lynx-id – A Python package for re-identifying lynx on images from camera traps

1 November 2024

Summary

Blabla.

Statement of need

The knowledge of animal population abundance is one of the cornerstones of scientific ecology (Williams et al. 2002). The acquisition of the necessary data for applying these abundance estimation methods has been revolutionized by the use of sensors in ecology, particularly through the advent of automatic imaging devices—camera traps—that passively capture tens, thousands, or even millions of images. The main appeal of camera traps is that they allow researchers to work on spatial and temporal scales that, until recently, were simply unrealistic with traditional sampling methods due to the high cost in time and money. However, the images captured by these sensors are not directly usable for statistical analysis as they still need to be sorted (e.g., removing images without animals) and annotated (typically, identifying an animal across multiple images).

Deep learning is used in computer vision problems with important applications in several scientific fields. In ecology, there is a growing interest in deep learning for automatizing repetitive analyses on large amounts of images, such as animal species identification (REFs). Here we focus the identification and re-identification of animals in camera-trap images. Identifying animals individually is required for the estimation of population size and demographic parameters. Extensive algorithms comparisons (benchmarks) have been carried out (e.g. Schneider et al. 2022), thanks to which we have a pretty good idea of which architectures we should use to ensure re-identification.

To date, the number of species for which visual identification of individuals without prior physical marking (such as rings, collars, etc.) is possible is limited. This mainly applies to species with "idiosyncratic" markings, meaning variable spots in number and shape visible on fur or skin, like those found on giraffes. Encouraging results have been obtained on giraffes by several members of the project (Miele et al. 2020). The goal here is to test these approaches on other species (zebras, lynxes, seahorses), and to compare them with other methods across these species. Specifically, we will use different methods that learn, from a small number of repetitions per individual, a representation space of images where the images of the same individual are grouped together, while those of different individuals are separated. The aim will be to compare different combinations of architectures (ResNet, Xception, MobileNet, EfficientNet) and loss functions (contrastive loss, Triplet loss, ArcFace, CosFace) known for their strong performance in human identification through video surveillance (Wang & Deng 2021). By learning a metric in these representation spaces that distinguishes images of the same individual from those of different individuals, we can develop an 'open' predictive model (open-set identification) that enables the re-identification of individuals not present in the initial training set.

Following these benchmarks, we have been working over the past few months on training a model for re-identification of lynx on images taken from camera traps (Joigneau 2023). We have made good progress using images from France. There is definitely room for improvement, and the main step to achieve better performances, we need to increase the number of images we use to train our model. The good news is that

CNRS (our research institute) has agreed to assign two engineers to continue working on that project for a year.

Computer vision is a field of artificial intelligence in which a machine is taught how to extract and interpret the content of an image (Krizhevsky, Sutskever, and Hinton 2012). Computer vision relies on deep learning that allows computational models to learn from training data – a set of manually labelled images – and make predictions on new data – a set of unlabelled images (Baraniuk, Donoho, and Gavish 2020; LeCun, Bengio, and Hinton 2015). With the growing availability of massive data, computer vision with deep learning is being increasingly used to perform tasks such as object detection, face recognition, action and activity recognition or human pose estimation in fields as diverse as medicine, robotics, transportation, genomics, sports and agriculture (Voulodimos et al. 2018). In ecology in particular, there is a growing interest in deep learning for automatizing repetitive analyses on large amounts of images, such as identifying plant and animal species, distinguishing individuals of the same or different species, counting individuals or detecting relevant features (Christin, Hervet, and Lecomte 2019; Lamba et al. 2019; Weinstein 2018). By saving hours of manual data analyses and tapping into massive amounts of data that keep accumulating with technological advances, deep learning has the potential to become an essential tool for ecologists and applied statisticians

We use dinov2 (Oquab et al. 2024)

No easy implementation for users.

Implementation

Package installation and use

How to use, plus example.

Conclusions

What it is, what it brings, how it can be useful.

Perspective. Other felids. Re-trained as new images arrive.

Acknowledgements

Thank you. Jean Zay. CNRS for engineer time. CNRS for funding.

References

Oquab, Maxime, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, et al. 2024. "DINOv2: Learning Robust Visual Features Without Supervision." https://arxiv.org/abs/2304.07193.