Received: 22 May 2018 | Accepted: 19 November 2018

DOI: 10.1111/2041-210X.13133

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

REVIEW

Past, present and future approaches using computer vision for animal re-identification from camera trap data

Stefan Schneider[1] | **Graham W. Taylor**[2] | **Stefan Linquist**[3] | **Stefan C. Kremer**[1]

[1]Department of Computer Science, University of Guelph, Guelph, ON, Canada

[2]Department of Engineering, University of Guelph, Guelph, ON, Canada

[3]Department of Philosophy, University of Guelph, Guelph, ON, Canada

Correspondence
Stefan Schneider
Email: sschne01@uoguelph.ca

Handling Editor: Robert B. O'Hara

Abstract

1. The ability of a researcher to re-identify (re-ID) an individual animal upon re-encounter is fundamental for addressing a broad range of questions in the study of ecosystem function, community and population dynamics and behavioural ecology. Tagging animals during mark and recapture studies is the most common method for reliable animal re-ID; however, camera traps are a desirable alternative, requiring less labour, much less intrusion and prolonged and continuous monitoring into an environment. Despite these advantages, the analyses of camera traps and video for re-ID by humans are criticized for their biases related to human judgement and inconsistencies between analyses.
2. In this review, we describe a brief history of camera traps for re-ID, present a collection of computer vision feature engineering methodologies previously used for animal re-ID, provide an introduction to the underlying mechanisms of deep learning relevant to animal re-ID, highlight the success of deep learning methods for human re-ID, describe the few ecological studies currently utilizing deep learning for camera trap analyses and our predictions for near future methodologies based on the rapid development of deep learning methods.
3. For decades, ecologists with expertise in computer vision have successfully utilized feature engineering to extract meaningful features from camera trap images to improve the statistical rigor of individual comparisons and remove human bias from their camera trap analyses. Recent years have witnessed the emergence of deep learning systems which have demonstrated the accurate re-ID of humans based on image and video data with near perfect accuracy. Despite this success, ecologists have yet to utilize these approaches for animal re-ID.
4. By utilizing novel deep learning methods for object detection and similarity comparisons, ecologists can extract animals from an image/video data and train deep learning classifiers to re-ID animal individuals beyond the capabilities of a human observer. This methodology will allow ecologists with camera/video trap data to reidentify individuals that exit and re-enter the camera frame. Our expectation is that this is just the beginning of a major trend that could stand to revolutionize the analysis of camera trap data and, ultimately, our approach to animal ecology.

KEYWORDS

animal reidentification, camera traps, computer vision, convolutional networks, deep learning, density estimation, monitoring, object detection

1 | INTRODUCTION

The ability to re-ID animals allows for population estimates which are used in a variety of ecological metrics including diversity, relative abundance distribution and carrying capacity (Krebs 1989). These contribute to larger, overarching ecological interpretations of trophic interactions and population dynamics (Krebs 1989). Ecologists have used a variety of techniques for re-ID including tagging, scarring, banding and DNA analyses of hair follicles or feces (Krebs 1989). While accurate, these techniques are laborious for the field research team, intrusive to the animal and often expensive for the researcher.

Compared to traditional methods of field observations, camera traps are desirable due to their lower cost and reduced workload for field researchers. Camera traps also provide a unique advantage by recording the undisturbed behaviours of animals within their environment. This has resulted in the discovery of surprising ecological interactions, habitat ranges and social dynamics, among other insights (Meek, Vernes, & Falzon, 2013; Scheel et al., 2017). These advantages have led to a 50% annual growth in publications using camera trap methods to assess population sizes between 1998 and 2008 and the trend has persisted until 2015 (Burton et al., 2015; Rowcliffe, Field, Turvey, & Carbone, 2008).

Despite their advantages, there are a number of practical and methodological challenges associated with the use of camera traps for animal re-ID. The discrimination of individual animals is often an expert skill requiring a considerable amount of training. Even among experienced researchers, there remains an opportunity for human error and bias (Foster & Harmsen, 2012; Meek et al., 2013). Historically, these limitations have restricted the use of camera traps to the re-ID of animals that bear conspicuous individual markings.

One strategy for overcoming these limitations has involved the use of computer vision to standardize the statistical analysis of animal re-ID. For decades, 'feature engineering' has been the most commonly used computational technique where algorithms are designed and implemented to focus exclusively on predetermined traits, such as patterns of spots or stripes, to discriminate among individuals. The main limitations of this approach surround its impracticality (Hiby et al., 2009). Feature engineering requires programming experience, sufficient familiarity with the organisms to identify relevant features, and lacks in generality where once a feature detection algorithm has been designed for one species, it is unlikely to be useful for other taxa.

Recent decades have witnessed the emergence of deep learning systems that make use of large data volumes (Zheng et al., 2015). Modern deep learning systems no longer require 'hard-coded' feature extraction methods. Instead, these algorithms can learn, through their exposure to large amounts of data, the particular features that allow for the discrimination of individuals. (LeCun, Bengio, & Hinton, 2015). These methods have been developed primarily outside the realm of ecology, first in the field of computer image recognition (Krizhevsky, Sutskever, & Hinton, 2012), and more recently in the security and social media industries (Zheng et al., 2015). Modern deep learning systems now consistently outperform feature engineered methods provided that they have access to large amounts of data (Lisanti, Masi, Bagdanov, & Del Bimbo, 2015; Martinel, Das, Micheloni, & Roy-Chowdhury, 2015).

In recent years, a handful of ecologists has begun utilizing deep learning systems for species and animal individual identification with great success (Brust et al., 2017; Carter, Bell, Miller, & Gash, 2014; Freytag et al., 2016; Loos & Ernst, 2013; Norouzzadeh et al., 2017; Schneider, Taylor, & Kremer, 2018). Our expectation is that this is just the beginning of a major trend that could stand to revolutionize the analysis of camera trap data and, ultimately, our approach to animal ecology. Here, we present a collection of works utilizing computer vision for animal re-ID. We begin with the earliest computer-aided and feature-engineered approaches followed by an introduction to deep learning relevant to animal re-ID. We then present recent works utilizing deep learning for animal re-ID and lastly conclude discussing its practical applications.

