Department of Conservation
Te Papa Atawhai

Maukahuka
PEST FREE AUCKLAND ISLAND

File note

DOC: 6306763

Filer: Lindsay Chan

Date: 06/07/2020

Analysis of the artificial intelligence and machine learning market in relation to trail camera footage review

Update 13/07/2020:

Since this paper was released, we've learnt that ClassifyMe has stopped development on its New Zealand version. Instead Landcare Research who was ClassifyMe's New Zealand liaison, is redirecting their AI efforts toward TrapsNZ.

Executive Summary

The ability to review imagery for a large-scale camera grid (1500+ cameras) is essential for the feasibility of eradicating cats as part of the Maukahuka Pest Free Auckland Island project. Trail cameras at 500m spacing are the only current detection tool that can feasibly saturate every cat home range across Auckland Island's vast (46,000 ha) and challenging landscape. Over the course of a year, the Maukahuka project team improved its manual footage review processing time from 300 hours to 51 hours and shared the labelled and reviewed data with machine learning (artificial intelligence; AI) developers for testing with their algorithms. Machine learning capabilities for trail camera footage review have improved drastically in the New Zealand market and internationally in only a few years. Developer results show promise in both object detection and cat classification in reducing the workload for the manual reviewer but are not a standalone tool at this stage. In an eradication project, 100% recall (detection of an object in an image) or as close to it as possible is required. There is scope for efficiencies and greater outcomes through collaboration, however this cannot be driven by the Maukahuka team because of its project going on hold. It is recommended that DOC: a) Produces user case requirements for predator control monitoring including the specific needs of the Maukahuka project; b) Standardizes its folder structure, labelling and footage review records for all its trial camera projects and stores this footage in the cloud where internal and external parties can easily access it; c) Appoints an individual or project to do this and keep the department accountable; d) Continue to follow this market and look for opportunities to collaborate on product development.

Glossary

Term	Definition
Machine learning	A form of artificial intelligence, where the model can learn and build on its knowledge
Deep learning	A sub field of machine learning focusing on the training of artificial neural networks, similar to the functions of neurons in the brain
Onboard processing	When the machine learning and image classification is done by the trail camera in the field (as opposed to downloading all raw data off the device for subsequent processing)
LoRa	A long range, remote sensing network that would be used for this project to transmit notifications from onboard processing out of the field
Bounding box	A frame that goes around any detected object in the image. This is part of the machine learning output. Some models include the object classified and confidence level.
False negative	When the machine learning model looks at an image and classifies it as empty when there is actually there is an object of interest in it. (e.g. a cat in the photo) This is the biggest risk for an eradication relying on machine learning and why many have refrained from using it.
False positive	When the machine learning model looks at an image and classifies it as having an object of interest, but in actuality it doesn't. For example, the model may classify the bait post as an animal.
True positive	When the machine learning model correctly classifies the image as having an object of interest.
True negative	When the machine learning model correctly classifies the image as being empty.
Precision	Measures the model's accuracy when looking solely at the objects the model has detected. Precision does not consider that the model might have missed objects of interest in the first place.
Recall	Measures how many cats the model detected compared to how many cats there really are. As an eradication project, this is critical to us. Recall can only be calculated when each individual image has been reviewed by a human.
Precision/Recall example 1	*Suppose a computer program for recognizing dogs in photographs identifies 8 dogs in a picture containing 12 dogs and some cats. Of the 8 identified as dogs, 5 actually are dogs (true positives), while the rest are cats (false positives). The program's precision is 5/8 while its recall is 5/12.* ("Precision and Recall," 2020)
Precision/Recall example 2	*When a search engine returns 30 pages only 20 of which were relevant while failing to return 40 additional relevant pages, its precision is 20/30 = 2/3 while its recall is 20/60 = 1/3. So, in this case, precision is "how useful the search results are", and recall is "how complete the results are".* ("Precision and Recall," 2020)

Introduction

Context

In recent years, artificial intelligence has made its way into conservation technology. Machine learning, a form of artificial intelligence, where the model builds its knowledge and learns from the training datasets, has begun being applied to automate review of infrared trail camera footage. A machine learning model can deliver comparable results in a fraction of the time of manual review. It has been applied internationally on standard trail camera footage review as evidenced through conservation technology websites Wildlabs.net and Wildlife Insights, conversations with Island Conservation and the Royal Society for the Protection of Birds (RSPB) as well as Microsoft AI for Earth. The software ClassifyMe has recently appeared in Australia and New Zealand and has been used in this research. In New Zealand, the Cacophony Project and Zero Invasive Predators (ZIP) are pursuing machine learning on thermal trail cameras using video with onboard processing.

As part of the proposed cat eradication on Auckland Island (46,000 ha), the Maukahuka project recommends trail cameras as the primary detection tool for island-wide surveillance of cats. This would require 1500 cameras (set at 500m spacing) for roughly two years.

The proposed methodology relies on cameras as they are the only detection tool that can saturate every home range. Utilisation of trail cameras at this scale is dependent on more efficient tools to process and review images. Machine learning provides the opportunity to develop such a tool but it must be able to reliably retain all images containing the target species so all individuals of a target species can be positively identified either manually or by artificial intelligence (AI) or a combination of both. The minimum pre-requisite capability for feasible use of cameras on Auckland Island is for AI to be able to triage empty images (false triggers) removing 80% of the manual processing work. The ultimate goal is 100% recall and precision from AI software for processing all imagery with minimal verification required.

