Using Remote Sensing Imagery and Machine Learning to Predict Poaching in Wildlife Conservation Parks

Rachel Guo
Harvard University
rguo@college.harvard.edu

Abstract

Illegal wildlife poaching is driving the loss of biodiversity. To combat poaching, rangers patrol expansive protected areas for illegal poaching activity. However, rangers often cannot comprehensively search such large parks. Thus, the Protection Assistant for Wildlife Security (PAWS) uses machine learning to help identify the areas with highest poaching risk. As PAWS is deployed to parks around the world, we recognized that many parks have limited resources for data collection and therefore have scarce feature sets. To ensure under-resourced parks have access to meaningful poaching predictions, we introduce the use of publicly available remote sensing data to extract features for parks. By employing this data from Google Earth Engine, we also incorporate previously unavailable dynamic data such as climate and primary production patterns to enrich predictions with seasonal trends. We automate the entire data-to-deployment pipeline and find that, with only using publicly available data, we recuperate prediction performance comparable to predictions made using features manually computed by park specialists. We conclude that the inclusion of satellite imagery creates for a robust system through which parks of any resource level can benefit from poaching risks for years to come.

Introduction

Illegal poaching of wildlife threatens endangered species, driving the loss of biodiversity and contributing to the climate crisis. To deter poaching, rangers search protected areas and remove snares laid out to trap animals. However, conservation parks have a limited number of rangers to search parks that scale to thousands of square kilometers. Thus, the Protection Assistant for Wildlife Security (PAWS) has been developed as a machine learning approach to predict areas of highest poaching risk based on historical poaching patterns and geospatial features. PAWS is currently being integrated with SMART, a leading conservation software, to become deployed to 800 parks around the world.

In a series of alpha tests of the PAWS integration with SMART (Fig. 1) involving over 20 parks, some park managers reported nonsensical predictions because they had just a single feature: park boundary. We recognized that an AI system would be useful only if we helped address these challenges of unavailable features, but seeing that some features require expensive aerial surveys to compute and GIS specialists to create, we were limited to processes that required virtually no effort or technical expertise from the parks.

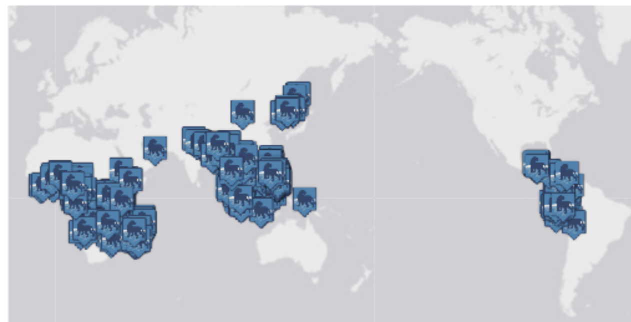
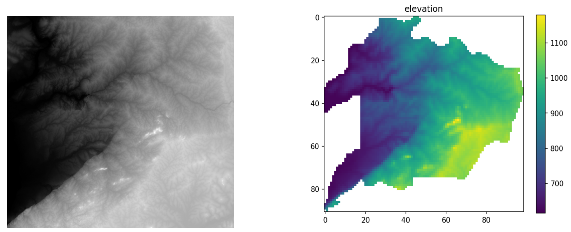


Figure 1: SMART is used by 800 protected areas around the world, many of which have limited access to geospatial data.

In order to make the PAWS system resilient, scalable, and deployable to parks around the world of varying resources, we introduce an automatic pipeline into PAWS that extracts remote-sensing data from Google Earth Engine (GEE), which ensures every park has access to the same abundant feature set. We discover that we recover almost all predictive performance using only GEE features on three test parks—Murchison Falls National Park (MFNP) and Queen Elizabeth National Park (QENP) in Uganda and Cross River National Park (CRNP) in Nigeria. We also find that dynamic data from GEE prove useful in predictive performance and that we can recreate some features provided by parks exactly using GEE imagery. This makes for a robust system on which PAWS can make poaching predictions virtually anywhere in the world for decades to come.

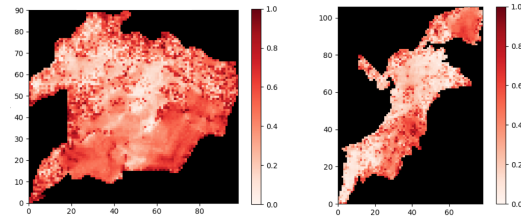
Related Work

PAWS was initially incepted as a predictive modelling approach to anti-poaching, and has since been iterated upon as it is deployed around the world (Gholami et al. 2018). While remote sensing data has been used in several applications in past decades (Kumar and Mutanga 2018), in the space of machine learning for wildlife protection, the use of remote sensing has been limited.



