REVIEW

Deep learning in terrestrial conservation biology

Zoltán Barta[1] 📍

Abstract

Biodiversity is being lost at an unprecedented rate on Earth. As a first step to more effectively combat this process we need efficient methods to monitor biodiversity changes. Recent technological advance can provide powerful tools (e.g. camera traps, digital acoustic recorders, satellite imagery, social media records) that can speed up the collection of biological data. Nevertheless, the processing steps of the raw data served by these tools are still painstakingly slow. A new computer technology, deep learning based artificial intelligence, might, however, help. In this short and subjective review I oversee recent technological advances used in conservation biology, highlight problems of processing their data, shortly describe deep learning technology and show case studies of its use in conservation biology. Some of the limitations of the technology are also highlighted.

Introduction

It is widely accepted that Earth entered a new geological epoch, the Anthropocene (Lewis and Maslin 2015). This era started as the "great acceleration"—an unprecedented increase of human influence on the environment (Steffen et al. 2015). One part of this new extreme dynamism is that more and more natural habitats are converted to agricultural land and cities because of the continuously growing human population. Another consequences is the global distribution network driven by human commerce; the level of connectedness turned the planet into a "small world" (Johnson et al. 2017), with the subsequent increases in biological invasions (Chapman et al. 2017). Biological invasion is also driven by another human-triggered effect, the dynamically changing climatic conditions that endanger the existence of the current local biota and create footholds for new, potentially invasive organisms. And last but not least, the above processes (change of land use, connectedness and climatic changes) also make existing habitats become fragmented (Fahrig 2003). As a result of these changes natural habitats on Earth are seriously endangered, biodiversity decreases at an unparalleled rate (Lewis and Maslin 2015; Hallmann et al. 2017) possibly turning into the sixth mass extinction (Turvey and Crees 2019). If we are to reverse these adverse processes, or at least dampen them, we have to act now.

One of the factors hampering our ability to fight against these devastating processes is the lack of up-to-date information about biodiversity at a large spatial scale (Tuia et al. 2022). Traditional ecological surveys can provide us with rather accurate information on the state of biota, but these studies usually concentrate on relatively small areas and they are very labour intensive, demanding a considerable workforce of highly trained individuals (taxonomists, ecologists). Consequently, they are considered as unsuitable to provide frequent, large scale updates on biodiversity. To cope with these limitations, biodiversity researchers have turned towards new technological advances (van Klink et al. 2022).

Emerging technologies in biodiversity research

Camera traps are remotely operated devices equipped with motion detectors. This setup allows to automatically trigger the camera and take photos or videos when something has moved in the vicinity of the device. Digitalisation has significantly contributed to the widespread use of camera traps

✉ Zoltán Barta
  barta.zoltan@science.unideb.hu

1 HUN-REN-DE Behavioural Ecology Research Group,
   Department of Evolutionary Zoology and Humanbiology,
   University of Debrecen, Debrecen, Hungary

by simplifying the camera structure itself (e.g. fewer moving parts) as well as the management of images (more efficient storage on the device, no film processing). Hence the camera trap became an effective tool for wildlife observation allowing non-invasive monitoring even at remote places. They are mainly used to monitor large bodied mammals, wildlife and their predators. Nowadays, millions of pictures and videos are taken by camera traps every day all over the world from polar areas to the Equator.

Animals cannot only be seen but can also be heard as they actively produce sound when they communicate or hunt (echolocation). Consequently, sound recordings can also be used to survey animal populations. The appearance of cheap, digital sound recording systems make this approach (passive sound monitoring, PAM) even more feasible than camera traps. These devices are small with long battery lives and hence allow convenient recording of animal sounds even in remote locations. Their programmable interface allows to set either continuous or periodic (e.g. for one minute in every 15 min) recording of environmental sounds. PAM is, at the first sight, comparable to camera traps as both are based on autonomously operated remote devices. As a data collecting method, PAM can, however, have several advantages over camera traps, including longer detection distance, fully circular detection angle and a more diverse set of target species, including birds, bats, amphibians or insects.

Aerial photography has been used for decades to survey wildlife in open habitats, like African savannas or Eurasian steppes. In the early days of these surveys, manned aeroplanes were used incurring substantial cost in resources and sometimes human lives (Tuia et al. 2022). The appearance of commercially available drones (unmanned aerial vehicles) changed the situation considerably. These drones are cheaper than manned aeroplanes both in terms of acquisition and operation and equipped with high resolution camera(s). These characteristics make them well suited for frequent small scale aerial surveys. Their high resolution cameras can be used for monitoring animal populations or mapping invasive plant species (e.g. Papp et al. 2021).

