

EarTip-Trap: On-Device Ear-Tip Recognition for Selective Capture of Unneutered Stray Cats

國立○○大學 ○○○○○○○○ ○○○

Abstract

Trap-Neuter-Return (TNR) programs mark sterilized cats by tipping one ear, yet conventional live traps capture every animal that enters, so neutered cats are trapped again and again. We propose EarTip-Trap, a cage trap that decides on-device whether to close its door. A YOLOv7-tiny detector running on the neural processing unit (NPU) of an AMB82-mini board localizes the cat face and classifies each ear as tipped or intact, and a two-stage decision layer (a per-frame structural gate followed by multi-frame evidence accumulation) turns detections into a capture decision. A preliminary field deployment of 116 visits shows that 69% of visits produced no valid frame, which frame-by-frame inspection traces to camera placement rather than model capacity. We will relocate the camera, retrain with more tipped-ear samples and hard negatives, and evaluate an abstention rule, reporting per-class recall against a majority-class baseline.

1. Introduction

Free-ranging domestic cats are efficient predators; in the United States alone they are estimated to kill 1.3–4.0 billion birds and 6.3–22.3 billion mammals every year [6], and strays are linked to zoonoses, accidents and bites [2]. In Taiwan, after the 2017 ban on shelter euthanasia, counties exceeding the legal shelter-veterinarian workload rose from 2 to 6 [10]. Among population-management interventions, fertility control is the most studied and shows the largest reported effect on population size [8], but only if the *unneutered* animals are the ones that get caught.

TNR workers tip one ear of every sterilized cat so that it can be recognized from a distance. The trap itself, however, cannot see this mark: any animal that steps in triggers the door. As a TNR program matures, neutered cats dominate the colony (about 80% at our test site, a site-level estimate), so most captures are repeat captures that must be released by hand, costing labor and stressing the animal. We therefore ask the trap to decide by itself: *capture only a cat whose ears are both intact*.

The cue is a small flat edge on one ear, the device is unattended and often offline, and the two errors are asymmetric: trapping a neutered cat cannot be undone, while a missed cat returns, so an uncertain decision should *not* close the door.

Prior work covers parts of this problem. Huang *et al.* [3] classify TNR ear-tip images with a YOLO ensemble and

majority voting and report mAP above 90%, but offline, on images taken around TNR surgery. Besala *et al.* [1] couple YOLOv7-tiny with a box trap for wild boars on a Jetson Nano at about 4 FPS, which only needs species-level recognition. WildFiT [7] shows that camera-trap models degrade under domain shift and must adapt in situ, and cloud offloading [5] assumes a stable link that a roadside feeding spot rarely has. To our knowledge, no system recognizes ear tips on a microcontroller-class NPU inside the trap and reports field data from real visits.

A preliminary deployment of two devices (116 visits) already exposes the real bottleneck. In 69% of visits *no* frame passed the structural gate, and the system never produced a *neutered* decision. Because every frame is logged, we could replay each visit (Fig. 1): the cat's face is well framed only while it walks down the middle of the cage; once it reaches the food bowl, where the camera is mounted, its head fills the lens and the detector sees nothing. Only 26% of gate rejections are low confidence; the rest are structural (face or ear count, overlap), so lowering thresholds cannot fix them. We contribute (1) a two-stage on-device decision layer, (2) a quantization-aware selection metric, and (3) a field protocol that is honest under class imbalance.

2. Method

2.1. System overview

An AMB82-mini (Realtek RTL8735B) runs the detector on its NPU with a 640×640 camera stream. An entrance infrared (IR) sensor starts observation, a rear IR sensor guards the door (closed by a relay-driven motor in 1.2 s), and every decision is saved to SD card before upload, so capture never depends on the network.

2.2. Detector and training data

YOLOv7-tiny [9] detects three classes: *cat_face*, *tipped ear* and *intact ear*. Training has two stages: general cat images (1,729 raw images) followed by fine-tuning on in-cage images (606 raw images, of which only 144 contain a tipped ear). Tipped-ear samples are therefore the main data bottleneck.

2.3. Two-stage decision layer

Gate. A frame is valid only if it contains exactly one face

and two ears, both ears overlap the face, not both ears are tipped (TNR tips one ear), and confidences reach 65 (face) and 50 (ears). *Evidence*. Valid frames at least 1 s apart are counted; the cat is judged neutered after 2 frames with tipped-ear confidence ≥ 75 or 3 frames ≥ 65 , within a 6 s window (measured frame interval 0.70 s, about 9 frames). Otherwise the current firmware treats the cat as unneutered and closes the door. We will add an *abstention* rule, evaluated first by offline replay of the logged frames: a visit with zero valid frames is labelled *uncertain* and the door stays open.

