International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Categorization of Species using Machine Learning

Veda R Babu[1], Priyanka M[2], Saurabh Kumar Basak[3], Manjunath S[4]

[1,2,3]Student, Department of Information Science & Engineering, Global Academy of Technology, Bengaluru, India
[4]Assoc.Professor, Department of Information Science & Engineering, Global Academy of Technology, Bengaluru, India

---

Abstract - *Efficient and reliable monitoring of wild animals in their natural habitats is essential to inform conservation and management decisions. Automatic covert cameras or "camera traps" are being an increasingly popular tool for wildlife monitoring due to their effectiveness and reliability in collecting data of wildlife unobtrusively, continuously and in large volume. However, processing such a large volume of images and videos captured from camera traps manually is extremely expensive, time-consuming and also monotonous. This presents a major obstacle to scientists and ecologists to monitor wildlife in an open environment. Leveraging on recent advances in deep learning techniques in computer vision, we propose in this paper a framework to build automated animal recognition in the wild, aiming at an automated wildlife monitoring system. In particular, we use a single-labelled dataset from Wildlife Spotter project, done by citizen scientists, and the state-of-the-art deep convolutional neural network architectures, to train a computational system capable of filtering animal images and identifying species automatically. Our experimental results achieved an accuracy at 96.6% for the task of detecting images containing animal.. This, in turn, can therefore speed up research findings, construct more efficient citizen science based monitoring systems and subsequent management decisions, having the potential to make significant impacts to the world of ecology and trap camera images analysis.*

Key Words: deep learning, convolutional neural networks, large scale image classification, animal recognition, wildlife monitoring, citizen science

1. INTRODUCTION

Observing wild animals in their natural environments is a central task in ecology. The fast growth of human population and the endless pursuit of economic development are making over-exploitation of natural resources, causing rapid, novel and substantial changes to Earth's ecosystems. An increasing area of land surface has been transformed by human action, altering wildlife population, habitat and behaviour. More seriously,. Many wild species on Earth have been driven to extinction, and many species are introduced into new areas where they can disrupt both natural and human systems [1]. Monitoring wild animals, therefore, is essential as it provides researchers evidences to inform conservation and management decisions to maintain diverse, balanced and sustainable ecosystems in the face of those changes. Various modern technologies have been developed for wild animal monitoring, including radio tracking [2], wireless sensor network tracking [3], satellite and global positioning system (GPS) tracking [4], [5], and monitoring by motion sensitive camera traps [6]. Motion-triggered remote cameras or "camera traps" are an increasingly popular tool for wildlife monitoring, due to their novel features equipped, wider commercial availability, and the ease of deployment and operation.

1.1 Wildlife Detection and Wildlife identification

Since the Wildlife Spotter dataset includes both animal and non-animal images, we divide the wild animal identifying automation into two subsequent tasks: (1) Wildlife detection, which is actually a binary classifier capable of classifying input images into two classes: "animal" or "no animal" based on the prediction of animal presence in images; and (2) Wildlife identification, a multiclass classifier to label each input image with animal presence by a specified species. The core of each task is essentially a deep CNN-based classifier, trained from prepared datasets manually labeled by volunteers. Several selected deep CNN architectures are employed to the framework for comparisons.

2. RELATED WORK

In this section we first briefly describe the CNN and its application to image classification. We then summarize various CNN architectures that have demonstrated the state-of-the-art performance in recent ImageNet Challenges. Finally we discuss existing approaches to a particular problem: animal classification in natural scenes from camera trap images.

A. Convolutional Neural Networks for Image

Classification Visual recognition is a relatively trivial task for human, but still challenging for automated image recognition systems due to complicated and varied properties of images. Each object of interest can alter an infinite number of different images, generated by variations in position, scale, view, background, or illumination. Challenges become more serious in real-world problems such as wild animal classification from automatic trap cameras, where most captured images are in imperfect quality as described previously in Section I. Therefore, for the task of image classifying automation, it is important to build models that

are capable of being invariant to certain transformations of the inputs, while keeping sensitivity with inter-class objects. Firstly proposed by LeCun et al., CNNs have been showing great practical performance and been widely used in machine learning in the past recent years, especially in the areas of image classification, speech recognition, and natural language processing. These models have made the state-of-the-art results that even outperformed human in image recognition task, due to recent improvements in neural networks, namely deep CNNs, and computing power, especially the successful implementations of parallel computing on graphical processing units (GPUs), and heterogeneous distributed systems for learning deep models in large scale such as Tensor Flow.

