Trends in Ecology & Evolution

CellPress
REVIEWS

Review

Uncovering Ecological Patterns with Convolutional Neural Networks

Philip G. Brodrick [1,2,*,@] Andrew B. Davies,[1,2,3,@] and Gregory P. Asner[1,2,@]

Using remotely sensed imagery to identify biophysical components across landscapes is an important avenue of investigation for ecologists studying ecosystem dynamics. With high-resolution remotely sensed imagery, algorithmic utilization of image context is crucial for accurate identification of biophysical components at large scales. In recent years, convolutional neural networks (CNNs) have become ubiquitous in image processing, and are rapidly becoming more common in ecology. Because the quantity of high-resolution remotely sensed imagery continues to rise, CNNs are increasingly essential tools for large-scale ecosystem analysis. We discuss here the conceptual advantages of CNNs, demonstrate how they can be used by ecologists through distinct examples of their application, and provide a walkthrough of how to use them for ecological applications.

Highlights

CNNs enable ecologists to identify biophysical components in high-resolution remotely sensed imagery by leveraging spatial context, and are particularly effective when ecological components have distinct shapes.

CNNs can be used for both object detection, where key components are identified throughout an image, and semantic segmentation, where each pixel is classified individually.

CNN accuracy is similar to human-level classification accuracy, but is consistent and fast, enabling rapid application over very large areas and/or through time.

The Value of Spatial Context for Ecology

A key goal in ecology is to understand the spatial organization of organisms and ecosystems, the processes these patterns reflect, and their underlying biotic and abiotic controls. However, the ability to explore ecological pattern is underpinned by an accurate spatial representation of organisms and other ecosystem components. From such observations, patterns can be interpreted, advancing our understanding of system interactions [1]. Remotely sensed imagery has enabled the identification of ecosystem components over vast scales, facilitating new ecological discoveries in complex systems [2–5].

A principal challenge for ecologists seeking to use remotely sensed data for classification at the landscape level is the large quantity of data to sort through. This often involves the mundane and laborious task of identifying massive numbers of similar features across a landscape. Not only is this time-consuming, but human subjectivity can generate inconsistencies in patterns and/or their interpretation. Recently, various machine learning techniques have been used to expedite these tasks over large areas with promising results [6,7].

Although most computational machine learning methods used by ecologists to analyze remotely sensed data rely on pixel-level information, and in many cases high accuracy can be achieved with these techniques [8–10], high spatial resolution images have driven the need for more automated approaches to consider multiple pixels during decision making [11]. This is because of the redundancy induced in images when spatial resolution is finer than the objects of interest, with multiple pixels being measured from the same component [12]. For instance, sub-meter spectral or light detection and ranging (LiDAR) data from large tree canopies includes many nearly identical pixels from the same individual tree crown. This redundancy of information imposes a barrier because pixel-level information is insufficient to identify biophysical components for two reasons: (i) the same information could represent multiple

[1]Center for Global Discovery and Conservation Science, Arizona State University, Tempe, AZ 85281, USA
[2]Department of Global Ecology, Carnegie Institution for Science, 260 Panama Street, Stanford, CA 94305, USA
[3]Present address: Department of Organismic and Evolutionary Biology, Harvard University, 22 Divinity Ave, Cambridge, MA 02138, USA
[@]Twitters: @pgbrodrick, @andrewbdavies, @greg_asner

*Correspondence: brodrick@alumni.stanford.edu (P.G. Brodrick).

 https://doi.org/10.1016/j.tree.2019.03.006

different facets of the same component, and (ii) a single component could be composed of multiple different pixel values. Additional **bands** (see Glossary) of information (e.g., spectral, topographic, or even geologic data) can help to meet this challenge, but there are limits to the degrees of freedom of current observational systems, and the problem therefore persists. Spatial context breaks this barrier by supporting an algorithm to consider more information by utilizing not only one pixel, but also the textures and patterns of the surrounding area.

During visual observation, humans instinctively make use of spatial context to understand content, without which we would be unable to disaggregate components within an image. Consider the example in Figure I(i) in Box 1, where a termite mound is clearly visible in a snapshot of a digital elevation model (DEM). Individual pixels that comprise the termite mound do not provide enough information for classification, evidenced by the equivalent values of the DEM on the mound and on the hillslope in the upper-left region of the image. Nevertheless, when considering the image as a whole, the termite mound is obvious. Multiple context-based methods, often referred to as geospatial object-based image analysis, exist and have seen success [13,14], but we focus here on an emerging class of algorithms known as **convolutional neural networks** (CNNs) that have shown breakthrough performance and interpretability in recent years [15]. We review and demonstrate how CNNs can account for spatial context and thus facilitate the identification of ecological features and patterns, yielding new insights.

How CNNs Use Spatial Context

CNNs provide an algorithmic means to bring spatial context to bear for both rapid and accurate interpretation of image texture, with broad implications for feature and pattern identification in ecological studies. CNNs are a subset of a class of machine learning algorithms, known as deep learning models (or deep artificial neural networks), that are widely used throughout biology and ecology [16]. Deep learning models work by passing data through a multitude of **neurons**, which are organized into a series of **layers**. Each neuron uses a linear combination of data from neurons in the previous layer that is then transformed through a nonlinear **activation function** (also known as a squashing function). The architecture of the algorithm (the number of neurons per layer and the arrangement of layers) is human-determined, whereas the path of data transmission through the algorithm, controlled by weights at each neuron, is determined by the deep learning model. As the depth of the layer stack increases, many linear and nonlinear transformations occur, enabling the representation of a wide variety of complex systems [17]. Early layers in the model tend to contain low-level information, which is aggregated and transformed into successively higher-level information until the desired inference can be made. Deep learning models are not specific to images, and can be used to analyze many types of data. CNNs, however, are specifically designed for spatial data, which are analyzed largely by leveraging **convolution layers**. Convolution layers use convolution matrices (or kernels) determined by the algorithm to aggregate spatial information from multiple pixels. With a sufficient number of layers, CNNs can then learn to interpret different textures within an image. Early layers in CNNs tend to recognize simple features such as edges, which progress to image subcomponents and to eventually high-level abstractions such as human or animal faces [18], or, in an ecological example (as in Figure 1), early layers can recognize general terrain characteristics which are ultimately aggregated to identify termite mounds. Since their initial introduction, many individual **network architectures** have been proposed, and the effectiveness of the methods has dramatically improved [19].