2.4. Quantization-aware model selection

The deployed model runs on the NPU after int16 quantization, and our firmware uses fixed confidence thresholds, so mAP alone is misleading; we stress-test candidates with stricter INT8 quantization [4]. We define *decision coverage* as the fraction of test images whose detections form a valid face-and-two-ears structure above threshold, before and after INT8 quantization (Table 1). RT-DETR-L [11] has the best tipped-ear AP, but quantization compresses its scores below threshold and coverage collapses to zero. YOLOv7-tiny keeps 84% of its coverage and is the only candidate the NPU supports.

Table 1. Candidate detectors on the test split (15 tipped-ear images; the bootstrap CI of YOLOv7-tiny tipped-ear AP50 is wide, 0.11–0.60).

Model	Tipped AP50	Coverage fp32→INT8	NPU
YOLOv7-tiny	0.377	0.47 → 0.40	yes
YOLO26n	0.493	0.27 → 0.29	no
RT-DETR-L	0.508	0.54 → 0.00	no
Faster R-CNN	0.505	0.52 → 0.48	no
SSDLite	0.123	0.01 → 0.00	no

2.5. Evaluation plan

Model: per-class AP with bootstrap intervals. *Visit*: each logged visit is labelled from its saved frames; we report per-class recall, the false-capture rate of neutered cats and the zero-evidence rate. Always answering *neutered* would already be about 80% accurate here, so accuracy is only reported beside this majority baseline. *System*: frame interval and upload success.

3. Expected Results

Figure 1 shows the failure we expect to remove. We move the camera from the food bowl to face the middle of the cage and expect (i) more gate-passing frames per visit and a clearly lower zero-evidence rate than the current 69%; (ii) the first *neutered* decisions, which make a full confusion matrix and per-class recall computable for the first time; (iii) zero captures without evidence under the abstention rule, versus 24 of 34 captures today; and (iv) more tipped-ear training samples and hard negatives

(empty cages) mined from the new logs. The final report will compare the same metrics before and after each change.

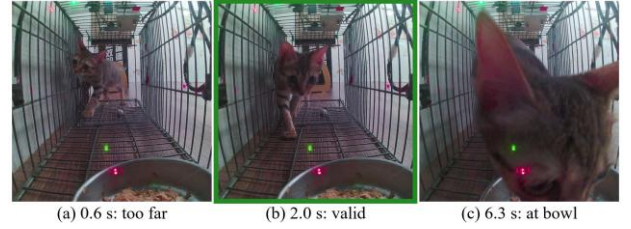


Figure 1. Three frames of one real visit (6 s window). (a) is rejected by the gate (ear count); (b), taken mid-cage, is the only valid frame; once the cat eats at the bowl beside the camera (c), nothing is detected.

References

- [1] Francis Ifaso Besala et al. Development of AI-based smart box trap system for capturing a harmful wild boar. *ROBOMECH Journal*, 12:4, 2025.
- [2] Gabriela C. Bicalho et al. Realities, perceptions, and strategies for implementation of an ethical population management program for dogs and cats on university campuses. *Frontiers in Veterinary Science*, 11:1408795, 2024.
- [3] Yi-Cheng Huang, Ting-Hsueh Chuang, and Yeong-Lin Lai. Classification of the trap-neuter-return surgery images of stray animals using Yolo-based deep learning integrated with a majority voting system. *Applied Sciences*, 11(18):8578, 2021.
- [4] Benoit Jacob et al. Quantization and training of neural networks for efficient integer-arithmetic-only inference. In *CVPR*, 2018.
- [5] Anis Koubaa et al. DeepBrain: Experimental evaluation of cloud-based computation offloading and edge computing in the Internet-of-Drones for deep learning applications. *Sensors*, 20(18):5240, 2020.
- [6] Scott R. Loss, Tom Will, and Peter P. Marra. The impact of free-ranging domestic cats on wildlife of the United States. *Nature Communications*, 4:1396, 2013.
- [7] Mohammad Mehdi Rastikerdar et al. WildFiT: Autonomous in-situ model adaptation for resource-constrained IoT systems. *arXiv:2409.07796*, 2024.
- [8] Lauren M. Smith et al. The effectiveness of dog population management: A systematic review. *Animals*, 9(12):1020, 2019.
- [9] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *CVPR*, 2023.
- [10] Tzu-Ying Yan and Kendy Tzu-yun Teng. Trends in animal shelter management, adoption, and animal death in Taiwan from 2012 to 2020. *Animals*, 13(9):1451, 2023.
- [11] Yian Zhao et al. DETRs beat YOLOs on real-time object detection. In *CVPR*, 2024.