Received: 16 March 2018 | Accepted: 30 July 2018
DOI: 10.1111/2041-210X.13075

REVIEW

Methods in Ecology and Evolution

Machine learning for image based species identification

Jana Wäldchen[1] | Patrick Mäder[2]

[1]Biogeochemical Integration, Max Planck Institute for Biogeochemistry, Jena, Thuringia, Germany

[2]Software Engineering for Safety-Critical Systems Group, Technische Universität Ilmenau, Ilmenau, Thuringia, Germany

Correspondence
Jana Wäldchen
Email: jwald@bgc-jena.mpg.de

Funding information
German Ministry of Education and Research (BMBF), Grant/Award Number: 01LC1319A and 01LC1319B; German Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety (BMUB), Grant/Award Number: 3514 685C19; Stiftung Naturschutz Thüringen (SNT), Grant/Award Number: SNT-082-248-03/2014.

Handling Editor: Natalie Cooper

Abstract

1. Accurate species identification is the basis for all aspects of taxonomic research and is an essential component of workflows in biological research. Biologists are asking for more efficient methods to meet the identification demand. Smart mobile devices, digital cameras as well as the mass digitisation of natural history collections led to an explosion of openly available image data depicting living organisms. This rapid increase in biological image data in combination with modern machine learning methods, such as deep learning, offers tremendous opportunities for automated species identification.
2. In this paper, we focus on deep learning neural networks as a technology that enabled breakthroughs in automated species identification in the last 2 years. In order to stimulate more work in this direction, we provide a brief overview of machine learning frameworks applicable to the species identification problem. We review selected deep learning approaches for image based species identification and introduce publicly available applications.
3. Eventually, this article aims to provide insights into the current state-of-the-art in automated identification and to serve as a starting point for researchers willing to apply novel machine learning techniques in their biological studies.
4. While modern machine learning approaches only slowly pave their way into the field of species identification, we argue that we are going to see a proliferation of these techniques being applied to the problem in the future. Artificial intelligence systems will provide alternative tools for taxonomic identification in the near future.

KEYWORDS
automated species identification, computer vision, convolutional neural network, deep learning, images

1 | INTRODUCTION

Accurate species identification is the basis for all aspects of taxonomic research and is an important component of workflows in biological research, including medicine, ecology and evolutionary studies. Many activities, such as studying the biodiversity richness of a region, monitoring populations of endangered species, determining the impact of climate change on species distribution and weed control actions depend on accurate identification skills. These activities are a necessity for e.g. farmers, foresters, pharmacologists, taxonomists, conservation biologists, technical personnel of environmental agencies or just fun for laypersons (Austen, Bindemann, Griffiths, & Roberts, 2016; Farnsworth et al., 2013). Automating the task and making it feasible for non-experts is highly desirable, especially considering the continuous loss of biodiversity (Ceballos et al., 2015) and of experienced taxonomists (Hopkins & Freckleton, 2002).

The fast development and ubiquity of relevant information technologies in combination with the availability of portable devices such as digital cameras and smartphones resulted in a vast number of digital images that are openly available in online databases. Examples highlighting the wealth of images captured by researchers and the public are the iNaturalist (675,000 images of 5,000 species, Van Horn et al., 2017) and the Zooniverse (1.2 million images of 40 species; Swanson et al., 2015) databases. Furthermore, the mass digitisation of natural history collections has become a major goal at museums around the world and has already resulted in large digital datasets. For example, the iDigBio portal, a nationally funded aggregator of museum specimen data, currently provides more than 1.8 million georeferenced images of vascular plant specimens (Willis et al., 2017). Such large image datasets in combination with the latest advances in machine learning technologies bring automating image based species identification to reality.

Considerable research in the field of computer vision and machine learning resulted in a plethora of papers proposing and comparing methods on automated species identification (Wäldchen & Mäder, 2018; Weinstein, 2018). Most of these studies published were conducted by technical specialists in computer vision, machine learning and multimedia information retrieval making the proposed methods difficult to access for biologists (Wäldchen & Mäder, 2018).

However, machine learning software becomes more and more user-friendly enabling people without substantial computer science background to individually apply the latest algorithms to their problems and datasets. Nonetheless, a basic understanding of the applied technologies and some time to get acquainted with them is still required. In this paper, we give a brief introduction into the state-of-the-art in machine learning techniques applicable for automated species identification. More specifically, we introduce the basic concepts, give an overview of existing machine learning frameworks, introduce the latest studies applying machine learning for species identification and discuss future research directions.

2 | MACHINE LEARNING IN COMPUTER VISION

Today, **machine learning** is the fastest growing field in computer science pervading fields as diverse as marketing, health care, manufacturing, information security and transportation. The main reason for this literal "explosion" of the technique is the availability and confluence of three things: (a) faster and more powerful computer hardware, i.e. massively parallel processors and general purpose graphics processing units (GP-GPUs); (b) software algorithms that take advantage of these computational architectures; and (c) almost unlimited amounts of training data for a given problem, such as digital images, digitized documents, social media posts or observations with geolocation. Machine learning is a form of artificial intelligence that can perform a task without being specifically programmed to solve it. Instead, it learns from previous examples of the given task during a process called training. After training, the task can be performed on new data in a process called inference (Mjolsness & DeCoste, 2001). Machine learning especially helps in extracting information from large amounts of continuously growing data and is particularly useful for applications

