

使用 FPN-PAN 架構的 CNN 自動化生態監測之鳥類辨識系統

Automated bird recognition system for ecological monitoring using FPN-PAN architecture CNN

鄭政文 Cheng-Wen Cheng, 吳俊傑 Jun-Jie Wu, 許哲晟 Che-Sheng Hsu, 丁襄龍 Xiang-Long Ding

國立高雄科技大學

資訊工程系

Department of Computer Science and Information Engineering

National Kaohsiung University of Science and Technology

Kaohsiung, Taiwan, R.O.C

{c109151104, c109151150, c109151162, c109151163 }@nkust.edu.tw

1 前言

近年來，隨著科技的不斷進步，深度學習技術在各個領域都展現出了巨大的應用價值。雖然已經取得了一定成果，但在生態監控方面仍然存在著較大的挑戰。傳統的生態監控方法主要依賴人工處理，例如通過架設攝影機來捕捉生物的影像，然後依靠人類觀察影片以判斷其中的生物種類。這種方法存在著效率低下、人力時間成本高昂，並且容易受到主觀因素的影響，導致辨識結果可能不夠一致。透過機器辨識，我們可以節省這些成本，並提高效率。機器辨識的準確性也較高，因為機器不受情緒和疲勞的影響，而且可以實現大範圍、長時間的監測，即使在偏遠或困難進入的地區也能進行監測，這樣就可以獲取更全面、準確的鳥類資料。其次，大量的鳥類數據需要處理和分析，這需要耗費大量時間和精力。透過機器學習，可以加速這一過程，提高處理效率。這樣我們就能更快地獲取、分析和應用鳥類數據，做出相應的決策和行動，對於保護生態環境也是極為重要的。鳥類在生態系統中扮演著重要的角色，它們的分佈和行為反映了生態系統的變化和健康狀況。通過監測鳥類，我們可以了解生態系統的健康狀況，進而保護重要的生態環境和物種。

我們的研究旨在開發一套鳥類辨識系統，以提高鳥類監測效率。通過增加訓練數據和引入深度學習技術，我們致力於提高系統的準確性和可靠性，確保監測結果的準確性。我們希望將這項技術應用於更廣泛的場景，

包括自然保護區的監測、外來種控制和農業害鳥防治等，提升其社會價值。通過更準確的鳥類監測，我們可以更好地感知生態環境變化，為生態保育提供更有力的支持，並提供相應的保育建議，增加對生態保育的關注度。我們的目標是建立長期的鳥類監測體系，持續追蹤鳥類群落變化，為生態保育政策提供科學依據，並進行生態系統健康狀態的長期監測。透過此次專案，我們希望開發易於操作且具有教育價值的鳥類辨識工具，普及鳥類知識、引導環境保護觀念，增加公眾對自然生態和生物多樣性的關注和了解。

2 相關研究

YOLO(You Only Look Once)(Redmon, Divvala, Girshick, & Farhadi, 2016)是一種單次前向傳播的卷積神經網路，它以高效率 and 快速性著稱。與傳統的目標檢測方法相比，如滑動視窗或候選區域的生成（如 R-CNN 系列），YOLO 將整個目標檢測過程簡化為一個單一的回歸問題。

YOLO 將圖像分成多個網格單元，每個單元負責預測對應的目標框和框中的目標類別。這意味著在一次前向傳播中，YOLO 可以直接預測圖像中的所有目標框和對應的類別概率。由於這種端到端的訓練方式，YOLO 具有非常高的處理速度，使其能夠實時進行目標檢測。然而，雖然 YOLO 在處理速度和簡化流程方面表現出色，但在檢測小目標和密集目標時效果不如其他算法。對於特別小或特別大的目標，可能存在準確性問題。