Regular spaceborne Earth observation started in the early 70s by the launch of the Landsat-1 satellite. Afterwards more and more satellites joined the network providing more and more detailed images. A significant step forward was achieved when many satellite images become openly available (Crowley and Cardille 2020). Satellite remote sensing has several advantages even over in situ observations (Kissling et al. 2018). They provide a reliable and consistent periodic sampling at a large (global) scale. They are not affected by wind conditions which might be a serious problems for alternative remote sensing methods, such as unmanned aircraft (Müllerová et al. 2017). Moreover, sampling is not limited by national borders or other political barriers (Kissling et al. 2018). They have a rather fine

spatial resolution (10–60 m, Crowley and Cardille 2020) useful if one considers the mapping of habitat patches or ecosystem structures. Furthermore, this resolution allows an efficient integration of the dense information obtainable by multispectral recording (Ball et al. 2017; Young et al. 2017). Remote sensing data are successfully used in landscape ecology (for a recent review, see Crowley and Cardille 2020), surveying biodiversity (e.g. Madonsela et al. 2017) or analysing ecosystem time series (e.g. Wang and Zhao 2019).

Citizen science, the participation of volunteers in scientific projects has a long history; a famous example is the unpaid naturalist on HMS *Beagle*, Charles Darwin (Silvertown 2009). The recent technological advances, like the spread of smart phones and purposeful mobile applications make the involvement of volunteers even easier and more valuable. Platforms, like *iNaturalist.org*, *eBird.org* or *izeltlabuak.hu* led to huge collections of documented (photographed) and georeferenced occurrence data (Van Horn et al. 2018; La Sorte and Somveille 2020). As a result of the Open Science Movement (e.g. Powers and Hampton 2019), even more, well curated, open datasets have become available (Kissling et al. 2018). These datasets, by utilising the advanced Web 2.0 technology, are collected and organised in internet-accessible data repositories, like the *Global Biodiversity Information Facility* (https://www.gbif.org, Edwards 2004) or the *OpenBioMaps* (https://openbiomaps.org, Bán et al. 2022). As a result, the amount of georeferenced publicly available location data have skyrocketed. For instance, *gbif.org* hosts more than 1.5 billion(!) occurrence records in more than 56,000 datasets. The Hungarian initiative, *OpenBioMaps* brings data transparency closer to the user and provides a number of opportunities for interaction between researchers and practical conservationists.

Humanity has entered the Zettabyte Era around mid 2010s when the amount of data produced and processed exceeded one zettabyte (10^{21} bytes) per year (Bottles, Begoli, and Worley 2014). Since then the amount of data has been exponentially increasing due to the wide spread use of smart image-registering tools, ranging from smartphones to digital cameras. These devices produce an incredible number of pictures and videos, a large part of which is accessible to the wider public through content providers (e.g. blogs), photo repositories (e.g. Flicker), video streamers (e.g. Youtube) and social media (e.g. Facebook or Instagram). Importantly, many of these images contain relevant metadata, like the time and location of their production. A large proportion of these recording were made in nature creating a treasure houses for biodiversity research. The newly emerged scientific field of *iEcology* (internet ecology) studies ecological processes using such Internet-based data sources (Jarić et al. 2020a, b). *iEcology* studies range from mapping species distributions, through establishing interactions between populations to explore behavioural repertoires. Another, related

field of science, *conservation culturomics* investigates the relation between humans and nature using internet data, mainly concentrating on social media (Jarić et al. 2020a, b).

Problems with new technologies

A common feature of the above new technologies penetrating ecology and biodiversity research is that they generate an incredible amount of data in a short time. This, on the one hand, is a very useful attributes as it allows large scale survey of biodiversity. On the other hand, handling huge piles of data is not trivial. Fortunately, current computer science is well prepared to manage large amount of data and the developed "big data" algorithms can help the handling of data produced by technologies mentioned above.

Second, most of the data primarily produced by these emerging technologies have to be processed before they can be used for inference. Camera traps usually produce many empty recordings because they were triggered by irrelevant environmental movements like wind. These recordings must be filtered out before any inference. Furthermore, animals must be identified in the remaining recordings.

Processing drone videos suffers from similar inconvenience; the large area surveyed makes it difficult to spot, identify and count objects (usually animals) of interest. The analysis of satellite imageries is a two-fold challenge. In many times it is not entirely clear what kind of information should be extracted from the images and then how to obtain the relevant information.

The data produced by PAM devices are also difficult to analyse. First, sound recordings are usually not triggered by environmental events, but devices operate either continuously or on a preset schedule; both produce a large volume of data. Second, sound identification is more difficult than the identification of those couple of large bodied mammals which can usually be found on camera trap recordings. Therefore, analysis and sample labelling require highly trained experts. Finally, environmental noise (like wind or heavy rain) is a bigger problem for PAM.

Analysing social media posts is also problematic because of their sheer volume and semantics. Currently, these pre-processing steps are mainly being done by humans; armies of well trained and enthusiastic students, researchers and citizen scientists spend long hours to classify camera trap images, recognise calls, extract features from satellite images or interpret tweets. This high demand for human work limits the wide applicability of these emerging technologies (Tuia et al. 2022).

The third problem with emerging technologies is inference. Data collected in this way are usually unsuitable for analysing by traditional statistical methods. For instance, the analysis of camera trap data is not trivial because it is unclear how these data fulfil the assumptions of current population biology methods. Another main difficulty is data dimensionality as several of these methods (e.g. satellite imageries, social media) can deliver many variables simultaneously.

As we have seen above, these promising, new biodiversity data collection methods still cannot deliver their full potential because of the difficulties mainly appearing during the pre-processing of these data. Fortunately, another newly emerging technology, deep learning, may show a way out.