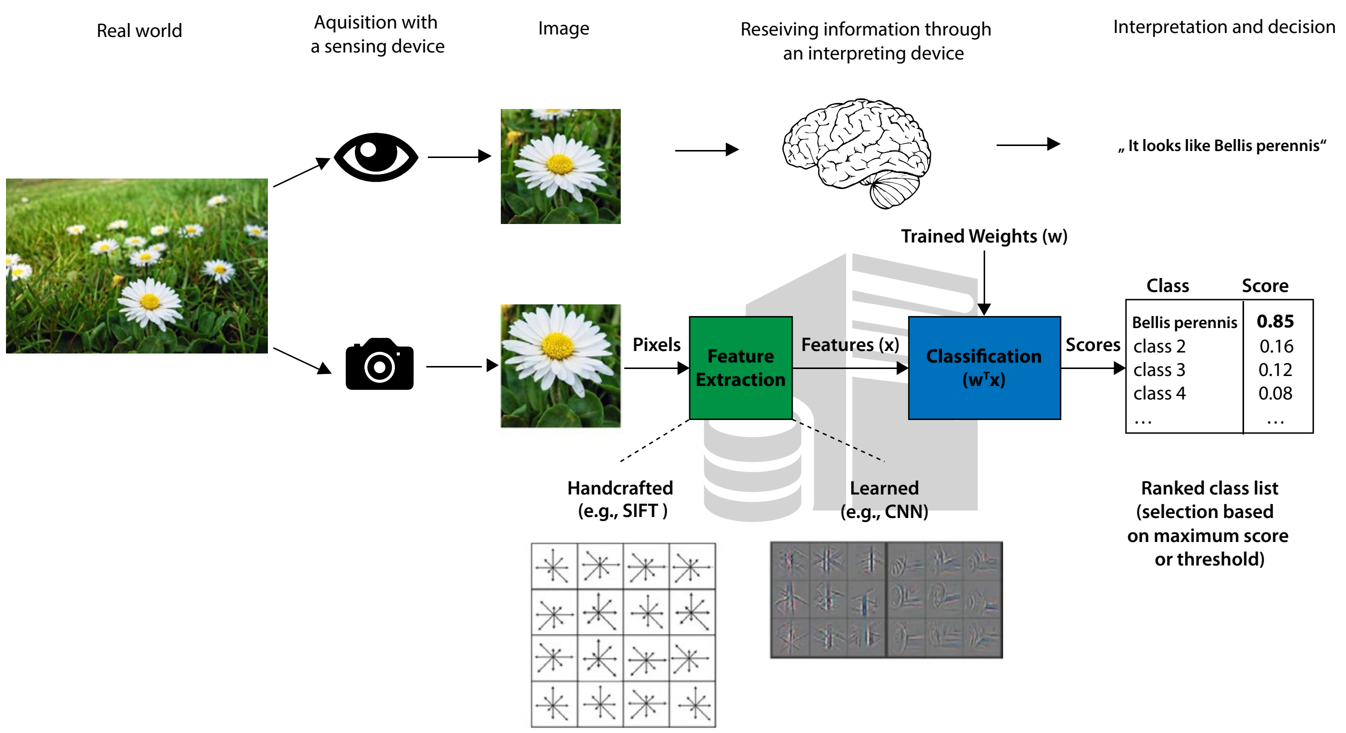

FIGURE 1 Typical human and computer vision pipeline for species identification. The machine learning platform takes in an image and outputs the confidence scores for a predefined set of classes

where the data is difficult to model analytically, for instance, analyzing image and video content.

Computer vision is a field of computer science that deals with gaining understanding and insights from digital images and videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do (Sonka, Hlavac, & Boyle, 2014). A computer vision machine learning pipeline consists of two phases: feature extraction and classification. Figure 1 shows a computer vision pipeline for plant species identification as an example explaining basic computer vision terminology. Solutions for image based species identification tasks are manifold and were comprehensively surveyed before, e.g. by Cope, Corney, Clark, Remagnino, and Wilkin (2012); Moniruzzaman, Islam, Bennamoun, and Lavery (2017); Seeland, Rzanny, Alaqraa, Wäldchen, and Mäder (2017); Wäldchen and Mäder (2018); Weinstein (2018).

2.1 | Feature extraction

Feature extraction transforms the raw data into meaningful representations for a given classification task. Images are typically composed of millions of pixels with associated colour information each. The high dimensionality of these images is reduced by computing abstract features, i.e. a quantified representation of the image retaining relevant information for the classification problem (e.g. shape, texture or colour information) and omitting irrelevant. Traditionally, features to be extracted were designed by domain experts in a typically long term and rather subjective manual process. For instance, it was observed that humans are sensitive to edges in images. Many well-known computer vision algorithms follow this pattern and use edge or gradient based features, e.g. the scale invariant feature transform (SIFT). SIFT is a widely adopted approach for object detection and image comparison that efficiently detects and describes characteristic and scale invariant keypoints within images that provided a huge improvement over earlier approaches (Lowe, 2004). In the following section, we refer back to feature extraction and discuss how it has evolved in the age of deep learning.

2.2 | Classification

The output of feature extraction is typically a vector (cp. x in Figure 1), which is than mapped to a score of confidence using a classifier. Depending on the application, the score is either compared to a threshold solely deciding whether an object is present or not (e.g. presence of a plant or animal in the image), or it is compared to other scores to distinguish object classes (e.g. species name). Prominent classification methods are machine learning algorithms such as support vector machines, Random Forest and artificial neuronal network (ANN).

3 | DEEP LEARNING NEURAL NETWORKS

The features extracted from images refer to what the model "sees about an image" and their choice is highly problem- and object-specific. In the past, manually deriving characteristic features was essential for classification performance, but also was a labour-intensive and subjective expert task. Furthermore a lot of feature cannot be extracted manually correctly so far. Therefore, a procedure allowing