Three principal categories of CNN applications exist today, and distinct model architectures are used for each application, including image classification [20–22], object detection [23–26], and semantic segmentation (also known as image segmentation) [27–29] (Box 1). There are several variants and hybrids (e.g., [30,31]) of these categories, and more are being developed, but

Glossary

Activation function: also known as a squashing function a nonlinear function applied to the output of a deep learning model layer.
Band: also known as a channel or feature, an input variable into a model. In data science these are referred to as features, in much of computer vision the term channel is used, whereas in remote sensing the term band is commonly favored.
Bounding box: used in object detection, a bounding box surrounds an object of interest (a response) within an image. Each bounding box is typically accompanied by a class label as well as the probability that an object is located within said box.
Channeling: a connection inside a CNN where encoding and decoding layers are concatenated after a decoding layer, allowing better-resolved semantic segmentation.
Convolution layer: the core concept in a CNN, a convolution layer is a step in the CNN that aggregates information via a kernel (or convolution) and passes that information to the next layer. The kernel has a user-specified size (typically ranging from 3 to 11 pixels), but the values of the kernel are determined during model training.
Convolutional neural network (CNN): a style of deep learning model that uses a series of convolution layers to incorporate image context to learn to identify images, find objects within images, and/or fully segment the contents of an image.
Decoding layer: a type of layer in a CNN where higher-order information is resolved toward the response of interest. Typically, these are later layers in a network, and resolution successively increases through encoding layers.
Encoding layer: a type of layer in a CNN where information is extracted at a higher level than the previous layer. Typically, these are earlier layers in a network, and resolution successively decreases through encoding layers.
Fully convolutional network (FCN): a specific type of CNN that does not include fully connected layers to preserve spatial information throughout the model; commonly used for semantic segmentation.

Box 1. Categories of CNN Applications

(i) Image Identification

In image classification, the original use of a CNN, the model output is designed to summarize the contents of an image through one or more labels. In this case, the entire image would simply be labeled as 'termite mound', as opposed to an image where no termite mound is present, which might be labeled 'other'. Image identification is not well suited for ecological applications in remote sensing because it requires arbitrary partitioning of the extent of interest.

(ii) Object Detection

Developed after image identification, object detection enables subcomponents within an image to be labeled. To date, this is typically achieved with **bounding boxes**, which locate and label the object of interest, and provide the likelihood that an object is within a generated box. Object detection is well suited for ecological applications in remote sensing because it does not depend on a specific extent. However, the bounding box can be limiting because it prevents consideration of shape.

(iii) Semantic Segmentation

Semantic segmentation is a more recent advance in CNN application whereby each pixel is labeled in a meaningful way. Probability maps can be generated for each class, or those maps can be condensed to a classified image (as shown here). Semantic segmentation is often the best choice for ecological applications in remote sensing because it allows complete partitioning of the extent of interest and enables arbitrarily shaped objects to be identified.

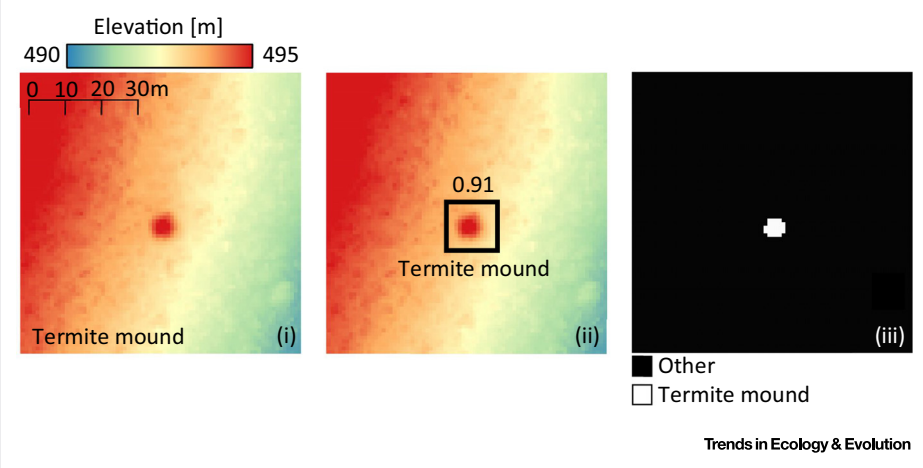


Figure I. Image Identification, Object Detection, and Semantic Segmentation.

Layer: a principal unit within a deep learning model. A convolution layer is an assembly of kernels, whereas a fully connected layer is an assembly of nodes. Layers are stacked in series (and sometimes in parallel) to successively transform information through the model.
Network architecture: also known as model architecture, the manner in which different components of a deep learning model are stacked together, including depth of each layer (either number of filters or number of neurons), the activation/squashing function used, the number of layers, and how each layer is connected to each other layer.
Neuron: a connection point within an artificial neural network. Neurons are connected to one another through linear weightings, and are grouped together with nonlinear functions in different ways depending on the network architecture.
Pooling: also known as pooling layers, the aggregation of pixels between convolution layers, inducing a reduction in the resolution of the next layer. This occurs by grouping neighboring pixels, commonly by taking the maximum value.
Response: the desired output from a model. In the case of image identification the response is an image label, for object detection the response is a series of bounding boxes, and for semantic segmentation the response is an image with each pixel labeled (or equivalent tensor structure).
Stride: the number of pixels a kernel moves when sliding through an image in a convolution layer. Higher strides step farther, and also reduce the resolution of the subsequent layer.

these three synthesize the range of CNNs used in ecology to date. Image classification, where whole images are assigned identifying labels, was the first application to become popular in 2012 [20] and is the most widely used in biology and ecology today, with applications including plant taxonomy [32–34] and animal identification in camera traps or aerial photographs [35–40]. Object detection began to see significant success in 2014 [23], and extends the idea of image classification by examining subcomponents within an image. For object detection, the CNN identifies regions (boxes) of interest within an image, each of which is labeled and given a probability of containing a component of interest. Object detection uses a similar model architecture to image classification, but includes a supplemental component at either the beginning [23–25] or the end [26,41–43] of the algorithm to facilitate object identification. In biology, object detection has been adopted to address problems such as cell counting [44–47]