Deep learning

The invention of the term "artificial intelligence" (AI) in 1956 marked the birth of a new academic field which aimed to create machines able to automate intellectual tasks (Roitblat 2020). The history of AI can be divided into three phases. In the first phase, AI was approached by formal manipulation of symbols—*symbolic AI*. These efforts culminated in the development of *expert systems* aiding, for instance, medical diagnoses. During the second phase *machine learning* became the determinative paradigm in AI research. Machine learning is related to mathematical statistics and it is based on the idea that instead of hard coding rules in computers (as symbolic AI did) one should develop algorithms which themselves are able to discover rules in the data. The third phase of AI is the development of *deep learning*. Interestingly, these three phases were separated by two long periods—*the AI winters*—when interest in and funding for AI diminished (Roitblat 2020).

Deep learning algorithms are the extensions of "classical" artificial neural networks (ANN, Chollet 2021). These networks are built of interconnected artificial neurons or *nodes*. A node summarises several input values into an output value which can then be the input value of other nodes. Nodes are organised into layers where the first layer, layer 1, supplies the original input values, while the last one, layer N, represents the final output, the prediction. Intermediate layers are called hidden layers. Nodes in layer i receive their inputs from layer $i-1$ and send their outputs to nodes in layer $i+1$. Nodes in the same layer do not usually communicate with each other. The strength of the connections (i.e. how strongly a node influences the state of the other, connected node) is controlled by their weights, high weight means strong influence while low ones implicate weak ones (Goodfellow, Bengio, and Courville 2016). Weights are represented within the algorithms as tables of numerical values (matrices). The training of a network basically means to find the combination of weights which maps the original input to the final output. During training the hidden layers "learn" a representation of the data supplied by previous layers. Training consists of iterations of alternated forward and backward passes (Chollet 2021). In the forward pass the

network calculates the final representation of the input, the prediction, and calculate the difference between the prediction and the true value. In the backward pass the weights are adjusted to decrease the difference between the prediction and the true value. After training, the network's weights represent what the network had learned from the data. We call a network deep if the number of hidden layers are large, usually more than ten. Another characteristic showing the size of the network is the number of trainable parameters. Advanced deep learning applications have millions or even billions of parameters (e.g. ChatGTP, currently considered as the most advanced language model, has more than seven billions parameters).

Layers form the building blocks of neural network architectures (Chollet 2021). Different network architectures are suitable for different tasks. Linearly arranged layers are amongst the simplest architectures (deep neural net, DNN) where a node of a given layer is connected to every node of the previous layer. Networks like this perform well in analysing numerical (tabular) data. A frequently used network architecture is convolutional neural network (CNN) which excels in computer vision tasks. In CNNs nodes in the convolutional layers are connected only to a subset of nodes in the previous layer. A node's subset corresponds to a geometrically well-defined region of the input layer, usually an image. These subsets can overlap each other but cover all of the input layer. In this way CNNs somehow imitate the work of vertebrate visual systems (LeCun et al. 2015). Another interesting ANN architecture is recurrent neural network (RNN, Chollet 2021). Neither DNNs, nor CNNs have memory, they view each input sample independently to others. Nevertheless, many natural (and human) data have an autocorrelated structure. These include speech, songs and any time series data in general. In these data the current sample is not independent from previous samples—what word comes in a sentence is largely influenced by what was spoken before. More generally, the current state of the system is predetermined by its previous state(s), a very valuable observation which are not utilised by DNNs or CNNs because of their lack of memory. To remedy this deficiency RNNs were invented. In this architecture the output of one sample is fed back to the network to modify its internal state, which, in turn, used to process the next sample. RNNs were promising candidates for successful natural language models, but it soon turned out that (i) they had short memories and hence were unable to process long texts and (ii) their architecture prevented efficient parallelisation of computations. To overcome these problems a new deep learning architecture was invented, the Transformer (Vaswani et al. 2017). Instead of sequentially processing sequences, the Transformer processes elements of the input sequences in parallel but "pays attention" to other input elements simultaneously. This "self-attention" conserves sequential information but allows heavy parallelisation resulting in massive improvement in training speed and the amount of text processed.

Currently, deep learning is mainly used for visual tasks, like object identification in images or face recognition (mainly CNNs) and for natural language processing, like in digital assistances (Apple's Siri, Amazons' Alexa or Google Assistant) and chatbots (OpenAI's ChatGTP, Google's BERT).

Developing deep learning models is surprisingly democratic in the sense that anyone with moderately advanced computer literacy can have access to cutting edge deep learning technology (Chollet 2021). Two main factors have contributed to the widespread possibility to use deep learning. One is the availability of non-expensive hardware: current commercial video cards which are very efficient at parallel matrix multiplications—the essence of deep learning algorithms—have computing power comparable to that of the supercomputers of the 1990s (Chollet 2021). The other factor is the availability of open source and hence free development platforms. Two main open-source frameworks dominate the field: *PyTorch* (https://pytorch.org, Paszke et al. 2017) and *TensorFlow* (https://www.tensorflow.org, Abadi et al. 2016). As a consequence, huge number of projects deploying deep learning in different areas are initiated. *Github* (a popular code sharing repository, https://github.com), for instance (as of 16 February 2023), houses more than 100,000 *PyTorch* and more than 130,000 *TensorFlow* based projects, and many of them are related to nature conservation.