2 | COMPUTER VISION FEATURE EXTRACTION METHODS FOR ANIMAL RE-IDENTIFICATION

To assess performance of computer vision classification tasks, a metric known as top-1 accuracy is used and will be referenced throughout. Top-1 accuracy describes the percentage of correct classifications a model outputs considering a test data set that was not used as part of training for the model. An overall summary of the reviewed studies can be found in Figure 1, Table 1 and Appendix S1.

The first use of computer vision for animal re-ID was introduced in 1990 by Whitehead, Mizroch et al. and Hiby and Lovell who published collectively in the same journal (Hiby & Lovell, 1990; Mizroch, Beard, & Lynde, 1990; Whitehead, 1990). Whitehead (1990) considered sperm whales *Physeter macrocephalus* using custom software to scan projector slides of a sperm whales fluke onto a digitizer tablet (Whitehead, 1990). The user would manually tap and save the location of a unique characteristic (such as a nick or a scratch) to a database. The software then considers the maximum sum of similarities of the descriptors and returns the most similar individual. Considering images collected from the Galapagos Islands, Ecuador containing 1,015 images of all unique individuals, 56 were considered for testing where the approach returned a 32% top-1 accuracy (Whitehead, 1990). Mizroch et al. (1990) considered humpback whales *Megaptera novaeangliae* re-identification using customized software where the user-labelled images of whale flukes considering a 14-sector fluke map selecting from a list of possible markings (Spots, pigment, etc.). Using a collection of 9,051 images of 790 individuals provided by the National Marine Mammal Laboratory, a similarity matrix was used to compare the ID of 30 individuals returning a top-1 accuracy of 43% (Mizroch et al., 1990). Re-ID from images of a whale's fluke remains a research interest today (Humpback whale identification challenge; Kaggle, 2018). These earliest approaches rely on qualitative descriptors which overall are too limited of a representation to capturing the nuanced detail required for animal re-ID. In 1990, Hiby and Lovell introduced the first feature engineered system considering the grey seal *Halichoerus grypus*. The user inputs numerous head captures of an animal from the same pose and the system renders a grid-based 3-D model of the animal where an individual's representation is captured in the summation of the greyscale pixel values of each grid cell (Hiby & Lovell, 1990). Considering a created dataset using photos taken from an undisclosed beach during grey seal breeding, a test set of 58 images of 58 individuals were re-identified with a 98% top-1 accuracy (Hiby & Lovell, 1990). This technique is limited to animals being in a specific orientation for the photos. In 1998, O'Corry-Crowe considered a feature engineering approach using a wavelet transformation to numerically represent the fluke of sperm whales using images provided by the Andenes Whale Center in Norway, receiving a 92.0% accuracy considering 56 images of eight individuals (OCorry-Crowe, 1998). The improved performance and unbiased results of these two feature extraction methods demonstrated their future potential for animal re-ID.

In the 2000s, Kelly, (2001) used the same 3-D model approach as Hiby & Lovell, (1990), comparing the collective similarity of pattern cells of cheetahs *Acinonyx jubatus*. Considering a catalogue of 10,000 images from the Serengeti National Park in Tanzania, Kelly (2001) achieved a 97.5% top-1 accuracy using a test set of 1,000 images of an undisclosed number of individuals (Kelly, 2001). In 2003, Hillman et al. (2003) developed an autonomous feature extraction method to capture unique nicks, scratches and markings from the dorsal fin of six whale and dolphin species. Named *Finscan*, the system localized 300 xy pairs from the dorsal fin of a variety of dolphin and shark species and determines individual similarity by calculating the minimum Euclidean distance of these pairs (Hillman et al., 2003). Hillman et al. (2003) report accurate classification results of 50% top-1 (Hillman et al., 2003) on their own datasets containing 500 images with on average 52 images of 36 individuals per species. This technique is ultimately limited by compressing an animal's representation into a single numeric value.

The complexity of feature representation increased in 2004 as Ravela and Gamble used a Taylor approximation of local colour intensities, multi-scale brightness histograms and curvature/orientation to re-ID marbled salamander *Ambystoma opacum* individuals (Ravela & Gamble, 2004). Ravela and Gamble (2004) constructed a custom housing for their camera trap to capture top–down images of salamanders as they entered. Using their collected images from traps in Massachusetts, USA, a database of 370 individuals was gathered and considering a test set of 69 images with an undisclosed number of individuals, they returned a top-1 accuracy of 72% (Ravela & Gamble, 2004). In 2005, Arzoumanian, Holmberg, and Norman (2005) explored whale shark *Rhincodon typus* re-ID by analysing the characteristic flank (front dorsal region) spot patterns using an astronomy algorithm for recognizing star patterns (Groth, 1986). The algorithm creates a list of possible triangles from white dots within an image using xy pairs and compares their orientation between images. Arzoumanian et al. (2005) used the ECOCEAN Whale Shark Photo-Identification library of 271 'datasets' to test 27 images of individuals from an undisclosed number of classifications

