

提升誘捕籠剪耳辨識之可靠度：從耳朵計數到結構化判定條件

朝陽科技大學 [學號] [姓名]

[組員 2 / 學校 / 學號 / 姓名]

[組員 3]

[組員 4]

摘要

TNR(捕捉—絕育—回置)以剪除單側耳尖標記已完成絕育之個體。本團隊已建置一套於微控制器上執行剪耳辨識、並自主決定門控之誘捕籠原型；現行邏輯於觀察窗內累加剪耳與未剪耳之偵測框，依兩者比例決定是否捕獲。實際部署顯現三項限制：累加單位為偵測框而非影像幀，同一姿態被重複計入，且不含結構性檢核；模型為單階段訓練，未經場域資料微調；剪耳正樣本結構性不足。本專題擬以兩階段遷移學習、訓練時資料增強與五項結構化檢查條件分別對應，使系統得以區辨「影像不足以判定」與「影像足夠但未觀察到剪耳」，並使每次判定具備可追溯之依據。若改良後仍無法取得足夠證據，即可確認瓶頸在拍攝配置而非辨識模型，此結論本身亦具價值。

1. 前言



同一隻貓：左耳完整、右耳已剪耳（TNR 僅剪除單側耳尖）

圖 1. 剪耳標記之實拍對照。TNR 僅剪除單側耳尖，同一個體必然一耳完整、一耳平截；差異僅在耳尖數公分之輪廓，即本專題須自動辨識之目標。

研究動機。臺灣自 2017 年停止以安樂死作為族群控制手段後，收容所容量即成為主要限制[1]，降低繁殖率遂成為主要管理途徑。TNR 是現有文獻中研究最完整、對族群規模影響最明確之介入方式[2]，其成效取決於能否捕獲未絕育個體，而現場唯一可用之辨識依據為耳尖是否被剪除（圖 1）[3]。傳統誘捕籠僅能觸發關門，凡進入籠內之動物均會被捕獲；於持續實施 TNR 之區域已絕育比例逐年累積，本團隊實測場域約八成個體已絕育，多數捕獲因而是重複捕獲，耗費人力並使動物承受不必要之緊迫。關門以鏈條驅動、關閉聲響明顯，鄰近貓隻亦可能習得迴避行為。

問題定義。現行流程為：單一 YOLO 模型每幀輸出貓臉、剪耳、未剪耳三類偵測框；狀態機偵測到貓臉

即進入觀察狀態，逐幀累加兩類耳朵之框數；窗期結束時若未剪耳累計數超過剪耳之一定倍數且達最低值，即判定未絕育並關門。此流程在影像清晰時可運作，但實際場域顯現三項限制。第一，累加單位為偵測框而非影像幀，同一姿態於連續多幀被重複計入，等同同一項證據投票多次。第二，流程不含任何結構性檢核：不檢查貓臉是否恰為一隻、耳朵是否恰為兩隻、耳框是否位於臉框內，亦不排除依 TNR 慣例不可能發生之「兩耳皆剪耳」；側臉或重複框選時數量條件仍可能成立，系統因而在證據不足下作成判定。第三，模型為單階段訓練，且同一畫面必然同時出現一隻剪耳與一隻未剪耳，剪耳樣本結構性地少於未剪耳，而剪耳又是判定已絕育之唯一依據，遂成瓶頸。

現有方法與其不足。既有智慧誘捕籠之研究著眼於物種層級之區分：Besala 等人之 AI 箱型陷阱判斷畫面中之動物是否屬於目標物種[5]。本專題待區分之兩類為同一物種之兩種狀態，差異僅在數公分之耳廓輪廓（圖 1）。運算位置配置已有延遲與能耗之研究[4]，惟多假設偵測器輸出可直接據以行動，未處理「裝置證據不足時應如何反應」。僅提升平均精度並不足夠，尚須判斷單一幀是否具備判定資格。

本專題之範圍。硬體、軟體與門控機構已完成建置並於場域運作，不在本專題範圍內；本專題只處理模型訓練方式與判定條件，皆以既有裝置與資料改良，不新增硬體。上述限制均經現行程式碼與訓練設定檔比對確認。

預期採用之方法。針對前述三項限制，本專題擬分別以兩階段遷移學習、訓練時資料增強與五項結構化檢查條件對應；細節見第 2 節。

2. 方法

2.1. 兩階段遷移學習。第一階段自 COCO 預訓練權重出發，以通用貓隻影像（約 1,700 張）學習貓臉與耳朵之一般特徵；第二階段以其最佳權重為起點，以籠內實拍之場域影像（約 600 張）微調，使模型適應部署之角度、光線與距離。場域資料不足以單獨訓練出穩定表示，通用資料雖多卻不含籠內條件，故須分階段使用。模型採 YOLOv7-tiny，輸入 640×640，以符合裝置端加速單元之限制。

2.2. 資料增強。資料集不預先增強，改於訓練中即時施加，以免重複樣本佔用資料量。擬採用 Mosaic 四張拼接（1.0）、水平翻轉（0.5）、HSV 色彩擾動（H 0.015/S 0.7/V 0.4）、平移與縮放（0.1/0.5）與少量

