Research Statement

Sara Beery

I develop computer vision methods that enable scientific understanding of life on earth, and I pioneer and solve novel challenges for computer vision. Working jointly with stakeholders, I deploy my methods to improve sustainability and conservation worldwide.

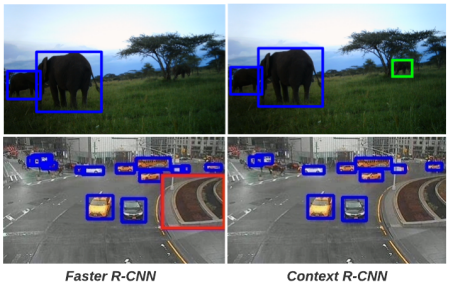
Advancements in computer vision (CV) have the potential to play a fundamental role in sustainability and conservation. We are currently witnessing an unprecedented loss of biodiversity [1], yet biodiversity is vital to sustainable development [2], public health [3], and mitigating climate change [4]. Earth observation, the gathering of information about the biological, physical, and chemical systems of the planet, is necessary for conservation and sustainability across spatial and temporal scales – micro to macro. I posit that CV, along with machine learning (ML) and data science (DS), will prove crucial to facilitate efficient extraction of scientific insight from quickly-growing repositories of natural world data.

In order to make a difference in the fields of conservation and sustainability we must shift the CV paradigm. Currently, CV research focuses on designing methods tailored to highly curated datasets. Such datasets frequently fail to capture the complexity of the real world, resulting in methods that fall short when deployed. Data captured in the wild presents a number of obstacles that current methods struggle to master, including strong spatiotemporal correlations, imperfect data quality, fine-grained categories, and long-tailed distributions. These challenges are shared by many real-world applications, and my methodological contributions confer benefit across domains [5].

My research program aims to solve the full range of problems to empower AI-assisted scientific discovery in a changing world. I ***collaborate with diverse stakeholders***, allowing me to ***identify universal challenges*** in conservation and sustainability. This understanding allows me to ***create benchmarks that matter*** – that truly capture the complexity of these real-world problems and enable the research community to come together and tackle them. Notable improvements on the benchmarks I develop translate to real-world gains, and I ***design novel methods*** that make progress on these challenges. I work with industry partners to ***build accessible, equitable Human-AI systems*** enabling the widespread use of my methods and models.

Past Research

Learning from imperfect, limited data *(CVPR, WACV)*
Data captured by human photographers, the norm in computer vision benchmarks, usually contains well-framed, in-focus objects of interest. In contrast, biodiversity monitoring data is often collected from sensors with limited intelligence, such as camera traps which collect data based on motion triggers. This leads to pictures where the animals are too close or too far from the sensor, low resolution, or obscured. Humans use temporal information, sometimes over long time horizons, to confidently ID species in challenging images. However, camera trap data is collected with a motion trigger and low frame



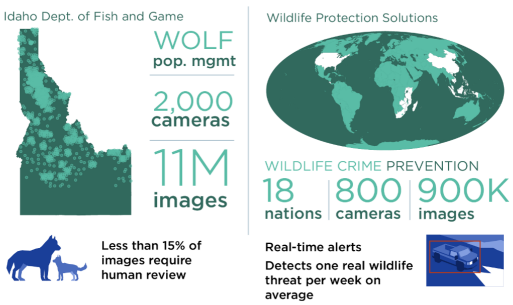
Context R-CNN [5] fixes challenging missed detections and removes salient false positives

rate, leading to failures from most video-based temporal approaches. We proposed Context R-CNN [5], which learns to attend to relevant context from up to a month of data at a single static sensor. ***Our model beats baseline methods by up to 17% mAP, and is able to correctly label cases with occlusion, poor lighting, even in severe fog.*** It is applicable to any static sensor – e.g. traffic cameras or home security systems. We are implementing Context R-CNN within the global-scale Wildlife Insights platform [6], where I serve as a core member of the AI team. Wildlife Insights provides data management and AI-assisted categorization for >900 users from 40 countries, and has ingested over 20M camera trap images globally.

The distribution of species worldwide is long-tailed: most observations are of common species, and the vast majority of species have few, if any, observations. This results in highly-imbalanced datasets, with insufficient data to learn rare species accurately [7]. Rare, at-risk species can be the most important to accurately categorize, but their rarity makes collecting additional training images for those species challenging – 14% of the International Union for Conservation of Nature (IUCN) Red List is considered data deficient [8]. We built a system to generate synthetic examples for rare species using modern game engines and used this synthetic data to ***decrease rare-class error by 70% without affecting performance on common species*** [9–11].