Applications of deep learning in conservation

Probably, one of the most important contributions of deep learning to nature conservation is providing tools for highly effective image processing.

Satellite imageries can allow efficient large scale surveys of huge areas especially if their effective handling is solved. At the moment, CNNs are the main workhorses for satellite imagery processing. Their success seems to crucially depend on their ability to consider high level spatial features (Rezaee et al. 2018).

A major threat for global biodiversity is the accelerated deforestation in the Amazonian Basin where the Earth's largest and most diverse tropical forest can be found. To readily protect this vast area, conservation agencies need an accurate picture about forest loss and, more importantly, a precise forecast of possible clearance in the near future which knowledge might then be used to design preventive actions, e.g. by anticipatory protection of identified endangered areas. Ball et al. (2021) developed a CNN based algorithm which was able to learn from short time

series formed from freely available databases and satellite imageries to predict possible locations of deforestation a year ahead with rather high accuracy. High resolution satellite images can even be used to identify large bodied mammals in open areas. A CNN counted elephants in heterogeneous landscapes with an accuracy comparable to humans' (Duporge et al. 2021) for a fraction of human effort. A novel and innovative application of CNNs is to survey genetic diversity by using high resolution satellite images. Kittlein et al. (2022) mapped microsatellite genotypes of a South American rodent, *Ctenomys australis*, living underground. Subsequently, they trained a CNN on high resolution satellite imagery to predict this genetic diversity. Currently, a variety of tools are developed to facilitate efficient satellite image processing like land use classification or object detection. For a comprehensive list of these projects see, for instance, Cole (2023).

The quick spreading of camera traps also provides a deluge of pictures to process (Tuia et al. 2022). CNNs were applied early to these data (Norouzzadeh et al. 2018) and nowadays several applications and ready-to-use pipelines help their processing. The web based service, *WildlifeInsights* (https://wildlifeinsights.org), provides a centralised platform where camera trap photos can be uploaded and analysed by state-of-the-art computer vision technologies (based on TensorFlow). The processed images and data are openly shared amongst members of the *WildlifeInsights* community. Another, decentralised approach is provided by the *MegaDetector* pipeline (Beery, Morris, and Yang 2019). This project consists of several tools for camera trap image processing, and each of them can be installed and operated locally. Amongst the tools is a pretrained CNN—also called *MegaDetector*—capable to recognise objects of interest like animals, humans and vehicles against a wide variety of different backgrounds (i.e. in pictures taken at different locations/environments). This pretrained CNN is not able to determine the species identity, but it still can save a lot of time and human effort by identifying and eliminating empty images—about 70% of camera trap pictures are usually empty (Beery, Morris, and Yang 2019). Furthermore, after reducing the image pool to pictures of interest, the training for species identification becomes more effective (Beery, Morris, and Yang 2019). Current CNN technology is able to achieve more than 90% of human accuracy of species identification in camera trap images if many labelled images (images with animals already identified by humans) are available for training (Norouzzadeh et al. 2018).

Somehow surprisingly, PAM data are usually analysed as pictures, i.e. the sound recordings are converted to spectrograms, a visual representation of sound, and then the standard computer vision toolset (like CNNs) is used to identify them (Sugai et al. 2019). New developments, based on the Transformer architecture, however, try to exploit the inherent time series structure of sound to improve accuracy (Stowell 2021).

Citizen scientists can substantially contribute to nature conservation efforts by providing a huge amount of observations on the occurrence of many species, from mushrooms through plants and butterflies to birds. The quality of these observations was, however, a concern for professional ecologists and conservation scientists (Brown and Williams 2019). This situation has considerably improved when citizen science projects started to request digital photographs to be submitted along the observations because this made verification possible (Wäldchen and Mäder 2018). This modification is also resulted in the building up of massive databases of labelled images, that can be used to train deep learning (usually CNN) algorithms (Van Horn et al. 2018; Wäldchen and Mäder 2018). Based on these treasure chests, citizen science projects can offer automated species identification services. One of the most popular of these services is *iNaturalist* (https://www.inaturalist.org) with more than 5 millions records. The *iNaturalist* dataset (Van Horn et al. 2018) clearly illustrates the difficulties that the training of deep learning algorithms faces: the images are of varying quality and the dataset contains a few species with many images and many species with only a few pictures. Despite of these obstacles the species identification algorithm of *iNaturalist* achives > 80% accuracy at the genus level identification and close to 80% at the species level (Wäldchen and Mäder 2018). *Merlin Photo-ID* (https://merlin.allaboutbirds.org/photo-id/) is another popular tool to identify birds. It is based on the huge image database collected and labelled by the citizen scientist community behind *eBird* (https://ebird.org). The tools available to citizen scientists are not restricted to pictures. The highly successful *Merlin Sound ID* smartphone application performs exceptionally well for bird songs (https://merlin.allaboutbirds.org/sound-id/). As a result of high accuracy of species identification offered by deep learning algorithms citizen science occurrence data become comparable to experts' data (e.g. Mahecha et al. 2021). Studies also indicate that even custom built CNNs are capable for reliable species identification in specific areas. For instance, Łysko et al. (2022) developed a CNN based algorithm to identify *Elatine* plants (a small genus consisting of difficult-to-identify, ephemeral aquatic species). Their method, based on photographs of *Elatine* seeds consistently outperformed traditional machine learning methods.

