The MegaDetector: Large-Scale Deployment of Computer Vision for Conservation and Biodiversity Monitoring*

Sara Beery[1] [0000-0002-2544-1844]

California Institute of Technology, Pasadena, CA, 91125, USA

Abstract. Biodiversity data is being collected at unprecedented scales, and is no longer able to be efficiently and sustainably processed with human effort alone. Camera traps, motion-activated static cameras used for long-term studies of ecosystems and the wildlife within them, are a prime example of this, collecting hundred of thousands of images per year for each survey. The camera trap community has increasingly turned to computer vision to aide in processing, but the challenges of domain generalization in CV led to the need for in-house ML engineers in order to train usable, reliable model for each project. This doesn't scale, so we performed a systematic analysis of generalization and sought to deploy a model that would serve the needs of diverse organizations worldwide by increasing human efficiency collaboratively with AI, as opposed to fully automating camera trap data processing. The focus on off-the-shelf generalizability and accessibility led to a a model that has been widely adopted, and is used by over 60 organization and NGOs globally. In this chapter, we discuss the development of the MegaDetector and introduce five diverse end users with different needs and target uses and discuss what made the MegaDetector accessible to them and how it has impacted their conservation and biodiversity work.

Keywords: Computer Vision · Camera Trapping · Biodiversity Monitoring · Data Science · ML Deployment.

1 Introduction

As the planet changes due to urbanization and climate change, biodiversity worldwide is in decline. We are currently witnessing an estimated rate of species loss that is up to 200 times greater than historical rates [12]. Monitoring biodiversity quantitatively can help us understand the connections between species decline and pollution, exploitation, urbanization, global warming, and conservation policy [10, 11].

* Supported by Microsoft AI for Earth and the Caltech Resnick Sustainability Institute.

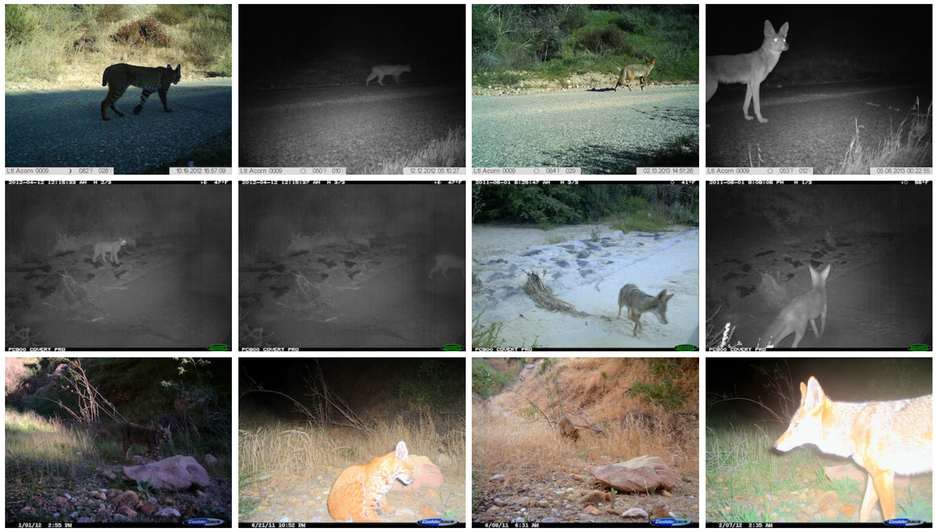


Fig. 1. Examples of camera trap images from three different locations in the Southwest United States. Each row is a different location and a different camera type. The first two cameras use IR, while the third row used white flash. The first two columns are images of bobcats, the next two columns are coyotes. This figure was originally published in [7] and is reproduced with permission.

We collaborate with researchers who study the effect of these factors on wild animal populations by monitoring changes in species diversity, population density, and behavioral patterns. In the past, much of this research was done via field studies, where biologists and zoologists would study animals in the wild by direct observation. However, once cameras became relatively inexpensive and easy to use, many scientists turned to motion-activated *camera traps* as an efficient, cost-effective, non-invasive method to collect experimental data. Networks of these static passive monitoring cameras are used to monitor changes in biodiversity, detect endangered and threatened species, build species distribution models, and estimate animal population sizes.

These cameras collect vast amounts of data, a single network of cameras can sometimes collect upwards of a million images per year. Manual processing of these images is a significant bottleneck to ecological monitoring and understanding in a reasonable timeframe. Our model, the **MegaDetector**, provides robust and geographically and taxonomically generalizeable animal, human, and vehicle detection in camera trap data. The model is widely adopted by the camera trap community, with use by over 60 different organizations and research groups worldwide, and has been found by organizations to drastically reduce their data processing time and costs, sometimes by up to 90%.

For this case study, we interviewed MegaDetector users from five different global organizations:

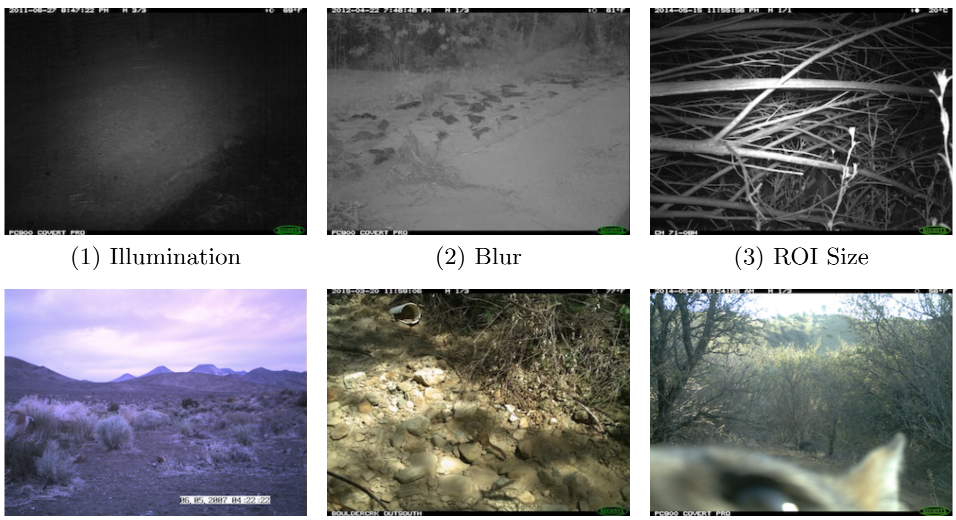
– Dylan Bergman, Point no Point Treaty Council, USA
– Tavis Forrester, Oregon Fish and Wildlife, USA
– Itai Namir and Ron Chen, Hamaarag, Israel
– Ben Pitcher, Brendan Altig, and Neil Jordan, University of New South Wales, Australia
– Damien Kerr and Ivory Lu, Australian Wildlife Council, Australia

Each organization reflected on their ecological use case for camera trap monitoring, how they got started using the MegaDetector, their current data processing workflow, and the impact the MegaDetector has had on their conservation efforts.

2 Problem Statement

Camera traps are one of the most widely used ecological monitoring sensors, with an At present, most camera trap images collected globally are annotated by hand, and the time required to sort images severely limits data scale and research productivity. Our collaborators estimate that they can annotate around 3 images/minute, and spend up to 500 hours per project on data annotation. Annotation of camera trap photos is not only time consuming, but it is also challenging. Because the images are taken automatically based on a triggered sensor, there is no guarantee that the animal will be centered, focused, well-lit, or an appropriate scale (they can be either very close or very far from the camera, each causing its own problems, see Figure 2 for more details) [25]. Further, up to 90% of the photos at any given location are triggered by something other than an animal, such as wind in the trees, a passing car, or a hiker.

Ideally, all camera trap data processing could happen automatically, including filtering out empty images, categorizing animals to species, counting individual animals across detection events, recognizing animal behaviors across taxa, detecting the spread of disease, etc. The computer vision community has tackled many of these challenges in the academic literature and shown highly promising results, with reported accuracy matching or surpassing human experts [24, 33, 31, 13, 19, 26, 34, 35, 20, 15, 32, 29, 22]. However, with some exceptions [30, 27, 21], many of the previous studies of automated data processing have used the same camera locations for both training and testing the performance of an automated system, thus failing to evaluate the ability of machine learning and computer vision models for camera trap data to *generalize* to new environments, new camera deployments, new sensor types, and new projects. This generalization to "out of distribution" (OOD) settings not seen during training is a known challenge for computer vision and machine learning (see Figure 3). In practice, we have found that most of these models have significantly overfit to their training datasets and that generalization performance is shockingly poor due to shifts in both visual and supopulation distributions between regions [7, 18, 25], compounded by data quality issues, challenges recognizing rare species with few training examples [28, 5], and the need to handle novel species never seen during training [25]. This leads to a lack of useability of models by new organizations without the need



(1) Illumination (2) Blur (3) ROI Size

(4) Occlusion (5) Camouflage (6) Perspective

Fig. 2. Common data challenges: (1) **Illumination**: Animals are not always salient. (2) **Motion blur**: common with poor illumination at night. (3) **Size of the region of interest** (ROI): Animals can be small or far from the camera. (4) **Occlusion**: e.g. by bushes or rocks. (5) **Camouflage**: decreases saliency in animals' natural habitat. (6) **Perspective**: Animals can be close to the camera, resulting in partial views of the body. This figure was originally published in [7] and is reproduced with permission.

(A) **Cow: 0.99**, Pasture: 0.99, Grass: 0.99, No Person: 0.98, Mammal: 0.98

(B) No Person: 0.99, Water: 0.98, Beach: 0.97, Outdoors: 0.97, Seashore: 0.97

(C) No Person: 0.97, **Mammal: 0.96**, Water: 0.94, Beach: 0.94, Two: 0.94

Fig. 3. Recognition algorithms generalize poorly to new environments. Cows in ‘common’ contexts (e.g. Alpine pastures) are detected and classified correctly (A), while cows in uncommon contexts (beach, waves and boat) are not detected (B) or classified poorly (C). Top five labels and confidence produced by ClarifAI.com shown. This figure was originally published in [7] and is reproduced with permission.

for custom model training, limiting the potential for scalable impact of many published computer vision models for camera trap data. In order to deploy automated, AI-based solutions that can be used across organizations, we need to build open source models that are generalizeable as well as provide accessible infrastructure to enable the use of those models easily without a background in data science or computer vision.

3 Method

Camera traps perfectly represent the challenge of *domain generalization* for computer vision [7]. Most benchmark datasets in machine learning and computer vision are what are considered ”in domain” - the training and test data is drawn from the same underlying distribution, and thus the assumption that the test data will be independently and identically distributed (IID) to the training data is a valid one. This assumption form the basis of much machine learning theory. However, the real world is seldom IID. If we consider specifically species identification from images, there is spatiotemporal structure to the underlying distribution of species which causes subpopulation shifts in the likelihood of seeing one or another species based on where the data was collected. Further, in the case of static passive monitoring sensors such as camera traps, there is additional correlation between the individual sensor location and what species are captured based on the behavior of individual nearby animals and their territories and habits, as well as visual correlations based on the type and placement of the sensor and the static background and microhabitat. All of these factors cause a notable drop in performance of state-of-the-art computer vision models when evaluated on sensor locations not seen during training, even if those sensors are located in regions that have been represented in the training set.