Measuring domain generalization under distribution shift *(ECCV, ICML).* Generalization to novel domains poses a fundamental challenge for computer vision. Near-perfect accuracy on benchmarks is now common [12], but these models do not work as expected when deployed. For example, models trained on a set of static cameras do not generalize to new cameras. We built a systematic framework and benchmark for analyzing generalization performance [7]. ***Our work promoted a paradigm shift in how CV researchers build image recognition benchmarks for real-world applications***, and our evaluation protocols have become the standard across data types including passive monitoring cameras [13], aerial surveys [14], and bioacoustics [15]. Challenges stemming from domain shift are ubiquitous in real-world problems, from medical diagnosis to code autocompletion. To better understand the generality of methods built to tackle these challenges *across domains*, we published the WILDS benchmark suite [16].

Domain-aware evaluation is crucial for understanding potential impact. Via our evaluation framework, we found class-agnostic animal detection generalizes far better to new deployments than species categorization [7]. Motivated by the community's need, we built a robust, generalizable animal detection model for camera trap data, the MegaDetector [17]. ***Our open-source API has processed over 100M images to date, and has been integrated into the wildlife monitoring workflows of over 40 organizations***, including The Nature Conservancy and San Diego Zoo Global. Our model is used as a key component in many computer vision papers for camera trap data [5, 9, 18, 19].



The MegaDetector is widely used, for everything from monitoring wolf populations in Idaho to detecting poaching threats across the globe.

Building human-AI systems for efficient, active, lifelong learning *(ME&E, CVPR, COMPASS).* Effective human-AI systems make human experts efficient, allowing them to extract scientific insight from large datasets with minimal manual labeling. The human in the loop provides quality

control, probing model performance in new regions and correcting mislabeled rare or out-of-sample categories. My work in this space focuses in three main directions, the first of which is active learning to efficiently categorize species in new static sensor deployments [19]. Our method matches fully supervised accuracy on a 3.2M image dataset with as few as 14K manual labels, ***decreasing manual labelling effort on new deployments by over 99.5%***. Second, when a user is provided with a CV-based automated solution, such as species identification, they immediately come back with more questions: How old is it? Is it healthy? What is it doing? We must build computer vision systems that can work with experts to efficiently answer *new questions*. Towards this goal, we ***compared supervised and self-supervised representations learned from 2.7M iNaturalist species observation images on 164 novel ecologically-relevant tasks*** [20]. Third, we built and deployed ElephantBook [21], an AI-assisted elephant re-identification system that combines robust contour matching, metric learning, and human-in-the-loop attribute labeling to enable ***long-term monitoring of the shifting, open-set elephant population in the Greater Mara Ecosystem***.

Connecting the AI and conservation research communities. *(Nature Comms, FGVC@CVPR, COMPASS).* Interdisciplinary communication and understanding is vital when developing automated methods for scientific fields. As interest in AI for conservation grew, I saw a need for a space where practitioners on both sides could share opportunities and find collaborators, so ***I launched a community, called AI for Conservation, that has grown to include over 600 interdisciplinary researchers and conservation technologists worldwide.*** Dan Morris, head of Microsoft AI for Earth, says *"the 'AI for Conservation' community that Sara Beery launched has become the de facto rallying point for energy in this area, and it's where we point everyone that comes to us asking how they can get involved."* To bring conservation challenges to the attention of the CV community, I designed four distinct iWildCam challenges for the Fine-Grained Visual Categorization Workshop at CVPR [22–25]. ***Each competition defined a difficult, multimodal task and over 500 teams of CV researchers have taken part to date.*** Bridging the gap between two communities requires an understanding of both, and the ability to translate fundamental concepts for both audiences. In the last year I have written a review of species distribution modeling aimed at machine learning practitioners [26], and a review of CV for biodiversity monitoring aimed at ecologists [27].

Future Directions

We require a real-time, modular earth observation system that unites efforts across research groups in order to provide the vital information necessary for global-scale impact in sustainability and conservation. The development of such systems requires collaborative, interdisciplinary approaches that translate diverse sources of raw information into accessible scientific insight. To that end, my research agenda will expand upon the strong foundation built by my past and current research. It will seek to make effective use of all available modalities of data, incorporate expert knowledge systematically, and ensure these systems are equitable and ethical – all fundamental and unresolved challenges for CV&ML.