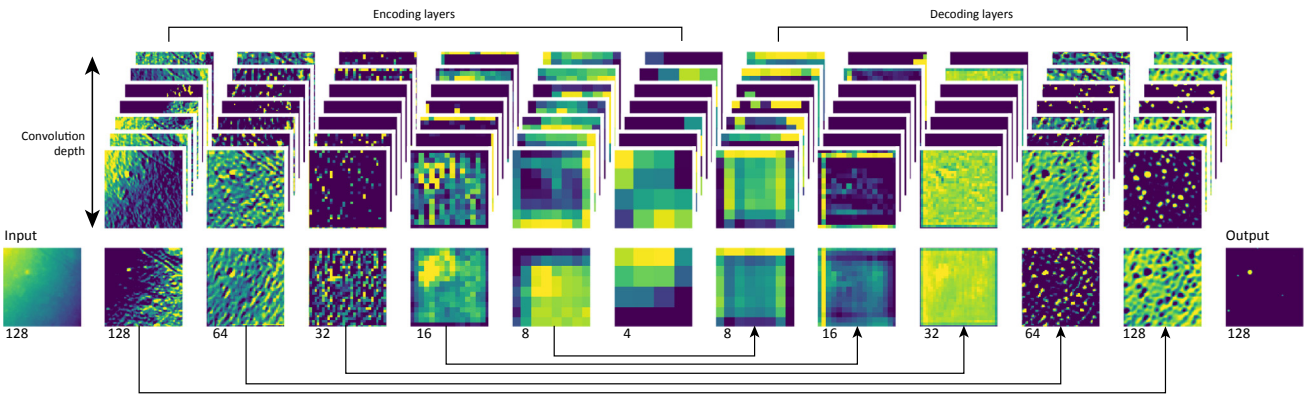


Figure 1. A Simple Example Convolutional Neural Network (CNN) Architecture for Semantic Segmentation. The convolution with the greatest variation is depicted in the foreground; the additional convolutions are shown in the background (eight are depicted in this example network). Each image shows the result of passing the input image from the far left through the network up to the given layer. Each layer uses a series of convolution matrices to connect adjacent layers. In encoding layers (left half of the network), resolution is decreased through pooling, and higher-level information is extracted. In decoding layers (right half of the network), this high-level information is then made spatially explicit into output classifications, aided through the process of channeling, depicted via arrows, where information is passed forward between equivalently sized encoder and decoder layers.

and animal identification [48]. In the final application, semantic segmentation, the concept of more locally refined classification is taken to its limit, and all pixels within an image are simultaneously classified. Semantic segmentation first became popular in 2015 [27], and has seen widespread adoption in biology, particularly in biomedical research [49–52], following some early advances [28]. Examples of each of these different types of applications on a sample termite mound image are shown in Box 1.

Extension to Remote Sensing

Most of the examples described above rely on stand-alone, spatially agnostic imagery taken by three-band red, green, and blue (RGB) cameras or laboratory equipment. Identifying ecological patterns across large scales, however, often requires the utilization of remotely sensed data, which differ from stand-alone images mostly due to their increased extent, but also because of the fixed-depth perspective (remotely sensed imagery is typically obtained at a relatively constant height above ground) and the addition of specific types of missing data from the images (e.g., clouds and nonuniform observational boundaries). Through either object detection or semantic segmentation, remotely sensed data can be treated as a continuous image, thereby alleviating concerns about boundary conditions from image subsets [53]. Several studies have demonstrated strong performance, with most emphasis to date being placed on land-cover classification [54–58]. Of these two application options, we emphasize that semantic segmentation is the better choice for ecological applications because it enables the delineation of nonrectangular objects in full detail, and object shape can have important consequences for ecological interpretation.

Combining CNNs and remotely sensed imagery for ecological applications remains an emerging trend, but several studies have successfully leveraged this combination. King *et al*. demonstrate the efficacy of several different CNNs for semantic segmentation to delineate multiple coral species, a sea fan, and several substrate types across a reefscape [59]. Rezaee *et al*. demonstrate the disaggregation of wetland categories (e.g., bog, fen, marsh) in a 700 km^2 study area in Canada [60]. Li *et al*. used CNNs to count individual oil palm trees from WorldView

satellite data [61], Csillik *et al*. segmented and counted citrus trees in multispectral drone imagery [62], and Wagner *et al*. mapped eucalyptus plantations, forest disturbance, and a specific tropical tree species [63]. Kellenberger *et al*. show how CNNs can be used with high-resolution drone imagery to census large wildlife in a 13 km^2 section of the Kuzikus Wildlife Reserve in Namibia [64], and Torney *et al*. used an object detection approach to count blue wildebeest in Serengeti National Park, Tanzania [65]. Ayrey *et al*. go so far as to demonstrate how CNNs can be used to interpret forest inventory attributes such as biomass estimates, tree counts, and needle-leaf tree fraction from LiDAR point clouds [66]. These studies are excellent early steps that demonstrate the potential of CNNs for ecological applications.

CNN Network Architecture Details: How They Work

The architecture of a CNN (the size, shape, and interconnection of different layers) ultimately determines the utility of the network, and we here provide an explanation of several crucial aspects of a CNN. We focus on semantic segmentation, that we suggest to be the most applicable for ecological studies, although object detection architecture is often also useful and is discussed in the supplemental information online. We provide an example of this architecture to highlight key components of the network, but emphasize that there are many adequate variations. Using a CNN with remotely sensed data requires the assembly and preprocessing of training/validation data, training the CNN, deploying the CNN to develop an output image, and assessing model accuracy. Each of these steps can be a potential barrier for new users, and we therefore provide both code and a walkthrough tutorial for managing these application steps in the supplemental information online.

Semantic segmentation with CNNs is performed using some variation of a **fully convolutional network** (FCN) [27], a subset of CNNs that has rapidly evolved since they were first introduced [28,29,67,68]. Many FCNs today follow an architectural style introduced as U-Net [28]. In this approach, a series of convolution layers successively coarsen the image spatial resolution while increasing in convolution depth, allowing data texture from one side of the image to influence the other side of the image. These layers are known as **encoding layers** because they aggregate information from throughout the spatial domain. The encoding layers are then followed by a series of **decoding layers** that successively increase in spatial resolution and decrease in convolution depth. This allows the model to make per-pixel predictions in the final layer at the same spatial resolution as the input image, while carrying forward knowledge gained from throughout the spatial domain by the encoding layers. Encoding layers reduce spatial resolution through a combination of **pooling**, which aggregates pixels within a window from the previous layer, and **stride** length, which determines how many pixels the kernel should slide during pooling. Decoding layers increase the spatial resolution through convolution layers that extrapolate features. This encoder/decoder style of model is enhanced by a concept known as **channeling**, whereby information from early encoding layers is carried forward to later points in the model, helping to produce an accurate and fine-scale segmentation of the final image. Many additional architecture components can play important roles in the network importance, which we do not discuss here for brevity [69–73].

A simple FCN model framework is illustrated in Figure 1. Encoding layers are added until the initial input 128 × 128 pixel image becomes less than an 8 × 8 pixel image, and then decoding layers are added until the spatial dimension matches the initial image resolution. Channeling connects each decoding layer with the last layer of matching spatial resolution (depicted by arrows in Figure 1). A useful property of channeling is that it decreases the risk of over-coarsening the image because particularly coarse layers can be bypassed if they provide little value. The network demonstrated in Figure 1 does not increase the depth of convolutions in the

center of the network, contrary to many common implementations. This reduced convolution depth sacrifices the possibility that large numbers of identifiable classes significantly reduce computation time, an effective approach for many ecological problems. Many CNNs that are used for semantic segmentation also use some form of post-processing to bring in additional information or further consider spatial context so as to refine the results. This post-processing often helps reduce isolated pixel misclassification. A common form of post-processing for semantic segmentation is the conditional random field (CRF) [74] that has been shown to be beneficial when used with CNNs [75–77].