YOLOv8 的結構(圖 1)由三個部分組成，其中骨幹網路(Backbone)負責多尺度卷積特徵擷取，主要作用是捕捉圖像中的細節，其 stage 2 會輸出低階特徵、stage 3 會輸出中階特徵、stage 4 中會經過 SPPF 後輸出高階特徵；頸部(Neck)目的是多尺度卷積特徵融合，呈現 FPN 和 PAN 結構，其中 FPN (feature pyramid networks)即是特徵金字塔網路，採用多尺度來對不同大小的目標進行檢測；PAN 為自底向上的特徵金字塔。這樣結合操作，FPN 層自頂向下傳達強語意特徵，而特徵金字塔則自底向上傳達強定位特徵，從不同的主幹層對不同的檢測層進行特徵融合；頸部連接到頭部(Head)共有三個路徑，分別藉由不同階層的特徵來偵測圖片中的小、中、大型物件，而在 v8 版本中有助解耦合頭(decoupled-head)

來個別偵測物件的位置和物件類別，來增加偵測的準確性。

- PAN-FPN：CV 中物件辨識的神經網路架構，它將特徵金字塔網路(FPN)與路徑聚合網路(PAN)結合，以提高物件辨識的準確性和效率；FPN 用於不同比例的影像特徵提取，PAN 用於跨網路的不同層聚合特徵。這允許網路檢測不同大小和分辨率的對象，並能處理具有更多對象的複雜場景。
- CSPLayer_2Conv：簡稱 C2F，透過跨層部分連接機制，在不同層之間建立直接連接，即是在卷積特徵始過 ConvModule 之後會被分成一半，一部分經過 n 個 DarknetBottleneck 進行深度的 CNN 運算，另一部分會跳過，並在最後會融合

