BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 5: e74047
doi: 10.3897/biss.5.74047

OPEN ACCESS

Conference Abstract

Reducing Manual Supervision Required for Biodiversity Monitoring with Self-Supervised Learning

Omiros Pantazis[‡], Gabriel J. Brostow[‡], Kate Jones[‡], Oisin Mac Aodha[§]

‡ University College London, London, United Kingdom
§ University of Edinburgh, Edinburgh, United Kingdom

Corresponding author: Omiros Pantazis (omiros.pantazis.16@ucl.ac.uk)

Received: 06 Sep 2021 | Published: 07 Sep 2021

Citation: Pantazis O, Brostow GJ, Jones K, Mac Aodha O (2021) Reducing Manual Supervision Required for Biodiversity Monitoring with Self-Supervised Learning. Biodiversity Information Science and Standards 5: e74047. https://doi.org/10.3897/biss.5.74047

Abstract

Recent years have ushered in a vast array of different types of low cost and reliable sensors that are capable of capturing large quantities of audio and visual information from the natural world. In the case of biodiversity monitoring, camera traps (i.e. remote cameras that take images when movement is detected (Kays et al. 2009) have shown themselves to be particularly effective tools for the automated monitoring of the presence and activity of different animal species. However, this ease of deployment comes at a cost, as even a small scale camera trapping project can result in hundreds of thousands of images that need to be reviewed.

Until recently, this review process was an extremely time consuming endeavor. It required domain experts to manually inspect each image to:

- determine if it contained a species of interest and
- identify, where possible, which species was present. Fortunately, in the last five years, advances in machine learning have resulted in a new suite of algorithms that are capable of automatically performing image classification tasks like species classification.

The effectiveness of deep neural networks (Norouzzadeh et al. 2018), coupled with transfer learning (tuning a model that is pretrained on a larger dataset (Willi et al. 2018), have resulted in high levels of accuracy on camera trap images.

However, camera trap images exhibit unique challenges that are typically not present in standard benchmark datasets used in computer vision. For example, objects of interest are often heavily occluded, the appearance of a scene can change dramatically over time due to changes in weather and lighting, and while the overall number of images can be large, the variation in locations is often limited (Schneider et al. 2020). These challenges combined mean that in order to reach high performance on species classification it is necessary to collect a large amount of annotated data to train the deep models. This again takes a significant amount of time for each project, and this time could be better spent addressing the ecological or conservation questions of interest.

Self-supervised learning is a paradigm in machine learning that attempts to forgo the need for manual supervision by instead learning informative representations from images directly, e.g. transforming an image in two different ways without impacting the semantics of the included object, and learn by imposing similarity between the two tranformations. This is a tantalizing proposition for camera trap data, as it has the potential to drastically reduce the amount of time required to annotate data. The current performance of these methods on standard computer vision benchmarks is encouraging, as it suggests that self-supervised models have begun to reach the accuracy of their fully supervised counterparts for tasks like classifying everyday objects in images (Chen et al. 2020). However, existing self-supervised methods can struggle when applied to tasks that contain highly similar, i.e. fine-grained, object categories such as different species of plants and animals (Van Horn et al. 2021).

To this end, we explore the effectiveness of self-supervised learning when applied to camera trap imagery. We show that these methods can be used to train image classifiers with a significant reduction in manual supervision. Furthermore, we extend this analysis by showing, with some careful design considerations, that off-the-shelf self-supervised methods can be made to learn even more effective image representations for automated species classification. We show that by exploiting cues at training time related to where and when a given image was captured can result in further improvements in classification performance. We demonstrate, across several different camera trapping datasets, that it is possible to achieve similar, and sometimes even superior, accuracy to fully supervised transfer learning-based methods using a factor of ten times less manual supervision. Finally, we discuss some of the limitations of the outlined approaches and their implications on automated species classification from images.

Keywords

computer vision, deep learning, camera traps

Presenting author

Omiros Pantazis

Presented at

TDWG 2021

References

- Chen T, Kornblith S, Norouzi M, Hinton G, et al. (2020) A Simple Framework for Contrastive Learning of Visual Representations. Proceedings of the 37th International Conference on Machine Learning 119: 1597-1607. URL: https://arxiv.org/abs/2002.05709
- Kays R, Kranstauber B, Jansen P, Carbone C, Rowcliffe M, Fountain T, Tilak S, et al. (2009) Camera traps as sensor networks for monitoring animal communities. 2009 IEEE 34th Conference on Local Computer Networks https://doi.org/10.1109/lcn.2009.5355046
- Norouzzadeh MS, Nguyen A, Kosmala M, Swanson A, Palmer M, Packer C, Clune J, et al. (2018) Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences 115 (25). https://doi.org/10.1073/pnas.1719367115
- Schneider S, Greenberg S, Taylor GW, Kremer SC, et al. (2020) Three critical factors affecting automated image species recognition performance for camera traps. Ecology and evolution 10 (7): 3503-3517. https://doi.org/10.1002/ece3.6147
- Van Horn G, Cole E, Beery S, Wilber K, Belongie S, Mac Aodha O, et al. (2021) Benchmarking Representation Learning for Natural World Image Collections. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition12884-12893.
- Willi M, Pitman R, Cardoso A, Locke C, Swanson A, Boyer A, Veldthuis M, Fortson L, et al. (2018) Identifying animal species in camera trap images using deep learning and citizen science. Methods in Ecology and Evolution 10 (1): 80-91. https://doi.org/10.1111/2041-210x.13099