Digital images of animals, however, cannot only be used for species identification alone, but—especially in the case of animals with variable patterns on their integument (e.g. skin, plumage or fur)—also to recognise individuals (Vidal et al. 2021). Individual identification is crucial for many ecological analyses. For instance, it makes possible to estimate survival rates, home ranges and migration rates, and to map movement patterns and social interactions.

Individual identification on the basis of photographs is, however, a complex process (Vidal et al. 2021). Nevertheless, several projects provide useful solutions, mainly for species with conspicuous patterns (Vidal et al. 2021). A highly successful attempt is *Wild Me* (https://www.wildme.org) which provides two open source platforms, the original *Wildbook* and the newly released *Codex*, that help to handle large volume animal images and eases the deployment of state-of-the-art deep learning algorithms in order to facilitate individual identifications and population assessments (Berger-Wolf et al. 2017). A permanent problem for machine learning aided individual identification is to obtain enough images of individuals (Vidal et al. 2021). *Wild Me* introduced a novel solution and automatically harvests pictures from social media platforms (Araujo et al. 2020). Sound recordings may also be a promising tool for individual identification (Stowell 2021).

As we have seen above *iEcology* also collects data from the world wide web to answer ecological and conservation biology questions. For instance, August et al. (2020) used the photo sharing social media site *Flickr* (https://flickr.com) and *Pl@ntNet* (https://plantnet.org) CNN service to survey the flora in urban and rural areas. The microblog service *Twitter* (https://twitter.com) is also an almost inexhaustible source of information about wildlife. Edwards et al. (2022) successfully applied the BERT natural language model (Devlin et al. 2019) to extract tweets about wildlife. *Conservation culturomics* (Correia et al. 2021), on the other hand, studies human-nature intersection by mining the vast resources provided by internet activities (visits, blogs, videos, social media etc.).

Deep learning is not only used to analyse images and text to help nature conservation. Zizka et al. (2020) used a self-developed deep learning algorithm to identify endangered orchid species. The algorithm was trained on data obtained from *GBIF* (https://www.gbif.org) and the International Union for Conservation of Nature (IUCN, https://iucn.org) assessments. They were able to assess the conservation status of nearly 14,000 species with a rather high accuracy. This is a large step forward given that the IUCN database previously contained only ca. 900 species. About 30% of these automatically assessed species were classified as possibly threatened and the authors were also able to identify several priority regions for orchids conservation (Zizka et al. 2020).

Limitations

The above examples illustrate the usefulness of deep learning in conservation biology. Nevertheless, a couple of limitations hinder to realise its full potential. One group of problems is technical. First, efficient training in deep learning requires a huge number of labelled records. For instance,

Norouzzadeh et al. (2018) trained their networks on millions of pictures, previously labelled by experts and volunteers. This means that to run a successful deep learning project in nature conservation one needs a huge initial investment in terms of labeling. A second technical limitation of deep learning is the lack of ability of deep learning models to generalise (Marcus 2018; Roitblat 2020; Tuia et al. 2022). The generalisation issue arises because deep learning models are paying attention to the whole picture (i.e. the background too), and not just the details of interest (i.e. the tracked animal). This inability to generalise prevents deep learning algorithms trained under a specific set of circumstances to be used under different circumstances (Marcus 2018). As a consequence, deep learning algorithms must be retrained for new projects. A third problem is related to the usual structure of natural data, which are very diverse and, more importantly, extremely imbalanced (a few species are represented by many images, while many species just by a few ones, see above, Van Horn et al. 2018). These inherent features make the training of deep learning NNs rather difficult (Tuia et al. 2022).

Fortunately, novel training approaches appear to overcome these technical problems. An especially promising one is active learning for processing camera trap data (Norouzzadeh et al. 2021). In this approach, a very generally trained object detection NN algorithm is run to separate empty pictures from those that contain animals. Empty pictures are excluded from further processing. Following this, humans are asked to label a small portion of the remaining, nonempty images. At the same time, image dimensionality is reduced by an embedding algorithm to create a low dimensional feature set for each picture. After labeling, a classification NN trained on the labelled feature sets to recognise species. Subsequently, the pipeline compares the remaining unlabelled images to the classified ones and selects for further human labelling those ones whose labeling can most improve the classification model. This process is repeated until the classification NN becomes accurate enough. As only a subset of images is labelled in this algorithm it can considerably reduce the labeling effort. For instance, Norouzzadeh et al. (2021) obtained the same accuracy as Norouzzadeh et al. (2018) on the same dataset, but with 99.5% less labeling effort. The data hungry characteristic of deep learning can also be tamed by either augmenting the available images or using deep learning algorithms to generate training data (Weinstein et al. 2020).