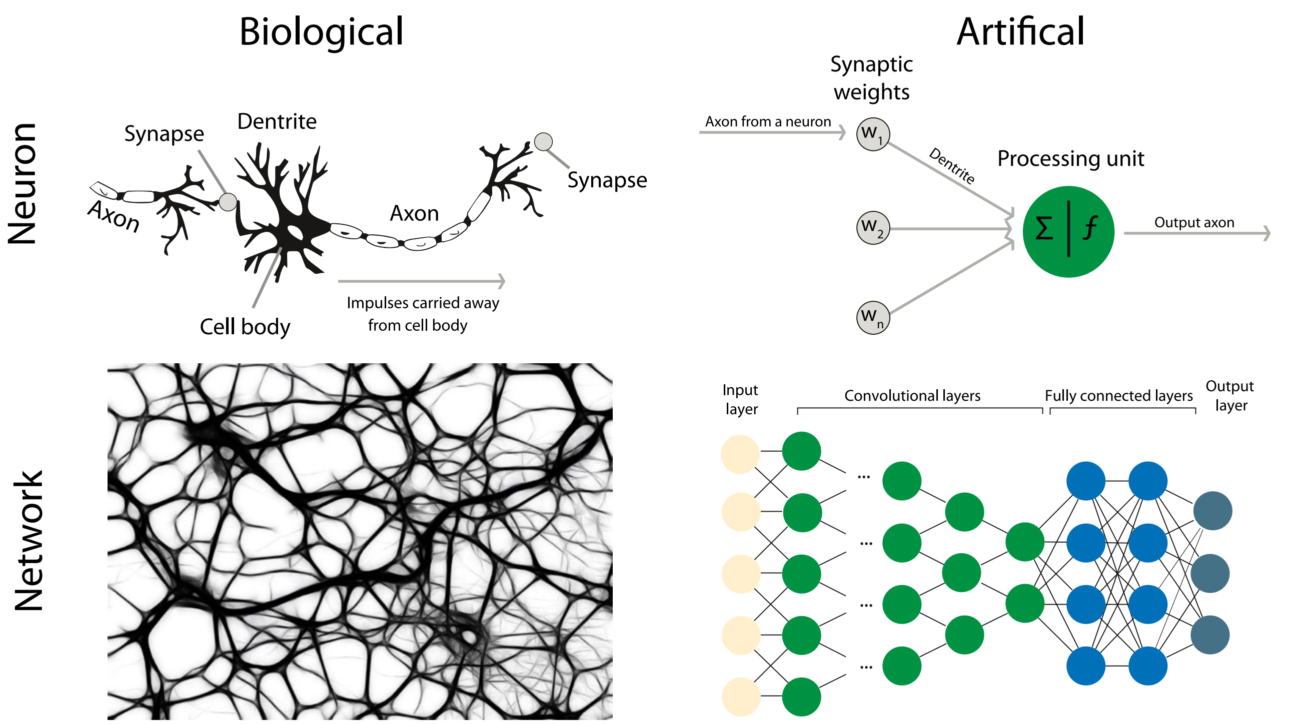

FIGURE 2 Comparison between biological and artificial neuron and networks

to automatically determining suitable features for a problem without a given logic was a sought-after for a long time. Artificial neural networks automate the feature extraction step by learning a suitable representation of the data from a collection of examples and developing a robust model themselves. This automated feature extraction proofs to be highly accurate for computer vision tasks and state-of-the-art models in image classification, object detection and image retrieval rely on it (Bengio, Courville, & Vincent, 2013).

Deep learning builds upon widely utilised ANN, which are mathematical models using learning algorithms inspired by biological neural networks, the central nervous systems of animals and in particular their brain. The human brain contains on average 86 billion neurons (Azevedo et al., 2009). Each biological neuron consists of a cell body, a collection of dendrites that bring electrochemical information into the cell and an axon that transmits electrochemical information out of the cell. A neuron produces an output along its axon, it fires when the collective effect of its inputs reaches a certain threshold. The axon from one neuron can influence the dendrites of another neuron across junctions called synapses. Some synapses will generate a positive effect in the dendrite encouraging its neuron to fire and others will produce a negative effect discouraging the neuron from firing (see Figure 2).

Artificial neural networks emulate this processing functionality of the brain. Unlike a biological brain, where any neuron can connect to any other neuron within a certain physical distance, artificial neural networks consist of a finite and predefined number of layers and connections (cp. Figure 2). Therefore, layers are made up of a number of interconnected nodes that contain an activation function. The leftmost layer of the depicted network is called the *input layer* and the rightmost layer is called the *output layer*. The layers in between are called *hidden layers*. While shallow learning neuronal networks consist of a single or at maximum two hidden layers, deep learning neuronal networks consist of multiple hidden layers, which together form the majority of the artificial brain. The leftmost layer in the stack is responsible for the collection of raw data. Each neuron of the leftmost layer stores information and passes it to the next layer of neurons and so on. When trained, the network forms a hierarchy of image features with increasing complexity starting with low-level image concepts close to the input layer to high-level image concepts close to the output layer. As data moves from the lowest layer to the highest layer more abstracted information is collected at increasing scales from small edges, to object parts and eventually to entire objects (LeCun, Bengio, & Hinton, 2015).

The network class, that is applicable to deep learning of images, is the convolutional neural network (CNN). CNNs are comprised of one or more convolutional layers followed by one or more fully connected layers as in a traditional multilayer neural network (see Figure 3). The architecture of a CNN is designed to take advantage of the 2D structure of an input image. Local connections and tied weights followed by some form of pooling result in translation invariant features.

Work on CNNs has been conducted since the early eighties (Fukushima & Miyake, 1982). CNNs saw their first successful real-world application in the LeNet (LeCun, Bottou, Bengio, & Haffner, 1998) for hand-written digit recognition. Despite these initial successes, the use of CNNs did not gather momentum until substantial improvements in parallel computing systems came together with various new techniques for their efficient training. The watershed was the contribution of Krizhevsky, Sutskever, and Hinton (2012) to the prestigious ImageNet Challenge (ILSVRC) in 2012. The proposed CNN, called AlexNet, won the competition by reducing the classification error from 26% to 15%. In the following years, architectures continuously evolved with the most prominent being VGGNet (Simonyan & Zisserman, 2014), GoogLeNet (Szegedy et al., 2015), and ResNet (He, Zhang, Ren, & Sun, 2016). Figure 4 shows how the classification error continuously decreased with ResNet being the first architecture to beat human accuracy in the given classification task (Russakovsky et al., 2015).

4 | GET STARTED WITH DEEP LEARNING

The difficulty in applying the latest machine learning algorithms has initially slowed down their application in ecology and taxonomy research. However, in the meantime, a large number of deep learning frameworks is publicly available allowing everybody with a basic

