

Smart Cat Trap: Recognizing Clipped Ear Tips

邱識瑋
朝陽科技大學
11530601

吳翊銘
朝陽科技大學
11430624

Abstract

This project uses image recognition to tell whether a stray cat has a clipped ear tip. In TNR (Trap–Neuter–Return), the clipped ear tip is the only visible sign that a cat is already neutered. Our smart trap runs YOLOv7-tiny on a microcontroller inside the cage and closes the door when no clipped ear is detected. The current system ran at a cat shelter for about four months: 86.1% of 509 decisions gave no valid result, and an offline re-evaluation on 111 images recognized only 19.6% of clipped ears.

From the field records we found three causes: too few clipped-ear samples, treating a single image as complete information, and no clear handling of insufficient evidence. Based on the machine learning and intelligent agent concepts of this course, we plan three improvements: more clipped-ear samples with transfer learning, accumulating evidence over several images, and an “undecidable” outcome that keeps the door open. We will test them with experiments during the semester.

1. Introduction

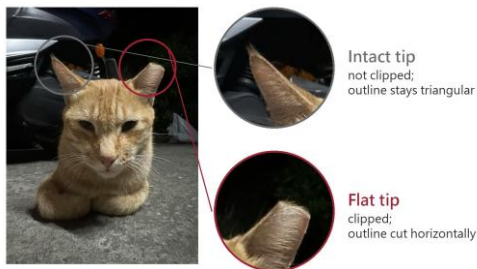


Figure 1. An intact ear and a clipped ear differ only at the tip.

Motivation

Since Taiwan stopped using euthanasia to control stray animal numbers in 2017, shelter capacity has become the main limit [1], so reducing reproduction is now the main approach. TNR is the most studied method with the clearest effect on population size [2], but it only works if un-neutered cats are caught, and the only cue in the field is whether the ear tip is clipped (Fig. 1) [3]. A conventional trap closes whenever a cat walks in, so in areas with long-running TNR most captured cats are already neutered.

Why this is hard

First, the difference is tiny: only a few centimeters at one ear tip. Second, the classes are naturally imbalanced: TNR clips only one ear, so even a neutered cat has one intact ear, and clipped ears are always the minority, yet they are the only evidence of neutering. Third, a single image is unreliable: the cat's pose and distance keep changing, most images are not good enough to decide, and this is where the current system gets stuck (Fig. 2). Existing smart traps separate species that look very different [4] and do not handle images that are insufficient for a decision.



Figure 2. The current decision rule and four months of field results.

Current problems

After a cat enters, the current system observes for about 7 s and takes about 9 images. It applies the same threshold to every image, counts how often a clipped ear appears, and then releases the cat or closes the door; if evidence is insufficient it retries, and when retries run out it lets the cat go (Fig. 2). Over four months with two devices at a cat shelter, 86.1% of 509 decisions gave no result, and an offline re-evaluation on 111 images recognized only 19.6% of clipped ears.

Causes

There are three causes. First, clipped-ear samples are too few, so the model does not learn them well. Second, the system treats a single image as complete information. It is a simple reflex agent that reacts to each image with a fixed rule, but the cage is a partially observable environment where one image cannot show the full situation [5]. Third, insufficient evidence has no clear handling: blurry and clear images count the same, and when stuck the system can only retry and release, leaving no record of what went wrong.

Our plan

Combining the course concepts with our current progress, we plan one improvement for each cause (Sec. 2, Fig. 3) and will adjust the details as we go.

2. Method

2.1. Model and device

We use YOLOv7-tiny [6] to detect three classes: cat face, clipped ear, and intact ear. After training, the model is compressed to integers so that it can run on the RTL8735B microcontroller in the cage, taking about 0.7 s per image without a server.

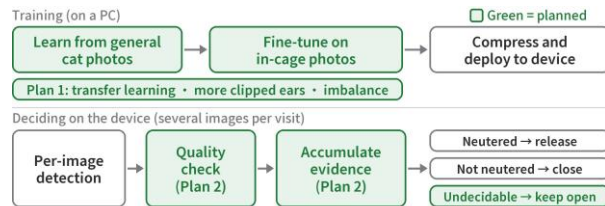


Figure 3. Planned improvements.

2.2. Plan 1: More clipped-ear samples

This follows the machine learning part of the course. We have about 1,700 general cat photos from the web and several hundred in-cage photos; clipped and intact ears are labeled at roughly 1:3. We will first train on general cat photos so the model learns faces and ears, then fine-tune on in-cage photos (transfer learning). We will also keep adding clipped-ear images from the cage and try imbalance methods such as repeat sampling [7] or focal loss [8]. Data will be split into training, validation, and test sets, and models will be chosen on the validation set only to avoid overfitting [5].

2.3. Plan 2: Accumulate evidence over several images

This follows the agent design in the course: in a partially observable environment, an agent should keep an internal state and decide from a sequence of percepts, not from the current one alone (model-based agent) [5]. We will decide in two steps. First, each image passes a quality check, for example whether the face and ears are complete and the confidence is high enough; failing images are ignored. Second, the passing images are accumulated, and a conclusion is made only when the clipped-ear evidence is enough.

2.4. Plan 3: Add an “undecidable” outcome

This follows the utility-based agent: a rational agent chooses the action with the highest expected utility [5]. Here the two errors have different costs: capturing a neutered cat cannot be undone, while a missed un-neutered cat will come back. So when evidence is insufficient, instead of guessing, the system answers “undecidable” and keeps the door open, which is the idea of a reject option [9]. It also records where each case got stuck and uses this as feedback, like the critic in a learning agent [5].

2.5. Evaluation

Following PEAS, we first define the performance measure [5]: recall on clipped ears, not overall accuracy. About 80% of the cats at the test shelter are already neutered, so always answering “neutered” would still give 80% accuracy, which means nothing. For the decision rules, we will replay the recorded images and compare the old and new rules.

3. Expected Results

We expect three outcomes. First, with more samples and transfer learning, clipped-ear recall will exceed the current 19.6%. Second, the 86.1% with no decision can be split into “images truly insufficient” and “actually decidable,” and we will know at which step each case failed. Third, when evidence is insufficient, the system will mark the case “undecidable” and record why, instead of retrying and letting the cat go. However, if the images never show the face or ears, a better model cannot fix it. Actual numbers and error analysis will be reported in later assignments.

References

- [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] 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.
- [5] Russell, S., and Norvig, P. *Artificial Intelligence: A Modern Approach*. Pearson, 4th edition, 2020.
- [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.