Scaling up this proposed surveillance method requires automated processing of image data to feasibly review a predicted 3.9 million images which would equate to 7800 hours or 3.75 FTE to review manually.

Two camera trials have now been completed on Auckland Island and the resulting footage has been manually classified and labelled. The New Zealand machine learning market has been approached to understand existing capability for auto analysis of footage, to compare how existing classification models perform (time, cost, accuracy) compared to the manual classification as well as to understand thermal and on-board processing capability. Developers got access to the verified data sets and contributed their time and shared results. The preliminary results are outlined in this document.

The primary capability of interest for Maukahuka is removing false triggers to reduce footage review to just 20% of initial workload. (See Table 1) How the model is designed can affect the other desired capabilities, so both essential (Table 1) and reach goals were communicated to the developers, allowing the consideration of future project intentions into the initial model design. Reach goals consisted of a machine learning model with onboard processing and ability to send output via a remote sensing network back to the field base. However, building such a system would be expensive, and potential maintenance complexity is risky for a large-scale remote island eradication project such as Maukahuka, where 100% confidence in detection and performance is required. A tool would need to be well proven before consideration.

Table 1: Minimum requirements communicated to developers for current trail camera technology for Auckland Island cat eradication feasibility

Importance	Prioritisation	Function
Essential	1.	Software separates images (captured by standard trail cameras) of false triggers from images containing animals (cats, pigs, mice, birds etc)
Essential	2.	Automatically labels and places footage into directories for easy reference
Highly desirable	3.	Software separates footage (captured by standard trail cameras) of cats from the rest of the data set

Purpose

The document outlines the field trial data analysis and process improvements, parties consulted (developers and users), results of industry engagement, lessons learnt, next steps and recommendations for trail camera footage review agency wide collaboration within DOC.

Methods

Evolution of Maukahuka footage review process

Trail camera footage from the same trail camera grid (65 cameras) was reviewed three times using three different processes in the span of 10 months. Each iteration provided lessons which led to significant improvements to the process, reducing effort required and improving accuracy.

The first review was completed in a backcountry hut during the Summer 2018/19 trials. The camera grid trial (1300 ha) consisted of 65 cameras at 500m spacings recording 147,107 images (in bursts of three) over four weeks. Over 80% of images were “false triggers” caused by moving vegetation. All images, including these false triggers, needed to be reviewed accurately as a missed cat detection would risk the success of an eradication attempt. The footage review goals of this trial were to have the data organised in a suitable format for Landcare Research’s Al Glen to undertake spatial analysis and detection probability. Footage review took place in the evenings after a day in the field or in between field days when the weather was poor. Key factors which contributed to the process being time intensive were individual cats were being identified, none of the team had led the data management of trail camera footage review before and the Bushnell Aggressor cameras used had limitations on file output naming conventions and folder organisation (didn't show the time or camera ID nor could be customised to).

Field staff used BR's (br-software.com) EXIFextracter (suggested by Al Glen), a free software download that extracts image metadata into an excel spreadsheet. However the camera created a new subfolder for every 1,000 images and BR's EXIFextracter can only process one folder at a time. This meant it had to be tediously run multiple times for each camera. Staff then reviewed each image and recorded what was observed (cat, collared cat, individual cat IDs, penguin, seal, tomtit, blackbird, etc.). This process was slow, laborious, and clunky, but was the best tool the team had at the time.

It was critical to label images to the second to enable unique identification and matching between the images and the excel list. An image renaming process was created through Safe FME, a data interpolation software to overcome the limitations of the camera outputs.

Complicating the process was the involvement of five different people downloading SD cards and reviewing footage on five different computers. Once reviewed, individuals had to copy their excel spreadsheets into the master. Countless duplicates eventuated as well as missing SD cards. This involved a long and tedious tidy up process upon return to the mainland. Lessons were captured in a trial debrief.

In preparation for the second trial (Winter 2019), a strict process was implemented for field staff to follow including; a physical inbox and outbox for SD cards, and a check list recording who collected the SD card, the day it was brought in from the field, what camera it came from and its stage in the trial. This checklist was viewable to everyone and filled out immediately upon return to the hut. A single "data champion" was also given an updated task specification, training and checklists to ensure consistency. Andre Wilfert created a program to replace BR EXIFextracter and the need to manually move and copy files. This reduced much of the manual process which was previously a source of errors and loss of data. This program reduced footage review and data management time from 300 hours to 105 hours. It should be noted the data came from the same camera grid as for the first trial, however footage was recorded as one image, a 10 second video then another single image rather than 3 images per 3 seconds. Field staff also identified individual cats where possible, a level beyond current AI capability.

The third review was required to create the training datasets (i.e. object specific folders) for the machine learning models. The Summer 2018/19 dataset was again used for this. Staff used Joris Tinnemans's MS Access Graphic User Interface (GUI) which he created for his camera grids in the Hawdon and Doubtful Valleys. This allowed each image to be viewed and records data via the simple click of a button (automatically renaming and refiling the image while simultaneously populating the spreadsheet), thus eliminating the majority of manual effort in the process. This reduced the total review time to 51 hours.