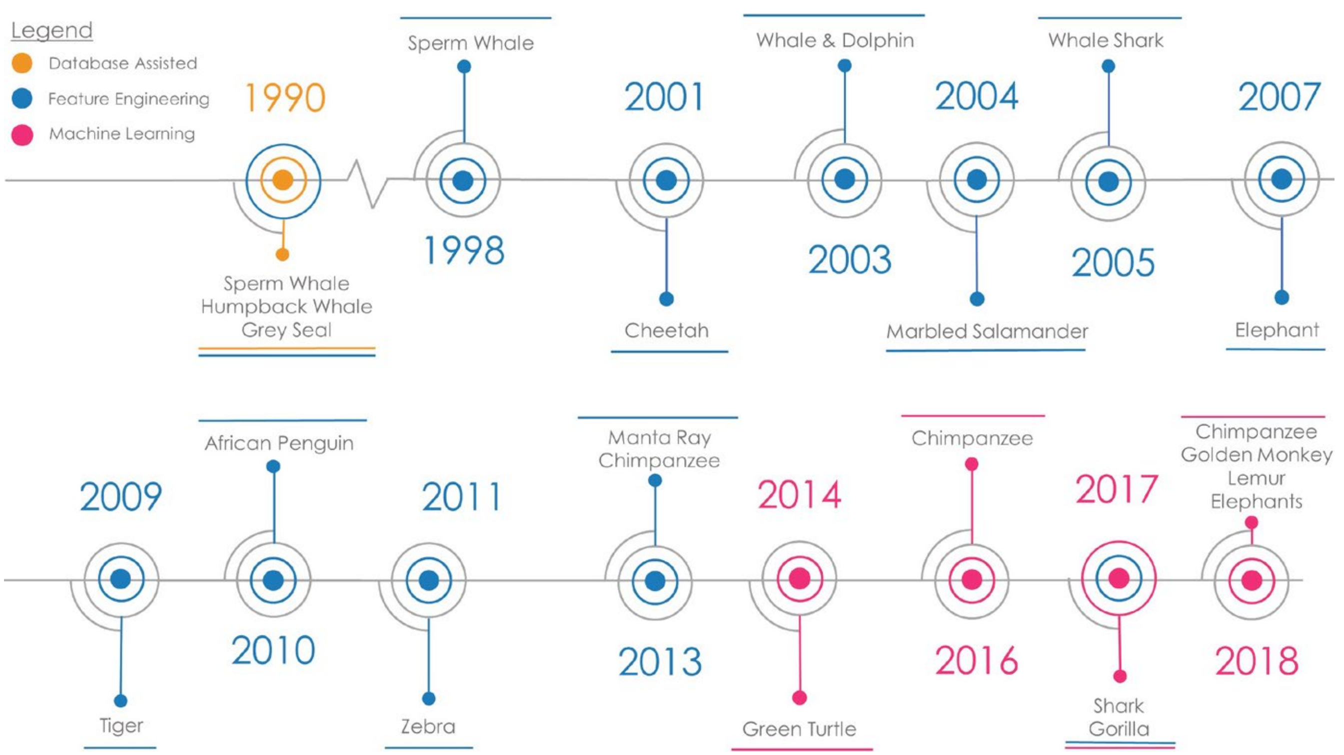

FIGURE 1 Timeline of animal re-ID methods discussed in this review segregated into categories: Database Assisted, Feature Engineering, and Machine Learning

TABLE 1 Summary of feature engineered and deep learning approaches for animal re-ID

Animal	Year	Methodology	Test size	Num. classes	Top-1 accuracy (%)
Sperm Whale	1990	Database similarity	56	1,015	59.0
Humpback Whale	1990	Database similarity	30	790	41.4
Grey Seal	1990	3-D pattern cell similarity	58	58	98.0
Sperm Whale	1998	Wavelet transformations	56	8	92.0
Cheetah	2001	3-D pattern cell similarity	1,000	NA	97.5
Whale/Dolphin	2003	XY pair euclidean distance	52	36	50.0
Marbled Salamander	2004	Pixel histogram and local Colours	69	NA	72.0
Whale Shark	2005	Star pattern recognition	27	NA	90.0
Elephant	2007	Polynomial multi-curve Matching	332	268	75.0
African penguin	2009	Per feature AdaBoost classifier	NA	NA	92.0-97.0
Tiger	2009	3-D pattern cell similarity	298	298	95.0
Manta Ray	2013	SIFT	720	265	51.0
Chimpanzee (C-Zoo)	2013	Support vector machine	478	120	84.0
Chimpanzee (C-Tai)	2013	Support vector machine	1,146	286	68.8
Green Turtle	2014	Feedforward network	180	72	95.0
Chimpanzee (C-Zoo)	2016	Convolutional network	478	120	92.0
Chimpanzee (C-Tai)	2016	Convolutional network	1,146	286	75.7
Shark	2017	Naive Bayes nearest neighbour	2,456	85	82.0
Gorilla	2017	Convolutional network	500	482	90.8
Elephant	2018	Support vector machine	2,078	276	59.0
Chimpanzee	2018	Siamese network	5,599	90	93.8
Lemur	2018	Siamese network	3,000	129	90.4
Golden Monkey	2018	Siamese network	241 videos	49	75.8

Computer vision animal re-identification techniques.

to report a top-1 accuracy of 90%. In 2007, Ardovini, Cinque, and Sangineto (2008) attempted to re-identify elephants *Loxodonta* spp. based on images using a multi-curve matching technique where a spline curve is fit to the mantle of an individual elephant. To re-ID an elephant, the known spline curves were overlaid atop the image of the elephant and the most similar was considered the same individual (Ardovini et al., 2008). Considering a database of elephants from Zakouma Ciad National Park consisting of 268 individuals and 332 test images, the researchers report a 75% top-1 accuracy. The methodology has since been used by the Centre for African Conservation (Ardovini et al., 2008). Overall, the studies so far are limited by their ability to capture nuanced details required for animal re-ID.

The first computer vision model for animal re-ID capable of generalizing across species was developed in 2007 by Burghardt and Campbell which extracted inherent singularities (spots, lines, etc.) of animal individuals (Burghardt et al., 2007). The technique involves three cameras pointed at a central location to capture a 3-D representation of the animal and extracts features using Haar-like (pixel difference) descriptors. For each feature, an AdaBoost classifier was trained to determine if the single feature was present and individual re-ID governed by an ensemble of these features (Burghardt et al., 2007). In 2010, Sherley, Burghardt, Barham, Campbell, and Cuthill (2010) used this approach on Robben Island, Africa, in one of the first works to perform fully autonomous population estimates solely from camera traps considering the African penguin *Spheniscus demersus*. The system returned an accuracy between 92% and 97% in comparison to records from humans on site (Sherley et al., 2010). Overall, this method is limited by the laborious task of training unique classifier for each possible feature on an animal.

In 2009, Hiby et al. (2009) collaborated with Karanath to explore their 3-D model representation technique considering a database of images from his long-term tiger population studies in Nagarhole and Bandipur tiger reserves. By again dividing individuals into pattern cells, re-ID was considered by the summation of the similarity of the cells which were considered using a Bayesian posterior probability estimate and tested considering live tigers and poached tiger rugs. Hiby et al. (2009) report a 95% top-1 accuracy considering 298 individuals from 298 possible individuals.

