

智慧誘捕籠：剪耳辨識

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

本專題要用影像辨識判斷一隻流浪貓有沒有被剪過耳朵。TNR(捕捉—絕育—回置)會把結紮過的貓剪掉一邊耳尖當記號，這是現場唯一看得出來的線索。我們的智慧誘捕籠把 YOLOv7-tiny 放在籠內的微控制器上，看到剪耳就放行、沒看到就關門。第二版在貓店實測 116 次進籠，判「已結紮」0 次，但現場約八成的貓已經結紮。本專題針對兩個卡點提出改進：剪耳樣本太少(約只有完整耳的三分之一)，用補樣本、重複抽樣和 focal loss 處理；照片不夠時規則仍硬下結論，改成多張累積並加入「無法判定」，不確定就不關門。我們以剪耳召回率和「零證據就決定關門」的次數，驗證這些方法有沒有用。

1. 前言



圖 1. 剪耳長什麼樣。同一隻貓一邊耳尖完整、一邊被剪平；兩類只差在耳尖那幾公分，這就是本專題要自動認出來的東西。

研究動機。臺灣 2017 年停止以安樂死控制族群數量後，收容所容量成為主要限制 [1]，降低繁殖率變成主要的管理方式。TNR 是文獻中研究最完整、對族群數量影響最明確的做法 [2]，成敗關鍵在於抓不抓得到還沒結紮的貓，而現場唯一的線索就是耳尖有沒有被剪掉(圖 1) [3]。傳統誘捕籠不會分辨，走進去就關門；在長期做 TNR 的地區，多數捕獲其實是重複抓到已結紮的貓。

這件事難在哪。第一，差別太細：兩類都是貓，只差在耳尖那幾公分的形狀。第二，數量天生不平衡：TNR 只剪一邊，已結紮的貓也是一耳剪、一耳完整，所以剪耳樣本一定比較少，偏偏它是判斷結紮的唯一依據。第三，單張照片不可靠：貓的角度和距離一直在變，一次進籠的照片多半不夠判斷(圖 2)。邊緣推論的研究多半假設辨識結果可以直接拿來用 [4]，現有

的智慧誘捕籠分的是長相差很多的物種 [5]，都沒有處理「照片根本不夠判斷」的情況。



圖 2. 同一隻貓一次進籠的 9 張照片中挑 3 張。只有走近時那張夠格，太近時耳朵數不對，低頭吃飯時完全認不到；這次 9 張裡只有 1 張能當證據。

我們已經改過一輪。第一版(到 2026 年 4 月)每張照片用同一個門檻(40 分)，剪耳出現兩次、平均 40 分以上就判已結紮；拿 111 張照片離線重跑，剪耳只認出 19.6%。第二版(2026 年 7 月起)做了三項改進：用兩階段遷移學習、以籠內照片重新訓練模型；每張照片先過五項檢查才算數；剪耳證據分兩條標準累積(見第 2 節)。

現況：第二版還是卡住(表 1)。兩台裝置在貓店實測 116 次進籠，判「已結紮」0 次，以現場約八成已結紮估計，約 93 次判錯。根本原因是證據不夠：1,125 張照片只有 5.9% 通過五項檢查，沒有任何一次進籠湊到兩張夠格的剪耳照片(標準要 2~3 張)；判定要關門的 34 次中，有 24 次完全沒有證據。照片被擋下的原因只有約四分之一是信心不足，所以光調門檻救不回來。

項目	結果
進籠判定次數	116
判「已結紮」	0
估計判錯	約 93
整段沒有一張可用的進籠	80 (69%)
判定要關門但完全沒有證據	24/34
通過五項檢查的照片	66/1,125 (5.9%)

表 1. 第二版實測結果(兩台裝置，貓店)。現場沒有逐筆核對真實答案，判錯次數以約八成已結紮估計。

本專題做什麼。針對兩個卡點：剪耳樣本太少，從資料和訓練下手；照片不夠時仍硬下結論，改判斷規則。一律用現有的裝置和資料，不動硬體。目的是驗證這些方法有沒有用，就算沒用也要說清楚卡在哪裡。流程見圖 3。

2. 方法

2.1. 模型與裝置。偵測模型是 YOLOv7-tiny [6]，輸入 640×640，辨識貓臉、剪耳、完整耳三類，從 COCO 預訓練權重開始訓練。訓練完量化成 int16 (把小數權重改用整數存，裝置才跑得動)，放進籠內的