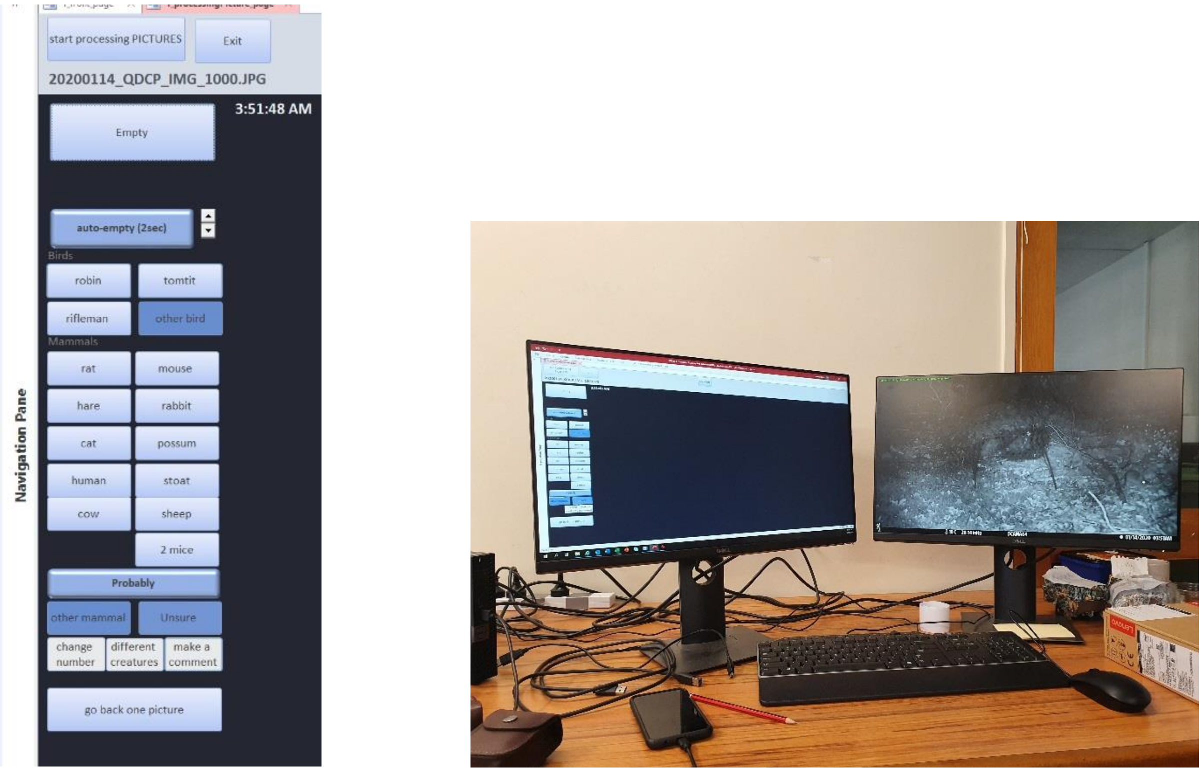

Figure 1: (a) Joris Tinnemans's MS Access GUI commands. (b) Set up with two screens and quick click commands to allow easy review and classifying in the GUI.

The third review was fastest and likely most accurate because it minimised manual processing and thus human errors, did everything in one place, rather than having to jump between windows and software, it was completed by staff dedicated specifically to footage review, and in the comfort of an office with two screens. In addition, it moved the data for staff into object specific folders, which was essential for developers to train their machine learning models.

Table 2: Summary of camera review process for each camera trial undertaken by the Maukahuka team including hours of effort.

Dataset	Review Method	Hrs	Review location
Summer 2019	BR's EXIFextracter, Excel, FME	300	Field hut and then office
Winter 2019	Andre's labelling/extracting script, Excel	105	Field hut
Summer 2019	Joris's Microsoft Access GUI	51	Office

Industry due diligence

Before starting it was understood that there were many efficiencies that could further be applied to machine learning and trail camera footage review such as onboard processing and the use of a remote sensing network to relay notifications back to a field base. There was also awareness that leading R&D parties in New Zealand conservation technology had already made the deliberate switch from trail cameras using infrared to thermal video.

Discussions with Phil Bell and John Wilks from ZIP took place in late 2019 along with a visit to their facilities at Lincoln to learn more about their research and development. ZIP have thoroughly investigated standard infrared trail cameras and explained the return on investment from start to finish of the footage review process, including equipment and staff. ZIP found that thermal cameras taking downward facing videos was the most viable option. ZIP plan to test these customized thermal video cameras with onboard machine learning and LoRa network on the West Coast. The potential and limitations of using the LoRa remote sensing network on Auckland Island were discussed including terrain limitations and need to rebait the cameras regularly. A follow up meeting is planned once the next round of ZIP's thermal trail camera testing is complete.

Shaun Ryan from the Cacophony Project provided advice and two of the Cacophany Project's thermal video camera prototypes were trialled during the Winter 2019 trip. Shaun demonstrated the web interface that housed the AI classified videos. The easy to use data management system, the seamless transition of data from the camera into the cloud and the potential of thermal cameras to remove false triggers were attractive features. Unfortunately, field staff were only able to get the cameras working the last week of the trial due to software update syncing issues – connectivity issues are an important consideration for technology in remote places. The prototypes in their current state were not user friendly and critical aspects like battery life require further to be useful in a remote island eradication situation. Developments of the Cacophony Project will be passively monitored.

In addition, investigation into ClassifyMe, a machine learning classification software, has been undertaken after gaining access to a free licence. ClassifyMe originates in Australia and so specialises in Australian pests and species. Their New Zealand model target kiwi, hedgehog, stoats, cats, birds and sheep is still in its early stages. It can't currently identify species frequently seen on Auckland Island like sea lions and penguins. One of its strengths is that it can be used offline with the intention of running one SD card at a time, not batch processing. Staff found it easy to use, and its project lead, Greg Falzon, was very responsive to queries and issues. Al Glen from Landcare Research is working with them on the New Zealand version. He has been gathering training data from various parties at DOC. ClassifyMe is the most well-known AI based classification software in New Zealand. We have included their results in Table 4 and 5.

