Title: Deep learning for ecological image analysis: an example using Optical Character Recognition.

Ben Weinstein1

1Department of Fisheries and Wildlife, Marine Mammal Institute, Oregon State University, 2030 Marine Science Drive, Newport, OR 97365, USA

Abstract

Introduction

To monitor biodiversity, ecologists and conservation managers need high quality temporal and spatial data on animal presence, movement, and behavior. The high costs, complex logistics, and significant expertise of human observation studies limit the ability to gather sufficient data collection for analysis. Ecologists are increasingly turning towards greater automation to facilitate biodiversity monitoring. The combination of decreased costs of image capture and growing image analysis tools makes automated biodiversity monitoring closer than ever before.

A major obstacle to the growth of image-based ecological analysis is the efficiency of scoring and extracting information from images. Annotating images with metadata, such as time, date and location, and species identity requires significant time investment. The emerging field of computer vision can decrease the time for image annotation, increase consistency among annotators, as well as engage less experienced observers in biodiversity monitoring. Computer vision is a field of image-based computer science that uses image pixels to mimic human perception based on image characteristics, shape and sequence.

While computer vision has made incredible strides in a variety of fields, its growth in ecology has been slowed by a lack of access to high level algorithms. Computer vision publications are often short, terse, and filled with jargon. Given the potential lucrative opportunities for new algorithms, source code is rarely made available. While there have tremendous advances in artificial intelligence capability, ecologists have largely been left out of this revolution. The challenge lies in adapting existing tools to meet the time and budget limitations of ecologists.

The recent unveiling of the Google Cloud Machine learning platform could be a quantum leap in access to machine learning tools. Released in trial form in December 2016, google cloudML gives users access to a pre-trained image classification model and a robust web service to classify images using convolutional neural networks (CNNs). CNNs uses image features to classify objects into known categories (LeCun *et al.* 2015). Described colloquially as ‘deep learning’ CNNs do x,y,z by x,and y. For a full treatment of the technical details of CNN see X,Y,Z.

While there have been applications of deep learning for ecology (Chen *et al.* 2014; Gomez *et al.* 2016), it is the availability of highly sophisticated tools for average users which holds revolutionary potential. Rather than building a model from scratch, user can ‘retrain’ pre-built models by teaching it new image classes, while maintaining the underlying architecture (known as ‘transfer learning’). This greatly reduces the time and expertise needed to implement image analysis solutions. The previous bottleneck was in each researcher developing his own solution independently. While the number of classes for state of the art solution is increasing, off-the-shelf models are unlikely to contain all labels that might benefit ecologists. For example, Inception may be able to differentiate a bear from a bird, but it may not have been taught the taxonomic resolution needed to contribute to ecological inventories. Through transfer learning…we have access to new tools.

A sample problem.

In 2013, my colleagues and I began placing time-lapse cameras in the montane cloud forests to monitor hummingbird-plant interactions. Our goal was to replace traditional human observation with image-based tools to increase the extent of spatial and temporal sampling. Ecological networks, and plant-pollinator networks in particular, are chronically undersampled, due to the rarity of many interacting partners, and the temporal pattern of plant phenology. By using cameras which turned on at 6am, turned off at 6pm, and took 1 photo per second, we can refine our understanding of the ecology of tropical communities and the co-evolution of plant-pollinator interactions. At our site at the Maquipucuna Ecolodge (), we currently have twenty cameras generating nearly 100 GB of video data per month. The videos are screened using computer vision software to find important candidate motion events, which reduces the total number of reviewed frames by >90%. These candidate frames are then scored by a human to record species identity, the time, date, and behavior of individuals. While the camera ID is encoded into the video, the timestamp of each frame is manually extracted by humans reviewing frames. Manually review of frames still is a large bottleneck in generating datasets, and limits the scalability of our approach to multiple geographic sites.



Figure 1. A sample image of a female violet-tailed sylph (aglaocercus colestris) visiting a Guzmania jaramolloi. At the bottom of each image is a camera ID, date and time stamp.

Methods

Our aim is test the performance of Google CloudML in extracting timestamp metadata from a set of 500 scored photos (Fig 1). This represents the first step in moving towards automated image analysis, and will provide a simple test case for the effectiveness and costs of this new service. Machine learning for optical character recognition (OCR) is a well-tested field of research and was a good choice for a first test of this new service.

Comparison to existing solutions

As a standard of comparison, I compared the performance of a re-trained Inception model on Google CloudML with tesseract, a highly regarded open-source tool for OCR. Tesseract is itself a machine learning tool developed at google, and is designed for transcribing text from visual images.

Results

Discussion

Acknowledgements.

The author acknowledges no conflict of interest and has no connection to Google Inc or its parent company.

Works Cited

Chen, G., Han, T.X., He, Z., Kays, R. & Forrester, T. (2014). Deep convolutional neural network based species recognition for wild animal monitoring. *2014 IEEE International Conference on Image Processing (ICIP)*, 858–862.

Gomez, A., Diez, G., Salazar, A. & Diaz, A. (2016). Animal Identification in Low Quality Camera-Trap Images Using Very Deep Convolutional Neural Networks and Confidence Thresholds. *Advances in Visual Computing: 12th International Symposium, ISVC 2016, Las Vegas, NV, USA, December 12-14, 2016, Proceedings, Part I*, pp. 747–756. Springer International Publishing.

LeCun, Y., Bengio, Y., Hinton, G., Y., L., Y., B. & G., H. (2015). Deep learning. *Nature*, **521**, 436–444.