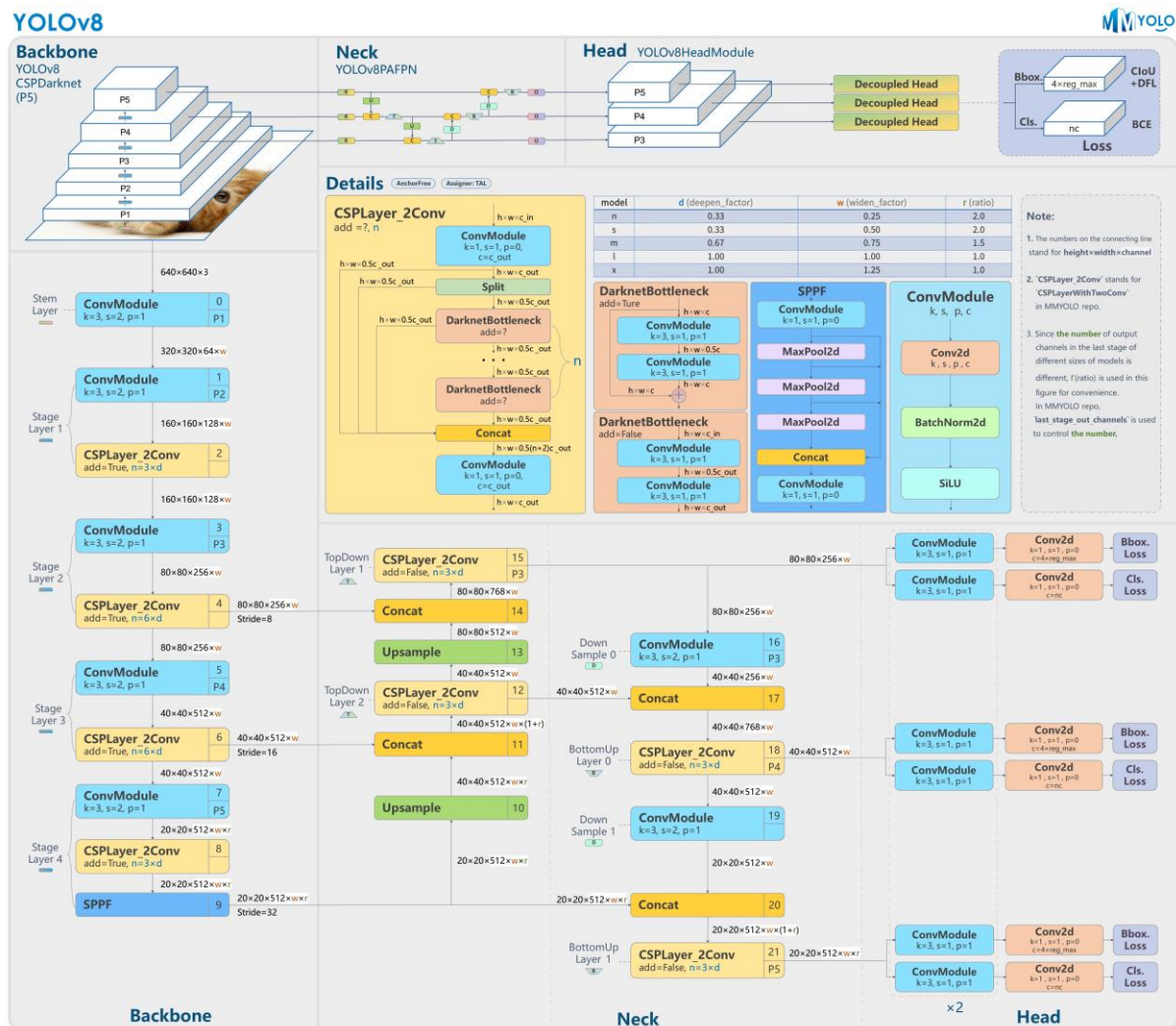


圖 1. YOLOv8 架構圖

起來再進經過 ConVModule，這樣可提高網路資訊流動性和特徵表達能力。

- DarknetBottleneck：該模塊包含兩個 ConVModule，並且有兩種類型，在 add = True 時會使用殘差連接，使其解決深度神經網路訓練過程中的梯度消失和梯度爆炸等問題，從而實現更深的網路架構。
- ConVModule：其包含一個二維的卷積層(Conv2d)、一個批次正規化、一個 SiLU 激勵函數，即是一個卷積塊。
- Upsample：用於在特徵金字塔網路(FPN)或解碼器部分增加特徵圖的尺寸，從而實現對不同尺度目標的檢測。常用得的方法有：最近鄰插值(Nearest neighbor interpolation)、雙線性插值(Bi-Linear interpolation)、雙立方插值(Bi-Cubic interpolation)、三線性插值(Trilinear Interpolation)、反池化(UnPooling)、反卷積(Deconvolution)。
- SPPF：其前身為 SPP，即空間金字塔池化，SPP 的主要思想是在不同大小的子區域上應用池化操作，然後將所有子區域的池化結果結合起來。這使得模型能夠處理任意大小的圖像，而不需要將它們調整為相同大小，可以避免發生固定池化層大小使其偵測尺度不容易調整的

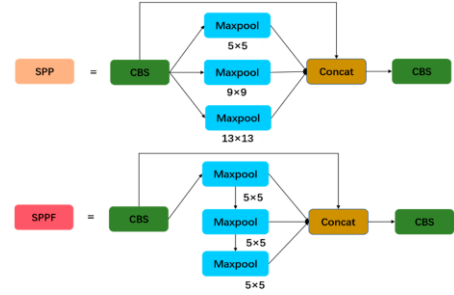


圖 2. SPP 和 SPPF 比較圖

問題。SPP 的一個主要缺點是計算量較大，因為它需要在不同大小的子區域上應用池化操作。而 SPPF 的理念是藉由小型池化層來取代大型池化層，像是(圖 3) K=13 可用三個 K=5 池化層取代，這樣可以減少運算量。

3 方法

本研究如圖 3 所示提出了一個系統架構，使用者可在 Web/Mobile 端註冊登入後，上傳圖片以取得鳥類辨識結果，且可在 BirdMap 中觀看曾經辨識過的鳥類資訊，如種類、拍照時間、拍照地點等，並可對 LLM 對話機器人提出鳥類與動保法相關疑問，來取得及時性地回答。Server 端的 YOLO 辨識模型訓練流程如圖 4 所示，首先，會從網路上取得不定數量的