(a) Raw GEE elevation TIF (b) Rasterized elevation

Figure 2: Raw elevation TIF extracted from GEE (90 m) discretized into 1×1 km cells in MFNP boundary (1000 m).



(a) MFNP (2017) (b) QENP (2016)

Figure 3: Predicted poaching risks for guide rangers to search areas with highest poaching risk (dark red).

Methods

Extracting remote sensing data We create an automated pipeline that extracts remote sensing data from Google Earth Engine (GEE), a catalog of satellite imagery available on a global scale in real time. We identified 16 features from GEE as potentially useful for poaching predictions, including static features such as land cover, rivers, surface water, and elevation, as well as dynamic data such as monthly temperature, precipitation, and net primary productivity.

Preprocessing We discretize the park boundary into 1×1 km cells for a 1000 m spatial resolution. We rasterize each feature into TIF files that match the spatial resolution of the park. For shapefiles provided by parks, which include roads, rivers, villages, animal density, elevation, patrol posts, and the park boundary, each 1×1 km cell represents distance to the nearest feature. High-resolution imagery from GEE is rasterized into a lower resolution of 1000 m. For dynamic data, we discretize time into three-month intervals.

Making predictions Our labels are a binary indicator of past illegal activity for each cell. Past illegal activity data is incomplete as rangers are unable to search the entire park; still, we aim to predict risk in areas not previously patrolled. Only 20%, 12%, and 47% of the the labeled data are positive labels in MFNP, QENP, and CRNP, respectively.

To address these challenges of uncertainty and class imbalance, we use iWare-E (imperfect-observation aWare Ensemble model), a bagging ensemble of decision trees as weak learners, training each weak learner from different bins of input data based on current patrol effort to account for uncertainty in negative labels (Gholami et al. 2018).

Results

We evaluate prediction accuracies for cases in which the feature set consists of: (i) only features provided by parks, (ii) only GEE features, and (iii) all features (both park and GEE features). We present the performance, assessed by AUC, in Table 1 for parks MFNP, QENP, and CRNP. For each test year, we train our model on the three years prior.

We achieve nearly the same AUCs using satellite imagery, without any of the several customized features from parks. Rangers can then use our poaching predictions to efficiently search parks (Fig. 3). We note the usefulness of dynamic data in learning patterns through their high feature importance. We are essentially able to recreate some features that parks provided using GEE remote sensing data.

		Baseline	GEE	All Features
MFNP	2014	0.707	0.696	**0.714**
	2015	**0.678**	0.652	0.671
	2016	0.678	0.647	**0.681**
	2017	0.683	0.665	**0.688**
	Avg	0.687	0.665	**0.689**
QENP	2014	0.710	0.697	**0.716**
	2015	0.710	**0.722**	0.720
	2016	0.715	0.699	**0.716**
	Avg	0.712	0.706	**0.717**
CRNP	2018	0.655	**0.693**	0.670
	2019	**0.741**	0.727	**0.741**
	Avg	0.698	**0.710**	0.706

Table 1: AUCs for predictions based on MFNP's 21 features, QENP's 19 features, and CRNP's 11 features (Baseline); only GEE features; and all features for each test year.

Conclusion and Future Work

Under-resourced parks without the resources to create their own features can benefit from extracting features from publicly available satellite imagery. We look towards exploring CNNs with a decision-focused learning approach.

Acknowledgments

We thank Lily Xu, Milind Tambe, Drew Cronin, Francis Okeke, Andrew Plumptre, WCS Nigeria, Uganda Wildlife Authority, NC Zoo, and SMART.

References

Gholami, S.; Ford, B. J.; Kar, D.; Fang, F.; Tambe, M.; Plumptre, A.; Driciru, M.; Wanyama, F.; Rwetsiba, A.; Nsubaga, M.; and Mabonga, J. 2018. Evaluation of Predictive Models for Wildlife Poaching Activity through Controlled Field Test in Uganda. In *AAAI Workshops*.

Kumar, L.; and Mutanga, O. 2018. Google Earth Engine Applications Since Inception: Usage, Trends, and Potential. *Remote Sensing* .