RTL8735B 微控制器，每張約 0.7 秒，不連伺服器。



圖 3. 方法流程。上半在電腦上訓練，下半在裝置上逐張辨識再累積判斷；綠色是本專題新增，灰色是第二版已做好的部分。

2.2. 資料與切分。資料有兩種（表 2）：網路蒐集的一般貓照，以及籠內相機實拍的照片。籠內照有 291 張是紅外線觸發但沒有貓的空籠，當作負樣本，教模型不要對背景亂框。兩種資料都依 70/15/15 切成訓練、驗證、測試，並按「含剪耳/只有完整耳/空籠」分層，讓三組的剪耳比例一致，亂數種子固定。增強只用在訓練組，挑模型只看驗證組，測試組留到最後只算一次。

	一般貓照	籠內照
用途	第一階段	第二階段微調
原始照片	1,729	606 (空籠 291)
含剪耳的照片	793	144
剪耳：完整耳標註	約 800 : 2,700 (1 : 3.3)	約 145 : 490 (1 : 3.4)

表 2. 訓練資料規模。剪耳是少數類別，標註數量約只有完整耳的三分之一。

2.3. 兩階段遷移學習與資料增強。第一階段用一般貓照學「貓臉和耳朵長什麼樣」；第二階段從第一階段最好的權重出發，用籠內照微調，適應籠內的角度、光線和距離——籠內照太少，單獨訓練學不穩。增強在訓練時即時做：Mosaic（四張拼一張）、MixUp（兩張疊合）、左右翻轉、色相/飽和/亮度擾動、平移與縮放、Paste-in（把物件貼到別張）。不做上下翻轉和旋轉，因為籠內的貓不會倒著出現。

2.4. 處理剪耳太少。籠內含剪耳的照片只有 144 張，我們從四個方向補：(1) 補資料：新版韌體會存下每次進籠的每一張照片，從中挑剪耳照片補標；(2) 重複抽樣：含剪耳的照片每輪多抽幾次 [7]；(3) focal loss [8]：降低「容易認的樣本」的權重，讓模型把力氣花在難認的剪耳上；(4) 換挑模型的標準：從三類平均分數改成剪耳召回率，因為判斷結紮只靠剪耳。

2.5. 多張累積再判斷。每次進籠觀察 6 秒、約 9 張。每張先過五項檢查才算「有效張」：貓臉恰好 1 個、耳朵恰好 2 個、不可兩耳都是剪耳（TNR 只剪一邊）、兩耳都和貓臉重疊、信心達門檻（貓臉 65、耳朵 50，滿分 100）；有效張至少間隔 1 秒，避免同一姿勢重複計算。剪耳 75 分以上累積 2 張，或 65 分以上累積 3 張，就判已結紮。本專題新增第三種結果「無法判定」：沒有任何有效張時不下結論、不關門，而不是當成「未結紮」直接關門。這是拒絕判斷 (reject option) 的

概念 [9]：錯抓已結紮的貓無法挽回，漏抓的貓下次還會再來，兩種錯的代價不對稱，理性的代理人應選預期損失最小的動作 [10]。

2.6. 怎麼驗證。模型層看測試組上每一類的 AP 與召回率，重點是剪耳。判斷層不必回現場：韌體會記下每張照片的分數和五項檢查結果，可以離線回放，讓新舊規則在同一批進籠上比較。不看整體正確率——現場八成已結紮，一律回答「已結紮」就有八成正確率，這個數字什麼都沒說明。

3. 預期結果

本專題會交出三樣東西：(1) 各種剪耳補強方法在測試組上的剪耳召回率對照；(2) 新規則的回放結果——「零證據就決定關門」應從 24 次降到 0 次，並算出「無法判定」佔多少；(3) 五項檢查各擋掉多少張，用來分辨問題出在模型還是拍攝。可以預見的限制是：照片被擋下的原因多半是臉或耳朵在畫面裡不完整，這部分光改模型可能救不回來。實驗結果、調參過程與錯誤分析會在後續作業報告。

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., Quinnell, 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.
- [6] Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *CVPR*, 2023.
- [7] Gupta, A., Dollár, P., and Girshick, R. LVIS: A dataset for large vocabulary instance segmentation. In *CVPR*, 2019.
- [8] Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. Focal loss for dense object detection. In *ICCV*, 2017.
- [9] Chow, C. K. On optimum recognition error and reject tradeoff. *IEEE Transactions on Information Theory*, 16(1):41 – 46, 1970.
- [10] Russell, S., and Norvig, P. *Artificial Intelligence: A Modern Approach*. Pearson, 4th edition, 2020.