In 2013, Town, Marshall, and Sethasathien (2013) compared individual similarity of manta ray species *Manta alfredi* and *Manta birostris* based on images of their ventral side enhancing local contrasts and then extracting features (i.e. spots) using the SIFT algorithm. Considering a test size of 720 images of 265 individuals from the Manta Ray & Whale Shark Research Centre Inhambane, Mozambique, Town et al. (2013) report a 51% accuracies commenting on the limitations of underwater photography. Also in 2013, Loos and Ernst attempt to re-ID the faces of chimpanzee (*Pan* spp.) considering features of pixel value gradient changes and pixel groupings to train a Support Vector Machine classifier, a classifier which maximizes a linear decision boundary between classifications (Hearst, Dumais, Osuna, Platt, & Scholkopf, 1998; Loos & Ernst, 2013). Loos and Ernst (2013) report an accuracy of 84.0% and 68.8% on their two labelled datasets, C-Zoo from the Leipzig Zoo, Germany, containing 598 images of 24 individuals and C-Tai from the Tai National Park, Cote d'Ivoire, Africa, containing 1,432 images of 71 individuals, using a randomly selected one-fifth of the data as test images. Loos and Ernst (2013) consider a promising approach but are limited by the capabilities of the SVM classifier. In 2017, Hughes and Burghardt use fin contours to extract features and use a local naive Bayes nearest neighbour nonlinear model for re-ID (Hughes & Burghardt, 2017). Using the dataset FinsScholl2456 containing 2456 images of 85 individuals, their approach returns a top-1 accuracy of 82% for correctly classified shark individuals demonstrating strong performance for re-ID by dorsal fin.

While these techniques have shown success, there remains improvement in the performance, ease of design, implementation and overall accessibility of these algorithms. Deep learning methods provide an opportunity to remedy each of these disadvantages.

3 | DEEP LEARNING AND ITS SUCCESS FOR HUMAN RE-IDENTIFICATION

Modern deep learning systems have shown great success learning the necessary features for re-ID from data and remove the need for feature engineering. Deep learning as concept originated in the 1940–1960s as cybernetics was rebranded as connectionism between 1980 and 1995, and starting in 2006 rebranded again as deep learning (Hinton, Osindero, & Teh, 2006; McCulloch & Pitts, 1943; Rumelhart, Hinton, & Williams, 1986). Despite its long history, there has been a rapid growth of interest in deep learning due to its success related to improved computational power and the availability of large datasets, both requirements for the model. In recent years, deep learning methods have dramatically improved performance levels in the fields of speech recognition, computer vision, drug discovery, genomics, artificial intelligence and others becoming the standard computational approach for problems with large amounts of data (LeCun et al., 2015). Here, we provide a brief description of the underlying mechanism of deep learning as it relates to computer vision and animal re-ID. The process that will be outlined in this section is an approach known as supervised learning and is the most common form of deep learning (LeCun et al., 2015).

Deep learning is a computational framework where a system's parameters are not designed by human engineers but trained from large amounts of data using a general-purpose learning algorithm (LeCun et al., 2015). Intuitively, deep learning systems are organized as a layered web/graph structure where labelled data are submitted as input, and many modifiable scalar variables (known as weights) are summed and multiplied by a nonlinear transformations to output a predicted label from a predefined number of possible choices (Goodfellow, Bengio, & Courville, 2016). Each training example is fed through the network providing an answer, and based on the results from an 'objective function' (i.e. the average answer accuracy across the dataset), the weight values are modified by an optimizer (e.g. gradient descent with backpropagation) in an attempt to improve accuracy (Goodfellow et al., 2016). Deep learning systems will often have millions of modifiable weight values and with enough data and computation, the underlying relationship between the data and output can be mapped to return accurate results (He, Zhang, Ren, & Sun, 2016; Krizhevsky et al., 2012; Simonyan & Zisserman, 2014; Szegedy et al., 2015). The general term for this framework is a neural network of which there are many architectures. The architecture described here is known as a feedfoward network (LeCun et al., 2015). In 1991, Hornik proved that feedforward neural networks are a universal approximator, capable of mapping any input to output if a relationship exists (Swanson et al., 2015).

Neural networks are best described as multiple processing layers where each layer learns a representation of the data with different levels of abstraction (LeCun et al., 2015). As the data representation passes through each transformation, it allows for complex representations to be learned (LeCun et al., 2015). Consider an example dataset of many black and white images of animal individuals. The images are initially unravelled into a single vector of pixel values between 0 and 255 and fed into a deep learning system. Using this raw input, the first layer is able to learn simple representations of patterns within the data, such as the presence or absence of edges at particular orientations and locations in the image. The second layer will typically represent particular arrangements of edges and open space. The third layer may assemble these edges into larger combinations that correspond to parts of familiar objects, such as the basics of a nose or eyes, and subsequent layers would detect objects as combinations of these parts, such as a face (LeCun et al., 2015). Based on the combination of larger parts, such as the ears, face or nose, a learning system is able to correctly classify different individuals with near perfect accuracy when given enough input data (Fukushima, 1979). A deep learning system can learn the subtle details distinguishing a unique classification, such as a differing antler structure, and ignore large irrelevant variations such as the background, pose, lightning and surrounding objects (Krizhevsky et al., 2012).

Many recent advances in machine learning have come from improving the architectures of a neural network. One such architecture is the Convolutional Neural Network (CNN), first introduced as the 'Neocognitron' in 1979 by Fukushima, which is now the most commonly used architecture for computer vision tasks today (Iandola et al., 2014; Krizhevsky et al., 2012). CNNs introduce 'convolutional' layers within a network which learn feature maps that represent the spatial similarity of patterns found within the image (e.g. the presence or absence of lines or colours within an image) (LeCun et al., 2015). Each feature map is governed by a set of 'filter banks', which are matrices of scalar values that are learned similar to the standard weights of a network. Intuitively, these feature maps learn to extract the necessary features for a classification task replacing the need for feature engineering. CNNs also introduce max pooling layers, a method that reduces computation and increases robustness by evenly dividing the feature map into regions and returning only the highest activation values (LeCun et al., 2015). A simple CNN will have a two or three convolution layers passed through non-linearity functions, interspersed with two or three pooling layers, ending with fully connected layers to return a classification output. Machine learning researchers continually experiment with modular architectures of neural networks and six CNN frameworks have standardized as well performing with differences generally considering computation cost/memory in comparison to accuracy. These networks include AlexNet, VGG16, GoogLeNet/InceptionNet, ResNet, DenseNet and CapsNet (He et al., 2016; Krizhevsky et al., 2012; Pan & Yang, 2010; Sabour, Frosst, & Hinton, 2017; Simonyan & Zisserman, 2014; Szegedy et al., 2015). These networks range from 7 to 152 layers. A common approach for training a network for a niche task like animal re-ID is to use the pretrained weights of one of these six network structures trained on a public dataset as initialization parameters, and then retraining the network using labelled data of animal individuals (Rodner et al., 2015). This approach is known as Transfer Learning and helps improve performance when training on limited amounts of data (Rodner et al., 2015).