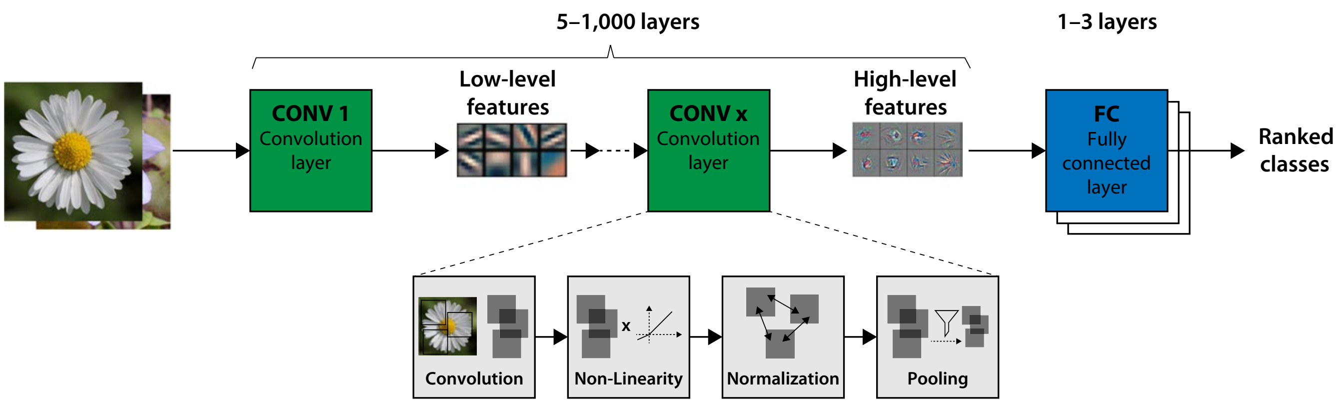

FIGURE 3 Basic architecture of convolutional neural network (CNN). CNNs are comprised of one or more convolutional layers followed by one or more fully connected layers

understanding of the machine learning concepts and some experience in script-based programming the training of new models. Due to a large variety of competing frameworks, a careful decision is suggested not only to select a package suitable to the problem at hand, but also to the background and skill set of the user. Fundamental conditions concern the supported operating system and the programming language a user prefers for setting up the training and analyzing the results. Since frameworks become more and more universal in these regards, other relevant points are the ease of prototyping and model tuning, mechanisms for the deployment of trained models, the size of the community supporting a framework and its scalability across multiple machines in order to reduce training times. All modern frameworks support CNN network architectures, provide pre-trained model zoos ready to be applied to individual problems, and support graphic processing units (GPU) in order to speed up training.

Table 1 lists the five currently most popular frameworks in terms of GitHub stars, a common method for measuring the relevance of

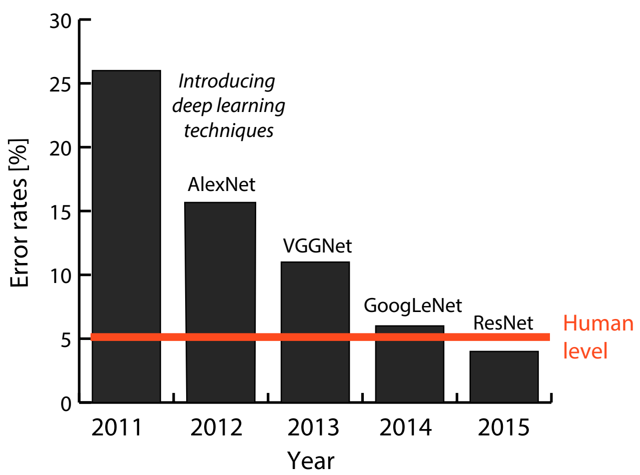

FIGURE 4 Top-5 Classification error rates of ImageNet Visual Recognition Challenge. In 2012, for the first time a deep neural network architecture (AlexNet) won the challenge

open-source software. While all five and many other frameworks are suitable to train a new model, we want to emphasise two of them specifically. If you are just starting out with deep learning, a solid selection is Keras on top of TensorFlow. Keras offers an additional graphical interface to TensorFlow and simplifies many steps. TensorFlow itself is being instructed in Python, is backed by Google, has a very good documentation, and there is a lot of educational material, such as tutorials and videos available on the internet guiding you in first steps. The second recommendation is MXNet as an alternative that supports the largest number of languages amongst the compared frameworks, e.g. R, PYTHON, C++ and MATLAB. If you are familiar with any of these languages this might be very helpful and simplify the training of a deep learning model. MXNet is backed by Amazon and also popular because it scales very well, i.e. to train a model with multiple GPUs and multiple computers, which makes it suitable for large-scale problems.

Another interesting development is cloud platforms that offer anything needed to get started with deep learning. For example, the Google Cloud Machine Learning platform released in 2016 gives users access to a web service for the training of models using TensorFlow. The service offers pretrained models of various architectures. Similar services are offered by Amazon Web Services (AWS), Wolfram Mathematica, Mathworks MATLAB and others.

Although CNNs are now clearly the top performers in most image based species identification tasks, the exact architecture is not the most important determinant in getting a good solution. When comparing studies that apply the same model architecture to a similar problem, they may report significantly differing results. A key aspect that is often overlooked is that expert knowledge about the task to be solved can provide advantages that go beyond “adding more layers to a CNN”. Researchers that obtain good performance when applying deep learning algorithms often differentiate themselves in aspects beyond the network architecture, like specific preprocessing and augmentation of training data that requires an in-depth understanding of the studied phenomena and data (Litjens et al., 2017).

TABLE 1 Overview of the most common software packages for the training of convolutional neural networks (CNNs).

Package	Interface Languages	Platform	GitHub Stars[a]
TENSORFLOW	PYTHON (Keras), C/C++, JAVA, GO, R	Linux, macOS, Windows	91,107
CAFFE2	PYTHON, MATLAB	Linux, macOS, Windows	30,471
KERAS	PYTHON, R	on top of MXNet, TensorFlow, CNTK	26,213
MICROSOFT COGNITIVE TOOLKIT	PYTHON (Keras), C++, COMMAND LINE, BRAINSCRIPT	Linux, Windows	13,951
APACHE MXNET	C++, PYTHON, JULIA, MATLAB, JAVASCRIPT, GO, R, SCALA, PERL	Linux, macOS, Windows, AWS, Android, iOS, JS	13,234

[a]Retrieved March 3, 2018. The Caffe2 stars include those of its predecessor.

TABLE 2 Synthesis of PlantCLEF identification challenge over the last 7 years.

Year	Life form	#Species	Image content	#Training images	#Testing images	mAP
2011	trees	71	leaves	4,004	1,432	0.47
2012		126		8,422	3,150	0.45
2013	trees, herbs	250	leaves, flowers, fruits, bark, branches	20,985	5,092	0.61
2014		500		47,815	8,163	0.45
2015*	trees, herbs, ferns	1,000		91,759	21,446	0.65
2016		1,000		113,205	4,633	0.82
2017		10,000		1,600,000	25,170	0.92

*In 2015, for the first time a deep neural network architecture won the challenge (mAP=mean average precision).