Another technical problem of deep learning, however, leads to ethical concerns when applying these algorithms (Roitblat 2020). This arises from the fact that it is immensely difficult to understand how a given deep learning algorithm works internally, and how it makes decisions. In this respect, it is a black box (Rudin 2019). It is also difficult to know what kind of implicit assumptions an algorithm is based

Table 1 The number of articles resulted from searching Clarivate Web of Science database by using the given search terms on 13th March 2023

Search terms	Alone	w/deep learning	%
"Deep learning"	132,231 (73%)	–	–
"Wildlife"	43,037 (19%)	196 (79%)	0.46
"Nature conservation"	3695 (25%)	19 (86%)	0.51
"Invasive plants"	1326 (26%)	16 (100%)	1.21
Search terms	Alone	w/deep learning	%
"Deep learning"	314 (68%)	–	–
"Wildlife"	303 (37%)	0	0.00
"Nature conservation"	82 (22%)	0	0.00
"Invasive plants"	18 (38%)	0	0.00

The searches were restricted for the period of 2020–2023 (inclusive). Search terms was run either alone or with "deep learning" (e.g. "deep learning" AND "wildlife"). The percentages after the article numbers show the proportion of papers published after 1st January 2020 to the total number of articles for the given search term. Column "%" gives the percentage of papers with deep learning and the given search term to the number of papers without the "deep learning" term. The top table gives the number of papers without any geographical restriction, while the bottom one gives the number of papers with at least one author with Hungarian address (search term of "address: Hungary")

on and, consequently, when an algorithm extrapolates outside the range of its training data (deep learning algorithms are notoriously bad in extrapolating, Rudin 2019; Roitblat 2020). Consequently, DNN algorithms can produce enormous results in unexpected ways. This is difficult to tolerate when human lives or the existence of critically endangered species are at stake (Wearn et al. 2019). Therefore, ethical issues of using deep learning must be considered and development of ethical guidelines must accompany the development of deep learning algorithms in conservation biology (Wearn et al. 2019).

The future

Deep learning is a quickly developing field as around three quarters of papers published in this subject in the world appeared just in the last 26 months (Table 1). The adaptation of deep learning into conservation biology is also a recent phenomenon. Currently, only a small fraction of papers related to conservation biology with deep learning. In Hungary, the study of deep learning lags behind the world (68 vs. 73% of papers are published recently). This is especially true for the application of deep learning in conservation biology, as no paper has yet been in the intersection of these fields (at least according to the *Web of Science* database). Based on these data it is expected that deep learning will penetrate conservation biology more deeply in the coming years. There are ample spaces for this kind of development in Hungary as well. Indeed, a research group has already been established at the University of Debrecen to use deep learning and satellite imagery to predict species distribution maps.

It seems that one of the most important factors hindering the quick adaptation of deep learning in conservation biology is the lack of appropriately trained experts in this truly interdisciplinary field. There are positive signs, however. The team of highly trained computer scientists and ecologists behind the *Wild Me* project, the foundation of the *Imageomics Institute* at the Ohio State University (USA, https://imageomics.osu.edu/) or the *Summer Workshop on Computer Vision Methods for Ecology* (https://cv4ecology.caltech.edu/) are promising examples. Nevertheless, we need more formal training, preferably at the PhD level.

Declarations

Conflict of interest The author declares no competing interest.

References

Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS et al (2016) TensorFlow: large-scale machine learning on heterogeneous distributed systems. arXiv https://doi.org/10.48550/arXiv.1603.04467

Araujo G, Ismail AR, McCann C, McCann D, Legaspi CG, Snow S, Labaja J, Manjaji-Matsumoto M, Ponzo A (2020) Getting the most out of citizen science for endangered species such as whale shark. J Fish Biol 96(4):864–867. https://doi.org/10.1111/jfb.14254

August TA, Pescott OL, Joly A, Bonnet P (2020) AI naturalists might hold the key to unlocking biodiversity data in social media imagery. Patterns 1(7):100116. https://doi.org/10.1016/j.patter.2020.100116

Ball JGC, Petrova K, Coomes DA, Flaxman S (2021) Using deep convolutional neural networks to forecast spatial patterns of Amazonian deforestation. Methods Ecol Evol 100:200. https://doi.org/10.1111/2041-210X.13953

Ball JE, Anderson DT, Chan CS (2017) Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. J Appl Remote Sens 11(4):042609. https://doi.org/10.1117/1.JRS.11.042609

Bán M, Boné GM, Bérces S, Barta Z, Kovács I, Ecsedi K, Sipos K (2022) OpenBioMaps—self-hosted data management platform and distributed service for biodiversity related data. Earth Sci Inform. https://doi.org/10.1007/s12145-022-00818-3

Beery S, Morris D, Yang (2019) Efficient pipeline for camera trap image review. https://doi.org/10.48550/arXiv.1907.06772

Berger-Wolf TY, Rubenstein DI, Stewart CV, Holmberg JA, Parham J, Menon S, Crall J, Van Oast J, Kiciman E, Joppa L (2017) Wildbook: crowdsourcing, computer vision, and data science for conservation. https://doi.org/10.48550/arXiv.1710.08880

Bottles K, Edmon B, Brian W (2014) Understanding the pros and cons of big data analytics