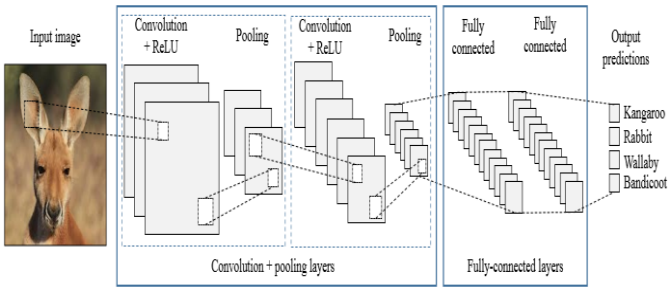

Figure 1: Illustration of a typical convolutional neural network architecture setup.

CNNs are basically neural network-based learning models specifically designed to take advance the spatial structure of input images, which are usually in 3-dimensional volume: width, height, and depth (the number of color channels). As illustrated in Figure 1, a CNN is essentially a sequence of layers which can be divided into groups each comprising of convolutional layer plus non-linear activation function, usually the Rectifier Linear Unit (ReLU), and pooling layer, mostly max pooling; ended by several fully-connected layers where the last one is the output layer with predictions. In the standard neural networks, each neuron is fully connected to all neurons in the previous layer and the neurons in each layer are completely independent. When applied to high dimensional data such as natural images, the total number of parameters can reach millions, leading to serious overfitting problem and impractical to be trained. In CNNs, by contrast, each neuron is connected only to a small region of the preceding layer, forming local connectivity. The convolution layer computes the outputs of its neurons connected to local regions in the previous layer, the spatial extent of this connection is specified by a filter size. In addition, another important property of CNNs, namely parameter sharing, dramatically reduces the number of parameters and so does computing complexity. Thus, compared to regular neural networks with similar size of layers, CNNs have much fewer connections and parameters, making them easier to train while their performance is slightly degraded. These three main characteristics – spatial structure, local connectivity and parameter sharing – allow CNNs converting input image into layers of abstraction; the lower layers present detail features of images such as edges, curves and corners, while the higher layers exhibit more abstract features of object.

B. Wildlife Classification

Monitoring wildlife through camera traps is an effective and reliable method in natural observation as it can collect a large volume of visual data naturally and inexpensively. The wildlife data, which can be fully automatic captured and collected from camera traps, however, is a burden for biologists to analyze to detect whether there exist animal in each image, or identify which species the objects belong to. But in this paper we overcome this.

Table -1: The most common and successful CNN architectures for image classification.

Model	Trainable layers	Main specifications
AlexNet	8	5 convolutional layers and 3 fully-connected layers.
VGG-16	16	13convolutional layers with 3x3 filters, and 3 fully-connected layers.
GoogLeNet	22	Developed an Inception Module that dramatically reduces the number of parameters while achieving high accuracy.
ResNet-50	50	A deep residual learning framework, skip connections and batch normalization. Much deeper than VGG-16 (50 compared to 16) but having lower complexity and higher performance.

Dramatically reduce a large amount of human resource and quickly provide research findings.

C. Citizen Science

Citizen science plays an important role in many research areas, particularly in ecology and environmental sciences. A citizen scientist is a volunteer who contributes to science by collecting and/or processing data as part of a scientific enquiry. Significant development in digital technique, especially the Internet and mobile computing, is one of key factors responsible for the great explosion of recent citizen science projects. Volunteers are now able to, remotely, take part to a project by using designated applications on their mobile phones or computers to collect data or process introduced data, and then enter them online into centralized, relational databases.

D. Wildlife Spotter Project

Wildlife Spotter is an online citizen science project undertaken by several Australian organizations and universities, taking crowd-sourcing approach to science by asking volunteers to help scientists classifying animals from millions of images collected from automatic trap cameras. These cameras, located in the nation wide: tropical rainforests, dry rangelands, and around the cities, set up to automatically snap color, high definition images day and night. To date, over 3 million images were completed. To deal with the enormous volume of images, the project invites volunteers playing as "citizen scientists" to join image analyzing. The main goal of the project is, through analyzing captured images, to assist researchers study Australian wildlife populations, behaviors and habitats to save threatened species and preserve balanced, diverse, and sustainable ecosystems.

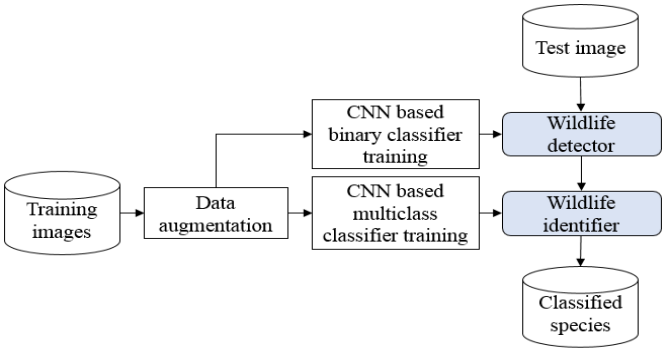

Figure 2: Key steps in the proposed framework for automated wild animal identification.