The University of Canterbury's Richard Green was introduced to us. He further introduced us to one of his postdoctoral researchers who had written an object detection model. We met with the researcher and shared our data with him, however his algorithm ran in to bugs. Communication eventually died off possibly due to the structure of the academic year, contracts and coronavirus. Not working with more academic institutions was a missed opportunity for collaboration and research.

Island Conservation and the RSPB, who also run eradication projects, suggested trialling Microsoft AI for Earth's (https://www.microsoft.com/en-us/ai/ai-for-earth) Megadetector, an open source classification model that is free for anyone to use and download from Github. Their model has been trained with millions of images already. The project lead, Dan Morris, is deeply ingrained in the artificial intelligence and machine learning trail camera world. Microsoft AI for Earth's results are included in results Table 4. Island Conservation and the RSPB use it for aiding the reviewer, not replacing the reviewer. A bounding box with classification and confidence level are an output on each image. This reduces the amount of searching and decision making the reviewer must do, turning the reviewer into a verifier.

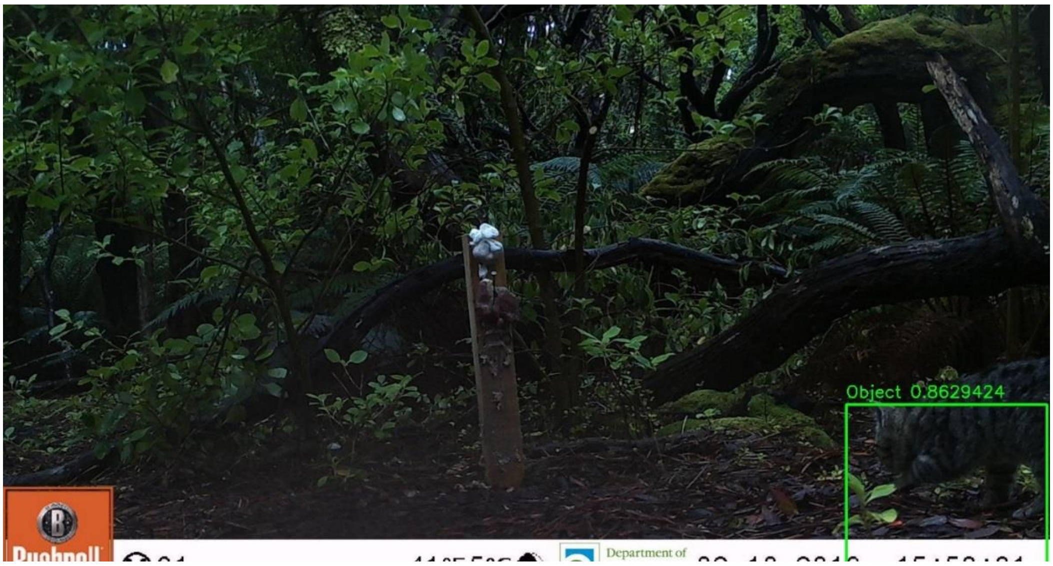

Figure 2 Bounding box. Object identified.

Internal investigations within DOC were undertaken to assess how many other projects were running trail camera grids and their methods. DOC has a trail camera standard operation procedure (SOP) drafted by Craig Gillies (DOC-5737005). DOC's internal L/Animal_Pest mailing list was used to find out what other projects were running camera grids, grid size, target species, footage review time, etc. Results showed each project has their software of choice, own methods for data management and footage review. There were at least 50 projects running trail camera grids. It was clear, rangers using trail camera grids wanted a faster more efficient process to review camera grid footage but didn't have the time, funding, managerial support or technical knowledge to pursue a solution. The list is available here - DOC-6161160

Machine learning developers

The purpose of engaging with the machine learning market was to understand current capability. To do this we shared our data and asked developers test it on their models.

About the developers

Developers were found via existing networks, mainly the DOC GIS and ISS channels and subsequently via the AI Forum (aiforum.org.nz). There was much enthusiasm and interest because of the challenges involved with a remote island eradication project, and the fact that the dataset was fully labelled (i.e. humans had already identified what is in each of the images). Participation was voluntary. Any developers who agreed to share their results would in turn be able to see other developers' shared results. Many of these developers had previously contracted to DOC through other projects.

