International Journal of Engineering and Modern Technology (IJEMT) E-ISSN 2504-8848
P-ISSN 2695-2149 Vol 10. No. 7 2024 www.iiardjournals.org Online Version

Using Deep Learning in the Ecology Field

Anouar DALLI
Ecole Nationale des Sciences Appliquées de Safi (ENSAS),
Université Cadi Ayyad, Marrakesh, Morocco,
Email : anouar_dalli@yahoo.fr, anouardalli@gmail.com
DOI: 10.56201/ijcsmt.v10.no4.2024.pg150.159

Abstract

To have an impact on the environment, deep learning currently occupies an important place in the progress of nature, science and the environment. Although it may not seem obvious at first glance to the public, deep learning technologies can make a positive contribution to ecology. In this article we present the use of deep learning to resolve ecological issues and the difficulties we encounter.

1. Introduction

Ecology is witnessing an exciting convergence. On the one hand, the scientific community massively produces image-type data, from photo/video traps, aerial images from drones or satellites, underwater photos/videos to laboratory images. This data comes from data acquisition campaigns across the entire ecological spectrum. On the other side in artificial intelligence, a new generation of mathematical and computer models has changed the vision by computer, that is to say the automation of the understanding of the content of an image by the machine. These so-called deep learning approaches [1] have revolutionized the field. The principle is simple: build a model which synthesizes a so-called training data set previously analyzed by hand – we will say that the images are annotated – so that the machine can use this model to carry out the processing of the data by itself. new data, mass if necessary. Used in many scientific disciplines (e.g. medical imaging) but also extremely “trendy” in the media, deep learning is a promising tool for the automated processing of images in ecology [2-3].

In this article, we will discuss this exciting synchrony between the data produced and the methodology available. In particular, the transfer of this methodology to our community requires the discovery of new terminology, new mathematical models and algorithms, as well as new programming techniques. However, this knowledge is rarely described in a way that is accessible to non-specialists. For the ecologist, discovering these techniques represents a certain effort, the challenge being to know if this effort can prove to be profitable.

I- The need of deep learning in the ecology field

Several image analysis problems in ecology come back to classic learning problems in vision. In many cases, it involves training the computer so that it can carry out a repetitive task that can be done by humans, e.g. identify a species of titmouse in a photo.

We are aiming here to save time to scale up to hundreds of thousands of images, to avoid the arduousness of manual processing or to offer a complement to citizen science approaches (participatory production) for image annotation [4].

The rise of camera traps led to the first applications of deep learning in ecology. This most often involves specific identification, in order to identify the species of an individual present (entirely or partially) in a photo, assuming here that there is only one species. This amounts to a classic classification problem where the algorithm must predict the class of each image analyzed (here, the species associated with each) with a confidence value (Figure 1). Ecologists, naturalists or citizens annotate as large a set of images as possible, specifying which species is present on which image, then the algorithms take over to learn for themselves a classification rule that can be used on new images. We will subsequently consider that we have on hand hundreds of images of flowers which correspond to three clades – the daisy, the iris and the tulip – and that we want to develop a computer program which will take as input an image and will automatically return the name of the flower found. This approach has been widely implemented, from seminal studies on large savannah mammals to the study of different insect species or fish [4-6].

