

TRUEBRANCH: ROBUST DEEP LEARNING-BASED VERIFICATION OF FOREST CONSERVATION PROJECTS

Simona Santamaría*, David Dao*

Department of Computer Science
ETH Zurich
Zurich, Switzerland
`{ssimona,david.dao}@inf.ethz.ch`

Björn Lütjens*

Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
Cambridge, USA
`lutjens@mit.edu`

Ce Zhang

Department of Computer Science
ETH Zurich, Zurich
`ce.zhang@inf.ethz.ch`

ABSTRACT

International stakeholders increasingly invest in offsetting carbon emissions, for example, via issuing Payments for Ecosystem Services (PES) to forest conservation projects. Issuing trusted payments requires a transparent monitoring, reporting, and verification (MRV) process of the ecosystem services (e.g., carbon stored in forests). The current MRV process, however, is either too expensive (on-ground inspection of forest) or inaccurate (satellite). Recent works propose low-cost and accurate MRV via automatically determining forest carbon from drone imagery, collected by the landowners. The automation of MRV, however, opens up the possibility that landowners report untruthful drone imagery. To be robust against untruthful reporting, we propose a robust deep learning-based algorithm that verifies the truthfulness of drone imagery from forest conservation projects. It aims to detect untruthfully reported drone imagery by matching it with public satellite imagery. Preliminary results suggest that nominal distance metrics are not sufficient to reliably detect untruthfully reported imagery. It averages a method from metric learning to create a feature embedding in which truthfully and untruthfully collected imagery is easily distinguishable by distance thresholding.

1 INTRODUCTION

Agriculture, forestry, and other land use is a key driver of climate change, accounting for 23% ($12.0 \pm 2.9 \text{ GtCO}_2\text{eq yr}^{-1}$) of total anthropogenic emissions of greenhouse gases during 2007–2016 (IPCC, 2019), largely driven by deforestation and forest degradation. Deforestation does not only release carbon (e.g., through slash-and-burn), but also destroys a multitude of other forest ecosystem services: preserving biodiversity, counteracting flooding and soil erosion, filtering water, and offering a livelihood for the local population.

The causes of deforestation are mostly economically driven: expansion of commercial or subsistence agriculture, logging, fuelwood collection, or livestock grazing (Hosonuma et al., 2012)). To counteract the economic incentives, payments for ecosystem services (PES) (Wunder, 2007) are increasingly (Donofrio et al., 2019) provided to forest conserving or restoring landowners by international stakeholders (e.g., through the governmental UN-REDD program (Gibbs et al., 2007) or the commercial voluntary carbon market (Donofrio et al., 2019)). However, current methods for monitoring, reporting, and verification (MRV) of the landowner-provided forest ecosystem services are either based on 1) on-ground inspection, which is too expensive (USD 20-30k), delayed (up to two years), corruptible, and biased (Gold Standard, 2017), 2) satellite, which is low-cost, but limited to the binary verification of forest/no-forest cover (Hansen et al., 2013), or 3) drones.

*Authors have contributed equally.

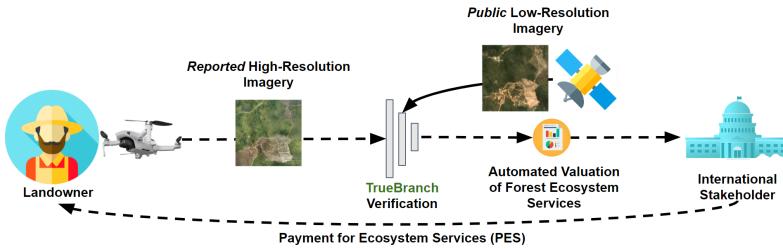


Figure 1: An overview of TrueBranch. A landowner takes a high-resolution drone image of their forest and reports the image, time stamp, and location. A robust deep learning-based algorithm, TrueBranch, verifies the submission with the corresponding low-resolution public satellite image. Another algorithm estimates the forest ecosystem service value based on the verified imagery and international stakeholders provide payments for ecosystem services (PES) to the landowner.



Figure 2: An overview of possible attack vectors in time, location, and value (left-to-right) that trick an automated forest valuation algorithm to detect high forest cover.