[8] formulated a systematic evaluation protocol and benchmark dataset, Caltech Camera Traps, to measure this ”generalization gap” in performance, based on evaluating models on both seen (cis) and unseen (trans) locations and comparing the results. As can be seen in Figure 4, there is a significant dropoff in performance on held-out camera locations, even when classifiying across sequences of frames taken in rapid succession, or classifying on close-cropped images of the animals in question as opposed to the entire image frame. This implies that building species classification models that will work off-the-shelf for new camera trap practitioners requires significantly more training data than building species classification models for a specific, fixed deployment, and that in-house training of custom species classification models for specific projects and deployments is still necessary, which is inaccessible for many ecologists working at NGOs or governmental agencies.

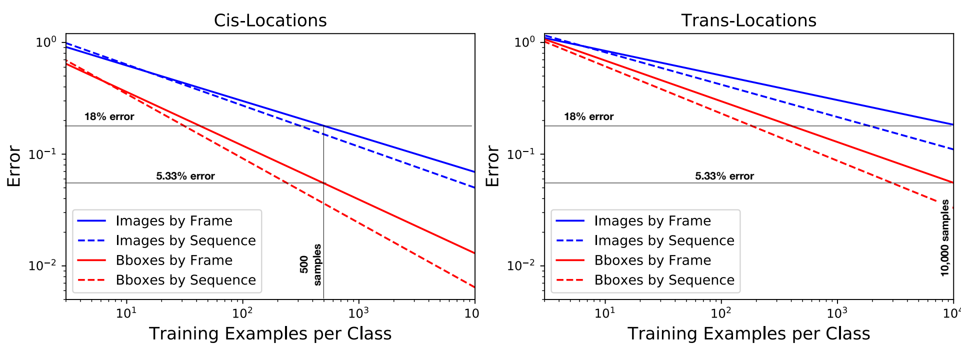


Fig. 4. Classification error vs. number of class-specific training examples. Error is calculated as 1 - AUC (area under the precision-recall curve). Best-fit lines through the error-vs-n.examples points for each class in each scenario (points omitted for clarity), with average $r^2 = 0.261$. As expected, error decreases as a function of the number of training examples. This is true both for image classification (blue) and bounding-box classification (red) on both cis-locations and trans-locations. However, trans-locations show significantly higher error rates. To operate at an error rate of 5.33% on bounding boxes or 18% on images at the cis-locations we need 500 training examples, while we need 10,000 training examples to achieve the same error rate at the trans-locations, a 20x increase in data. This figure was originally published in [7] and is reproduced with permission.

However, the silver lining that was discovered in [7] is that while species categorization still struggles to generalize, that class-agnostic animal detection in fact generalizes quite well. Figure 5 shows the surprising lack of generalization gap for two different object detection architectures trained on the CCT-20 benchmark. This is promising, as robust and generalizeable animal detection has the capacity to significantly reduce human effort in processing camera trap data by reliably filtering out empty images which can be 70-90% if the data collected at any given sensor. Though not able to fully automate camera trap data processing, we could see the potential for impact as a generalizeable model would not need to be re-trained by an expert for every project. Thus the MegaDetector project [6] was born.

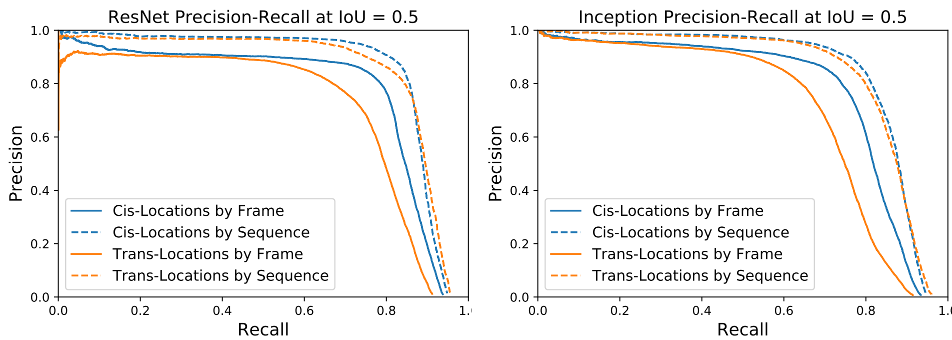


Fig. 5. Faster-RCNN precision-recall curves at an IoU of 0.5, by frame and by sequence, using a confidence-based approach to determine which frame should represent the sequence. This figure was originally published in [7] and is reproduced with permission.

The first step in MegaDetector development was the curation of a more diverse training and evaluation dataset, with examples of camera trap data far beyond the American Southwest (the location of Caltech Camera Traps). Partnering with multiple institutions, including the Snapshot Safari project, the Wildlife Conservation Society, the Nature Conservancy, the Idaho Department of Fish and Game, and others, data was curated and labeled for detector training with bounding boxes around each animal. When possible, the curated data from each organization, along with the bounding box labels, was published on LILA.science, an open repository for machine learning training datasets for biodiversity, ecology, and conservation applications. An evaluation dataset made up of held-out camera locations from each region was carefully constructed to ensure the generalization seen on CCT would hold, even for regions and species not seen during training.