One niche task of deep learning research focuses on improving performance on extremely similar classifications, known as fine-grained visual recognition. Species identification tasks, such as classifying 675 similar moth species, have been an active research area for testing this problem domain and applicable for animal re-ID (Rodner, Simon, Fisher, & Denzler, 2016). Two techniques have been demonstrated to improve performance. The first is data augmentation, a universally recommended approach when training computer vision networks. This involves randomly flipping, cropping, rotating, blurring, shifting and altering colour/light images during each iteration of training (Souri & Kasaei, 2015). A second is an approach known as ‘object localization’, where classification predictions are made for each unique parts of the body of an animal (i.e. head, beck, ears, wings, etc.) that are then all considered in combination for a final classification (Ren, He, Girshick, & Sun, 2017).

The deep learning methods described so far are limited to returning one animal classification per image; however, this is suboptimal for the analysis of camera trap images. In order to identify multiple objects (i.e. animals), researchers train object detectors which segregate the image into regions that are passed through a CNN. Three approaches for object detection have recently grown in popularity. The first is Faster Region-CNN which segregates the image into *c*. 2000 proposal regions and passes each through a CNN (Redmon, Divvala, Girshick, & Farhadi, 2016). The second is YOLO (You-Only-Look-Once) which divides an image into a grid, and passes each grid cell through the network considering a series of predefined ‘anchors’ relevant to the predicted shape and size classifications of interest (Liu et al., 2016). Lastly is Single Shot Multibox Detector, which considers a set of default boxes and makes adjustments to these boxes during training to align over the objects of interest (Nowozin, 2014). Object detection methods have an additional objective function evaluation metric known as Intersection over Union (IOU), which returns performance as the area of overlap of the true and predicted regions divided by the entire area of the true and predicted regions (Taigman, Yang, Ranzato, & Wolf, 2014).

In 2015, two research teams, Lisanti et al. (2015) and Martinel et al. (2015), demonstrated the successful capabilities of CNNs on human re-ID using the ETHZ dataset, a dataset composed of 8580 images of 148 unique individuals taken from mobile platforms, where CNNs were able to correctly classify individuals from the test set with 99.9% accuracy after seeing five images of an individual. In 2014, Taigman et al. (2014) introduced Deepface, a method of creating a 3-dimensional representation of the human face to provide more data to a neural network which returned an accuracy of 91.4% on the YouTube faces dataset containing videos of 1,595 individuals (Bromley, Guyon, LeCun, Säckinger, & Shah, 1994). Considering traditional CNNs for re-ID requires a large number of labelled data for each individual and re-training the network for every new individual sighted, both of which are infeasible requirement for animal re-ID. In 1993, Bromley et al. (1994) introduced a suitable neural network architecture for this problem, titled a Siamese network, which learns to detect if two input images are similar or dissimilar (Schroff, Kalenichenko, & Philbin, 2015). Once trained, Siamese networks require only one labelled input image of an individual in order to accurately re-identify if an individual in a second image. In practice, one would train a Siamese network to learn a species’ similarity and compare individuals from a known database. In 2016, Schroff et al. (2015) introduced FaceNet which considers three input images in the Siamese format to train similarity: a standard image of an individual, an image of the same individual appearing vastly different, and an image of a different individual that appears similar (Körschens, Barz, & Denzler, 2018). This structure demands additional performance from the model and FaceNet currently holds the highest accuracy on the YouTube Faces dataset with a 95.12% top-1 accuracy (Körschens et al., 2018).

Based on the results for human re-ID, deep learning systems show promise as a generalizable framework for animal re-ID eliminating the biases related to human judgement and reducing the requirement for human labour to analyse images. With enough data, deep learning systems can be trained for any species which eliminates the need for domain expertise and species-specific feature extraction algorithms. Deep learning systems also show promise to exceed human level performance when re-identifying animal individuals without obvious patterns and markings.

4 | DEEP LEARNING FOR CAMERA TRAP ANALYSIS OF SPECIES AND ANIMAL RE-IDENTIFICATION

Despite the success of deep learning methods for human re-ID, few ecological studies have utilized its advantages. In 2014, Carter et al.

(2014) published one of the first works using neural networks for animal re-ID, a tool for green turtle *Chelonia mydas* re-ID. The authors collected 180 photos of 72 individuals from Lady Elliot Island in the southern Great Barrier Reef, both nesting and free swimming considering an undisclosed number of testing images. Their algorithm preprocesses the image by extracting a shell pattern, converting it to greyscale, unravelling the data into a raw input vector, and then training a simple feedforward network (Carter et al., 2014). Each model produces an output accuracy of 80–85% accuracy, but the authors utilize an ensemble approach by training 50 different networks and having each vote for a correct classification. The ensemble approach returns an accuracy of 95%. Carter et al. (2014)'s work has been considered a large success and is currently used to monitor the southern Great Barrier Reef green turtle population.