Examples of Ecological Applications of CNNs

To demonstrate the potential of utilizing CNNs to identify ecological patterns, we constructed three brief examples, and used the same CNN architecture to identify and map important biophysical components in each. These examples include very distinct input data and types of **responses**, but are nevertheless all mapped well with CNNs. These examples stop short of investigating the processes that link system facets together, but demonstrate how, by facilitating large-scale mapping of ecosystem components, subject experts can use these data to gain insights into landscape-level patterns.

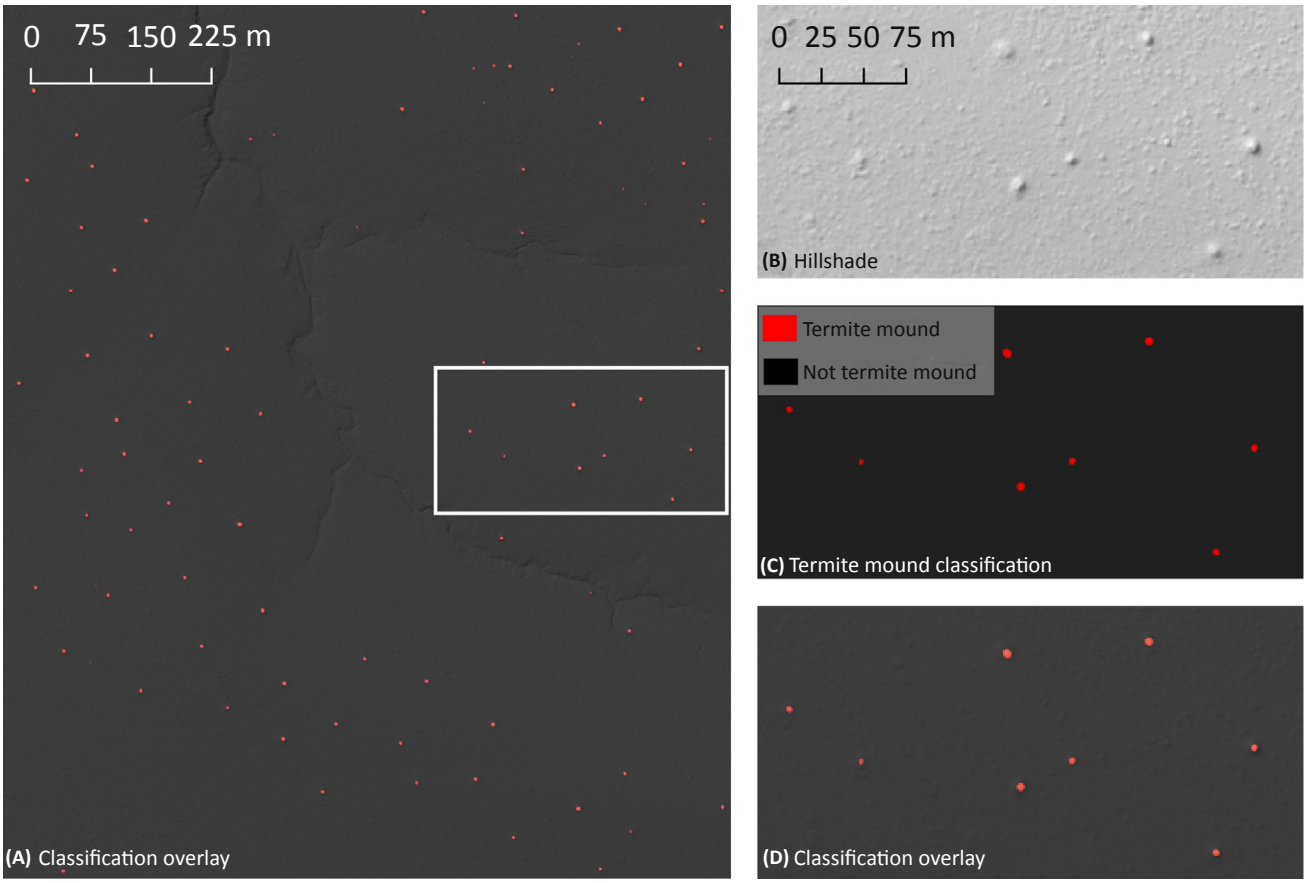
Termite Mound Identification

Termites have long been known to perform important functional roles in savanna ecosystems through their foraging and nesting behavior. By concentrating nutrients and moisture in central nesting locations, termites create spatial heterogeneity in the form of nutrient-rich mounds [78]. However, before the landscape patterning and distribution of termite mounds was understood through the use of remote sensing imagery, it was unclear how extensive the influence of termites was across large landscapes. Both high-resolution satellite and LiDAR data have been used to map termite mounds and reveal the importance of their spatial patterning across landscapes [79–82], but identifying individual mounds has remained a laborious and manual task, limiting the extent of possible analysis.

Using LiDAR data collected with the Carnegie Airborne Observatory (CAO) [83] in Kruger National Park, South Africa, we demonstrate that CNNs can facilitate the rapid and accurate identification of large numbers of termite mounds. LiDAR data were processed to a digital elevation model (DEM) at 1 m ground-level spatial resolution, as in Davies *et al*. [80], following which termite mounds were identified by hand using a hillshade model of the DEM, as in Figure 2A. Although the general classification of a particular mound by hand might be valid, the placement of mound centers can vary by several meters from the actual mound center owing to human error, making the accuracy of manual mound identification less than ideal. Because only individual points were identified during the preparation of the training dataset, we assumed that all termite mounds were square and 5×5 m (which is certainly not true). Nevertheless, the CNN produced a classification image at an accuracy rate equivalent to the error rates of our human-generated training data, where termite mounds reflected their true circular shape, and were aligned with the actual mound center (Figure 2B). Although a large dataset (>10 000 identified mounds) was available for training, in the supplemental information online we demonstrate how even a very small set of mounds can be used effectively.

Coral Reef Classification

Coral reefs are some of the most threatened ecosystems on the planet, facing deleterious impacts from increasing ocean temperatures, human use, and ocean acidification [84,85]. Furthermore, human activity on land influences coral distribution and health in the sea through