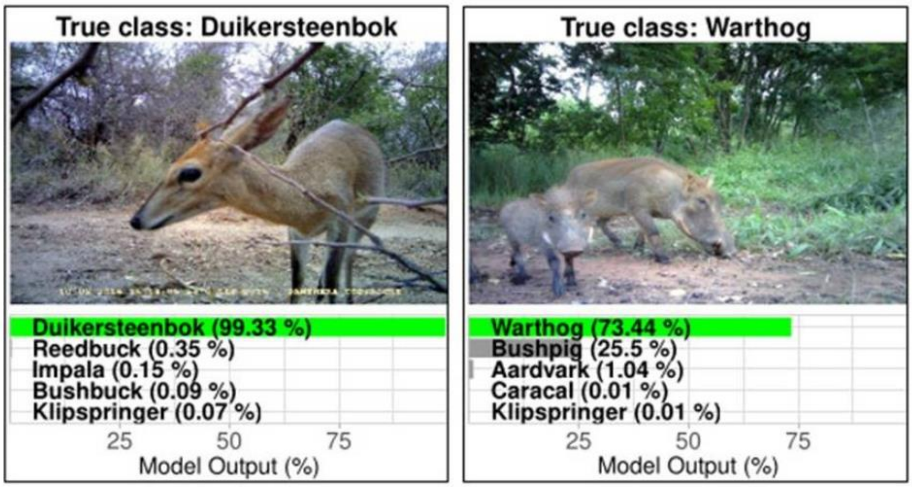

Figure 1 : Specific identification on camera trap images (Willi et al., 2019). Image reproduced with permission of the author

However, an image can be heterogeneous, it contain several elements, possibly from several classes of interest (e.g. several species). Object detection algorithms are then regularly used. Their principle is to detect the element and frame it with a rectangle (Figures 2 and 3), such as each animal present in a photo trap image . In our example, this approach would prove necessary if an image could contain both irises and tulips that would need to be detected. This principle is also used to detect a sub-part of an organism (we then speak of cropping), for example the flanks of giraffes [2-3] or the faces of chimpanzees . Again, the ecologist constitutes the training dataset with images for which the elements of interest are manually framed with the graphical interface of software [6-8].

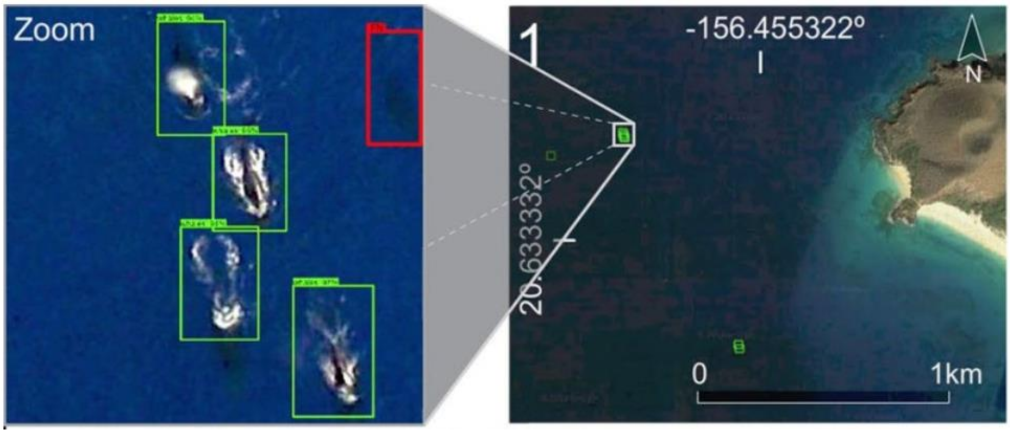

Figure 2 : Whale counting from aerial images (Guirado et al., 2019).

The same idea applies when it comes to counting from an image. This approach has been implemented in different contexts, to count wildebeests or whales (Figure 2) in aerial or satellite images, sea urchins in underwater images. -marines, but also any other enumerable element taking into account the resolution of the image (Figure 3) . Note that additional difficulties arise when the elements to be counted are only visible on a few pixels [2-3].

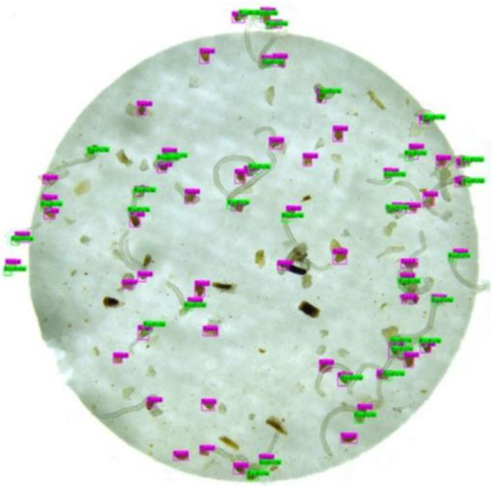

Figure 3 : Counting seeds in a gel.