Augmentation and preprocessing are, of course, not the only contributors to accurate and robust models. For example, designing architectures incorporating unique task-specific properties can obtain better results than straightforward CNNs. Two examples are multi-view and multi-scale networks. Cropping different parts of an image will provide different contributing information. Patches with small scale areas can provide details of the organism (e.g. teeth or vein structure of plant leaves), while patches depicting large-scale areas can provide information surrounding the organism. Other, often underestimated, parts of network design are the network input size and receptive field (i.e. the area in input space that contributes to a single output unit). Input sizes should be selected considering for example the required resolution and context to solve a problem (Litjens et al., 2017).

5 | RECENT RESEARCH STUDIES USING DEEP LEARNING FOR SPECIES IDENTIFICATION

Recent works on automated species identification can be divided into two categories: lab-based investigations and field-based investigations (Martineau et al., 2017). In a lab-based condition there is a fixed protocol for image acquisition. This protocol governs the sampling, its placement and the material used for the acquisition. Lab-based setting is often used by biologist that brings the specimen (e.g. insects or plants) to the lab for inspecting them, to identify them and mostly to archive them. In this setting, the image acquisition can be controlled and standardised. In contrast to field-based investigations, where images of the specimen are taken in-situ without a controllable capturing procedure and system. For field-based investigations, typically a mobile device or camera is used for image acquisition and the specimen is alive when taking the picture (Martineau et al., 2017). In the field of insect identification, the lab-based setting is still the most widely used setup and the positioning of the insects is made mostly manual with a constrained pose. Also automated plankton identification is mostly done with microscopic images in the lab. In contrast, automated mammal and fish identification are done by taking images under field site condition (Weinstein, 2018). Automated plant identification is required under field as well as under lab conditions (Wäldchen, Rzanny, Seeland, & Mäder, 2018).

How dramatically deep learning has improved classification accuracy is impressively demonstrated in the results of the PlantCLEF challenges, a plant identification competition hosted since 2011 as an international evaluation forum (http://www.imageclef.org/). At each year, a larger and more complex image dataset stimulates the evaluation of competing methods uncovering strengths and weaknesses (Joly et al., 2016). Table 2 synthesises the results of the seven preceding challenges. Identification performance improved year after year despite the task becoming more complex by increasing the number of species, adding images and introducing image perspectives (from single leaves to flowers, fruits, barks, branches). Additionally, a tremendous gain in classification accuracy is visible in 2015, which is attributed to the adoption of deep learning CNN's.

Deep learning neural networks have also greatly improved model performance across a wide variety of animal taxa, from underwater marine specimen classification, such as fishes, zooplankton and corals over insects, such as moths and ants to large mammals. Table 3 aggregates recent studies using deep learning techniques for image based animal species identification. Though ecologists are collecting vast amounts of high-quality data, image datasets of different animal groups are still rarely available making the lack of labelled data the major obstacle in applying the latest machine learning techniques.

Deep learning could also improve the digitalization workflow of historical collections and herbaria. For example, herbaria all over the world have invested large amounts of money and time in collecting plant samples. Today, over 3,000 herbaria across 165 countries possess over 350 million specimens, collected in all parts of the world and throughout multiple centuries (Willis et al., 2017). Currently, many herbaria are undertaking large-scale digitisation projects to simplify access and to preserve delicate specimens. For example in the USA, more than 1.8 million imaged and georeferenced vascular plant specimens are digitally archived in the iDigBio portal, a nationally funded primary aggregator of museum specimen data (Willis et al., 2017). This activity is likely going to be expanded over the coming decade. We can look forward to a time when there will be huge repositories of taxonomic information, represented by specimen images, accessible publicly through the internet. Coupling these data with the classification capabilities of CNNs will unlock more of

TABLE 3 Recent studies using deep learning for animal identification. Accuracy is reported for the best performing model per paper without manually preprocessing

Taxa	#Taxa	#Images	Architecture	Accuracy	Study
Mammals	26	32,240	ResNet-101	69.0%	Gomez Villa, Salazar, and Vargas (2017)
Mammals	48	3,200,000	ResNet-101	93.8%	Norouzzadeh et al. (2018)
Ant genra	57	150,088	AlexNet	83.0%	Marques et al. (2018)
Aquatic macro-invertebrate	29	11,832	AlexNet	85.6%	Raitoharju et al. (2016)
Fishes	23	27,370	–individual–	98.6%	Qin, Li, Liang, Peng, and Zhang (2016)
Fishes	15	29,000	–individual–	94.0%	Salman et al. (2016)
Insects	10	550	ResNet-101	98.7%	Cheng, Zhang, Chen, Wu, and Yue (2017)

the rich potential of natural history collections (Schuettpelz et al., 2017).

As a first publication in this direction, Carranza-Rojas, Goeau, Bonnet, Mata-Montero, and Joly (2017) apply deep learning methods to large herbarium image datasets. In total, more than 260,000 scans of herbarium sheets representing more than 1,204 species were analysed. The approach reaches a top-1 species identification accuracy of 80% and a top-5 accuracy of 90%. Given the accurate results, the classifier could be used to create a semi- or fully automated system helping taxonomists in annotation, classification and revision work at herbaria. However, the researchers also found that it is currently impossible to directly transfer trained knowledge from a herbarium to identification in the field due to the greatly varying visual appearance (e.g. strong colour variation and the transformation of 3D objects after pressing like fruits and flowers).

6 | NEXT GENERATION FIELD GUIDES USING DEEP LEARNING TECHNIQUES

Despite intensive and elaborate research on automated species identification, only very few approaches resulted in usable tools. Available applications are typically the product of close interactions between computer scientists and end-users, such as ecologists, educators at schools, land managers and the general public (Table 4).