The first MegaDetector model was trained in the summer of 2018 and both model and code was open-sourced in the Microsoft Camera Traps Github repository. That first model included only "animal" detection, and was trained to ignore humans and vehicles as background classes. Based on community interest and after careful evaluation of generalization, subsequent models added a "human" and a "vehicle" class. The categories supported in MegaDetector v2-v5 can be tracked in Figure 6. All versions of the MegaDetector before v5 were based on the Tensorflow Object Detection API [17] implementation of Faster R-CNN with an Inception backbone, while v5 switched to a Pytorch implementation of

YoloV5 for both easier training and evaluation, since Pytorch is a more accessible ML language, and more efficient inference with a 5x speedup in inference time from MDv4.

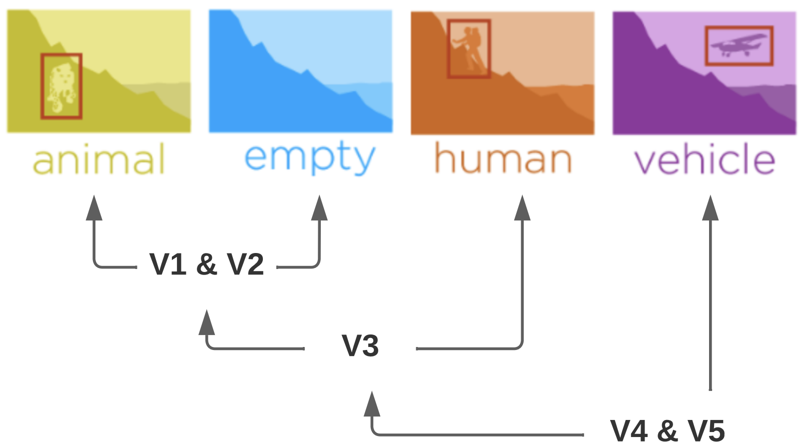


Fig. 6. MegaDetector has expanded its coverage over time, with 5 versions released so far.

Beyond just open-sourcing the training code and model versions, considerable effort was put into accessibility of use and community engagement. In order to support community users and stakeholders at different levels of python and machine learning literacy, individual image and batch APIs on Azure were also provided free of charge by Microsoft from 2019-2021, with all API code open sourced. We also worked with the developers of popular camera trap data management tools such as TimeLapse [2] and Camelot [1] to provide MegaDetector interactability within those tools, such as the ability to import MegaDetector results for easy visualization and correction in TimeLapse [16]. These efforts, along with the off-the-shelf usefulness of the model, led to large-scale adoption with over 60 NGOs, governmental agencies, and research groups using the model globally. Two end use cases are shown in Figure 7, emphasizing the breadth of end-use applications of the model. MegaDetector has also been a key aspect of numerous academic publications in computer vision [23, 4, 9] and ecology [14, 21], as well as incorporated into several open-source tools and GUIs for camera trap data processing such as TrapTagger [3].

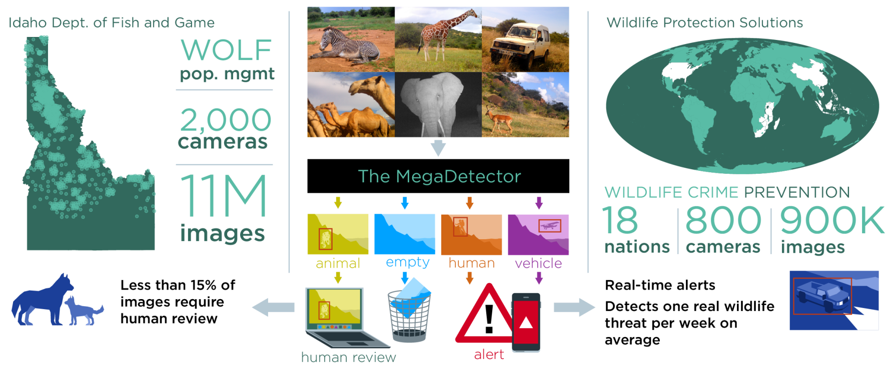


Fig. 7. The MegaDetector has been used to process data for over 60 NGOs, governmental agencies, and research groups worldwide, and was used to process over 100 million images in 2021 alone.

4 Risk Mitigation

MegaDetector is used by organizations to process data used to set wildlife management policy, stop the spread of invasive predators that are devastating to local wildlife, and make decisions on how to allocated limited resources for conservation. Systematic biases in the model learned from the training data, which are an inherent challenge for all current machine learning and computer vision systems, carry risk in this real-world scenario where model results are integral to decision-making. In order to mitigate these risks, the MegaDetector team works with organizations to determine whether MegaDetector is sufficiently accurate for their use case, and to set operating points for model confidence based on the risks of false positives vs false negatives for their specific use case. For example, when detecting invasive rodents on islands, which is how Island Conservation uses MegaDetector, a missed image of a rodent (a *false negative*) can have disastrous consequences for local nesting bird populations, so they use a lower detection confidence threshold and manually filter more images. In contrast, for the Idaho Department of Fish and Game, which run a network of over 2000 camera traps run on a timelapse setting across Idaho and collect tens of millions of images per year, mostly of elk and deer, a few missed elk is not a significant concern vs. the increased manual processing effort needed to filter through *false positives* when running the model at a lower confidence threshold. Our team works with different stakeholders to help them determine the optimal tradeoff between resources and risk for their use case. This is done primarily qualitatively, though if the organization has previously labeled data we can calculate quantitative metrics such as precision-recall curves. Without previously-labeled data, we sample a set of data from the organization, usually 10K images, and generate a simple html page that visualizes detections at different operating points on different pages. This way an ecologist can scroll through examples of that tradeoff between false positives and false negatives and build up an intuitive sense of model performance for their specific cameras and their specific use case before determining how to use the model. We have found that this initial communication of risk and model performance specific to an individual organization is key to the safe adoption of the model. In some cases, MegaDetector has been determined to be insufficient for a given use. For example, MegaDetector v4 struggled to correctly detect animals at camera traps that use bait stations to attract wildlife, as the animals were often at strange, out-of-distribution angles to the camera and the bait trap itself could result in high rates of false positives. We added a large amount of labeled data from bait traps to MegaDetector v5 training, which mitigated the issue. We also have a separate postprocessing script that can be run per-camera to detect possible repeated, static false positives caused by objects such as bait traps based on box position over time, which can then be easily visualized by an ecologist and removed en masse if found to be in error. We also encourage users to continue to monitor performance of the models over time with periodic quality control checks and to revisit the optimal operating point for their use, to ensure that model performance isn't drifting.