In 2016, Freytag et al. (2016) trained the CNN architecture AlexNet on the isolated faces of chimpanzees considering the C-Zoo and C-Tai datasets. Freytag et al. (2016) report an improved accuracy of 92.0% and 75.7% in comparison to the original feature extraction method of 84.0% and 68.8% (Freytag et al., 2016; Loos & Ernst, 2013). In 2017, Brust et al. (2017) trained the object detection method YOLO to extract cropped images of Gorilla *Gorilla gorilla* faces from 2,500 annotated camera trap images of 482 individuals taken in the Western Lowlands of the Nouabal′e -Nodki National Park in the Republic of Congo. Once the faces are extracted, Brust et al. (2017) followed the same procedure as Freytag et al. (2016) to train the CNN AlexNet achieving a 90.8% accuracy on a test size of 500 images. The authors close discussing how deep learning for ecological studies show promises for a whole realm of new applications if the fields of basic identify, spatio-temporal coverage and socioecological insights. (Brust et al., 2017; Freytag et al., 2016)

In 2018, Koerschens considered a completed pipeline of object detector and classifier to re-identify elephant individuals considering a dataset from *The Elephant Listening Project* from the Dzanga-Sangha reserve of which 2,078 images of 276 different individuals are present (Deb et al., 2018). Koerschens use a YOLO object detector trained using the ResNet50 architecture to localize the head of an elephant from an image and pass each head through a train Support Vector Machine classifier (Deb et al., 2018). Their model returns a top-1 accuracy of 59% (Deb et al., 2018). Most recently, Deb et al. (2018) tested the re-ID capabilities of deep learning systems on three primate species: chimpanzees, lemurs *Lemutttttriformes* spp., and golden monkeys *Cercopithecus kandti* (Zhou, 2017). Deb et al. (2018) consider three metrics for re-ID: verification (two image comparison), closed-set identification (individual is known to exist within the data), open-set identification (individual may or may not exist within the data) (Zhou, 2017). For chimpanzees, Deb et al. (2018) combined the C-Zoo and C-Tai datasets to create the *ChimpFace* dataset containing 5,599 images of 90 chimpanzees. For lemurs, they consider a dataset known as *LemurFace* from the Duke Lemer Center, North Carolina containing 3,000 face images of 129 lemur individuals from 12 different species. For golden monkeys, they extracted the faces of 241 short video clips (average 6 seconds) from Volcanoes National Park in Rwanda where 1,450 images of 49 golden monkey faces were cropped and extracted (Zhou, 2017). Deb et al. (2018) use a custom Siamese CNN containing four convolutional layers, followed by a 512 node fully connected layer (Zhou, 2017). Deb et al. (2018) report verification, closed-set and open-set accuracies, respectively, for lemurs: 83.1%, 93.8% and 81.3%; golden monkeys: 78.7%, 90.4% and 66.1%; and chimpanzees: 59.9%, 75.8% and 37.1% (Zhou, 2017).

5 | NEAR FUTURE TECHNIQUES FOR ANIMAL RE-IDENTIFICATION

Ecologists familiar with techniques in computer vision have exhibited considerable ingenuity in developing feature extraction methods for animal re-ID, but only recently have researchers considered modern deep learning methods, such as CNNs or Siamese networks. By considering modern deep learning approaches, ecologists can utilize improve accuracies without the requirement of handcoded feature extraction methods by training a neural network from large amounts of data.

We foresee the greatest challenge for deep learning methods being the creation of large labelled datasets for animal individuals. From our review, we suggest that 1,000+ images are required to train deep networks. Our proposed approach for data collection would be to utilize environments with known ground truths for individuals, such as national parks, zoos or camera traps in combination with individuals being tracked by GPS, to build the datasets. We would recommend using video wherever possible to gather the greatest number of images for a given encounter with an individual. We encourage researchers with images of labelled animal individuals to make these datasets publicly available to further the research in this field. In addition to gathering the images, labelling the data is then also a laborious task, especially when training an object detection model where bounding boxes are required. One approach for solving this problem is known as weakly supervised learning, where one provides object labels to a network (i.e. zebra) and the network returns the predicted coordinates of its location (Simpson, Page, & De Roure, 2014). An alternative approach is to outsource the labelling task to online services, such as Zooniverse which can be time saving for researchers, but introduces inevitable error and variability (Holzinger, 2016).

While deep learning approaches are able to generalize to examples similar to those seen during training, we foresee various environmental, positional and timing-related challenges. Environmental difficulties may include challenging weather conditions, such as heavy rain, or extreme lighting/shadows. A possible solution to limit these concerns may be to re-ID only during optimal weather conditions. A positional challenge may occur if an individual were to enter the camera frame at extremely near or far distances. To solve this, one could limit animals to a certain range from the camera before considering it for re-ID. Lastly, a challenge may be if an individual's appearance was to change dramatically between sightings, such as being injured or the rapid growth of a youth. While a network would be robust to such changes given training examples, this would require examples be available as training data. To account for this issue, we would consider having a 'human-in-the-loop' approach, where a human monitor results and relabels erroneous classifications for further training to improve performance (Holzinger, 2016).

While today fully autonomous re-ID is still in development, researchers can already use these systems to reduce manual labour for their studies. Examples include training networks to filter images by the presence/absence of animals or species classifications (Norouzzadeh et al., 2017; Schneider et al., 2018; iWildcam 2018 camera trap challenge). Soon deep learning systems will accurately perform animal re-ID at which time one can create systems that autonomously extract from camera traps a variety of ecological metrics such as diversity, evenness, richness, relative abundance distribution, carrying capacity and trophic function, contributing to overarching ecological interpretations of trophic interactions and population dynamics.

6 | CONCLUSION

Population estimates are the underlying metric for many fundamental ecological questions and rely on the accurate re-identification of animals. Camera and video data have become increasingly common due to their relatively inexpensive method of data collection; however, they are criticized for their unreliability and bias toward animals with obvious markings. Feature engineering methods for computer vision have shown success re-identifying animal individuals and removing biases from these analyses; however, these methods require algorithms designed for feature extraction. Deep learning provides a promising alternative for ecologists as it learns these features from large amounts and has shown success for human and animal re-ID. By utilizing deep learning methods for object detection and similarity comparison, ecologists can utilize deep learning methods to autonomously re-identify animal individuals from camera trap data. Such a tool would allow ecologists to automate population estimates.