Table 3 Participating developer's list

Participant name	Website	About
Andrew Digby (DOC)	https://twitter.com/takapodigs	• Lead scientist for the takahe and kakapo programmes • Also works with lots of trail camera footage • Made model using CreateML from Apple's Developer kit on his MacBook Pro
BECA	https://www.beca.com/ai	• Traditionally an engineering firm • Created model that can be taken offline • Spoke at Leading Emerging Technologies workshop for SLT • Keen on working more with conservation & environment
ClassifyMe	https://classifymeapp.com/	• Software already made before this project • Designed to be a scouting tool • A collaboration between the New South Wales Department of Primary Industries and University of New England • Created for the field worker (offline, normal laptop) • New Zealand model still in early stages • Connection with Landcare Research/ Al Glen
Lynker	https://www.lynker-analytics.com/	• Specialist expertise in data science, geospatial analytics, and machine learning • Has contracted for DOC GIS team before • Made good recommendations for how to further develop model and software/hardware to use • Efficient processing and turnaround time
Microsoft AI for Earth	https://www.microsoft.com/en-us/ai/ai-for-earth	• Recommended by Island Conservation and RSPB • Very active internationally in artificial intelligence/ machine learning trail camera footage community • Open source model • Offers grants and training • https://agentmorris.github.io/camera-trap-ml-survey/ • Current model not capable of cat classification • Fastest results turn around
Nelson Artificial Intelligence Institute	https://www.nai.org.nz/	• Emphasized previous work in environment/ conservation • Provincial Growth Fund recipient • Has worked with NASA, Mars and other remote environments • Keen on working more with conservation/environment
Sagar Soni (Orbica)	https://orbica.world/	• Participated as an independent, not through his company • Has done AI contract work for DOC before • Keen on open source and collaboration

Process

A short context document about the data, how to access it and the objectives (DOC-6084653) was distributed. All developers were asked to use the Winter 2019 data set for training and Summer 2018/19 data set for testing. The desired outcomes were (in order of importance): a) removing false triggers; b) relabelling and moving into folders; and c) classifying cats (Table 1).

Developers were asked to send their results in spreadsheet form, advise time frames, detail hardware and software used as well as any feedback or recommendations. Many developers came back with questions and results from a first "go" before sending through their final results. Communication was available throughout the process. Timeframes varied as new developers were introduced to us over several months.

Data storage and sharing

The footage was divided into a training folder and testing folder. The training folder contained 15,558 images and 7,291 videos (98 GB) organised into object specific subfolders. The testing folder contained 150,120 images (81 GB) organised by CameraID.

The data was stored in the DOC Amazon Web Services (AWS) S3 Bucket, an ISS approved cloud storage system with adjustable permission levels, which makes it accessible to parties outside of DOC.

Results

We compared developer results to those manually reviewed by the Maukahuka field staff and Joris's office staff. BECA, Sagar Soni (from Orbica) and the ClassifyMe software reported results for both object detection and cat classification. Most developers only used a subset of data. Lynker and NAI were the only developers to run the entire data set.

Object detection

Three developers focused on object detection. The primary objective was to remove images without objects of interest. Two of the developers used Microsoft Azure, one used Tensorflow. ClassifyMe's is based on darknet and YOLOv2.

The Maukahuka field team took the longest time to process images at 7.34 seconds for trial 1. This is an average per image which includes renaming, moving folders, extracting metadata and putting reviewed data into the spreadsheet. When looking at this figure it is important consider that the Maukahuka field staff were identifying cats down to individuals, whereas object detection models by the developers were not asked to do this. Although still requiring review by humans, Joris's GUI reduced average time per image by 6 seconds. BECA had the fastest processing time at 0.16 seconds per image.

ClassifyMe had the highest precision rate by far (99%) but lowest recall (45%), and BECA the highest recall rate (88%), but just one percent higher than Sagar (87%). Manually reviewed footage would be near 100% precision and recall rate because each individual image was checked by humans (not allowing for human error). That is the rate a near perfect model would have and what is required for an eradication project. The lower the precision is, the more images that must be reviewed, which in eradication project, is acceptable if that means minimizing risk of missing a cat.

Comprehensive results can be found in the Appendix.

Table 4 Summary of results for object detection (i.e. removal of false triggers/empties from data set)

Source	Method	Software	Processing time in seconds per image	Precision	Recall
DOC -Maukahuka team	Manual in the field	Excel	7.34 seconds	100%	100%
DOC - Joris team	Manual in the office	Microsoft Access GUI	1.25 seconds	100%	100%
Microsoft AI for Earth	Machine learning	Azure/ Megadetector	0.80 seconds	44%	78%
BECA	Machine learning	Azure Custom Vision	0.16 seconds	62%	88%
Sagar (Orbica)	Machine learning	Tensorflow (Transfer learning with YOLO)	0.37 seconds	93%	87%
ClassifyMe	Machine learning	ClassifyMe	1.60 seconds	99%	45%

Cat classification

Five developers plus the software ClassifyMe produced results for cat classification. This answers the targeted question, *Is there a cat in the image?* For cat classification, Andrew Digby from the kakapo and takahe team also created a model using CreateML from his MacBook Pro. Three of the developers used TensorFlow and one used Microsoft Azure. ClassifyMe's is based on darknet and YOLOv2.

Nelson Artificial Intelligence Institute (NAI) had the fastest processing time at 0.01 seconds as well as the highest recall rate 96%. Lynker had the second fastest processing time at 0.05 seconds. In general, processing times for cat classification were faster than for object detection. Sagar from Orbica had the highest precision rate at 87%. ClassifyMe had the lowest recall and precision rate at 6% and 3%. ClassifyMe was the only generic model listed below that was not custom made based off of the Auckland Island data. It would be interesting to compare what ClassifyMe's cat training data set looks like in comparison to the Auckland Island cat training set. Important to note is ClassifyMe was not able to process the intended amount of images and threw errors when we tried other variations of camera data. As a result a small, random subset of 500 images were used for its testing. ClassifyMe had a high accuracy rate in identifying birds.

Table 5 Summary of results for cat classification (ie Is there a cat in the image?)