One example is the Pl@ntNet application, developed in a collaboration of the four French research organizations Cirad, INRA, Inria and IRD. Pl@ntNet started in 2010 as a joint project with the Tela Botanica social network in order to collect plant images and to train a visual identification tool. The application allows to submit one or several images of an unknown plant and retrieves a list of the most similar species. The application was initially restricted to a fraction of the European flora (in 2013) and has then incrementally been extended to the Indian Ocean and South America in 2015, and to North Africa in 2016. Since June 2015, Pl@ntNet utilizes deep learning techniques, i.e. a CNN pretrained on the ImageNet dataset and periodically fine-tuned on the growing own dataset. Before that date, the service used a traditional workflow based on a combination of hand-crafted visual features (Affouard, Goeau, Bonnet, Lombardo, & Joly, 2017). In 2014, Joly et al. evaluated the traditional approach on about half of the French plant species (2,200 species) showing mediocre top-5 identification rates of up to 69% for single image observations. An evaluation after introducing deep learning techniques is not yet available for comparison. However, Affouard et al., 2017 reported that Pl@ntNet's user ratings on the Google Play Store, which distributes the Android app, increased from 3.2 to 4.3 after the deep learning based approach was integrated. Pl@ntNet describes itself as being "an image sharing and retrieval application for the identification of plants". One of the main features of Pl@ntNet is that the image training set is collaboratively enriched and revised. This means user can share their observation with the community on a web platform called IdentiPlante. Each time a user identifies a plant using Pl@ntNet and shares this observation, the database of plant images grows thereby enriching the training set for the classifier. The result is a large-scale image collection publicly available under a Creative Common Attribution-ShareAlike 2.0 license.

Another application demonstrating the potential of machine learning techniques for species identification is Merlin Bird ID app with the Photo ID function. Merlin Photo ID is a joint project of Visipedia, the Cornell Laboratory of Ornithology, Cornell Tech and Caltech aiming to identify 650 of North America's most common bird species based on images. The imaged-based identification algorithm uses Google's TensorFlow deep learning platform, as well as citizen science data from the eBird platform to generate a potential species lists. The user inputs date, location and an image of the unknown bird and a suggestion of the most likely candidate appears. Similar to Pl@ntNet the image can be resubmitted and is then used to improve the algorithm (http://merlin.allaboutbirds.org/photo-id/).

Another popular app for the automated identification of animals and plants at species level was launched by iNaturalist.org in summer 2017. Initially, iNaturalist solely offered crowd-sourced species identification. Users post an image of a plant or animal and a community of scientists and naturalists identifies it. A taxon is elevated to "research grade" once more than 2/3 of the involved identifiers agree in their identification of an observation (https://www.inaturalist.org/pages/help#quality). iNaturalist recently passed five million observations, with 2.5 million of them having reached research grade. In the meantime, the app also offers automated identification trained on the database of "research grade" observations. The more images are uploaded by users and identified by experts the better. In favour of accuracy, the approach

TABLE 4 Prominent examples of automated species identification applications using mobile devices and deep learning techniques

App	Plattform	Organism	Cite
iNaturalist	Android/iOS	All kind of species from the iNaturalist database	https://www.inaturalist.org/
FloraIncognita	Android/iOS	Plants	https://floraincognita.com
Pl@nNet	Android/iOS	Plants	https://identify.plantnet-project.org/
Merlin Photo ID	Android/iOS	Birds	http://merlin.allaboutbirds.org/

aims to give a confident response about a species' genus and a more cautious response about the species by listing the top ten possibilities. An evaluation shows that genus classification reaches 86% accuracy, while top-10 species accuracy is at 77% (www.inaturalist.org).

FloraIncognita is an app for automated plant species identification of Germany's 2,770 wild flowering taxa launched in spring 2018. The app originates from a joint project between the Technische Universität Ilmenau and the Max Planck Institute for Biogeochemistry in Jena, Germany. In 2017, the same researchers released the FloraCapture app supporting interested hobbyist and experts in capturing high-quality training observations for FloraIncognita. FloraCapture requests contributors to photograph plants from at least five precisely defined perspectives. Expert botanists then identify depicted species and share their classification with the contributors. Since its launch in 2017, more than 15,000 observations containing more than 80,000 images of about 1,000 wild growing plant species of Germany were collected and are used to train the machine learning classifier of the FloraIncognita App. This classifier uses currently a cascade of CNNs to automatically identify unknown species based on two individual images of an unknown plant and its geolocation (www.floraincognita.com).

So far, there are no "in-field applications" that carry out the identification process semi-automatically. To ensure and concretize the results it would be useful not only relying on the purely automatic identifications. Combining inter-active identification keys with computer vision should be a future alternative to the above mentioned identifications apps. This requires very high implementation effort and expert knowledge, but achieves more accurate results in the end, especially for species that are easy to confuse.

7 | OUTLOOK

Humans are able to widely abstract relationships between concepts and make accurate decisions based on very little information. In contrast, deep learning algorithms are still narrow in their abstraction and reasoning capabilities, i.e. they need large quantities of precisely labelled information in order to deliver accurate results. The hurdles towards fully automated species identification remain high due to the immense labour required for accumulating and labelling the necessary datasets. New data collection opportunities through data mining and citizen scientists will broaden the potential sources of labelled ecological data (Weinstein, 2018). For example, projects like Zooniverse, iNaturalist, Pl@ntNet and Flora Incognita demonstrate how to engage user communities. In exchange, automatic species identification could also support existing citizen science projects. Offering nature enthusiasts the possibility of automatically identifying species based on photos they have taken as part of an observation and afterwards reviewing these observation by experts could enhance the completeness and quality of the observation database.

We argue that the quality of an automated identification system crucially depends not only on the amount, but also on the quality of the available training data. There are no investigations on the amount and characteristics of training data required for an equivalent classifier so far. Although existing CNN methods are able to model a suitable feature representation from images of an organism, they still lack the capability to model global relationships between different views, e.g. organs, depicting the same individual. Existing CNN based approaches were designed to reason based on single images, focusing on capturing the similar region-wise patterns within an image but not the structural patterns of an organism seen from multiple views with one or more of its organs (Lee, Chan, & Remagnino, 2018). Only few studies investigated the potential gain in accuracy by combining different organs (e.g. leaves, fruits, flower for plants) or perspectives (e.g. head, dorsum and profile for insects) (Lee et al., 2018; Marques et al., 2018). Analogous to a biologist that generally tries identifying an organism by observing several organs or a similar organ from different viewpoints, an important research direction is analyzing how to increase accuracy by combining different perspectives in an automated identification. Although cameras allow us to create persistent copies of real-world objects the projection to a planar image sensor inherits a loss of major information about the 3D structure of the scene and especially its absolute dimensions. Knowing an object's absolute dimensions can be beneficial for manifold image based species identification tasks. Size information can improve accuracy significantly. Analyzing contextual information such as species size in images should find more attention in the future (Hofmann, Seeland, & Mäder, 2018). While funding organizations are willing to support research into the direction of automated species identification and nature enthusiasts are helpful by contributing images, these resources are limited and should be efficiently utilised. Researchers as well as enthusiasts miss guidance on how to acquire suitable training images in an efficient way (Rzanny, Seeland, Wäldchen, & Mäder, 2017).