ACKNOWLEDGEMENTS

This work would not be possible without the interdisciplinary environment offered by the University of Guelph. I thank my supervisor Stefan Kremer for his constant support, guidance and grounding of my sometimes outlandish ideas. I thank Stefan Linquist for his passion and enthusiasm surrounding the study of ecology and ethology and the use of computer science techniques to help improve these disciplines. And I thank Graham Taylor for his incredible wealth of knowledge, his ability to offer meaningful advice and knowledge of possible next steps and how he always makes time for his students no matter how busy his schedule.

Lastly, I thank all the researchers listed within this review for their hard work and outside-the-box thinking to combine the disciplines of computer science and ecology.

AUTHORS' CONTRIBUTIONS

S.S. is a PhD candidate in the department of Computational Science and was the primary author for the work responsible for the written content, structure, papers selected, tables and figures. G.W.T. who is an associate professor of Engineering and a member of the Vector Institute for Artificial Intelligence provided feedback on presentation, content and additional near future techniques for animal re-ID. S.S.L. is an associate professor in Philosophy and Integrative Biology who provided editorial comments on both structure and content. S.C.K. who is a professor of Computer Science specializing in machine learning and bioinformatics, provided editorial comments on both structure and content.

DATA ACCESSIBILITY

All the data that we collected and used in this work are contained in a Microsoft Excel XML file found on the University of Guelphs Portal Web-server at: https://doi.org/10.5683/sp2/n8gkwt.

REFERENCES

Ardovini, A., Cinque, L., & Sangineto, E. (2008). Identifying elephant photos by multi-curve matching. *Pattern Recognition*, *41*(6), 1867–1877. https://doi.org/10.1016/j.patcog.2007.11.010

Arzoumanian, Z., Holmberg, J., & Norman, B. (2005). An astronomical pattern-matching algorithm for computer-aided identification of whale sharks *Rhincodon typus*. *Journal of Applied Ecology*, *42*(6), 999–1011. https://doi.org/10.1111/j.1365-2664.2005.01117.x

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., & Shah, R. (1994). Signature verification using a "siamese" time delay neural network. In G. Tesauro, D. S. Touretzkym & T. K. Leen (Eds.), *Advances in neural information processing systems* (pp. 737–744). Cambridge, UK: MIT Press.

Brust, C.-A., Burghardt, T., Groenenberg, M., Käding, C., Kühl, H. S., Manguette, M. L., & Denzler, J., (2017). Towards automated visual monitoring of individual gorillas in the wild, In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2820–2830). New York, NY: IEEE.

Burghardt, T., Barham, P., Campbell, N., Cuthill, I., Sherley, R., & Leshoro, T. (2007). A fully automated computer vision system for the biometric identification of African penguins *Spheniscus demersus* on Robben Island, In *Abstracts of oral and poster presentations*, *6th international penguin conference*. Birds Tasmania, Hobart

Burton, A. C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J. T., ... Boutin, S. (2015). Wildlife camera trapping: A review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, *52*(3), 675–685. https://doi.org/10.1111/1365-2664.12432

Carter, S. J., Bell, I. P., Miller, J. J., & Gash, P. P. (2014). Automated marine turtle photograph identification using artificial neural networks, with application to green turtles. *Journal of Experimental Marine Biology and Ecology*, *452*, 105–110. https://doi.org/10.1016/j.jembe.2013.12.010

Deb, D., Wiper, S., Russo, A., Gong, S., Shi, Y., Tymoszek, C., & Jain, A. (2018) Face recognition: Primates in the wild. arXiv preprint arXiv:1804.08790.

Foster, R. J., & Harmsen, B. J. (2012). A critique of density estimation from camera-trap data. *The Journal of Wildlife Management*, *76*(2), 224–236. https://doi.org/10.1002/jwmg.275

Freytag, A., Rodner, E., Simon, M., Loos, A., Kühl, H. S., & Denzler, J. (2016). Chimpanzee faces in the wild: Log-euclidean cnns for predicting identities and attributes of primates, In B. Rosenhahn & B.

Andres (Eds.), *German conference on pattern recognition* (pp. 51–63). Berlin, Germany: Springer.

Fukushima, K. (1979). Neural network model for a mechanism of pattern recognition unaffected by shift in position- neocognitron. *Electron. & Commun. Japan*, *62*(10), 11–18.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA: MIT press.

Groth, E. J. (1986). A pattern-matching algorithm for two-dimensional coordinate lists. *The Astronomical Journal*, *91*, 1244–1248. https://doi.org/10.1086/114099

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). New York, NY: IEEE.

Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their applications*, *13*(4), 18–28. https://doi.org/10.1109/5254.708428

Hiby, L., & Lovell, P. (1990). Computer aided matching of natural markings: A prototype system for grey seals. *Report of the International Whaling Commission*, *12*, 57–61.

Hiby, L., Lovell, P., Patil, N., Kumar, N. S., Gopalaswamy, A. M., & Karanth, K. U. (2009) A tiger cannot change its stripes: Using a three-dimensional model to match images of living tigers and tiger skins. *Biology Letters*, *5*(3), 383–386.

Hillman, G., Wursig, B., Gailey, G., Kehtarnavaz, N., Drobyshevsky, A., Araabi, B., … Weller, D. (2003). Computer-assisted photo-identification of individual marine vertebrates: A multi-species system. *Aquatic Mammals*, *29*(1), 117–123. https://doi.org/10.1578/016754203101023960

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*(7), 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527

Holzinger, A. (2016a). Interactive machine learning for health informatics: When do we need the human-in-the-loop? *Brain Informatics*, *3*(2), 119–131. https://doi.org/10.1007/s40708-016-0042-6

Hughes, B., & Burghardt, T. (2017). Automated visual fin identification of individual great white sharks. *International Journal of Computer Vision*, *122*(3), 542–557. https://doi.org/10.1007/s11263-016-0961-y

Humpback whale identification challenge, https://www.kaggle.com/c/whale-categorization-playground, accessed: 2018-05-15

Iandola, F., Moskewicz, M., Karayev, S., Girshick, R., Darrell, T., & Keutzer, K. (2014). Densenet: Implementing efficient convnet descriptor pyramids. arXiv preprint arXiv:1404.1869

Kaggle (2018). Humpback whale identification challenge. Retrieved from https://www.kaggle.com/c/whale-categorization-playground (accessed 15 May 2018).