5 Resource Requirements

Many of the datasets used to train the MegaDetector, along with the bounding boxes collected for MegaDetector training, are publicly available at LILA.science. There are a few additional training datasets that were provided by organizations that were not able to share their data publicly, and because of ethical issues around lack of consent, all images found to contain humans are removed from datasets before publication on LILA. As a result, MegaDetector performance unfortunately cannot be replicated based on public data alone. All model versions were trained on Azure cloud GPUs based on open-source object detection repositories (Tensorflow Object Detection API for v1-4, Detectron for v5), with camera-trap-specific data preprocessing that is open-sourced along with all other training, evaluation, inference, and data management code on the Microsoft/CameraTraps github repository. Prior to 2022, Microsoft AI for Earth provided free MegaDetector inference via an Azure batch API. Though all the code to launch that API is open source, meaning that organizations can run their own instance of the API easily on Azure, Microsoft no longer financially supports the public use of the model. The MegaDetector team has launched a small nonprofit that processes data free of charge through the MegaDetector for smaller-scale organizations, but is unable to support significantly large-scale use on their GPU cluster. The need to pay for cloud GPU usage has proven to be cost-prohibitive for some larger organizations, and has resulted in a transition to the purchase of in-house GPUs and local processing to remove the need for reliance on industry to maintain sustainable and needed access to AI for their conservation applications. MegaDetector v5 was released specifically to reduce the computational overhead, and thus time and money, needed to run MegaDetector, as we found that YOLOv5 was able to replicate performance of the larger

Inception-backbone Faster R-CNN model trained in previous versions. We hope this lighter-weight model will increase the accessibility of MegaDetector by reducing computational and cost overheads of cloud-based use.

6 Field Evaluation

In this section we will highlight 5 different end users of MegaDetector, how they use the model, and the impact it has had on their organization. We conducted informational interviews with members of each organization, and found several main themes. First, though the models are intended to be as accessible as possible, most organizations worked directly with MegaDetector team members to get started using the model. Second, organizations with in-house data scientists have an advantage, as they are able to set up scalable data pipelines and increase the turnaround time for individual scientists within the organization. Third, that all users found that the MegaDetector was a game changer for their conservation efforts, not only enabling them to analyze data much more efficiently but also changing the way they thought about monitoring and what was possible. Fourth, that working with the MegaDetector gave each organization a much more nuanced understanding of machine learning as a conservation technology tool.

6.1 Point no Point Treaty Council

Dylan Bergman is a wildlife biologist with the Point no Point Treaty Council in Washington State, USA. The Council is a natural resource consortium that works for two Western Washington tribes, the Port Gamble and Jamestown S'Klallam Tribes, that are natural resource co-managers in Washington State. The tribes set their own hunting seasons and are legally empowered by the Point No Pont Treaty to help with Washington State resource management decisions. The tribes employ their own in-house wildlife biologists to help set wildlife policy, which enables wildlife conservation research and data-informed wildlife management for both the tribes and state agencies. Dylan primarily focuses on wildlife population estimation for deer, elk, and predator species such as bobcats, and sets the tribal hunting seasons for deer and elk. He has several camera trap grids that collect a timelapse photo every five minutes, on the order of a million images per season, and has been a MegaDetector user since 2019. He uses the model to filter out empty images and localize animals, which he then categorizes manually and uses to build space-to-event and time-to-event population models for deer, elk, bear, coyotes, bobcats, and mountain lions. His use of AI is heavily supported by the MegaDetector team, he uploads images to Azure to be processed via the batch API and then does verification and categorization through TimeLapse, and he can process up to 10K images a day by himself with the model in the loop. He estimates MegaDetector has contributed least a 50% speedup in his data processing workflow, and has enabled him to significantly expand his camera trap monitoring efforts without needing additional manpower for processing.