Brown ED, Williams BK (2019) The potential for citizen science to produce reliable and useful information in ecology. Conserv Biol 33(3):561–569. https://doi.org/10.1111/cobi.13223

Chapman D, Purse BV, Roy HE, Bullock JM (2017) Global trade networks determine the distribution of invasive non-native species. Glob Ecol Biogeogr 26(8):907–917. https://doi.org/10.1111/geb.12599

Chollet F (2021) Deep learning with python, 2nd edn. Manning Publications, Shelter Island

Cole RM (2023) Satellite-Image-Deep-Learning. https://github.com/robmarkcole/satellite-image-deep-learning

Correia RA, Ladle R, Roll U (2021) Advancing conservation culturomics introduction. Conserv Biol 35(2):395–397. https://doi.org/10.1111/cobi.13700

Crowley MA, Cardille JA (2020) Remote sensing's recent and future contributions to landscape ecology. Curr Landsc Ecol Rep 5(3):45–57. https://doi.org/10.1007/s40823-020-00054-9

Devlin J, Chang MW, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. https://doi.org/10.48550/arXiv.1810.04805

Duporge I, Isupova O, Reece S, Macdonald DW, Wang T (2021) Using very-high-resolution satellite imagery and deep learning to detect and count African elephants in heterogeneous landscapes. Remote Sens Ecol Conserv 100:200. https://doi.org/10.1002/rse2.195

Edwards JL (2004) Research and societal benefits of the global biodiversity information facility. Bioscience 54(6):485–486. https://doi.org/10.1641/0006-3568(2004)054[0486:RASBOT]2.0.CO;2

Edwards T, Jones CB, Corcoran P (2022) Identifying wildlife observations on twitter. Eco Inform 67:101500. https://doi.org/10.1016/j.ecoinf.2021.101500

Fahrig L (2003) Effects of habitat fragmentation on biodiversity. Annu Rev Ecol Evol Syst 34:487–515

Goodfellow I, Bengio Y, Courville A (2016) Deep learning. Adaptive computation and machine learning. The MIT Press, Cambridge

Hallmann CA, Sorg M, Jongejans E, Siepel H, Hofland N, Schwan H, Stenmans W et al (2017) More than 75 percent decline over 27 years in total flying insect biomass in protected areas. PLoS ONE 12(10):e0185809. https://doi.org/10.1371/journal.pone.0185809

Jarić I, Correia RA, Brook BW, Buettel JC, Courchamp F, Di Minin E, Firth JA et al (2020a) iEcology: harnessing large online resources to generate ecological insights. Trends Ecol Evol 35(7):630–639. https://doi.org/10.1016/j.tree.2020.03.003

Jarić I, Roll U, Arlinghaus R, Belmaker J, Chen Y, China V, Douda K et al (2020b) Expanding conservation culturomics and iEcology from terrestrial to aquatic realms. PLoS Biol 18(10):e3000935. https://doi.org/10.1371/journal.pbio.3000935

Johnson CN, Balmford A, Brook BW, Buettel JC, Galetti M, Guangchun L, Wilmshurst JM (2017) Biodiversity losses and conservation responses in the anthropocene. Science 356(6335):270–275. https://doi.org/10.1126/science.aam9317

Kissling WD, Ahumada JA, Bowser A, Fernandez M, Fernández N, García EAlonso, Guralnick RP et al (2018) Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. Biol Rev 93(1):600–625. https://doi.org/10.1111/brv.12359

Kittlein MJ, Mora MS, Mapelli FJ, Austrich A, Gaggiotti OE (2022) Deep learning and satellite imagery predict genetic diversity and differentiation. Methods Ecol Evol 13(3):711–721. https://doi.org/10.1111/2041-210X.13775

van Klink R, August T, Bas Y, Bodesheim P, Bonn A, Fossøy F, Høye TT et al (2022) Emerging technologies revolutionise insect ecology and monitoring. Trends Ecol Evol 37(10):872–885. https://doi.org/10.1016/j.tree.2022.06.001

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444. https://doi.org/10.1038/nature14539

Lewis SL, Maslin MA (2015) Defining the anthropocene. Nature 519:171–180. https://doi.org/10.1038/nature14258

Łysko A, Popiela A, Forczmański P, Attila MV, Lukács BA, Barta Z, Maćków W, Wolski GJ (2022) Comparison of discriminant methods and deep learning analysis in plant taxonomy: a case study of *Elatine*. Sci Rep 12(1):20450. https://doi.org/10.1038/s41598-022-24660-1

Madonsela S, Cho MA, Ramoelo A, Mutanga O (2017) Remote sensing of species diversity using Landsat 8 spectral variables. ISPRS J Photogramm Remote Sens 133(N):116–127. https://doi.org/10.1016/j.isprsjprs.2017.10.008

Mahecha MD, Rzanny M, Kraemer G, Mäder P, Seeland M, Wäldchen J (2021) Crowd-sourced plant occurrence data provide a reliable description of macroecological gradients. Ecography 44(8):1131–1142. https://doi.org/10.1111/ecog.05492

Marcus G (2018) Deep learning: a critical appraisal. arXiv:1801.00631 [Cs, Stat]