Source	Method	Software	Processing time in seconds per image	Precision	Recall
DOC -Maukahuka team	Manual in the field	Excel	7.34	100%	100%
DOC – Joris team	Manual in the office	Microsoft Access GUI	1.25	100%	100%
DOC - Andrew Digby	Machine learning	CreateML	0.14	48%	46%
BECA	Machine learning	Azure Custom Vision	0.10	21%	88%
Lynker	Machine learning	TensorFlow	0.05	25%	86%
Nelson Artificial Intelligence Institute	Deep learning	Tensorflow	0.01	17%	96%
Sagar (Orbica)	Machine learning	Tensorflow (Transfer learning with YOLO)	0.36	87%	81%
ClassifyMe	Machine learning	ClassifyMe	1.64	6%	3%

Developer results and reports are available for DOC staff to review here - DOC-6293363.

Discussion

The review into optimizing the manual footage review process was a success because the project reached out to its network both internal and external to DOC. The increase of capability and speed in footage review using Joris Tinneman's Microsoft Access GUI raises the potential benefits for a higher data allocation for satellite internet to send footage back to the mainland and have it reviewed there.

The research into the application of machine learning on trail camera footage review would have been more effective if we had had someone with artificial intelligence and machine learning expertise leading it. Timeframes were longer than anticipated due to the need to prioritise this work within a broader work programme and substantial ongoing learning as data was prepared, shared and results received from developers. Reaching out internally within DOC proved fruitful and it was evident there is a want and need for footage review processing to be streamlined and the agency as a whole could benefit from centralized collaboration, rather than individual projects reinventing the wheel.

Data format

The desired objectives were known but the developer's requirements were only understood at a high level and many intricacies emerged such as how to share the data within the constraints of the Department. Should all the cat photos be in one folder? What if there's only a tail? Should the images be in sequential order because the images have been taken in 3 second bursts?

In addition, data preparation prior to sharing with developers took much longer than anticipated and didn't receive enough importance because its ramifications weren't fully understood at the time. Both the Summer 2018/19 and Winter 2019 datasets were reviewed twice. It was subsequently learned that even after office staff had moved all the cat images into a cat specific folder for developer training data, it would have been more useful to the developers to have a subfolder within the cat folder segregating cat_full and cat_partial photos. This folder also should have excluded any empty images, however there were a handful of empty images considered as containing cats because they were part of a three burst sequence where a cat was seen. Training a young model on what a cat tail is when it does not understand what a cat is, is not useful.

In the end, many of the developers downloaded cat images from the internet and used these as training data. There are many trail camera repositories accessible for free online for this very purpose.

Figure 3: Image included in cat training set

A crucial shortfall was the lack of sufficient training data. Training datasets for machine learning should at the very least be 3 parts training to 1 part testing, preferably 4:1. This issue was raised by multiple developers.

The number of images in the Winter 2019 dataset were far less than the Summer 2018/19 dataset because the former were collected as 1 image, 10 second video, 1 image vs 3 second burst. The priority for the winter trial was to see how the cats interacted with the bait, and this was best observed by video, not image. The models created by the developers were all image based and did not process videos. Videos could be converted into images, frame by frame. At present, the plan for the eradication is to use images only.

In hindsight, contacting developers and sharing data should have been delayed until the Summer 2018/19 dataset was reviewed again and that data should have become the training set. Winter 2019 should have been the testing set.

Multiple developers raised concerns that most of the cat training images were taken at night time, and in contrast, most of the test images with cats were from the daytime. Because this wasn't like for like, this prolonged the process. In order to use the same model for both types of images, some developers reprocessed the images to increase colour contrast. Developers had to put more emphasis on teaching the model what an image without a cat looked like and not rely solely on teaching it what a cat is. This may be why some developers chose to classify object vs no object rather than going into the further detail of cat vs no cat.

Lesson learnt – *Training to test dataset must be 4:1.*

Sharing data with external parties

It was not easy to find and get access to a DOC approved portal to share the large datasets externally. The pathway was not clearly defined. Eventually access was given to the Amazon Web Services (AWS) S3 Bucket. Most of the developers were not familiar with the AWS interface causing significant additional effort and a quick drop off in interest.

Lesson learnt - *The S3 Bucket is a steep learning curve for those unfamiliar with AWS. This will likely prevent any non-AWS (ie Microsoft Azure) users from working with DOC if work is being done on a pro-bono basis.*

Communication with developers

Requirements were not clearly articulated at the outset, both because staff did not know what/how the project wanted the results and the layered nature of the problem making it difficult to balance the requirements with the pro-bono nature of the work. This lack of clarity was apparent when the results were received from the developers who also have their own development priorities.

Results were received in various formats and metrics, which meant results were not like for like. Some groups only ran algorithms for detecting cat vs no cat, while others focused on object vs no object, sometimes precision/recall was included, other times not.

All of the developers ended up testing a different number of images of their own selection. Lynker and NAI were the only developers to run through the entire testing data set. This had been requested of all developers and outlined in the context document. This is likely due to the training and testing dataset not being at least 3:1, cat folders containing full cat shots, tails and empties.

More tidying and standardizing of the results was required than anticipated to be able to compare all results.

Lesson learnt – *Clearly define the minimum requirements and output format*

Model shortfalls

Microsoft AI for Earth reported issues with its Megadetector model continuously classifying the bait post as an animal, however issues similar to this were recurring in datasets unrelated to the Maukahuka project. Microsoft AI for Earth had already developed another model that ignores the bait post. Another reported shortfall by a different developer was the model not being sophisticated enough to differentiate between a cat and a pig. This may not be a problem for the Maukahuka project however as the pig eradication is intended to take place prior to the cat eradication.