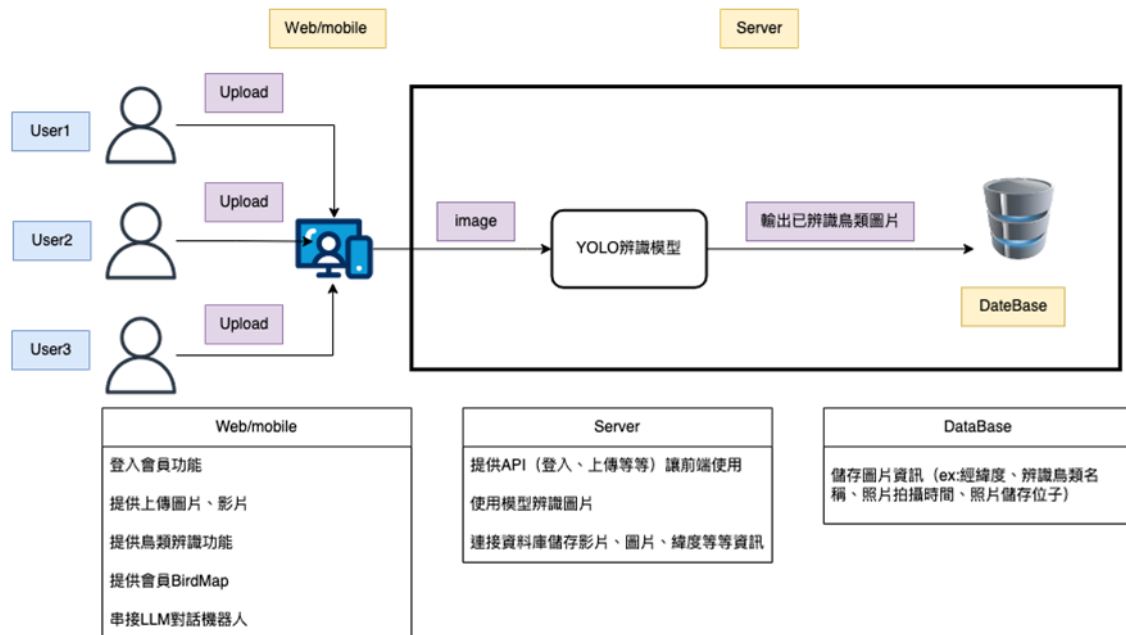


圖 3. 本研究所提系統架構

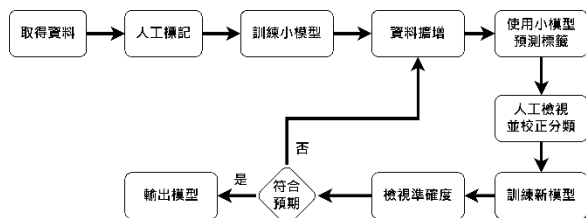


圖 4. 模型訓練流程

數種鳥類的圖片，並人工對照片中所有鳥類圖像如圖 5 所示進行框選並標記類別，接下來，將會訓練用於幫助標記資料的小模型，將模型訓練成能夠有效框選出鳥類圖像的能力後，會再將資料類別與數量進一步擴增，以小模型對擴增資料框選後再人工對部分框選結果與類別作修正。完成擴增後會將此資料重新訓練成新的模型用於正式的鳥類辨識任務上，若模型效能不符合預期將會把該模型作為幫助標記資料之小模型用於資料擴增任務上，若模型效能符合預期則輸出該模型。



圖 5. 圖像框選範例

4 實驗

本系統的所有資料集皆從 eBird Taiwan (eBird Taiwan, 2015) 收集，並將其以八比二比例分割為訓練及與驗證集，除此之外本系統無使用任何其他公開或非公開數據。下載圖片使用了 One Click Image Download (Google Chrome Web Store, n.d.)，人工標記圖片則使用了 Make Sense (MakeSense AI, n.d.)。

4.1 小模型之訓練資料與效能

本系統用於訓練標記資料之小模型的資料分布如圖 6 所示，共有 13 種類別。使用的基底模型為 YOLOv8 (Ultralytics, n.d.)，模型的參數設定如下：epochs 為 200、batch size 為 50、image size 為 256、mask ratio 為 4、learning rate 為 0.01、momentum 為 0.937。此模型將著重於讓模型擁有辨別鳥類的能力，並不在意分類