Learning from everything: reasoning across non-homogeneous data. Data is increasingly accessible in large volumes, collected from multitudes of diverse sensors and platforms. These modalities are complementary: no one data collection method can capture the entire picture. Valuable information is captured in everything from text-based historical records to social media posts to satellite imagery. There have been amazing recent successes in multimodal CV, particularly with video+audio and images+text. However, these methods barely scratch the surface of what is possible, focusing primarily on highly correlated pairs of modalities. There has also been extensive work on multimodal

data fusion in domains like diagnostic medical imagery and land cover prediction, focusing on accurate co-registration of spatial data and generating interpretably fused imagery. I seek to expand the scope, building methods that ***reason about data across modalities despite non-homogeneous structure and vastly inconsistent spatial and temporal scales of sampling.***

Ecosystem monitoring across modalities at global scale is an exciting and important testbed for extracting scientific insight from diverse, non-homogeneous data. I have built the foundation of a multi-year research program in this space – collaborating with researchers at Google, I am undertaking the first large-scale study of tree species categorization in urban forests. Our study covers 25 cities across North America and over 5M trees so far. We are combining satellite imagery from Google Earth Engine with on-the-ground data from Google Street View, iNaturalist, and local tree censuses, and are developing novel methods which use cross-modal agreement over time as self-supervision to efficiently adapt to unseen cities. I plan to augment this work by collaborating with local ecologists and community scientists to investigate a combination of self-supervision, anomaly detection, and active learning to enact efficient validation and model adaptation via community science "bio-blitzes". In the future we will expand our methods to wild forests, using data from the National Ecological Observatory Network (NEON) and Wildlife Insights.

Incorporating knowledge systematically into learning. There is a considerable amount of domain-specific knowledge and theory, which is almost completely ignored by current CV methodology in pursuit of "pure" data-centric approaches. This causes real harm: biases in data are propagated through to systems without a priori understanding of domain-specific risk. Further, results from black-box models are uninterpretable, making these systematic errors difficult to catch without domain experts carefully probing models with known high-risk corner cases.

General methods provide opportunity for outsize impact when designed carefully, keeping in mind the diverse set of potential use cases and risks [17]. We need to ***understand the tradeoffs between generality and domain specificity in our methods in order to develop rigorous ways to think about designing impactful end-to-end solutions*** – computer vision systems that are general purpose but optimal for each stakeholder. I have begun to study this tradeoff in large-scale systems where each user has a unique and specific goal for their ecological study, such as Wildlife Insights and our ongoing project on sustainable fisheries management [28]. I will investigate methods that can efficiently and systematically capture domain expertise as a model is training, such as risk-aware CV and data programming, or at inference time via active model adaptation.

Equitable, ethical technology. Conservation and sustainability are time-sensitive global challenges: to make rapid progress we must ***build and deploy solutions that are accessible to all stakeholders***, from academic research groups to policymakers to on-the-ground conservationists. I witnessed firsthand the gap between cloud-based methods and user need when I deployed a network of 100 camera traps in Kenya. The fastest, most cost-effective way to extract information from the raw imagery is to mail terabytes of data to the US, where it can be quickly uploaded to the cloud and analyzed at-scale with our CV models. The cost of computation, data storage, and data movement make many automated solutions inaccessible to all but the most privileged researchers. I seek to make CV equitable by developing methods which (1) increase efficiency of training, inference and data storage, and (2) incorporate and expand federated learning to enable models to learn from multiple organizations without data needing to be centralized, vital in cases where data privacy must be preserved (as we show in [29]). I further seek to increase the accessibility of technical knowledge used to build and use AI-based solutions, which I discuss in my teaching statement.

References

[1] REA Almond, Moonique Grooten, and T Peterson. *World Wildlife Fund Living Planet Report 2020-Bending the curve of biodiversity loss*. 2020.

[2] ODDS Cf. Transforming our world: the 2030 agenda for sustainable development. *United Nations*, 2015.

[3] World Health Organization et al. Connecting global priorities: biodiversity and human health. World Health Organization and Secretariat of the Convention on Biological . . . , 2015.

[4] Kathy J Willis and Shonil A Bhagwat. Biodiversity and climate change. *Science*, 2009.

[5] **Sara Beery**, Guanhang Wu, Vivek Rathod, Ronny Votel, and Jonathan Huang. 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*, 2020.

[6] wildlife insights.

[7] **Sara Beery**, Grant Van Horn, and Pietro Perona. Recognition in terra incognita. In *Proceedings of the European conference on computer vision (ECCV)*, 2018.

[8] SSC IUCN. The iucn red list of threatened species. *Version 3*, 2017.

[9] **Sara Beery**, Yang Liu, Dan Morris, Jim Piavis, Ashish Kapoor, Neel Joshi, Markus Meister, and Pietro Perona. Synthetic examples improve generalization for rare classes. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020.

[10] Edoardo Lanzini and **Sara Beery**. Image-to-image translation of synthetic samples for rare classes. *The Computer Vision for Animals Workshop at CVPR*, 2021.

[11] Tuhin Das, Robert-Jan Bruintjes, Attila Lengyel, Jan van Gemert, and **Sara Beery**. Domain adaptation for rare classes augmented with synthetic samples. *arXiv preprint arXiv:2110.12216*, 2021.