As depicted in Figure 4, our proposed recognition system consists of two CNN-based image classification models corresponding to the two addressed tasks. First a CNN-based model is designed to train a binary classifier, namely Wildlife detector; then another CNN-based model is created to train a multi-class classifier, namely Wildlife identifier.

1) CNN Architectures: Three CNN architectures with different depths are employed to our proposed framework, namely Lite AlexNet, VGG-16, and ResNet-50. We use a simplified version of AlexNet and call it Lite AlexNet, with less hidden layers and feature maps at each layer. In particular, the Lite AlexNet comprises of three 2-D convolutional layers with ReLU activations and MaxPooling, followed by two fully-connected layers: one with ReLU nonlinear activation plus Dropout for reducing overfitting, the output layer with sigmoid activation for binary classification in detecting task and softmax activation for multiclass classification in recognizing task. All convolutional layers have small filter size of 3_3, while all max-pooling layers have window size of 2_2 pixels. VGG-16 and ResNet-50 are two representatives of the state-of-the-art CNN architectures that not only showed excellent performances on the ILSVRC, but also generalized well to other datasets . The input to all CNN architectures is a fixed-size 224_224 image in RGB color.

2) Image Processing: The Wildlife Spotter dataset contains high resolution images of 1920 _ 1080 and 2048 _ 1536 pixels, while the input of CNN models must be in fixed dimension. Therefore, in our experiments all original images were downscaled to 224_224 pixels for training. In this process was carried out by firstly rescaling the shorter side of image to the fixed length, then applying center cropping the image with the same length. In this work, for simplicity, we rescale both image width and height simultaneously, which may result in image distortion. Pixel intensities are normalized into the range of [0;1]. Data quality, which can be enhanced by augmentation techniques, is a key to data-driven machine learning models; however in this work a few data augmentation processes, shearing and zooming, were applied to training images.

3) Training Deep Networks: Our implementation is in Keras , a high-level neural networks API, with TensorFlow backend .Adam optimizer, the first-order gradient-based optimization based on adaptive estimates of lower-order moments, was employed for training all networks [38]. A small minibatch size of 16 was set to all experiments. We train our models on four NVIDIA Titan X GPUs, each network takes three to five days to finish training. For each task we train CNN models in two scenarios: imbalanced and balanced datasets. We compute classification accuracy for both cases. In case of dataset imbalance, Fmeasure is employed in addition to accuracy, to test the robustness of the proposed system. Accuracy on the validation set is used as performance metric. To evaluate transfer learning, we carry out training Task 2 – Wildlife identification, in two scenarios: training model from scratch and fine-tuning with available ImageNet pre-trained models. Fine-tuning techniques leverage a network pre-trained on a large dataset, in this case is the ImageNet, based on the assumption that such network would have already learned useful features for most computer vision problems, thus could reach better accuracy

than a model trained on a smaller dataset. Our fine-tuning process follows three steps: firstly the convolutional blocks are instantiated, then the model will be trained once on new training and validation data, finally the fully-connected model with fewer specified classes will be trained on top of the stored features IRJET sample template format ,

3. CONCLUSION

This work shows that Deep Learning techniques can be used towards problems in many areas, and they can help extract large amounts of information out of large amounts of data. This is one use case where Deep Learning is more than useful and can help experts like biologists and ecologists in their work towards studying and conserving wildlife.

REFERENCES

[1] P. M. Vitousek, H. A. Mooney, J. Lubchenco, and J. M. Melillo, "Human domination of Earth's ecosystems," Science, vol. 277, no. 5325, pp. 494– 499, 1997.
[2] G. C. White and R. A. Garrott, Analysis of wildlife radio-tracking data. Elsevier, 2012.
[3] R. Szewczyk, A. Mainwaring, J. Polastre, J. Anderson, and D. Culler, "An analysis of a large scale habitat monitoring application," in Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems, 2004, pp. 214–226.
[4] B. J. Godley, J. Blumenthal, A. Broderick, M. Coyne, M. Godfrey, L. Hawkes, and M. Witt, "Satellite tracking of sea turtles: Where have we been and where do we go next?" Endangered Species Research, vol. 4, no. 1-2, pp. 3–22, 2008.