Figure 4 Misclassified bait post

Developer recommendations

The developers were asked for their feedback and suggestions for how to improve the process and make it work in a remote island location. Along with the discussion above, some developers said they would have liked to train a daytime model and a night time model for each individual camera location. Much more images were needed for training, and of these, the developers would have liked to have more cat images from different angles.

Microsoft Azure models, which can be created online and then downloaded for offline use was suggested by one developer. Using edge device (a mobile phone) to process the footage at the camera, a familiar piece of equipment and cost-effective option was suggested by more than one developer. This is informative feedback to be considered when designing eradication methodology.

Diving deeper into conversations with the developers, led to questioning of the approach. The appendix contains a table with more detailed developer results, which includes the factors, true positive, true negative, false positive and false negative, that calculate the precision and recall. More than one developer pointed out, even if recall was a few percentages shy of 100%, there's still a short fall of possibly one or two cats missing. There is also the known risk that manual footage review is not 100% accurate because humans make mistakes too. What's the fastest way to check those photos to find the possibly missing cats? There isn't a fastest way. It would still require staff to manually review each image in the entire dataset to find two cats that may or may not be there. However, manual review has improved significantly in the course of the work (6-fold).

For eradication, one missed cat can be the difference between a success and a failure. One missed image of a cat isn't necessarily a failed eradication. If a cat is crossing the camera screen to reach the bait, using the three burst image methodology and the machine is reviewing the images sequentially, the cat is likely to be picked up. Cats are cryptic and some are skittish around the camera, approaching it, but not entering the entire screen. Sometimes only a tail or a limb can be seen. How will the Maukahuka team counteract that? Is the Maukahuka team approaching machine learning for trail camera footage review from the wrong angle? The argument for thermal trail cameras with video may not be the solution for the Maukahuka project however. Both ZIP and the Cacophany Project who are investigating this technology are targeting rodents and mustelids, smaller pests with different movements.

Eradication methodology relies on overlapping tools. Every cat must be put at risk. Therefore, every cat home range must have at least one detection device in it. Trail cameras are one of many monitoring tools that can contribute to making our monitoring methods more efficient. Island Conservation and the RSPB both run their camera grid images through the machine learning models, which output the images with a bounding box, classification and confidence percentage pertaining to the classification. Staff still look through each image, however this speeds up their processing time. Staff check if the image is wrong. Staff don't have to search for the object and only have to verify if the classification is correct rather than make the classification themselves.

Further development

A strategic approach is required by DOC to provide guidance on development (accountability, internal vs external), requirements (reliability, quality, cost, maintenance, island vs mainland, species), timeframes and budget. Many at DOC are sharing trail camera footage externally, although all are doing it individually. There is scope for efficiencies and greater outcomes through collaboration, however this cannot be driven by the Maukahuka team because of its project going on hold.

Opportunities external to DOC include:

- Al Glen at Landcare Research is working with ClassifyMe. While there is extensive anticipation about this project as a possible solution, no results have yet been shared and timeframes for the next version are unknown.

- ZIP and The Cacophony Project both have given up on infrared trail cameras and are pursuing trail cameras with thermal videos with onboard processing and machine learning, however via different routes. Their target species are rodents and mustelids, which differ from the Maukahuka project's target species.

- The Biological Heritage National Science Challenge has been suggested as an external means to continue this development work.

Next steps/ recommendations

The following recommendations are a result of conversations with developers, industry due diligence, machine learning's application to trail camera footage review and the developer results.

1. DOC should standardize its trail camera footage system. All projects should use the same folder structure (species, partial_, full_, testing, training, etc.), record sheet and file naming convention. This should be added to the DOC trail camera SOP. Lindsay Chan, Joris Tinnemans and Andrew Digby will discuss this and propose a standard to Craig Gillies, the lead for the DOC trail camera SOP. Input to the proposed standard can be sent out via the L/Pest distribution list.

2. DOC should save all its trail camera footage, using the format above, in a central, easily accessible location in the cloud. This enables all parties within DOC to access other footage, contribute and easily share with external parties such as universities and developers for training their algorithms. The *doc-trail-camera-footage* bucket exists in the Amazon S3 Bucket. This has been set up as a repository that has capacity for all other trail camera footage at DOC. Each project would require a new sub folder (created by ISS, who can also give staff AWS logins to access the bucket).

3. An individual or group should be appointed and be responsible for championing the new standards and keeping DOC abreast with the industry both in New Zealand and internationally, especially the research and development DOC has helped fund. These successes and failures should be made transparent. They should further access the size of the need, identify user case requirements for AI, coordinate trail camera knowledge sharing and keep DOC standards up to date with international development and protocols.

These first three recommendations would provide DOC with an extensive, standardised and useable training data set, significantly benefitting the creation of a machine learning model to identify target species. The ability to teach a model what an empty image/false trigger is would have cut out 80% of the images reviewed in the Summer 2018/19 data set.

Similar gains could be expected across other camera networks and projects realising large savings for DOC. The cost to implement these recommendations is low resulting in an extremely high return on investment.

4. Documentation of a user case to capture the need for an Auckland Island specific solution to automating trail camera footage review. This will be completed as part of the Maukahuka Project wrap up, so lessons are captured and knowledge can be passed on to enable progression of Maukahuka objectives prior to project initiation.

Conclusion

There was a lot of interest in and buy-in to the Maukahuka project and its interest in applying machine learning to trail camera footage review. Developers love a good challenge, especially in things that have a greater good and are generally interested in further involvement.