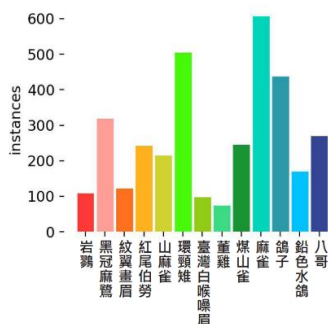


圖 6. 小模型資料分布

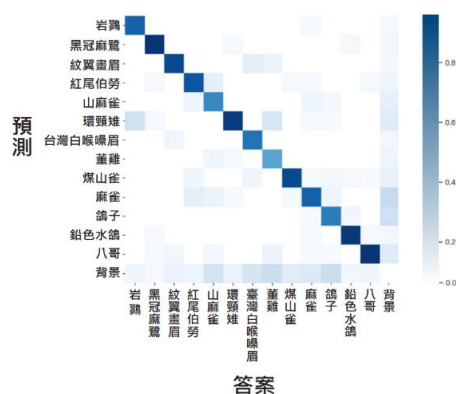


圖 7. 小模型預測準確度

的準確度，所以此處的資料量不多且不平均。模型效能如圖 7 所示，X 軸為正確答案，Y 軸為模型預測答案，圖中斜對角方格的藍色越深代表預測正確率越高，總平均正確率為 86%。由結果得知，此模型對該資料集已有些許成效，已能夠有效辨認出圖片中的鳥，故以此做為幫助標注資料之模型。

4.2 擴增資料

擴增原先資料後的類別為 21 種，所有類別之圖像數目也增加至一千以上，擴增資料分布圖如圖 8 所示。由於本系統之目標為生態監測，故此處有再將類別細分為台灣特有種(圖中藍色字體類別)與非特有種。

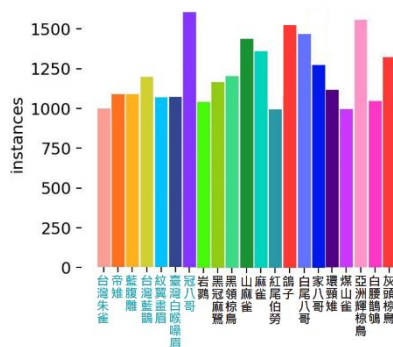


圖 8. 擴增資料分布圖

4.3 最終模型效能

此模型也為最終在系統上所使用模型，參數中的 image size 設定為 512，基底模型與其餘參數設置與小模型相同。模型效能如圖 9 所示，總平均正確率為 93%。

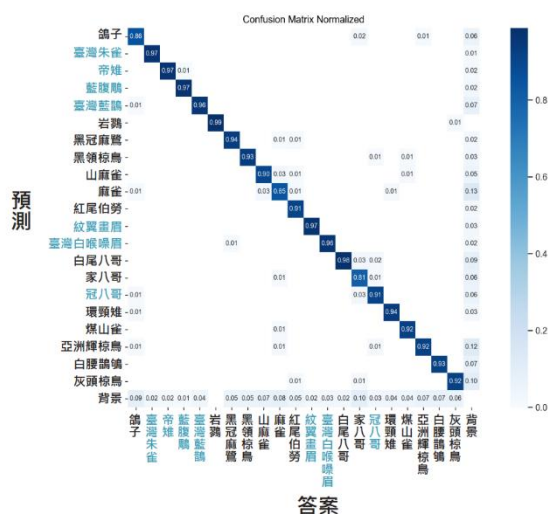


圖 9. 最終模型預測準確度

4.4 系統頁面

本研究所提系統實際頁面如圖 10 所示，並大致分為兩個主畫面，左：鳥類辨識相關，右：LLM 對話機器人。

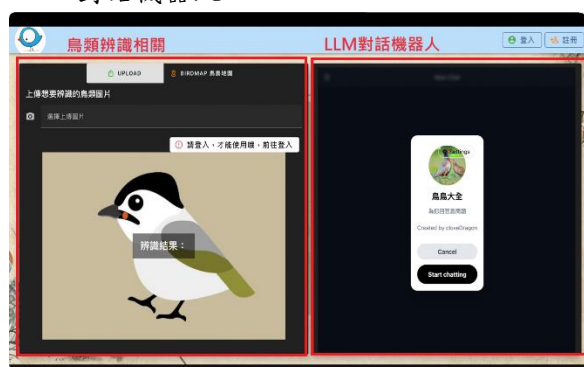


圖 10. 系統實際頁面

鳥類辨識相關功能使用流程如下(圖 11)：首先，使用者上傳欲辨識之圖片，接下來系統會顯示鳥類圖像與名稱，並把該次照片拍攝時間、經緯度以及辨識出類別紀錄在 BirdMap 供使用者觀看。