On heterogeneous images, it can also be a matter of demarcating different areas in the image, such as different types of soil, vegetation or habitats (Figure 4).

The problem comes down to image segmentation where each pixel is classified by the learning algorithm into one of the classes of interest. Of course, without this framework, the effort of annotating the training data by the ecologist turns out to be very substantial since each image must be segmented manually.

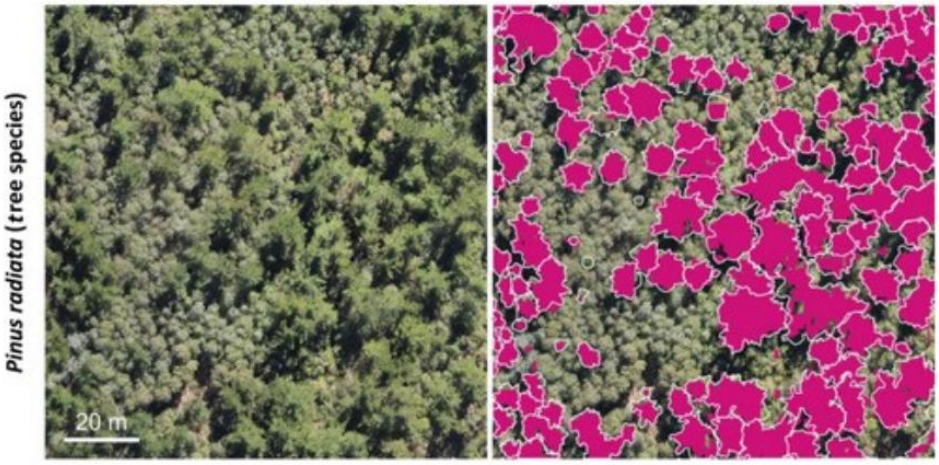

Figure 4 : Vegetation segmentation.

Furthermore, the detection of points of interest in an image can make it possible to study the morphological profile of an individual. For example, the algorithm can learn to predict the junction points between the limbs of an animal. Thus, recent approaches have made it possible to estimate the position of different parts of the body of a mouse or a fruit fly [4]. We then speak of pose estimation. These approaches open perspectives for the large-scale study of body movement and animal behavior.

Deep learning also makes it possible to consider individual animal recognition by image (Figure 5). Traditionally, this recognition has been based on the search for characteristics such as the shape of spots in the coat of giraffes. Recent work has borrowed facial recognition algorithms to adapt them to animal facial processing: these algorithms, applied to individual human recognition, however, raise ethical questions about respect for private life. Other studies have reformulated the problem of classification into a problem of individual recognition: the classes of interest are the different individuals, e.g. individual whales , giraffes or birds [2-3]. It should be noted, however, that these methods run into the problem of identifying individuals not known a priori.

(i.e. not present in the training game), necessary for the concrete implementation of a photographic capture-recapture approach.

Figure 5 : Individual recognition of chimpanzees (Schofield et al., 2019).

Although deep learning techniques are mainly dedicated to images, sounds can also be analyzed using these techniques, particularly in ecology. Sound can be directly processed by a specific model previously trained on a huge sound bank (e.g. all of YouTube) to classify soundscapes from different ecosystems [5]. But several studies have proposed working on sounds seen as images in a time/frequency/intensity representation.