When comparing the processing time for machine learning models to person hours/cost of manual review, the results show significant efficiencies in processing time, saved dollars on field worker time and allows for mitigating health and safety risks.

Manual processing can be improved considerably with the aid of software to sort and label images and with the use of quick keys for image allocation. This knowledge needs to be shared department wide and a system should be set in place to make this sharing easier.

The results show that at the moment using machine learning to review the footage rather than staff to manually review trail camera footage is not *the* solution. Rather, it could be used as part of a broader solution to reduce field staff manually reviewing each image by pre-processing and adding in the bounding frame around objects picked up in the image as well as classifying and adding confidence rates. It could also compliment other detection tools but is not a standalone tool.

Considering Moore's Law, the theory that technology will exponentially improve, which in turn increases user demand and therefore decreases price, it would be wise to follow machine learning developments, especially in the case of advancements to work offline in remote areas and in the use of thermal cameras.

This technology is key to the Maukahuka cat programme and has potential to provide extensive added value to many other projects in DOC, with Predator Free New Zealand and further afield. How to effectively apply machine learning to trail camera footage review in an island eradication scenario and biosecurity in general has not yet been championed in the conservation world. Is it because the technology is not there yet? Is it because there isn't enough collaboration between camera grid staff and machine learning developers? Could it be an opportunity for DOC? A strategic approach to continue development of this tool would have a large return on investment for DOC.

Acknowledgements

We would like to acknowledge Joris Tinnemans and Andre Wilfert who created data management scripts that have saved us hours of work. Thank you for automating the process!

We would like to thank the participating developers for giving their time to help us learn about AI and machine learning. Not only did they share their results with us, but also included write ups, met with us in person, and patiently explained how the process works. We are grateful for their time in assisting the advancement of technology in conservation, and hopefully future predator control and eradication projects.

Reference List

Precision and Recall. (2020, June 9). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Precision_and_recall

Appendix

Table 1: Important links accessible internally at DOC

Title	Link
Task Assignment	DOC-6108668
List of trail camera projects at DOC	DOC-6161160
List of contacts	DOC-6306762
Context document shared w/ developers	DOC-6084653
Result reports from developers	DOC-6293363
Footage review tools used by Maukahuka	DOC-6293355
DOC trail camera footage	S3 Bucket
List of AI/ML trail camera resources	https://agentmorris.github.io/camera-trap-ml-survey/
Original file note	DOC-6127504

Cat classification table - detailed

Source	Method	Software	Training images	Training time	Testing images	Testing time per image in seconds	Testing time	True positive	True negative	False positive	False negative	Precision	Recall
DOC - Maukahuka team	Manual using excel	Field staff	N/A	N/A	147,107	7.34	300 hours	4448	N/A	N/A	N/A	100%	100%
DOC - Joris team	Manual using Microsoft Access GUI	Office staff	N/A	N/A	146,531	1.25	51 hours	4250	N/A	N/A	N/A	100%	100%
DOC - Andrew Digby	Machine learning	CreateML	15070	42 minutes	18,913	0.14	45 min					48%	46%
BECA	Machine learning	Azure Custom Vision	1975 (663 cat, 1312 empty)	2 hrs for a human, a few seconds to create the model	61,388	0.10	1 hr 45 min	13037 (61.71%)	38455 (95.51%)	8090 (38.29%)	1806 (4.49%)	21%	88%
Lynker	Machine learning	TensorFlow	Total: 1600 Cat: 400 No cat: 1200	12 hours	144,778	0.05	2 hours	317	143525	936	50	25%	86%
Nelson Artificial Intelligence Institute	Deep learning	Tensorflow	Cat: 3603 Empty: 7206	30 min	84,195	0.01	11.96 min	4254	94884	21520	172	17%	96%
Sagar (Orbica)	Machine learning	Tensorflow	150	2 hrs	5000	0.36	30 min	2196	1986	304	514	87%	81%
ClassifyMe	Machine learning	ClassifyMe	assumed millions from		513	1.64	14 min					6%	3%

Object detection table - detailed

Source	Method	Software	Training images	Training time	Testing images	Testing time per image per second	Testing time	True positive	True negative	False positive	False negative	Precision	Recall
DOC - Maukahuka team	Manual using excel	Field staff	N/A	N/A	147107	7.34	300 hours	16627		126968		100%	100%
DOC - Joris team	Manual using Microsoft Access GUI	Office staff	N/A	N/A	146531	1.25	51 hours	16647		129884		100%	100%
DOC - Andrew Digby	Machine learning	CreateML	N/A	N/A	N/A	N/A	N/A					N/A	N/A
Microsoft AI for Earth	Machine learning	Azure/ Megadetector	millions from prev projects	around 10 days	5000	0.80	0.80 s per image	769	4231	N/A	N/A	44%	78%
BECA	Machine learning	Azure Custom Vision	Total: 2172 Animal: 860 Empty: 1312	2 hrs for human, a few seconds to create the model	61388	0.16	2 hrs 45 min	3722 (21.21%)	43316 (98.81%)	13829 (78.79%)	521 (1.19%)	62%	88%
Sagar	Machine learning	Tensorflow (Transfer learning with YOLO)	300	3 hrs	4900	0.37	30 min	2236	2180	164	320	93%	87%
ClassifyMe	Machine learning	ClassifyMe	assumed millions from other projects		513	1.64	14 min					99%	45%