Müllerová J, Brůna J, Bartaloš T, Dvořák P, Vítková M, Pyšek P (2017) Timing is important: unmanned aircraft vs. satellite imagery in plant invasion monitoring. Front Plant Sci. https://doi.org/10.3389/fpls.2017.00887

Norouzzadeh MS, Morris D, Beery S, Joshi N, Jojic N, Clune J (2021) A deep active learning system for species identification and counting in camera trap images. Methods Ecol Evol 12(1):150–161. https://doi.org/10.1111/2041-210X.13504

Norouzzadeh MS, Nguyen A, Kosmala M, Swanson A, Palmer MS, Packer C, Clune J (2018) Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proc Natl Acad Sci 115(25):E5716–E5725. https://doi.org/10.1073/pnas.1719367115

Papp L, van Leeuwen B, Szilassi P, Tobak Z, Szatmári J, Árvai M, Mészáros J, Pásztor L (2021) Monitoring invasive plant species using hyperspectral remote sensing data. Land 10(1):29. https://doi.org/10.3390/land10010029

Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L, Lerer A (2017) Automatic differentiation in PyTorch

Powers SM, Hampton SE (2019) Open science, reproducibility, and transparency in ecology. Ecol Appl 29(1):e01822. https://doi.org/10.1002/eap.1822

Rezaee M, Mahdianpari M, Zhang Y, Salehi B (2018) Deep convolutional neural network for complex wetland classification using optical remote sensing imagery. IEEE J Sel Top Appl Earth Obs Remote Sens 11(9):3030–3039. https://doi.org/10.1109/JSTARS.2018.2846178

Roitblat HL (2020) Algorithms are not enough: creating general artificial intelligence. The MIT Press, Cambridge

Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1(5):206–215. https://doi.org/10.1038/s42256-019-0048-x

Silvertown J (2009) A new dawn for citizen science. Trends Ecol Evol 24(9):467–471. https://doi.org/10.1016/j.tree.2009.03.017

La Sorte FA, Somveille M (2020) Survey completeness of a global citizen-science database of bird occurrence. Ecography 43(1):34–43. https://doi.org/10.1111/ecog.04632

Steffen W, Broadgate W, Deutsch L, Gaffney O, Ludwig C (2015) The trajectory of the anthropocene: the great acceleration. Anthropocene Rev 2(1):81–98. https://doi.org/10.1177/2053019614564785

Stowell D (2021) Computational bioacoustics with deep learning: a review and roadmap. https://doi.org/10.48550/arXiv.2112.06725

Sugai LS, Silva TS, Ribeiro JW Jr, Llusia D (2019) Terrestrial passive acoustic monitoring: review and perspectives. Bioscience 69(1):15–25. https://doi.org/10.1093/biosci/biy147

Tuia D, Kellenberger B, Beery S, Costelloe BR, Zuffi S, Risse B, Mathis A et al (2022) Perspectives in machine learning for wildlife conservation. Nat Commun 13(1):792. https://doi.org/10.1038/s41467-022-27980-y

Turvey ST, Crees JJ (2019) Extinction in the anthropocene. Curr Biol 29(19):R982–R986. https://doi.org/10.1016/j.cub.2019.07.040

Van Horn G, Mac Aodha O, Song Y, Cui Y, Sun C, Shepard A, Adam H, Perona P, Belongie S (2018) The inaturalist species classification and detection dataset, pp 8769–78. https://openaccess.thecvf.com/content_cvpr_2018/html/Van_Horn_The_INaturalist_Species_CVPR_2018_paper.html

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: Guyon I, Von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R (eds) Advances in neural information processing systems, vol 30. Curran Associates Inc.

Vidal M, Wolf N, Rosenberg B, Harris BP, Mathis A (2021) Perspectives on individual animal identification from biology and computer vision. arXiv:2103.00560 [Cs, q-Bio]

Wäldchen J, Mäder P (2018) Machine learning for image based species identification. Methods Ecol Evol 9(11):2216–2225. https://doi.org/10.1111/2041-210X.13075

Wang C, Zhao H (2019) Analysis of remote sensing time-series data to foster ecosystem sustainability: use of temporal information entropy. Int J Remote Sens 40(8):2880–2894. https://doi.org/10.1080/01431161.2018.1533661

Wearn OR, Freeman R, Jacoby DMP (2019) Responsible AI for conservation. Nat Mach Intell 1(2):72–73. https://doi.org/10.1038/s42256-019-0022-7

Weinstein BG, Marconi S, Aubry-Kientz M, Vincent G, Senyondo H, White EP (2020) DeepForest: a python package for RGB deep learning tree crown delineation. Methods Ecol Evol 11(12):1743–1751. https://doi.org/10.1111/2041-210X.13472

Young NE, Anderson RS, Chignell SM, Vorster AG, Lawrence R, Evangelista PH (2017) A survival guide to landsat preprocessing. Ecology 98(4):920–932. https://doi.org/10.1002/ecy.1730

Zizka A, Silvestro D, Vitt P, Knight TM (2020) Automated conservation assessment of the orchid family with deep learning. Conserv Biol. https://doi.org/10.1111/cobi.13616