Kelly, M. J. (2001). Computer-aided photograph matching in studies using individual identification: An example from Serengeti cheetahs. *Journal of Mammalogy*, *82*(2), 440–449. https://doi.org/10.1644/1545-1542(2001)082<0440:CAPMIS>2.0.CO;2

Körschens, M., Barz, B., & Denzler, J. (2018). Towards automatic identification of elephants in the wild. In *AI for wildlife conservation workshop (AIWC).* arXiv preprint, arXiv:1812.04418

Krebs, C. J. (1989). *Ecological methodology*, (2nd ed.). New York, NY: Harper & Row New York, Technical report.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks, In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (pp. 1097–1105). Cambridge, UK: MIT Press.

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

Lisanti, G., Masi, I., Bagdanov, A. D., & Del Bimbo, A. (2015). Person re-identification by iterative re-weighted sparse ranking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(8), 1629–1642. https://doi.org/10.1109/TPAMI.2014.2369055

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21–37). New York, NY: Springer.

Loos, A., & Ernst, A. (2013). An automated chimpanzee identification system using face detection and recognition. *EURASIP Journal on Image and Video Processing*, *2013*(1), 49. https://doi.org/10.1186/1687-5281-2013-49

Martinel, N., Das, A., Micheloni, C., & Roy-Chowdhury, A. K. (2015). Re-identification in the function space of feature warps. *IEEE transactions on pattern analysis and machine intelligence*, *37*(8), 1656–1669. https://doi.org/10.1109/TPAMI.2014.2377748

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*(4), 115–133. https://doi.org/10.1007/BF02478259

Meek, P. D., Vernes, K., & Falzon, G. (2013). On the reliability of expert identification of small- medium sized mammals from camera trap photos. *Wildlife Biology in Practice*, *9*(2), 1–19.

Mizroch, S. A., Beard, J. A., & Lynde, M. (1990). Computer assisted photo-identification of hump- back whales. *Report of the International Whaling Commission*, *12*, 63–70.

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M., Packer, C., & Clune, J. (2017). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. In *Proceedings of the National Academy of Sciences of the United States of America*. Washington: United States National Academy of Sciences.

Nowozin, S. (2014). Optimal decisions from probabilistic models: The intersection-over-union case, In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 548–555). New York, NY: IEEE.

OCorry-Crowe, G. (1998). Identification of individual sperm whales by wavelet transform of the trailing edge of the flukes. *Marine Mammal Science*, *14*(1), 143–145.

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Ravela, S., & Gamble, L. (2004). On recognizing individual salamanders In K.-S. Hong & Z. Zhang (Eds.), *Proceedings of asian conference on computer vision*, (pp. 742–747). Jeju, Korea.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection, In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779–788). New York, NY: IEEE.

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Rodner, E., Simon, M., Brehm, G., Pietsch, S., Wägele, J. W., & Denzler, J. (2015). Fine-grained recognition datasets for biodiversity analysis. arXiv preprint arXiv:1507.00913, 2015.

Rodner, E., Simon, M., Fisher, R. B., & Denzler, J. (2016). Fine-grained recognition in the noisy wild: Sensitivity analysis of convolutional neural networks approaches. arXiv preprint arXiv:1610.06756.

Rowcliffe, J. M., Field, J., Turvey, S. T., & Carbone, C. (2008). Estimating animal density using camera traps without the need for individual recognition. *Journal of Applied Ecology*, *45*(4), 1228–1236. https://doi.org/10.1111/j.1365-2664.2008.01473.x

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back- propagating errors, *Nature*, *323*(6088), 533. https://doi.org/10.1038/323533a0

Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (pp. 3856–3866). Cambridge, UK: MIT Press.

Scheel, D., Chancellor, S., Hing, M., Lawrence, M., Linquist, S., & Godfrey-Smith, P. (2017). A second site occupied by octopus tetricus at high densities, with notes on their ecology and behavior. *Marine and Freshwater Behaviour and Physiology*, *50*(4), 285–291. https://doi.org/10.1080/10236244.2017.1369851

Schneider, S., Taylor, G., & Kremer, S. (2018). "Deep learning object detection methods for ecological camera trap data," Conference on Computer and Robot Vision, to appear.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 815–823). New York, NY: IEEE.

Sherley, R. B., Burghardt, T., Barham, P. J., Campbell, N., & Cuthill, I. C. (2010). Spotting the difference: Towards fully-automated population monitoring of african penguins spheniscus demersus. *Endangered Species Research*, *11*(2), 101–111. https://doi.org/10.3354/esr00267

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

Simpson, R., Page, K. R., & De Roure, D. (2014). Zooniverse: Observing the world's largest citizen science platform. In *Proceedings of the 23rd international conference on world wide web. ACM* (pp. 1049–1054). New York, NY: IEEE.

Souri, Y., & Kasaei, S. (2015). Fast bird part localization for fine-grained categorization. In *Proc. 3rd workshop fine-grained vis. categorization (FGVC3) CVPR, Boston, MA*.

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.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … Rabinovich, A. (2015). Going deeper with convolutions, In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–9). New York, NY: IEEE.

Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1701–1708). New York, NY: IEEE.

Town, C., Marshall, A., & Sethasathien, N. (2013). Manta Matcher: Automated photographic identification of manta rays using keypoint features. *Ecology and Evolution*, *3*(7), 1902–1914. https://doi.org/10.1002/ece3.587

Whitehead, H. (1990). Computer assisted individual identification of sperm whale flukes. *Reports of the International Whaling Commission*, *12*, 71–77.

"iWildcam 2018 camera trap challenge," https://www.kaggle.com/c/iwildcam2018, accessed: 2018-07-11.

Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., & Tian, Q. (2015). Scalable person re-identification: a benchmark, In *Proceedings of the IEEE international conference on computer vision* (pp. 1116–1124). New York, NY: IEEE.

Zhou, Z.-H. (2017). A brief introduction to weakly supervised learning. *National Science Review*, *5*(1), 44–53.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Schneider S, Taylor GW, Linquist S, Kremer SC. Past, present and future approaches using computer vision for animal re-identification from camera trap data. *Methods Ecol Evol*. 2019;00:1–10. https://doi.org/10.1111/2041-210X.13133