6.2 Oregon Department of Fish and Wildlife

Tavis Forrester is one of 8 wildlife biologists embedded in research units at the Oregon Department of Fish and Wildlife (ODFW). His research unit is housed in a Forest Service Pacific Northwest research lab, and their research is primarily wildlife management driven with a focus on predator/prey interactions and their ecosystem effects. They manage community science run camera grids across 6 states, as well as 200 in-house-managed cameras that collect around 1.5 million images a year. This 200-camera grid is being used to study cattle, deer, and elk species interaction inside an outside a fenced experimental range. They also have several GPS-collared animals, and analyze habitat use from both collars and cameras. ODFW got started with MegaDetector in 2020 after seeing the impact of its successful adoption of a MegaDetector/Timelapse workflow by the Idaho Department of Fish and Game. They adopted the same workflow, based on MegaDetector pre-processing to filter empty images and localize animals and quality control+species identification in TimeLapse, and have seen it reduce their camera trap data processing time by 2/3. They had a 4 year backlog when they started using MegaDetector, and now they have only a 1 year backlog and plan to be caught up before the data comes off the cameras this year, in 2022. Before working with MegaDetector they found that with the resources they had available it was impossible to process their data in a meaningful timeframe, so the impact has been invaluable in their ability to effectively manage wildlife populations. Similar to Point no Point Treaty Council, their use is supported by the MegaDetector team through Azure data upload and use of the batch processing API. Their use of MegaDetector has been informative of the significant value of AI in a human/AI interaction context, as opposed to the often-touted "AI do it for me" scenario that is seen hyped in the literature but rarely translates to practice. Tavis sees more and more value in systems that "handle the easy stuff and make humans more efficient", and has noted that people get really upset if the AI makes a single mistake, wheras before they were used to many mistakes coming from the inexpert data labelers they were forced to use due to lack of time and resources. One of the most aspects of incorporating MegaDetector is that ODFW can now label data efficiently enough that expert biologists can do the labeling, which significantly reduces human error in species identification previously coming from inexperienced undergraduates and technicians.

6.3 HaMARAAG

Itai Namir and Ron Chen work for HaMARAAG, a web of different ecological research groups in Israel established in the mid-2000s. They run a national wildlife monitoring program that started in 2013, motivated by a need to provide scientific information to decision-makers about natural ecosystems. They're now monitoring 9 ecosystems in Israel, with their main focus on a set of bioindicator taxa: birds, reptiles, plants, medium to large mammals, and arthropods. HaMARAAG also works with the Israeli Nature and Parks Authority to assess natural hazards such as oil spills and chemical spills, the Ministry of Agriculature, and many others. They are in the process of stablishing a complementary community science center to curate existing data and target the collection of new data, and are building a community science platform on top of Living Atlas. MegaDetector was recommended to them through eMammal, a Smithsonian community-science-driven camera trapping project, and they have processed over 1.2 million images with MegaDetector so far. They started with a workflow that was highly dependent on the MegaDetector team, uploading large batches of data to the cloud as they were collected from the field and processing via the batch API. They are currently in the development of their own in-house data processing workflow, and are purchasing GPUs to enable them to run the MegaDetector and perhaps also their own self-trained classification models going forward without relying on corporate cloud credits, which they hope will provide long-term sustainability for their ecological monitoring efforts. They find MegaDetector saves them a significant amount of time and therefore money, particularly filtering empty images. They also appreciate the reproducibility of an AI-based workflow as opposed to human labeling, though they have had concerns around the reproducibility and consistency of MD results as versions change over time which they have devoted resources to evaluate.

6.4 Australian Dingo Project

Ben Pitcher and Neil Jordan work with the Australian Dingo Project (ADP), and are affiliated with McMasters University and the University of New South Wales, respectively. The ADP is an offshoot of the Taronga Conservation Society, a governmental agency modeled after the Wildlife Conservation Society, that runs two public zoos about 600km apart. These zoos are the hands-on animal conservation arm of the Australian government, meaning they are responsible for wildlife management programs such as breed-for-release, and they have a small research organization which collaborates with academic partners. ADP runs several conservation projects, one of which is the Myall Lakes Dingo Project: a long-term dingo ecology project that involves local council and the Australian National Parks and uses MegaDetector as a key aspect of their monitoring efforts. They first heard about MegaDetector through a WILDLABS community Tech Tutorial, and leveraged the open-source Colab-based demo to run initial tests. When those were successful they moved to local-based processing at first on university GPUs which they used to tune the MegaDetector confidence threshold vs a manual test set, then got funding to develop their own cloud-based, large-scale processing pipeline in AWS which they used to process 1.5 million images. Out of that large set of images they found 48K images with animals that needed to be manually labeled, reducing their manual labeling effort by 97%. Their processing workflow uses MegaDetector to filter empty images and images containing humans or vehicles, and then uploads animal images to Zooniverse to be labeled to species by community scientists. They've also learned a lot via MegaDetector about what aspects of camera placement were leading to increased false positives, and have changed their placement strategies to avoid roads and paths, avoided angles with direct sun exposure, and started doing active vegetation clearing to avoid false positives from waving foliage.