MixUp 及 Paste-in (各 0.05)；垂直翻轉與旋轉不予採用，因不符合籠內實際拍攝姿態。本專題將以增強前後之逐類別精度檢驗其對剪耳類別之助益。

2.3. 結構化檢查條件。本專題擬以五項條件取代數量比對，每幀須全部通過方成為有效證據：① 貓臉數量恰為 1；② 耳朵數量恰為 2；③ 兩耳不可皆判為剪耳；④ 兩耳之偵測框均須與貓臉框重疊；⑤ 三類別之信心值各自達門檻。條件③ 依據 TNR 僅剪除單側耳尖：兩耳皆剪耳實際上不會發生，出現即表示分類錯誤。條件①、②、④ 檢核影像結構，與門檻無關；此一分離使系統得以區辨兩種不同性質之失敗，並使失敗原因可被統計。累加單位亦由「偵測框個數」改為「有效幀數」，並設最短計數間隔。

2.4. 證據累積。有效幀依剪耳信心值分入兩條並行軌道：高信心軌道 (≥ 75) 累積 2 幀即判定已絕育，低信心軌道 (≥ 65) 則需 3 幀。兩軌並行使系統於影像清晰時可提早判定，品質普通時仍保有判定能力。計數間隔至少 1 秒，門檻與幀數集中定義於設定檔以便調整。

2.5. 評估方式。本專題將逐幀記錄推論結果、各結構條件之通過與否與當時生效之門檻值，據此計算三項指標：有效證據幀之比例、各結構條件之拒絕分布，以及逐類別之召回率。第二項尤為關鍵：拒絕若主要來自信心門檻，問題在模型精度；若主要來自結構條件，問題在拍攝配置，兩者改善方向完全不同。

2.6. 部署限制。所有改良皆須能於既有之 RTL8735B 微控制器上執行。模型須可量化為裝置端之執行格式，判定邏輯須能以既有韌體結構實作。任何無法滿足此條件之方案，即使離線評估表現較佳，亦不納入。

3. 預期結果

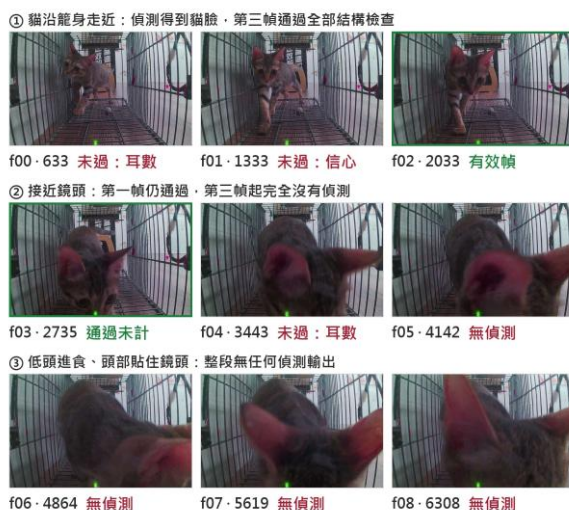


圖 2. 一次觀察窗內之連續影像 (九幀，時間為自觀察起算之毫秒)。九幀僅一幀成為有效證據，其餘或因結構條件未過、

或因動物貼近鏡頭而無偵測；現行做法無法區分這兩者，本專題預期將前者明確標示。

項目	現行做法	本專題目標
判定依據	耳朵框數比例	五項結構檢查 + 雙軌累積
累加單位	偵測框個數	有效幀數 (含節流)
訓練方式	單階段	兩階段遷移學習
不可用影像	無法辨識	標示並記錄拒絕原因

表 1. 現行做法與本專題目標之對照。

評估將以逐類別召回率呈現，而非整體正確率：在已絕育約佔八成之場域，「一律判為已絕育」即可得約八成正確率，該數字不具參考價值。預期產出有三：有效證據幀比例之提升幅度、各結構條件之拒絕分布，以及剪耳類別在兩階段訓練與增強前後之精度變化。負面結果仍屬可能，但拒絕分布仍可指出瓶頸所在，此即本專題之實質貢獻。

參考文獻

- [1] Yan, T.-Y., and Teng, K. T.-y. Trends in animal shelter management, adoption, and animal death in Taiwan from 2012 to 2020. *Animals*, 13:1451, 2023.
- [2] Smith, L. M., Hartmann, S., Munteanu, A. M., Dalla Villa, P., Quinell, R. J., and Collins, L. M. The effectiveness of dog population management: A systematic review. *Animals*, 9:1020, 2019.
- [3] Agriculture Knowledge Portal, Taiwan. Introduction to TNR. 2020.
- [4] Koubaa, A., Ammar, A., Alahdab, M., Kanhouh, A., and Azar, A. T. DeepBrain: Experimental evaluation of cloud-based computation offloading and edge computing in the Internet-of-Drones for deep learning applications. *Sensors*, 20:5240, 2020.
- [5] Besala, F. I., Niimoto, R., Lee, J. H., and Okamoto, S. Development of AI-based smart box trap system for capturing a harmful wild boar. *ROBOMECH Journal*, 12:4, 2025.