The principal challenge in automated species identification arises from the vast number of potential species. However, most species are not evenly distributed throughout a larger region as they require more or less specific combinations of biotic and abiotic factors and resources to be present for their development. Therefore, species can be encountered within their specific ranges. Using range maps as they appear in-field guides to support manual species identification has been state-of-the-art for quite some time (Wittich, Seeland, Wäldchen, Rzanny, & Mäder, 2018). In contrast analyzing metadata like the location or the time of the image and combining these information with the image recognition results has been neglected so far. An important research direction would be to tackle the problem of image classification with location and time context.

Beyond taxa identification, machine learning could also automate trait recognition, such as leaf position, leaf shape, vein structure and flower colour from herbaria and natural images. Trait data could be gained on a large scale from digital images for taxa which are already known but for which no trait data are available so far. Linking traits inferred by a deep learning algorithm to databases such as the TRY Plant Trait Database can yield powerful new datasets for exploring a range of questions in studies of plant diversity (Soltis, 2017). Automated trait recognition and extraction using machine learning techniques is an open and unexplored research direction.

8 | CONCLUSION

Building accurate knowledge of the identity, the geographic distribution and the evolution of living species is essential for a sustainable development of humanity as well as for biodiversity conservation. Traditionally, identification has been based on morphological diagnoses provided by taxonomic studies. Today, mostly experts such as taxonomists and trained technicians can identify taxa accurately, because it requires special skills acquired through extensive experience. However, the number of taxonomists and identification experts is drastically decreasing. Consequently, the need for alternative and accurate identification methods applicable by non-experts is constantly increasing. Finding automatic methods for such identification is an important topic with high expectations.

While modern machine learning approaches only slowly pave their way into the field of species identification, we argue that in the near future we are going to see a proliferation of these techniques being applied to the problem. Although taxa identification by experts would be the preferred way, artificial intelligence systems will provide alternative tools for identification task. In a few years, we will rely on machine algorithms routinely to advance our core science while reducing the number of routine identifications performed by taxonomist allowing them to focus on experimental setup and data analytic. Furthermore, these technologies allow for a larger community observing our nature and contributing to develop an increasing interest of the society in their environment.

ACKNOWLEGEMENTS

We are funded by the German Ministry of Education and Research (BMBF) Grants: 01LC1319A and 01LC1319B; the German Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety (BMUB) Grant: 3514 685C19; and the Stiftung Naturschutz Thüringen (SNT) Grant: SNT-082-248-03/2014.

AUTHORS' CONTRIBUTIONS

J.W. and P.M. were writing the review.

DATA ACCESSIBILITY

We did not analyse any data.

ORCID

Jana Wäldchen http://orcid.org/0000-0002-2631-1531

REFERENCES

Affouard, A., Goeau, H., Bonnet, P., Lombardo, J. C., & Joly, A. (2017). Pl@ntNet app in the era of deep learning. In ICLR 2017 Workshop Track-5th International Conference on Learning Representations.

Austen, G. E., Bindemann, M., Griffiths, R. A., & Roberts, D. L. (2016). Species identification by experts and non-experts: Comparing images from field guides. *Scientific Reports*, *6*, 33634. https://doi.org/10.1038/srep33634

Azevedo, F. A., Carvalho, L. R., Grinberg, L. T., Farfel, J. M., Ferretti, R. E., Leite, R. E., … Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, *513*(5), 532–541. https://doi.org/10.1002/cne.21974

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, *35*(8), 1798–1828. https://doi.org/10.1109/TPAMI.2013.50

Carranza-Rojas, J., Goeau, H., Bonnet, P., Mata-Montero, E., & Joly, A. (2017). Going deeper in the automated identification of Herbarium specimens. *BMC Evolutionary Biology*, *17*(1), 181. https://doi.org/10.1186/s12862-017-1014-z

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. *Science Advances*, *1*(5), e1400253. https://doi.org/10.1126/sciadv.1400253

Cheng, X., Zhang, Y., Chen, Y., Wu, Y., & Yue, Y. (2017). Pest identification via deep residual learning in complex background. *Computers and Electronics in Agriculture*, *141*, 351–356. https://doi.org/10.1016/j.compag.2017.08.005

Cope, J. S., Corney, D., Clark, J. Y., Remagnino, P., & Wilkin, P. (2012). Plant species identification using digital morphometrics: A review. *Expert Systems With Applications*, *39*(8), 7562–7573. https://doi.org/10.1016/j.eswa.2012.01.073

Farnsworth, E. J., Chu, M., Kress, W. J., Neill, A. K., Best, J. H., Pickering, J., … Ellison, A. M. (2013). Next-generation field guides. *BioScience*, *63*(11), 891–899. https://doi.org/10.1525/bio.2013.63.11.8

Fukushima, K., & Miyake, S. (1982). Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In S. Amari & M. A. Arbib (Eds.), *Competition and cooperation in neural nets* (pp. 267–285). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-46466-9

Gomez Villa, A., Salazar, A., & Vargas, F. (2017). Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. *Ecological Informatics*, *41*, 24–32. https://doi.org/10.1016/j.ecoinf.2017.07.004

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770-778.

