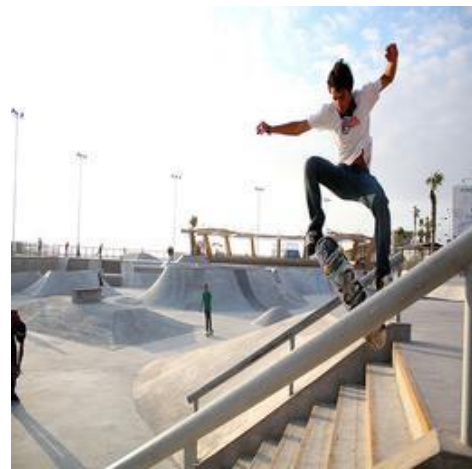


圖 5-1

此時進行第二次測試，一樣從訓練資料集隨機挑出肉眼看似安全的照片做為測試圖片，當使用者輸入圖 5-2 時，此時模型會產出「<n>」，當模型認為圖片不具危險時，就會產出<n>。



圖 5-2

重複挑選不同的訓練集進行測試後，發現擁有一定的準確度後，於是便開始在校園中拍攝具有危險及不危險的照片做為測試圖片，當輸入圖 5-3 時，照片場景中為「一個人坐在椅子上，拿著手機向前方拍照」，最後，模型會產出「<n>」，告訴使用者這個場景不具有危險性，因此不做任何防護用語的提醒。



圖 5-3

然而，挑選圖 5-4 做測試圖片時，圖中為「一位同學在樓梯間使用手機」，當將此圖片輸入至模型中，模型則會產出「**正在講電話的同學，請注意前面樓梯！**」。



圖 5-4

## 五、結論

此模型還有許多部份仍可以繼續再改進，舉例如下：

### (1) 圖片來源的多樣性

因本次資料來源都來自公開圖片庫，因此有著各式各樣的圖片，但其中具有危險的照片量過少，導致模型的準確度相對沒那麼高，屆時再訓練時，如果都是具有危機因素的照片時，此模型的準確度，一定能大幅提升。

### (2) 自然語言生成

此次的警示與防護用語，僅針對挑選出來的圖片，做人工撰寫，若能在圖片與警示與防護用語間再加一層 LSTM 模型，針對此場景特徵向量產生的文字描述，此時的提醒用語將更像人類般的即時回饋，達到提醒及警示的作用，降低災害發生。



### (3)混合使用不同模型

為了達到成精準的場景辨識與自然語言生成，若混和使用不同的熱門影像辨識模型，如

AlexNet、ResNet、Inception v3 等模型進行反覆訓練此機器人，將能夠使機器的判斷日漸純熟，以增加判斷與說話的穩定度。根據相關文獻指出，ResNet 模型擁有殘差塊的結構，造成模型中擁有較多分析層，使訓練誤差較小，故相較於其他辨識模型更加精準，所以 ResNet 模型為本次研究主要選擇的圖像辨識模型之原因。

若以上敘述加以改善後，人工智慧監測器感知同學們在有危機前兆時，分析當時場景狀況，除了立即做出安全提醒用語外，亦可產生出更具臨場感之安全防護提醒用語，透過機器人幽默的口吻，讓同學的注意力轉移至此機器所傳述的事情上，有助於降低校園災害的發生，使校園成為更加安全的學習場所。

此次訓練的模型皆已校園安全為主軸，但是其技術也能應用在各行各業，例如：辨識工地的工人們，是否都有依規定配戴安全帽或安全配備；充當初學者的健身教練，當它感測使用者在操作器材時，若是動作不正確，會立即提醒或講述一些簡單的動作教學；同事匆忙上班卻忘了打卡，若公司安裝此感測器，就能夠立即提醒那位同事，使公司能減少後面補打卡的繁瑣程序。

因此，有此機器的出現，我們團隊相信，除了能使校園變得

更加安全外，若是能夠依照不同場域巧妙結合相關機器，勢必能夠對社會增添一些福祉。

### 六、參考文獻

1. Yeh James, 卷積神經網絡介紹  
<https://medium.com/@yehjames/%E8%B3%87%E6%96%99%E5%88%86%E6%9E%90-%E6%A9%9F%E5%99%A8%E5%AD%B8%E7%BF%92-%E7%AC%AC5-1%E8%AC%9B-%E5%8D%B7%E7%A9%8D%E7%A5%9E%E7%B6%93%E7%B6%B2%E7%B5%A1%E4%BB%8B%E7%B4%B9-convolutional-neural-network-4f8249d65d4f>
2. wiki, 卷積神經網路  
<https://zh.wikipedia.org/wiki/%E5%8D%B7%E7%A7%AF%E7%A5%9E%E7%BB%8F%E7%BD%91%E7%BB%9C>
3. 陳仲銘, 全卷積網絡詳解  
<https://zhuanlan.zhihu.com/p/30195134>
4. 中研院, 中文斷詞系統  
<http://ckipsvr.iis.sinica.edu.tw/>
5. 李宏毅教授, 深度學習  
<https://zhuanlan.zhihu.com/p/32085405>
6. CKIP Lab, 中研院中文斷詞系統  
<https://www.google.com/url?sa=i&source=images&cd=&ved=2ahUKEwj-hb3dtbrfAhUD6bwKHQmYCdAQjRx6BAGBEAU&url=http%3A%2F%2Fckip.iis.sinica.edu.tw%3A8080%2Flicense%2F&psig=A0vVaw3tYBagrTEda97Y6FCTVuDx&ust=1545807872571218>