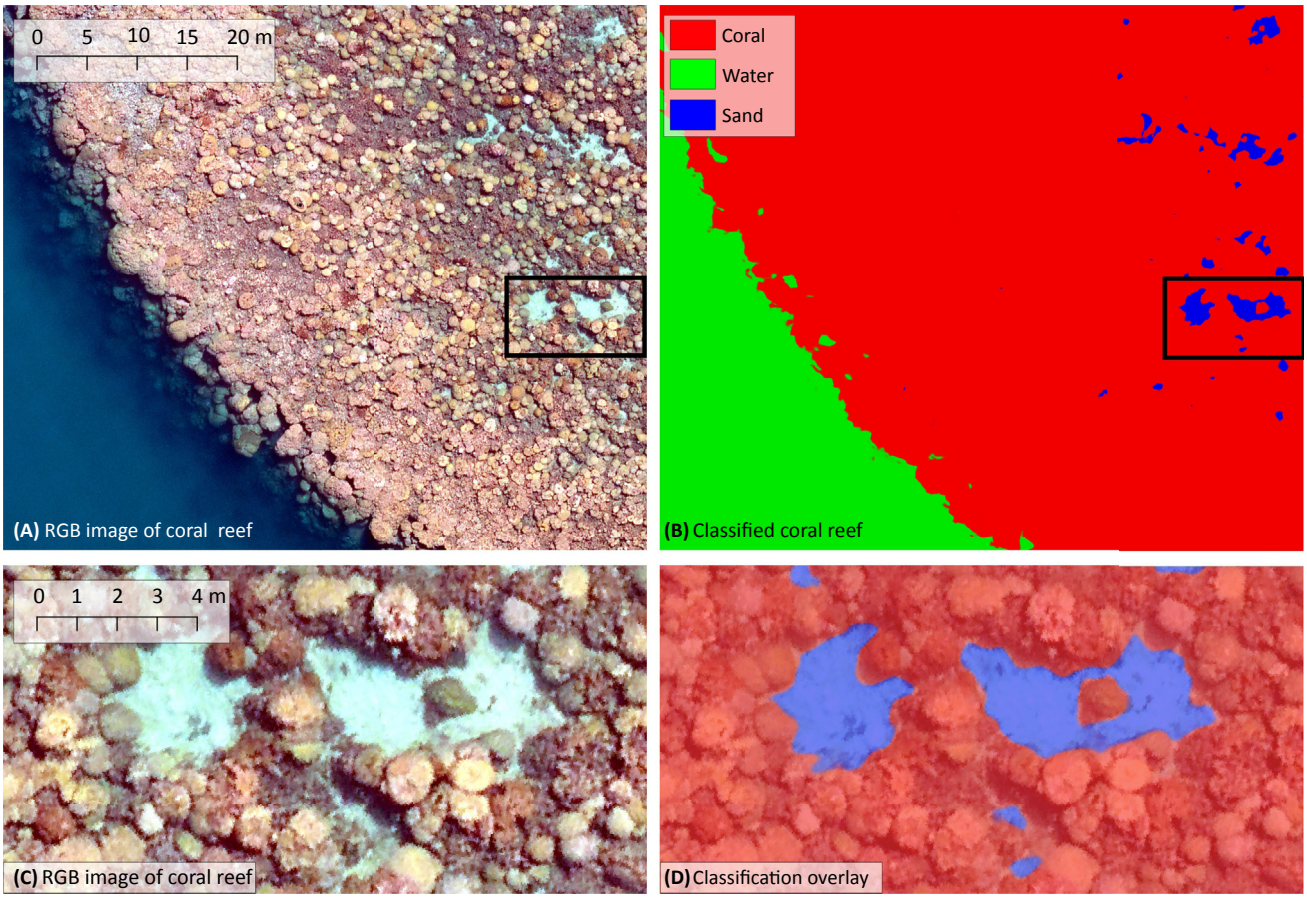


Trends in Ecology & Evolution

Figure 2. Segmentation of Individual Termite Mounds Across a Landscape. A hillshade map overlain with termite mound identification (A) makes the termite mounds visible over areas with elevation changes. Hillshade is shown for an image subset with a greater zoom (B), together with the corresponding termite mound classification obtained by the convolutional neural network (CNN) (C) and for the hillshade again overlain with termite mound classification (D).

runoff and groundwater discharge [86,87]. Better understanding of reefscapes and their connections with the land surface will shape how we characterize coral resilience, and will influence the prioritization of conservation efforts. However, understanding these interactions requires accurate representations of the distribution of corals, and of how coral cover changes through time.

Using three-band, 40 cm resolution RGB images from a high-resolution camera mounted on the CAO [83] (Figure 3A), we manually outlined a ~0.2 ha area (~1 million pixels) of ocean, sand, and coral reef in an independent image subset. We then trained and applied a CNN to accurately classify these ecological components at scale (Figure 3B). Segmenting this image into these components would be largely ineffective if only pixel-level information is considered; redundant colors would lead to misidentified pixels scattered throughout the image, challenging its utility for subsequent spatial analyses of coral distribution. By considering spatial context, however, the separation becomes obvious, as demonstrated by the overlay in Figure 3C.



Trends in Ecology & Evolution

Figure 3. Major Component Segmentation of a Reefscape. An RGB image of a small coral reef area containing open water, fore coral, and patch coral (A) was segmented into coral, water, and sand. We found that the classification resolves well locally (C and D) as well as globally (A and B). Abbreviation: RGB, red, green, and blue.

Oil Palm Classification

Close to 20% of all remaining forest worldwide lies within 100 m of a forest boundary [88,89]. This incredible degree of forest fragmentation, frequently caused by human activities such as the construction of oil palm plantations, can influence ecosystems up to several kilometers away from the forest edge [90,91]. To better understand these edge effects, however, we need more accurate mapping of the extent and edges of oil palm plantations, and at high resolutions this can again be a laborious process, particularly given the massive spatial extents in question [92].

We again draw on LiDAR data collected by the CAO [83] over oil palm plantations on the island of Borneo, in Sabah, Malaysia. We use here the top of canopy height (TCH), the height of vegetation as measured by the difference between first and last LiDAR pulse returns, calculated as in Asner *et al*. [93]. Without spatial context, TCH is a poor oil palm classifier given that intact forests (shown on the right-hand side of Figure 4A) have many canopy gaps with equal tree height to the oil palm plantations shown on the left-hand side of Figure 4A. Training data were constructed by manually delineating oil palm over a ~6 500 ha area (~16 million pixels) of TCH