II- Technical solutions for ecology

The various applications of image recognition in ecology have benefited from the revolution that has recently taken place in the field of deep learning. What is it about ? To understand it, we must first discuss the semantic subtleties of this field, from artificial intelligence to automatic (machine learning) and deep learning (Figure 6). Long before becoming a buzzword, artificial intelligence is first and foremost a well-identified scientific disciplinary field

[6], which focuses on techniques allowing a computer/system to perform a task or solve a problem that usually requires human intelligence. There are sub-disciplines such as robotics and natural language processing, as well as computer vision and machine learning which concern us here. This last term characterizes a set of methods relating to statistics, mathematics, computer science and algorithms: any algorithm that has the capacity to automatically learn structures in the data without having to worry about it can be qualified as machine learning. they must necessarily be made explicit. Said more simply: we do not tell the algorithm what to look for, we let it find structures of interest by itself in the data

In the following, we will restrict ourselves to so-called supervised learning techniques. It involves learning to solve a problem from a sample of data (the training set) for which the solution is known, i.e. learning a prediction function from examples annotated. The linear model, well known to ecologists, can appear in this category. For example, we can learn to recognize an iris from the morphology of the flowers of the three clades studied.

with a multinomial linear model giving the probability of classifying an image as an iris based, for example, on the angles between the sepals, the number of petals, etc. But let's return to the deep learning revolution for images. What is interesting here is that many computer vision problems will be able to be formulated as machine learning problems, so that two sub-disciplines of artificial intelligence come together.

Among these learning techniques, neural networks are the option that interests us here. Their general spirit is to imagine that by applying several waves of transformations to the data, one after the other, we must be able to immerse the data in a mathematical space which will facilitate the learning of prediction/classification rules. With this idea, we can imagine a wave which consists of transforming an image of a flower into a vector of the frequencies of the yellow, green and

purples. Then it will be easy to estimate a prediction rule on this vector to discriminate daisies and irises. Developed in the 1950s, this approach is based on the concept of an artificial neuron, a small box that transforms a data vector into a digital value. A neural network will be composed of several layers of neurons, each layer acting on the result of the previous layer (we find the idea of waves of transformations). Those in the first layer take the raw data as input, and those in the last layer will be the predictions (scores in the case of classification).

Our multinomial linear model is a particular example of a single-layer neural network composed of a linear neuron. Behind this complicated architecture, it is a question of imagining that we can optimize each neuron so that the model can predict as best as possible despite the presence of highly non-linear structures in the data: this is the learning phase.

An evolution of neural networks, convolutional neural networks (CNN) then come into play. Developed in the 1980s, but then little used, they were taken up by Alex Krizhevsky in 2012 during a recognition competition of images that he outrageously dominated [7]. These CNNs are based on an additional concept, convolution, which consists of feeding neurons not with the entire image but with each small piece of this image (for example, 9x9 pixels). Two images of daisies may look very different overall, but small pieces of these images will turn out to be very similar in specific areas such as the ends of the petals.

In practice, CNNs are composed of a large number of layers of neurons, so that we then speak of deep neural networks. By contraction of the various terminological elements, we speak of deep learning. In our example, each flower image

will be traversed by neurons which will transform it into numerical values. These will be used to estimate the best classification rule which will make it possible to predict the clade present on the images for which the solution is known (i.e. daisy, iris or tulip).

It can be explained by a conjunction of societal and scientific elements:

- the availability of masses of data (due to the Internet; in ecology, participatory science initiatives have also made it possible to create large databases);
- manual annotations carried out on a large scale due to a liberalized work organization [2-3];
- improving learning algorithms (e.g. data augmentation techniques);
- the involvement of giant digital companies (Facebook and Google in the lead);

- the diffusion of open source codes (e.g. via GitHub) and the emergence of a huge community of developers;
- massive parallel computing capabilities on graphics processors.

III- Limitation of using deep learning in the ecology field

In the above, we have outlined an overview of the applications of deep learning in ecology, then we have explained the principle of the methods. In what follows, we focus on the difficulties linked to their implementation in practice.

a. The size and completeness of the training set

To estimate the parameters of a CNN, millions of annotated images are needed. This is how the millions of parameters of the best-known models (ResNet, Inception, MobileNet) were estimated.

This amount of data is of course beyond the reach of ecologists. Fortunately, there is the technique of transfer learning, which consists of relying on a model

already estimated: in a way, we adapt an existing model to the question of interest, without having to build a model from scratch [2-3]. It is then possible to modify the parameters of the existing model by training with a few hundred images per class of interest for very simple cases.