[12] Mohammad Sadegh Norouzzadeh, Anh Nguyen, Margaret Kosmala, Alexandra Swanson, Meredith S Palmer, Craig Packer, and Jeff Clune. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, 2018.

[13] Robin C Whytock, Jędrzej Świeżewski, Joeri A Zwerts, Tadeusz Bara-Słupski, Aurélie Flore Koumba Pambo, Marek Rogala, Laila Bahaa-el din, Kelly Boekee, Stephanie Brittain, Anabelle W Cardoso, et al. Robust ecological analysis of camera trap data labelled by a machine learning model. *Methods in Ecology and Evolution*, 2021.

[14] Benjamin Kellenberger, Diego Marcos, Sylvain Lobry, and Devis Tuia. Half a percent of labels is enough: Efficient animal detection in uav imagery using deep CNNs and active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 2019.

[15] Félix Gontier, Vincent Lostanlen, Mathieu Lagrange, Nicolas Fortin, Catherine Lavandier, and Jean-François Petiot. Polyphonic training set synthesis improves self-supervised urban sound classification. *The Journal of the Acoustical Society of America*, 2021.

[16] Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, Tony Lee, Etienne David, Ian Stavness, Wei Guo, Berton Earnshaw, Imran Haque, **Sara Beery**, Jure Leskovec, Anshul Kundaje, Emma Pierson, Sergey Levine, Chelsea Finn, and Percy Liang. Wilds: A benchmark of in-the-wild distribution shifts. *In Proceedings of the International Conference on Machine Learning*, 2021.

[17] **Sara Beery**, Dan Morris, and Siyu Yang. 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 for KDD Earth Day**.

[18] Omiros Pantazis, Gabriel Brostow, Kate Jones, and Oisin Mac Aodha. Focus on the positives: Self-supervised learning for biodiversity monitoring. *Proceedings of the European Conference in Computer Vision*, 2021.

[19] Mohammad Sadegh Norouzzadeh, Dan Morris, **Sara Beery**, Neel Joshi, Nebojsa Jojic, and Jeff Clune. A deep active learning system for species identification and counting in camera trap images. *Methods in Ecology and Evolution*, 2021.

[20] Grant Van Horn, Elijah Cole, **Sara Beery**, Kimberly Wilber, Serge Belongie, and Oisin Mac Aodha. Benchmarking representation learning for natural world image collections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[21] Peter Kulits, Jake Wall, Anka Bedetti, Michelle Henley, and **Sara Beery**. Elephantbook: A semi-automated human-in-the-loop system for elephant re-identification. *Proceedings of the 4th ACM SIGCAS Conference on Computing and Sustainable Societies*, 2021.

[22] **Sara Beery**, Grant Van Horn, Oisin MacAodha, and Pietro Perona. The iWildCam 2018 challenge dataset. *The Fifth Fine-Grained Visual Categorization Workshop at CVPR*, 2018.

[23] **Sara Beery**, Dan Morris, and Pietro Perona. The iWildCam 2019 challenge dataset. *The Sixth Fine-Grained Visual Categorization Workshop at CVPR*, 2019.

[24] **Sara Beery**, Elijah Cole, and Arvi Gjoka. The iWildCam 2020 competition dataset. *The Seventh Fine-Grained Visual Categorization Workshop at CVPR*, 2020.

[25] **Sara Beery**, Arushi Agarwal, Elijah Cole, and Vighnesh Birodkar. The iWildCam 2021 competition dataset. *The Eighth Fine-Grained Visual Categorization Workshop at CVPR*, 2021.

(* indicates co-first authorship)

[26] **Sara Beery***, Elijah Cole*, Joseph Parker, Pietro Perona, and Kevin Winner. Species distribution modeling for machine learning practitioners: A review. *Proceedings of the 4th ACM SIGCAS Conf. on Computing and Sustainable Societies*, 2021.

[27] Devis Tuia*, Benjamin Kellenberger*, **Sara Beery***, Blair R. Costelloe*, Silvia Zuffi, Benjamin Risse, Alexander Mathis, Mackenzie W. Mathis, Frank van Langevelde, Tilo Burghardt, Roland Kays, Holger Klinck, Martin Wikelski, Iain D. Couzin, Grant Van Horn, Margaret C. Crofoot, Charles V. Stewart, and Tanya Berger-Wolf. Seeing biodiversity: perspectives in machine learning for wildlife conservation. *Nature Communications (to appear)*.

[28] Peter Kulits, Angelina Pan, Grant Van Horn, **Sara Beery**, Erik Young, and Pietro Perona. Automated salmonid counting in sonar data.

[29] **Sara Beery*** and Elizabeth Bondi*. Can poachers find animals from public camera trap images? *CV for Animals Workshop at CVPR*, 2021.