Recent works have proposed low-cost and accurate MRV via drones (Dao et al., 2019b;a; Lütjens et al., 2019). Specifically, these works propose algorithms that estimate forest ecosystem services (e.g., stored forest carbon) from drone imagery that was reported by landowners. Replacing on-ground inspection with remote assessment via drones, however, opens up the possibility of untruthfully reported imagery. Given that the landowner is financially incentivized by PES to report higher forest ecosystem services value (e.g., higher forest cover or biodiversity), the possibility of false reporting is high, but not addressed by previous works. We hypothesize that the landowner could report untruthful imagery by altering the image 1) in location, 2) in time, or 3) with adversarial perturbation (e.g., via PGD (Madry et al., 2018)), as displayed in Figure 2.

We propose TrueBranch, a robust deep learning-based algorithm that verifies the trustworthiness of reported drone imagery of forest conservation projects. Specifically, TrueBranch aims to verify the truthfulness of drone images via matching them with public satellite images. Matching is proposed to be done in a deeply learned feature space that, ideally, 1) allows for easy distinction of images with different value of forest ecosystem services, 2) is robust to adversarial perturbation of the drone image, and 3) generalizes the verification of drone images to other ecosystems (e.g., mangroves or peatlands). Figure 1 shows how TrueBranch is embedded in the scheme of automated MRV systems to achieve low-cost, accurate, and trustworthy MRV, which will promote international investments in forest conservation.

2 ATTACK VECTORS

Landowners are financially incentivized to report untruthful drone imagery that displays forest with higher ecosystem services value to receive higher PES. To reason about an algorithm that detects untruthful imagery, we classify common attack vectors with examples, as displayed in Figure 2:

1. Altering drone image **location**: The landowner has land with 50% forest cover and reports imagery from a neighbouring land with 80% forest cover to receive higher valuation.
2. Altering drone image **time stamp**: The landowner reports imagery (e.g., with altered time stamp metadata) from previous flights, before their land has been logged or cleared.
3. Altering drone image **values**: The landowner tricks a neural network-based forest valuation algorithm into estimating higher ecosystem value by altering the image values with

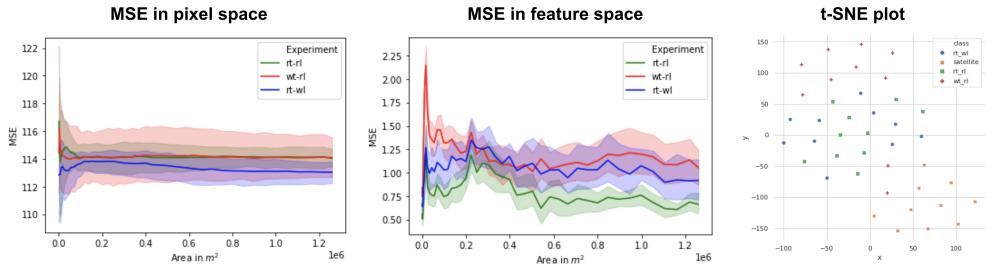


Figure 3: Preliminary results show that neither the MSE metric in pixel space nor in RESISC-45 (Neumann et al., 2019) feature space are sufficient to reliably distinguish right time/right location (rt-rl) images from their attack vectors (wrong time/right location (wt-rl) and right time/wrong location (rt-wl)). The t-SNE plot shows that the drone images are naturally clustered together and difficult to separate. The plots have been generated using the 10 images from Figure 5.

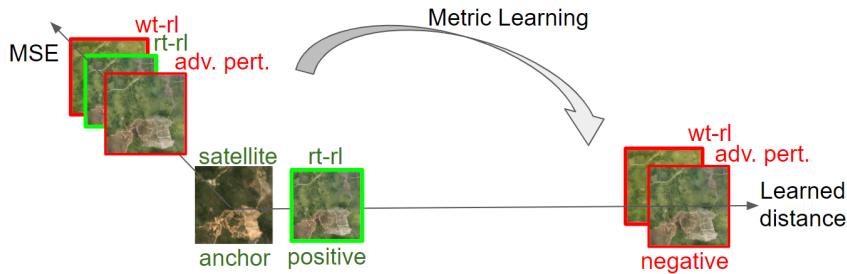


Figure 4: Illustration of metric learning. Truthful (green), untruthful (red), and satellite imagery are all close to each other, as measured by MSE distance. Metric learning pulls the satellite (anchor) and corresponding drone imagery (positive) even closer together and pushes the attack vectors (negative) away from the satellite imagery.

sophisticated attacks such as human-imperceptible **adversarial perturbation** (e.g via PGD (Madry et al., 2018)).