6.5 Australian Wildlife Council

Damien Kerr and Ivory Lu work for the Australian Wildlife Council (AWC), a nonprofit that runs 30 wildlife sanctuaries covering over 12.5 million hectares of land across Australia which are formed via partnerships with indiginous and paturalist communities. They put over 100km-long feral-predator-proof fencing around the sanctuaries and remove feral cats and other feral predators, which have been found to eat or destroy up to a dozen native animals per day. They have camera traps monitoring all of their sanctuaries, but processing before using the MegaDetector was time consuming as 90% of their images are empty. Some sanctuaries are using camera traps to help clear a fenced area of feral predators, they have cameras out continually and need as close to real-time processing as possible, and only care about sightings of the feral predators they are attempting to eliminate. For them, false negatives are a significant issue, and they use MegaDetector with a lower confidence threshold to reduce the risk of missing a predator. Other sanctuaries are working on research around conservation outcomes within the fenced sanctuaries and run grids of 100 cameras, they want to detect and classify every species they see. For this use case they want fewer false positives and high accuracy, and are particularly interested in rare species. Only in the last 5-6 years has there been a national broadband network in Australia, before that the ecologists at each sanctuary did all their data classification manually and the data was held within site-based servers. If data needed to be centralized it was put on on a hard drive, driven 10 hours to the nearest post office, and then picked up by a mail plane which could sometimes take up to a week. Three years ago the broadband speeds finally increased to the point for it to be viable to move their large datasets to the cloud directly from the sanctuaries, which made cloud-based computer vision data processing a viable option for them. AWC had already collected over 300K labeled camera trap images from Australia that they shared with the MegaDetector team to be able to train MDv3. We labeled many of the images with bounding boxes for detector training, particularly from Northeast Australia which was an ecosystem we found the v2 model struggled with (lush, tropical rainforest). They started using the MegaDetector v3 model on their data in 2019 via a grant from Microsoft AI for Earth, however when they tried to use the model on arid landscapes they found it struggled on reptiles, so they again sent data to help expand the coverage and variability seen during training for the v4 model and saw a significant improvement. They have found that MegaDetector gives them an 89-90% speedup in processing their data. At one sanctuary they were able to process 500K images in about 70 hours, "This is a game changer for our ecologists." The ecologists have built up trust in the models over time based on systematic manual verification over time. They check MegaDetector's empty filtering based on confidence thresholds and process as they go, and over time have streamlined this verification process. They currently sample 1000 images at seeral different detection confidence operating points to verify performance of MD for each large batch. AWC previously used the batch API with some assistance from the MegaDetector team. Now they use their own centralized cloud-based datastore, and Ivory has built out documentation to help ecologists better use the MegaDetector API to get their own data processed in the cloud. AWC sanctuaries currently run their own data through the MegaDetector batch API on Azure, then they get the results back and use Timelapse locally at each sanctuary to ID to species.

7 Lessons learned

The key aspects of the MegaDetector that enable it's widespread adoption, use, and impact are (1) that it works reliably off-the-shelf for many organizations globally, (2) that it fills a significant need in the community (filtering empty images and localizing animals of interest), and its use saves organizations significant time and money when processing their camera trap data, and (3) that the model and code are open-source and many different access points to use are provided based on the skillset of the ecologist, from setting up data pipelines and running the model on their own in-house GPUs to recieving model results back from a batch API after a simple data upload that can be easily visualized and corrected within their existing camera trap processing workflows, incorporated within tools that are already widely used by the community such as TimeLapse. This demonstrates how rigorous and representative evaluation that carefully avoids potential overfitting or data poisoning due to strong spatiotempoal correlations between training and test data is vital to understanding the generalizability and thus useability of machine learning. The choice to reduce the granularity of the class set to just humans, animals, and vehicles based on systematic evaluation enabled the model to generalize more robustly and thus facilitated that off-the-shelf use. Further, this demonstrates the value of community engagement and direct collaboration with domain experts in order to understand and address domain-specific and project-specific risks, and to understand how to meet the community where they are and make automated approaches accessible for organizations and stakeholders with varied backgrounds, skillsets, and needs.

One significant lesson learned is that AI alone is far from enough. A good model is nothing without data infrastructure, including a scalable pipeline for data to migrate from the field to the model and tools for experts to interact with and correct imperfect model results. Another lesson learned is around the sustainability of industry collaboration: after Microsoft AI for Earth stopped covering the costs of the MegaDetector API in 2022, both AWC and HaMARAAG, high-throughput users, found API use to be cost-prohibitive at their scale of use and are in the process of switching to running on their own local servers of GPUs. Though this does require upfront investment, empowering conservation organizations to be self-reliant and in control of their use of AI-based solutions moves power back into the hands of stakeholders. Though in-house camera trap data processing with the MegaDetector is not accessible to all users, it is exciting to see some conservation organizations take this leap and even go beyond to training their own in-house species classification models for their fixed camera networks. Finally, ecological expertise is invaluable when moving beyond raw data processing to decision making, policy determination, and resource allocation. Enabling ecologists to be efficient with their time and interact with AI models in flexible ways is key to impactful and well-grounded AI solutions to conservation and biodiversity challenges.

References

1. Camelot. https://camelotproject.org/, accessed: 2022-7-8
2. Timelapse. https://saul.cpsc.ucalgary.ca/timelapse/, accessed: 2022-7-8
3. Traptagger. https://wildeyeconservation.org/traptagger/, accessed: 2022-7-8
4. Beery, S., Liu, Y., Morris, D., Piavis, J., Kapoor, A., Joshi, N., Meister, M., Perona, P.: Synthetic examples improve generalization for rare classes. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 863–873 (2020)
5. Beery, S., Liu, Y., Morris, D., Piavis, J., Kapoor, A., Meister, M., Perona, P.: Synthetic examples improve generalization for rare classes. arXiv preprint arXiv:1904.05916 (2019)
6. Beery, S., Morris, D., Yang, S.: Efficient pipeline for camera trap image review. In the Data Mining and Artificial Intelligence for Conservation Workshop at Knowledge Discovery in Databases (KDD) (2019 *selected to be featured at KDD Earth Day)
7. Beery, S., Van Horn, G., Perona, P.: Recognition in terra incognita. In: The European Conference on Computer Vision (ECCV) (September 2018)
8. Beery, S., Van Horn, G., Perona, P.: Recognition in terra incognita. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 456–473 (2018)
9. Beery, S., Wu, G., Rathod, V., Votel, R., Huang, J.: Context r-cnn: Long term temporal context for per-camera object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13075–13085 (2020)
10. Butchart, S.H., Walpole, M., Collen, B., Van Strien, A., Scharlemann, J.P., Almond, R.E., Baillie, J.E., Bomhard, B., Brown, C., Bruno, J., et al.: Global biodiversity: indicators of recent declines. Science p. 1187512 (2010)
11. Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., et al.: Biodiversity loss and its impact on humanity. Nature **486**(7401), 59 (2012)
12. CBD, S.: Global biodiversity outlook 1. In: Convention on Biological Diversity (2017)
13. Chen, G., Han, T.X., He, Z., Kays, R., Forrester, T.: Deep convolutional neural network based species recognition for wild animal monitoring. In: Image Processing (ICIP), 2014 IEEE International Conference on. pp. 858–862. IEEE (2014)
14. Fennell, M., Beirne, C., Burton, A.C.: Use of object detection in camera trap image identification: Assessing a method to rapidly and accurately classify human and animal detections for research and application in recreation ecology. Global Ecology and Conservation **35**, e02104 (2022)
15. Giraldo-Zuluaga, J.H., Salazar, A., Gomez, A., Diaz-Pulido, A.: Camera-trap images segmentation using multi-layer robust principal component analysis. The Visual Computer pp. 1–13 (2017)