However, that's not all: a CNN synthesizes the information contained in the training game, but not beyond. The higher the variability in CNN use cases, the larger the training set will need to be to include this variability when training the model. For example, to predict that an image contains a species based on any subpart of an animal,

it is necessary to include a large quantity of images with sub-parts of the animal when learning.

Another example is the prediction of “empty” images, in other words without the elements of interest, which requires a large number of training images associated precisely with the “empty” configuration [4]. For our example, a CNN model trained only on images of the three flower clades that interest us will not be able to recognize an image without a flower and will predict one of the three species.

The variability of the image background and environmental conditions (night/day, habitats, seasons) can also be decisive. For example, one study shows that a cow on a beach is not recognized by a model that was trained with images of alpine cows [5].

The training set must also cover the expected configurations in the images to be subsequently processed, which represents a challenge for ecologists. In summary, the more bias there is in the training game [6], the more learning techniques will produce errors when they are used in cases concrete on new images [7].

b. The difficult control of model performance

It is difficult to ensure the predictive power and generalizability [8] of a CNN. In other words, it is difficult to know how effective a CNN model will be on a completely new dataset, for example in a new location [9]. Indeed, the metrics used to monitor performance are calculated on so-called test data which, generally, are acquired under the same conditions as those of the training dataset [11]. Furthermore, annotation errors are possible, especially since it is necessary to constitute large datasets [10] and, as with many other methods , this aspect can lead to poor performance evaluation. For example, if our flower recognition algorithm was trained on a dataset including photos of tulips annotated as photos of iris, the performance of the model would certainly suffer.

c. The black box syndrome

It is currently impossible to know a priori the ideal architecture of a CNN dedicated to this or that task. Empiricism reigns here, extremely costly in terms of calculation time [12], with everyone testing the performance of different CNNs and choosing the best option.

Fortunately, ecologists will often restrict themselves to predefined model architectures (e.g. ResNet-50), in the context of transfer learning.

Furthermore, it is very difficult to interpret the predictive capacity of CNNs which are readily described as black boxes [13]. Could there be criteria understandable by humans that led to this ability? Any areas of the images that are particularly useful [14]? Or would there be unwanted and implicit elements contained in the context or background [15] that would make the model less generalizable? If, in our example, a daisy is predicted to be a tulip, then there is no easy element to query to understand the reasons for this error. To date, there are a few methods that attempt to interpret model parameters [16] but this still remains a subject of research.

d. The black box syndrome

The machine learning community, including those from digital giants, provides numerous open access codes that are reusable and modifiable. This represents an excellent opportunity for ecologists, except that these codes are mainly written in the Python language while ecologists favor the R language. On an optimistic note, it should be noted that the discovery of Python is affordable (numerous online resources available).

Then, the level of sophistication of the algorithms is such that the use of software libraries in the field (TensorFlow, Keras, PyTorch) is essential: additional effort to master these software libraries is therefore required. Finally, the rise of deep learning is linked to the rise of computing on graphics processors (GPUs), hardware made up of thousands of small computing units, initially dedicated to video games.

The matrix calculations associated with convolution are perfectly adapted to the constraints of GPUs and speed gains of around x100 (in our experiments, from a few minutes to a few hours respectively) are to be expected between a graphics card and a conventional computer. But, this equipment is difficult to install and manage without IT skills. It is also rarely present in

ecology laboratories. Inexpensive equipment, it consumes a lot of electricity and the installation of a graphics processor on the computer station of each ecologist would be anachronistic in a context of reducing greenhouse gas emissions. We recommend here to turn to the various academic centers for pooling calculation resources (e.g. Jean Zay national CNRS calculator or GriCad regional center).

IV- **Conclusion**

Deep learning techniques for computer vision arouse interest tinged with legitimate curiosity. We call here for a pragmatic approach aimed at consciously evaluating their strengths and weaknesses with regard to use in ecology. We still remind you that the human eye remains essential given the annotation (upstream) and validation (downstream) stages of a predictive model. The massive use of machines should also be coupled with ethical discussions on the potential dangers involved [17]. For example, the dissemination of information on the positioning of certain species could potentially be exploited by poachers [18]. Finally, in the context of participatory production programs, the desire to be part of an educational approach in order to raise awareness among citizens can also lead to not turning to the machine.