4. A combination of the above.

3 APPROACH

3.1 CHALLENGES OF DETECTING ATTACK VECTORS WITH NOMINAL METRICS

We ran preliminary experiments to investigate if we can detect attack vectors (i.e., untruthfully reported imagery). Figure 5 shows all used publicly available drone (0.3m/px) and satellite (4m/px) imagery from OpenForest and PlanetLabs, respectively. Figure 3 suggests that nominal distance metrics (here, MSE in pixel or RESICS-45 feature space (i.e., the activation layer of a classification network, trained on satellite imagery (Neumann et al., 2019))) are not sufficient to reliably separate truthful from untruthful drone imagery by using the corresponding low-resolution satellite imagery.

3.2 DETECTING ATTACK VECTORS WITH A LEARNED METRIC

Based on our preliminary results, we design TrueBranch to leverage the triplet loss function (Schroff et al., 2015) from the field of metric learning. The loss function is designed to pull truthful images closer to the true satellite image and push untruthful images away from it. The result of training with triplet loss should be a deeply learned feature space (i.e., learned metric) in which truthful and untruthful imagery is separated by a large margin, as displayed in Figure 4. In comparison to the nominal metrics from section 3.1, the learned metric is also designed to be robust to adversarially perturbed images (sec. 2.3), as shown by Mao et al. (2019). As metric learning requires large amounts of imagery and high-resolution drone imagery is scarce, future works could augment drone imagery by satellite imagery, enhanced by super resolution algorithms (Deudon et al., 2020). Note that the metric learning-approach enables TrueBranch to be extended from the verification of im-

agery of forest conservation projects to a multitude of other ecosystem conservation projects (e.g., mangroves, peatlands, wetlands, etc.) by adding the respective imagery.

4 CONCLUSION

Truthfully reported imagery is a vital requirement to issue Payments for Ecosystem Services (PES) based on trusted and automatic valuation of ecosystem services (e.g., stored forest carbon). In this paper, we list possible attack vectors and propose TrueBranch, a robust deep learning-based algorithm that uses public satellite data and metric learning to distinguish truthfully reported from untruthfully reported imagery.

ACKNOWLEDGMENTS

The authors are thankful for the guidance and advise by the mentors (Prof. Dava Newman, Forrest Meyen, Adam de Sola Pool), support from the local community (Sandro Pimentel, La Niebla Forest), academic collaborators (Prof. Pedro Brancalion, Prof. Paulo Guilherme Molin), non-governmental institutions (WWF Peru and Brazil) and CONAF (Daniel Montaner, Cesar Mattar, Jose Antonio Prado). Part of this research has been developed as part of the OpenSurface platform and a real-world pilot in Chile, which was launched at the COP25 United Nation’s Climate Summit. OpenSurface is funded by IDBLab and EIT Climate-KIC.

REFERENCES

- David Dao, Catherine Cang, Clement Fung, Ming Zhang, Nick Pawlowski, Reuven Gonzales, Nick Beglinger, and Ce Zhang. GainForest: Scaling Climate Finance for Forest Conservation using Interpretable Machine Learning on Satellite Imagery. *ICML Climate Change AI workshop 2019*, 2019a.
- David Dao, Johannes Rausch, and Ce Zhang. Geolabels: Towards efficient ecosystem monitoring using data programming on geospatial information. In *2019 NeurIPS Workshop on Tackling Climate Change with AI*, 2019b.
- Michel Deudon, Alfredo Kalaitzis, Md Rifat Arefin, Israel Goytom, Zhichao Lin, Kris Sankaran, Vincent Michalski, Samira E Kahou, Julien Cornebise, and Yoshua Bengio. Highres-net: Multi-frame super-resolution by recursive fusion, 2020. URL <https://openreview.net/forum?id=HJxJ2h4tPr>.
- Stephen Donofrio, Patrick Maguire, William Merry, and Steve Zwick. Financing emissions reductions for the future. In *State of the Voluntary Carbon Markets 2019*. Ecosystem Marketplace; A Forest Trends Initiative, 2019.
- Holly K Gibbs, Sandra Brown, John O Niles, and Jonathan A Foley. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. *Environmental Research Letters*, 2(4):045023, 2007.
- Gold Standard. Afforestation/reforestation (a/r) ghg emissions reduction & sequestration methodology, July 2017.
- M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160):850–853, 2013.
- Noriko Hosonuma, Martin Herold, Veronique De Sy, Ruth S De Fries, Maria Brockhaus, Louis Verchot, Arild Angelsen, and Erika Romijn. An assessment of deforestation and forest degradation drivers in developing countries. *Environmental Research Letters*, 7(4):044009, 2012.
- IPCC. 2019: Summary for policymakers. In P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.-O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. Portugal Pereira, P. Vyas, E. Huntley, K. Kissick, M. Belkacemi, and J. Malley (eds.), *Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*, pp. 7–11. 2019.
- Björn Lütjens, Lucas Liebenwein, and Katharina Kramer. Machine learning-based estimation of forest carbon stocks to increase transparency of forest preservation efforts. In *2019 NeurIPS Workshop on Tackling Climate Change with AI*, 2019.
- Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. In *International Conference on Learning Representations*, 2018.
- Chengzhi Mao, Ziyuan Zhong, Junfeng Yang, Carl Vondrick, and Baishakhi Ray. Metric learning for adversarial robustness, 2019.