Hofmann, M., Seeland, M., & Mäder, P. (2018). Efficiently annotating object images with absolute size information using mobile devices. *International Journal of Computer Vision*. https://doi.org/10.1007/s11263-018-1093-3

Hopkins, G., & Freckleton, R. (2002). Declines in the numbers of amateur and professional taxonomists: Implications for conservation. *Animal Conservation*, *5*(3), 245–249. https://doi.org/10.1017/S1367943002002299

Joly, A., Bonnet, P., Goëau, H., Barbe, J., Selmi, S., Champ, J., … Barthélémy, D. (2016). A look inside the pl@ntnet experience. *Multimedia Systems*, *22*(6), 751–766. https://doi.org/10.1007/s00530-015-0462-9

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *In Advances in Neural Information Processing Systems.*, Retrieved fromhttps://papers.nips.cc/book/advances-in-neural-information-processing-systems-25-2012

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

Lee, S. H., Chan, C. S., & Remagnino, P. (2018). Multi-organ plant classification based on convolutional and recurrent neural networks. *IEEE Transactions on Image Processing*, *27*(9), 4287. https://doi.org/10.1109/TIP.2018.2836321

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., … Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, *42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

Marques, A. C. R., M Raimundo, M., B Cavalheiro, E. M., FP Salles, L., Lyra, C., & J Von Zuben, F. (2018). Ant genera identification using an ensemble of convolutional neural networks. *PLoS ONE 13*(1), e0192011. https://doi.org/10.1371/journal.pone.0192011

Martineau, M., Conte, D., Raveaux, R., Arnault, I., Munier, D., & Venturini, G. (2017). A survey on image-based insect classification. *Pattern Recognition*, *65*, 273–284. https://doi.org/10.1016/j.patcog.2016.12.020

Mjolsness, E., & DeCoste, D. (2001). Machine learning for science: State of the art and future prospects. *Science*, *293*(5537), 2051–2055. https://doi.org/10.1126/science.293.5537.2051

Moniruzzaman, M., Islam, S. M. S., Bennamoun, M., & Lavery, P. (2017) Deep learning on underwater marine object detection: A survey. In: J. Blanc-Talon, R. Penne, W. Philips, D. Popescu & P. Scheunders (Eds.) *Advanced concepts for intelligent vision systems*. ACIVS 2017. Lecture Notes in Computer Science(Vol. *10617*, pp. 150–160). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-70353-4_13

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Packer, C., & Clune, J. (2018). Automatically identifying wild animals in camera trap images with deep learning. *Proceedings of the National Academy of Sciences 115*:E5716–E5725. *arXiv preprint arXiv:1703.05830*. https://doi.org/10.1073/pnas.1719367115

Qin, H., Li, X., Liang, J., Peng, Y., & Zhang, C. (2016). DeepFish: Accurate underwater live fish recognition with a deep architecture. *Neurocomputing*, *187*, 49–58. https://doi.org/10.1016/j.neucom.2015.10.122

Raitoharju, J., Riabchenko, E., Meissner, K., Ahmad, I., Iosifidis, A., Gabbouj, M., & Kiranyaz, S. (2016). Data enrichment in fine-grained classification of aquatic macroinvertebrates. In *2nd Workshop on Computer Vision for Analysis of Underwater Imagery (CVAUI), 2016* (pp. 43–48). Cancun: IEEE. https://doi.org/10.1109/cvaui.2016.020

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., … Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*(3), 211–252. https://doi.org/10.1007/s11263-015-0816-y

Rzanny, M., Seeland, M., Wäldchen, J., & Mäder, P. (2017). Acquiring and preprocessing leaf images for automated plant identification: Understanding the tradeoff between effort and information gain. *Plant methods*, *13*, 97.

Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J., & Harvey, E. (2016). Fish species classification in unconstrained underwater environments based on deep learning. *Limnology and Oceanography: Methods*, *14*(9), 570–585.

Schuettpelz, E., Frandsen, P. B., Dikow, R. B., Brown, A., Orli, S., Peters, M., … Dorr L. J. (2017). Applications of deep convolutional neural networks to digitized natural history collections. *Biodiversity Data Journal*, *5*, e21139. Advance online publication. https://doi.org/10.3897/bdj.5.e21139

Seeland, M., Rzanny, M., Alaqraa, N., Wäldchen, J., & Mäder, P. (2017). Plant species classification using flower images—A comparative study of local feature representations. *PLoS ONE*, *12*(2), e0170629. https://doi.org/10.1371/journal.pone.0170629

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Soltis, P. S. (2017). Digitization of herbaria enables novel research. *American Journal of Botany*, *104*(9), 1281–1284. https://doi.org/10.3732/ajb.1700281

Sonka, M., Hlavac, V., & Boyle, R. (2014). *Image processing, analysis, and machine vision* 4th edn. Stamford, CT: Cengage Learning.

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., & Packer, C. (2015). Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific Data*, *2*, 150026. https://doi.org/10.1038/sdata.2015.26

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … Rabinovich, A. (2015). Going deeper with convolutions. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Van Horn, G., Mac Aodha, O., Song, Y., Shepard, A., Adam, H., Perona, P., & Belongie, S. (2017). The iNaturalist Challenge 2017 Dataset. arXiv preprint arXiv:1707.06642.

Wäldchen, J., & Mäder, P. (2018). Plant species identification using computer vision techniques: A systematic literature review. *Archives of Computational Methods in Engineering*, *25*(2), 507–543. https://doi.org/10.1007/s11831-016-9206-z

Wäldchen, J., Rzanny, M., Seeland, M., & Mäder, P. (2018). Automated plant species identification—Trends and future directions. *PLoS Computational Biology*, *14*(4), e1005993. https://doi.org/10.1371/journal.pcbi.1005993

Weinstein, Ben G. (2018). A computer vision for animal ecology. *Journal of Animal Ecology*, *87*(3), 533–545. https://doi.org/10.1111/1365-2656.12780

Willis, C. G., Ellwood, E. R., Primack, R. B., Davis, C. C., Pearson, K. D., Gallinat, A. S., … Sparks, T. H. (2017). Old plants, new tricks: Phenological research using herbarium specimens. *Trends in Ecology & Evolution*, *32*(7), 531–546. https://doi.org/10.1016/j.tree.2017.03.015

Wittich, H. C., Seeland, M., Wäldchen, J., Rzanny, M., & Mäder, P. (2018). Recommending plant taxa for supporting on-site species identification. *BMC Bioinformatics*, *19*, 190.

How to cite this article: Wäldchen J, Mäder P. Machine learning for image based species identification. *Methods Ecol Evol*. 2018;9:2216–2225. https://doi.org/10.1111/2041-210X.13075