Here we venture some tips for taking advantage of deep learning:

- ✓ Fully understand the principle of machine learning;
- ✓ Start with simple and well-marked cases, most of the time to automate an easy but tedious task with the mass of data;
- ✓ Several current needs can be resolved, at least partially (i.e. for example in preprocessing before manual analysis), by simple approaches which are textbook cases for deep learning developers.
- ✓ Promote interdisciplinarity by forming working groups with different scientific sensibilities, from ecologists to specialists in image analysis and IT;
- ✓ The progressive mastery of the different concepts and techniques by as many people as possible will allow a collective evaluation of these approaches and the merits of their use in ecology.

V- References

[1]. Abrams, J. F., A. Vashishtha et al., 2019. Habitat-Net: Segmentation of habitat images using deep learning. *Ecological informatics* 51, 121-128.

[2]. Anses, 2020. La saison de cueillette des champignons commence : restez vigilants face aux risques d'intoxications !

[3]. Ärje, J., C. Melvad et al., 2020. Automatic image-based identification and biomass estimation of invertebrates. *Methods in Ecology and Evolution*.

[4]. Baraniuk, R., D. Donoho & M. Gavish, 2020. The science of deep learning. *Proceedings of the National Academy of Sciences* 117, 30029-30032.

[5]. Beery, S., D. Morris & S. Yang, 2019. Efficient pipeline for camera trap image review. *arXiv preprint arXiv:1907.06772*.

[6]. Beery, S., G. Van Horn & P. Perona. Recognition in terra incognita. *Proceedings of the European Conference on Computer Vision (ECCV)*, 456-473.

[7]. Bellman, R., 1966. Dynamic programming. *Science* 153, 34-37.

[8]. Bogucki, R., M. Cygan et al., 2019. Applying deep learning to right whale photo identification. *Conservation Biology* 33, 676-684.

[9]. Bolger, D. T., T. A. Morrison et al., 2012. A computer-assisted system for photographic mark--recapture analysis. *Methods in Ecology and Evolution* 3, 813-822.

[10]. Brodrick, P. G., A. B. Davies & G. P. Asner, 2019. Uncovering ecological patterns with convolutional neural networks. *Trends in ecology and evolution* 34, 734-745.

[11]. Charpentier, M. J. E., M. Harté et al., 2020. Same father, same face: Deep learning reveals selection for signaling kinship in a wild primate. *Science Advances* 6, eaba3274-eaba3274.

[12]. Christin, S., É. Hervet & N. Lecomte, 2019. Applications for deep learning in ecology. *Methods in Ecology and Evolution* 10, 1632-1644.

[13]. Duporge, I., O. Isupova et al., 2020. Using very-high-resolution satellite imagery and deep learning to detect and count African elephants in heterogeneous landscapes. *Remote Sensing in Ecology and Conservation*.

[14]. Dutta, A. & A. Zisserman, 2019. The VIA annotation software for images, audio and video Proceedings of the 27th ACM International Conference on Multimedia:2276-2279.

[15]. Ferreira, A. C., L. R. Silva et al., 2020. Deep learning-based methods for individual recognition in small birds. *Methods in Ecology and Evolution*.

[16]. Graving, J. M., D. Chae et al., 2019. DeepPoseKit, a software toolkit for fast and robust animal pose

[17]. estimation using deep learning. *eLife* 8, e47994-e47994. Guirado, E., S. Tabik et al., 2019. Whale counting in satellite and aerial images with deep learning. *Scientific Reports* 9, 1-12. Huang, H., H. Zhou, X. Yang, L. Zhang, L. Qi & A.-Y. Zang, 2019. Faster R-CNN for marine organisms detection and recognition using data augmentation. *Neurocomputing* 337, 372-384.

[18]. Jin, H., Q. Song & X. Hu. Auto-keras: An efficient neural architecture search system. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*,1946-1956