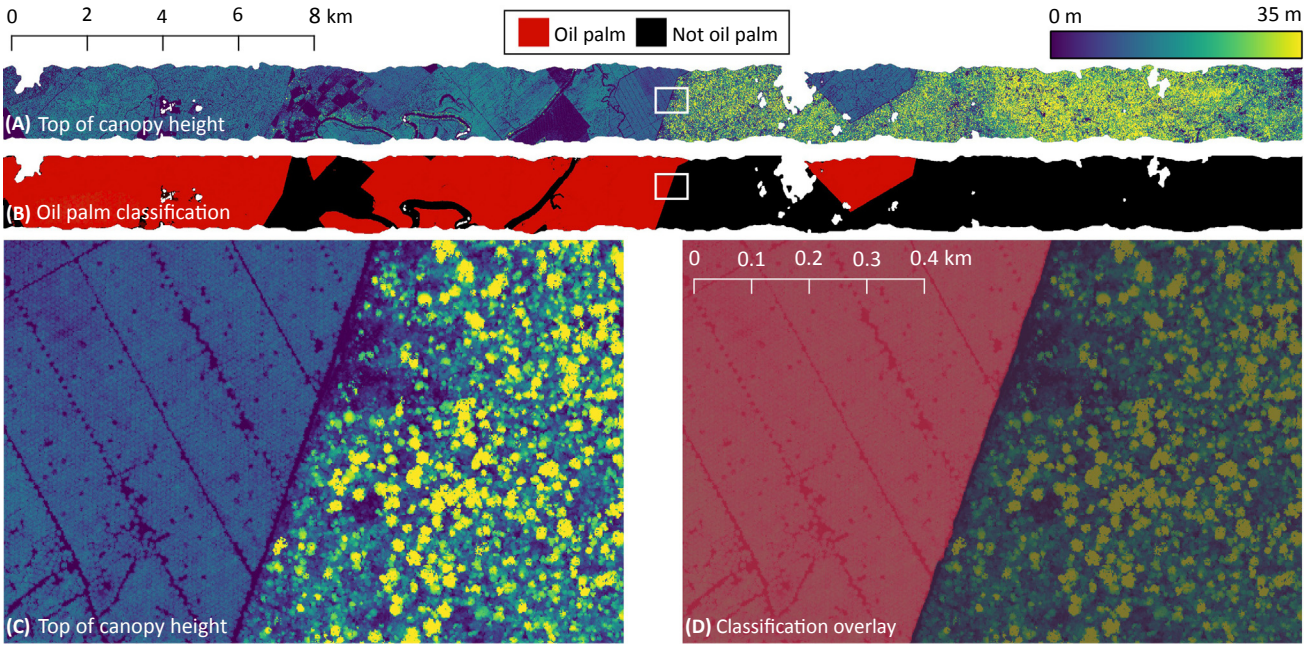


Figure 4. Separating Intact Tropical Forest and Oil Palm from Tree Canopy Height. A map of tree canopy height (A) shows that landscape texture leads to accurate classification of plantations, show in red in (B). The zoomed figures (C) and (D) demonstrate that, despite the large application area, the segmentation remains highly locally resolved.

data. Although broadly accurate, boundaries were difficult to accurately designate by hand. However, the CNN demonstrates clean disaggregation of oil palm, even in the presence of multiple different types of land cover, including intact forest, impacted forest, and riparian zones (Figure 4B).

From Classification to Ecology

The examples above stop short of the pursuit of ecological questions: simply identifying where features of interest are in a landscape does not inherently lead to an understanding of system interactions. However, the ability to rapidly identify biophysical components over large areas is a significant first step towards enabling investigation of ecosystem processes and changes through time. Environmental gradients and landscapes with varying land-use histories can provide natural laboratories, the monitoring of which can enable insights that are otherwise difficult and costly to establish at large scales. Manipulation experiments, for example, are a standard method for understanding mechanisms underlying ecological patterns, but they can require years or decades of monitoring to be able to discern processes [4], and are usually restricted to small scales. Augmenting actual experiments with observation-based ecology, however, requires extensive scale to control for nested multiscale variation, and this is where automation becomes invaluable. Automated, large-scale monitoring of ecosystem properties also has the potential to enable new insights into ecosystem patterns and processes that are typically not observable through more traditional manipulative and/or small-scale observation work. To meet this goal, a trained CNN can also be used to track fine-scale changes in ecological properties through both broad spatial scales as well as through time.

Concluding Remarks and Future Perspectives

CNNs have provided enormous breakthroughs in image analysis in recent years. When coupled with high-resolution remote sensing data, these methods provide a powerful tool to classify objects and segment landscapes over large geographic areas. Crucially, CNNs leverage spatial context, considering not only single pixels but also the surrounding landscape to classify biological phenomena, even with relatively few training input data features. To date, the use of CNNs in ecology is sparse, although it is rapidly increasing. The adoption of CNNs by the ecological community has the potential to significantly increase the characterization of spatial and temporal biological patterns through automated mapping over large extents and through time. This in turn can lead to the discovery of the processes that govern these patterns and thereby yield new insights into how ecosystems function at scales that were impossible to consider before.

Specific CNN architectures will continue to evolve in the coming years, further increasing model accuracy, decreasing training times, and expanding the types of data that can be incorporated. Advances that combine semantic segmentation with object detection to simultaneously generate per-pixel classes while also cleanly separating occurences of individual objects, known as instance segmentation, will continue to develop [31,94,95], and may ultimately provide an ideal choice for ecology. The generation of training data still requires significant effort, but semisupervised [96] and unsupervised learning [97,98] with CNNs, where labeled training data are either limited or even completely unnecessary for the algorithm to learn relevant landscape features, is under development and could eventually facilitate the use of CNNs for new ecological discovery (see Outstanding Questions). As CNNs evolve, ecologists will need to continue experimenting with the latest methods, but the fundamental principle of context consideration will remain the same. In the coming years, we not only encourage ecologists to use CNNs as a tool to identify ecological properties at new spatial and temporal scales, but even more importantly to innovate with the use of CNNs in the characterization of pattern and process in ecosystems.

Outstanding Questions

Can CNNs be extended to directly identify ecological patterns, rather than only biophysical components that ecologists can then link together as patterns?

How can unsupervised CNNs, as they develop, facilitate the identification of ecologically relevant patterns and processes without predefined classes?

How can CNN architectures be modified to incorporate temporal data to encapsulate both spatial and phenological context?

How best can multiple scales and/or resolutions of remotely sensed data be integrated to match the scale of different biophysical components?

Acknowledgments

The authors thank O. Csillik, K.D. Chadwick, and N. Fabina for their helpful comments on this manuscript. This review features example datasets made available by the Carnegie Airborne Observatory, which is made possible by grants and donations to G.P. Asner from the Avatar Alliance Foundation, the Grantham Foundation for the Protection of the Environment, the Gordon and Betty Moore Foundation, the John D. and Catherine T. MacArthur Foundation, the W. M. Keck Foundation, the Margaret A. Cargill Foundation, Mary Anne Nyburg Baker and G. Leonard Baker, Jr, and William R. Hearst III.

Supplemental Information

Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.tree.2019.03.006.

References

1. Kerr, J.T. and Ostrovsky, M. (2003) From space to species: ecological applications for remote sensing. *Trends Ecol. Evol.* 18, 299–305
2. Kellner, J.R. and Hubbell, S.P. (2018) Density-dependent adult recruitment in a low-density tropical tree. *Proc. Natl. Acad. Sci. U. S. A.* 115, 11268–11273
3. Roughgarden, J. *et al.* (1991) What does remote sensing do for ecology? *Ecology* 72, 1918–1922
4. Pringle, R.M. and Tarnita, C.E. (2017) Spatial self-organization of ecosystems: integrating multiple mechanisms of regular-pattern formation. *Annu. Rev. Entomol.* 62, 359–377
5. Anderson, C.B. (2018) Biodiversity monitoring, earth observations and the ecology of scale. *Ecol. Lett.* 21, 1572–1585
6. Phillips, S.J. *et al.* (2006) Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231–259
7. Elith, J. *et al.* (2008) A working guide to boosted regression trees. *J. Anim. Ecol.* 77, 802–813
8. Fagan, M. *et al.* (2015) Mapping species composition of forests and tree plantations in Northeastern Costa Rica with an integration of hyperspectral and multitemporal Landsat imagery. *Remote Sens.* 7, 5660–5696
9. Vaughn, N.R. *et al.* (2018) An approach for high-resolution mapping of Hawaiian Metrosideros Forest mortality using laser-guided imaging spectroscopy. *Remote Sens.* 10, 502
10. Paz-Kagan, T. *et al.* (2017) What mediates tree mortality during drought in the southern Sierra Nevada. *Ecol. Appl.* 27, 2443–2457