16. Greenberg, S.: Automated image recognition for wildlife camera traps: Making it work for you. Tech. rep., Science (2020)
17. Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S., et al.: Speed/accuracy trade-offs for modern convolutional object detectors. In: IEEE CVPR (2017)
18. Koh, P.W., Sagawa, S., Marklund, H., Xie, S.M., Zhang, M., Balsubramani, A., Hu, W., Yasunaga, M., Phillips, R.L., Gao, I., et al.: Wilds: A benchmark of in-the-wild distribution shifts. In: International Conference on Machine Learning. pp. 5637–5664. PMLR (2021)
19. Lin, K.H., Khorrami, P., Wang, J., Hasegawa-Johnson, M., Huang, T.S.: Foreground object detection in highly dynamic scenes using saliency. In: Image Processing (ICIP), 2014 IEEE International Conference on. pp. 1125–1129. IEEE (2014)
20. Miguel, A., Beery, S., Flores, E., Klemesrud, L., Bayrakcismith, R.: Finding areas of motion in camera trap images. In: Image Processing (ICIP), 2016 IEEE International Conference on. pp. 1334–1338. IEEE (2016)
21. Norouzzadeh, M.S., Morris, D., Beery, S., Joshi, N., Jojic, N., Clune, J.: A deep active learning system for species identification and counting in camera trap images. Methods in ecology and evolution **12**(1), 150–161 (2021)
22. Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Packer, C., Clune, J.: Automatically identifying wild animals in camera trap images with deep learning. arXiv preprint arXiv:1703.05830 (2017)
23. Pantazis, O., Brostow, G.J., Jones, K.E., Mac Aodha, O.: Focus on the positives: Self-supervised learning for biodiversity monitoring. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10583–10592 (2021)
24. Ren, X., Han, T.X., He, Z.: Ensemble video object cut in highly dynamic scenes. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. pp. 1947–1954. IEEE (2013)
25. Schneider, S., Greenberg, S., Taylor, G.W., Kremer, S.C.: Three critical factors affecting automated image species recognition performance for camera traps. Ecology and evolution **10**(7), 3503–3517 (2020)
26. Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., Packer, C.: Snapshot serengeti, high-frequency annotated camera trap images of 40 mammalian species in an african savanna. Scientific data **2**, 150026 (2015)
27. Tabak, M.A., Norouzzadeh, M.S., Wolfson, D.W., Sweeney, S.J., VerCauteren, K.C., Snow, N.P., Halseth, J.M., Di Salvo, P.A., Lewis, J.S., White, M.D., et al.: Machine learning to classify animal species in camera trap images: Applications in ecology. Methods in Ecology and Evolution **10**(4), 585–590 (2019)
28. Van Horn, G., Perona, P.: The devil is in the tails: Fine-grained classification in the wild. arXiv preprint arXiv:1709.01450 (2017)
29. Villa, A.G., Salazar, A., Vargas, F.: Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. Ecological Informatics **41**, 24–32 (2017)
30. Whytock, R.C., Świeżewski, J., Zwerts, J.A., Bara-Słupski, T., Koumba Pambo, A.F., Rogala, M., Bahaa-el din, L., Boekee, K., Brittain, S., Cardoso, A.W., et al.: Robust ecological analysis of camera trap data labelled by a machine learning model. Methods in Ecology and Evolution **12**(6), 1080–1092 (2021)
31. Wilber, M.J., Scheirer, W.J., Leitner, P., Heflin, B., Zott, J., Reinke, D., Delaney, D.K., Boult, T.E.: Animal recognition in the mojave desert: Vision tools for field biologists. In: Applications of Computer Vision (WACV), 2013 IEEE Workshop on. pp. 206–213. IEEE (2013)

32. Yousif, H., Yuan, J., Kays, R., He, Z.: Fast human-animal detection from highly cluttered camera-trap images using joint background modeling and deep learning classification. In: Circuits and Systems (ISCAS), 2017 IEEE International Symposium on. pp. 1–4. IEEE (2017)
33. Yu, X., Wang, J., Kays, R., Jansen, P.A., Wang, T., Huang, T.: Automated identification of animal species in camera trap images. EURASIP Journal on Image and Video Processing **2013**(1), 52 (2013)
34. Zhang, Z., Han, T.X., He, Z.: Coupled ensemble graph cuts and object verification for animal segmentation from highly cluttered videos. In: Image Processing (ICIP), 2015 IEEE International Conference on. pp. 2830–2834. IEEE (2015)
35. Zhang, Z., He, Z., Cao, G., Cao, W.: Animal detection from highly cluttered natural scenes using spatiotemporal object region proposals and patch verification. IEEE Transactions on Multimedia **18**(10), 2079–2092 (2016)