圖 11. 鳥類辨識相關功能流程

圖 12 為 LLM 對話機器人的使用範例，該對話機器人採用了 CohereForAI/c4ai-command-r-plus (CohereForAI, n.d.)，並通過 Hugging Chat 進行串接。在本系統中，使用者可以直接向該對話機器人發送問題，以快速獲取有關動物的相關資訊。

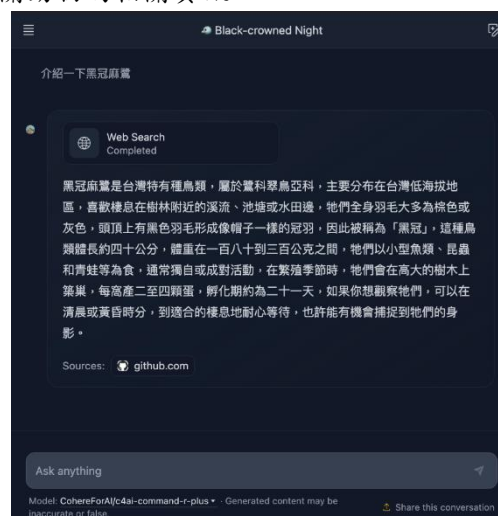


圖 12. 對話機器人互動內容

5 結論與未來工作

本研究開發了一套基於 FPN-PAN 架構的卷積神經網路 (CNN) 自動化生態監測之鳥類辨識系統。通過深度學習技術，系統在處理一般使用者的鳥類辨識需求提高了效率和準確性。使用者可以通過 Web 或 Mobile 端上傳鳥類圖片，並快速獲得辨識結果和相關資訊。本系統還整合了 LLM 對話機器人功能，提供了與鳥類及動物保護相關的即時問答服務。我們認為未來可再進一步深入的目標如下：

- 擴展更多類別：未來計劃將系統擴展至更多鳥類種類，增加訓練數據的多樣性，進一步提高模型的泛化能力和辨識精度。
- 加入實時辨識影像功能：開發實時影像處理功能，使系統能夠即時分析監控視頻中的鳥類，提升監測的及時性和覆蓋範圍。
- 族群數量與棲息地範圍分析：利用系統累積的後端數據，對鳥類族群數量和棲息地範圍進行深入分析，提供詳細的生態研究報告，支持生態保育決策。
- 提升用戶體驗：進一步優化系統界面和交互功能，提升使用者的操作體驗，促進系統在更多生態監測場景中的應用。

通過這些未來工作，我們希望進一步提升系統的應用價值，推動生態監測技術的發展，為保護生態環境和生物多樣性做出更大的貢獻。

致謝

本專題由國立高雄科技大學資訊工程系深度學習授課教授陳俊豪教授的耐心指導與建議以及助教的協助，讓本專題計畫得以成功製作。

References

- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 779-788).
- Bird Taiwan. (2015). eBird Taiwan. <https://ebird.org/region/TW>
- Google Chrome Web Store. (n.d.). One-click image downloader. Chrome Web Store. <https://chromewebstore.google.com/detail/one-click-image-downloader/djacobamaplcmhmaocomnkfdbcoiggepo?pli=1>
- MakeSense AI. (n.d.). MakeSense AI. Retrieved June 23, 2024, <https://www.makesense.ai/>
- Ultralytics. (n.d.). Ultralytics/yolov8: YOLOv8. GitHub. <https://github.com/ultralytics/ultralytics>
- CohereForAI. (n.d.). C4AI Command R+. Hugging Face. <https://huggingface.co/CohereForAI/c4ai-command-r-plus>