11. Blaschke, T. *et al.* (2014) Geographic object-based image analysis – towards a new paradigm. *ISPRS J. Photogramm. Remote Sens.* 87, 180–191
12. Blaschke, T. (2010) Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* 65, 2–16
13. Drăguţ, L. *et al.* (2014) Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS J. Photogramm. Remote Sens.* 88, 119–127
14. Csillik, O. (2017) Fast segmentation and classification of very high resolution remote sensing data using SLIC superpixels. *Remote Sens.* 9, 243
15. Everingham, M. *et al.* (2015) The PASCAL visual object classes challenge: a retrospective. *Int. J. Comput. Vis.* 111, 98–136
16. Webb, S. (2018) Deep learning for biology. *Nature*, 554, 555–557
17. LeCun, Y. *et al.* (2015) Deep learning. *Nature* 521, 436–444
18. Le, Q.V. *et al.* (2013) Building high-level features using large scale unsupervised learning. In *29th International Conference on Machine Learning*, pp. 8595–8598, ICML
19. Mishkin, D. *et al.* (2017) Systematic evaluation of convolution neural network advances on the Imagenet. *Comput. Vis. Image Underst.* 161, 11–19
20. Krizhevsky, A. *et al.* (2012) ImageNet classification with deep convolutional neural networks. In *International Conference on Neural Information Processing Systems 25*, pp. 1097–1105, Curran Associates
21. Simonyan, K. and Zisserman, A. (2015) Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations* 1–14
22. Szegedy, C. *et al.* (2015) Going deeper with convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, IEEE
23. Girshick, R. *et al.* (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, IEEE
24. Girshick, R. (2015) Fast r-cnn. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1440–1448, IEEE
25. Ren, S. *et al.* (2015) Faster r-cnn: towards real-time object detection with region proposal networks. In *International Conference on Neural Information Processing Systems 28*, pp. 91–99, Curran Associates
26. Redmon, J. *et al.* (2016) You only look once: unified, real-time object detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788, IEEE
27. Long, J. *et al.* (2015) Fully convolutional networks for semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, IEEE
28. Ronneberger, O. *et al.* (2015) U-net: convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted intervention*, pp. 234–241, MICCAI Society
29. Badrinarayanan, V. *et al.* (2017) SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2481–2495
30. Takeki, A. *et al.* (2016) Detection of small birds in large images by combining a deep detector with semantic segmentation. In *IEEE International Conference on Image Processing*, pp. 3977–3981, IEEE
31. He, K. *et al.* (2017) Mask R-CNN. In *IEEE International Conference on Computer Vision*, pp. 2980–2988, IEEE
32. Lee, S.H. *et al.* (2015) Deep-plant: plant identification with convolutional neural networks. In *2015 IEEE International Conference on Image Processing*, pp. 452–456, IEEE
33. Younis, S. *et al.* (2018) Taxon and trait recognition from digitized herbarium specimens using deep convolutional neural networks. *Bot. Lett.* 8107, 1–7
34. Wäldchen, J. and Mäder, P. (2018) Machine learning for image based species identification. *Methods Ecol. Evol.* 9, 2216–2225
35. Weinstein, B.G. (2018) A computer vision for animal ecology. *J. Anim. Ecol.* 87, 533–545
36. Willi, M. *et al.* (2019) Identifying animal species in camera trap images using deep learning and citizen science. *Methods Ecol. Evol.* 10, 80–91
37. Gomez, A. *et al.* (2016) Animal identification in low quality camera-trap images using very deep convolutional neural networks and confidence thresholds. In *International Symposium on Visual Computing*, pp. 747–756, Springer
38. Bowley, C. *et al.* (2016) Detecting wildlife in uncontrolled outdoor video using convolutional neural networks. In *12th International Conference on e-Science*, pp. 251–259, IEEE
39. Gray, P.C. *et al.* (2019) A convolutional neural network for detecting sea turtles in drone imagery. *Methods Ecol. Evol.* 10, 345–355
40. Qin, H. *et al.* (2016) DeepFish: accurate underwater live fish recognition with a deep architecture. *Neurocomputing* 187, 49–58
41. Redmon, J. and Farhadi, A. (2017) YOLO9000: better, faster, stronger. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7263–7271, IEEE
42. Redmon, J. and Farhadi, A. (2018) YOLOv3: an incremental improvement. *arXiv* Published online April 8, 2018. https://arxiv.org/abs/1804.02767
43. Liu, W. *et al.* (2016) SSD: single shot multibox detector. In *European Conference on Computer Vision*, pp. 21–37, Springer
44. Zhang, J. *et al*. (2016) Cancer cells detection in phase-contrast microscopy images based on faster R-CNN. In *Ninth International Symposium on Computational Intelligence and Design*, pp. 363–367, Publisher
45. Poostchi, M. *et al.* (2018) Malaria parasite detection and cell counting for human and mouse using thin blood smear microscopy. *J. Med. Imaging* 5, 1
46. Gupta, A. *et al.* (2019) Deep learning in image cytometry: a review. *J Quantitative Cell Sci* 95, 366–680
47. Hung, J. and Carpenter, A. (2017) Applying faster R-CNN for object detection on malaria images. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 56–61, IEEE
48. Chen, C.H. and Liu, K.H. (2017) Stingray detection of aerial images with region-based convolution neural network. In *IEEE International Conference on Consumer Electronics*, pp. 175–176, IEEE
49. Xie, Y. *et al.* (2016) Spatial clockwork recurrent neural network for muscle perimysium segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 185–193, Springer
50. Ghafoorian, M. *et al.* (2016) Non-uniform patch sampling with deep convolutional neural networks for white matter hyperintensity segmentation. In *IEEE 13th International Symposium on Biomedical Imaging*, pp. 1414–1417, IEEE
51. Wang, J. *et al.* (2016) A deep learning approach for semantic segmentation in histology tissue images. In *Medical Image Computing and Computer-Assisted Intervention*, pp. 176–184, Springer
52. Falk, T. *et al.* (2019) U-Net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* 16, 67–70
53. DeFries, R.S. and TownShend, J.R.G. (1994) NDVI-derived land cover classifications at a global scale. *Int. J. Remote Sens.* 15, 3567–3586
54. Marcos, D. *et al.* (2018) Land cover mapping at very high resolution with rotation equivariant CNNs: towards small yet accurate models. *ISPRS J. Photogramm. Remote Sens.* 145, 96–107
55. Mnih, V. (2013) *Machine Learning for Aerial Image Labeling*, University of Toronto
56. Maggiori, E. *et al.* (2017) Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Trans. Geosci. Remote Sens.* 55, 645–657
57. Zhu, X.X. *et al.* (2017) Deep learning in remote sensing: a comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* 5, 8–36
58. Volpi, M. and Tuia, D. (2018) Deep multi-task learning for a geographically-regularized semantic segmentation of aerial images. *ISPRS J. Photogramm. Remote Sens.* 144, 48–60