Maxim Neumann, Andre Susano Pinto, Xiaohua Zhai, and Neil Houlsby. In-domain representation learning for remote sensing, 2019.

Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

Sven Wunder. The efficiency of payments for environmental services in tropical conservation. *Conservation Biology*, 21(1):48–58, 2007.

A APPENDIX

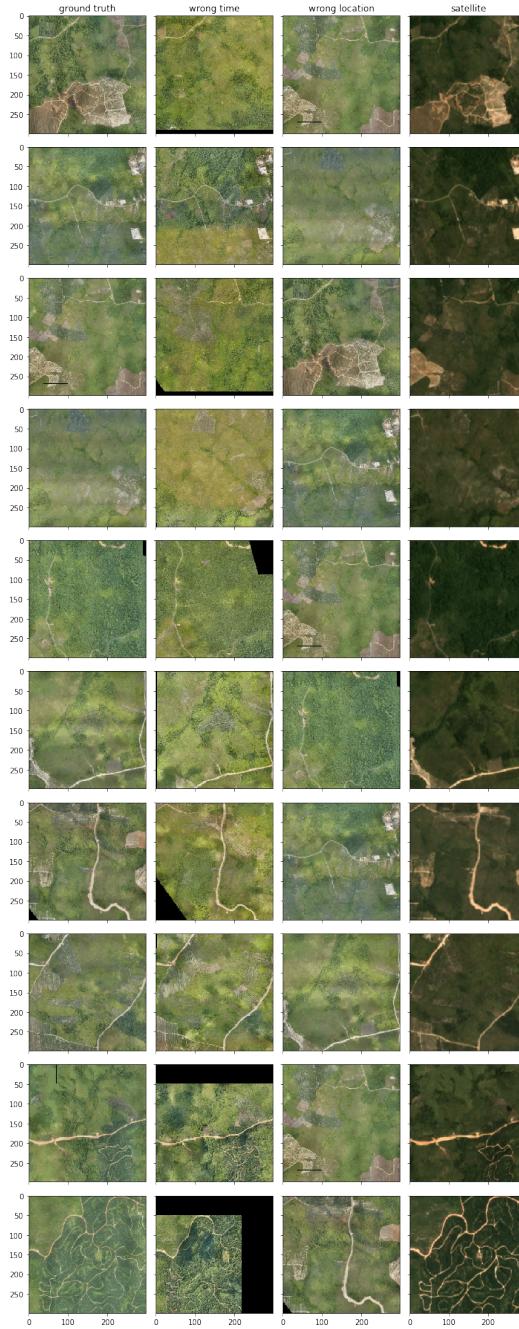


Figure 5: Drone and satellite images. The columns show drone imagery (from left to right), that is recorded 1) at the right time/right location, 2) wrong time/right location, 3) right time/wrong location, and 4) the ground truth satellite imagery. The x- and y-axis are in pixels. It is noteworthy that wrong time imagery only occasionally has higher forest cover (e.g., 1st or 3rd row), shows artifacts (e.g., last row), and generally has higher intensity than ground truth imagery. The wrong location images are the same as ground truth images, but in a different order. The satellite images have generally lower intensity and resolution than the ground truth imagery.