59. King, A. *et al.* (2018) A comparison of deep learning methods for semantic segmentation of coral reef survey images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1394–1402, IEEE
60. Rezaee, M. *et al.* (2018) Deep convolutional neural network for complex wetland classification using optical remote sensing imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 11, 3030–3039
61. Li, W. *et al.* (2016) Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sens.* 9, 22
62. Csillik, O. *et al.* (2018) Identification of citrus trees from unmanned aerial vehicle imagery using convolutional neural networks. *Drones* 2, 39
63. Wagner, F.H. *et al.* (2019) Using the U-net convolutional network to map forest types and disturbance in the Atlantic rainforest with very high resolution images. In *Remote Sens. Ecol. Conserv.*. http://dx.doi.org/10.1002/rse2.111
64. Kellenberger, B. *et al.* (2018) Detecting mammals in UAV images: best practices to address a substantially imbalanced dataset with deep learning. *Remote Sens. Environ.* 216, 139–153
65. Torney, C.J. *et al.* (2019) A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. *Methods Ecol. Evol.* http://dx.doi.org/10.1111/2041-210X.13165
66. Ayrey, E. *et al.* (2018) The use of three-dimensional convolutional neural networks to interpret LiDAR for forest inventory. *Remote Sens.* 10, 649
67. Yu, F. and Koltun, V. (2016) Multi-scale context aggregation by dilated convolutions. *International Conference on Learning Representations* 1–13
68. Zhao, H. *et al.* (2017) Pyramid scene parsing network. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2881–2890, IEEE
69. Srivastava, N. *et al.* (2014) Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1929–1958
70. Ioffe, S. and Szegedy, C. (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. *arXiv* Published online February 11, 2015. http://arxiv.org/abs/1502.03167
71. He, K. *et al.* (2016) Identity mappings in deep residual networks. In *European Conference on Computer Vision*, pp. 630–645, Springer
72. He, K. *et al.* (2016) Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, IEEE
73. Yu, F. *et al.* (2017) Dilated residual networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 472–480, IEEE
74. Krähenbühl, P. and Koltun, V. (2011) Efficient inference in fully connected crfs with gaussian edge potentials. In *Advances in neural information processing systems, 24*, pp. 109–117, Curran Associates
75. Chen, L-C. *et al.* (2015) Semantic image segmentation with deep convolutional nets and fully connected CRFs. In *International Conference on Learning Representations*, pp. 1–14, ICLR
76. Chen, L-C. *et al.* (2018) DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* 40, 834–848
77. Chen, L-C. et al. (2017) Rethinking atrous convolution for semantic image segmentation. *arXiv* Published online June 17, 2017. https://arxiv.org/abs/1706.05587
78. Sileshi, G.W. *et al.* (2010) Termite-induced heterogeneity in African savanna vegetation: mechanisms and patterns. *J. Veg. Sci.* 21, 923–937
79. Davies, A.B. *et al.* (2016) Termite mounds alter the spatial distribution of African savanna tree species. *J. Biogeogr.* 43, 301–313
80. Davies, A.B. *et al.* (2014) Spatial variability and abiotic determinants of termite mounds throughout a savanna catchment. *Ecography* 37, 852–862
81. Levick, S.R. *et al.* (2010) Regional insight into savanna hydrogeomorphology from termite mounds. *Nat. Commun.* 1, 65
82. Pringle, R.M. *et al.* (2010) Spatial pattern enhances ecosystem functioning in an African savanna. *PLoS Biol.* 8, e1000377
83. Asner, G.P. *et al.* (2012) Carnegie Airborne Observatory-2: increasing science data dimensionality via high-fidelity multi-sensor fusion. *Remote Sens. Environ.* 124, 454–465
84. Roff, G. and Mumby, P.J. (2012) Global disparity in the resilience of coral reefs. *Trends Ecol. Evol.* 27, 404–413
85. Beyer, H.L. *et al.* (2018) Risk-sensitive planning for conserving coral reefs under rapid climate change. *Conserv. Lett.* 11, e12587
86. Fabricius, K.E. (2005) Effects of terrestrial runoff on the ecology of corals and coral reefs: review and synthesis. *Mar. Pollut. Bull.* 50, 125–146
87. Bainbridge, Z. *et al.* (2018) Fine sediment and particulate organic matter: a review and case study on ridge-to-reef transport, transformations, fates, and impacts on marine ecosystems. *Mar. Pollut. Bull.* 135, 1205–1220
88. Haddad, N.M. *et al.* (2015) Habitat fragmentation and its lasting impact on Earth's ecosystems. *Sci. Adv.* 1, e1500052
89. Brinck, K. *et al.* (2017) High resolution analysis of tropical forest fragmentation and its impact on the global carbon cycle. *Nat. Commun.* 8, 14855
90. Broadbent, E.N. *et al.* (2008) Forest fragmentation and edge effects from deforestation and selective logging in the Brazilian Amazon. *Biol. Conserv.* 141, 1745–1757
91. Chaplin-Kramer, R. *et al.* (2015) Degradation in carbon stocks near tropical forest edges. *Nat. Commun.* 6, 1–6
92. Qie, L. *et al.* (2017) Long-term carbon sink in Borneo's forests halted by drought and vulnerable to edge effects. *Nat. Commun.* 8, 1966
93. Asner, G.P. *et al.* (2018) Mapped aboveground carbon stocks to advance forest conservation and recovery in Malaysian Borneo. *Biol. Conserv.* 217, 289–310
94. Pont-Tuset, J. *et al.* (2017) Multiscale combinatorial grouping for image segmentation and object proposal generation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 128–140
95. Maninis, K.-K. *et al.* (2018) Convolutional oriented boundaries: from image segmentation to high-level tasks. *IEEE Trans. Pattern Anal. Mach. Intell.* 40, 819–833
96. Fried, O. *et al.* (2017) Patch2Vec: globally consistent image patch representation. *Comput. Graph. Forum* 36, 183–194
97. Jean, N. *et al.* (2018) Tile2Vec: unsupervised representation learning for spatially distributed data. *arXiv* Published online May 8, 2018. https://arxiv.org/abs/1805.02855
98. Xia, X. and Kulis, B. (2017) W-Net: a deep model for fully unsupervised image segmentation. *arXiv* Published online November 22, 2017. https://arxiv.org/abs/1711.08506