Automatically Identifying Animals Using Deep Learning

Parinita Badre
Dept. of Computer Engineering
Rajiv Gandhi Institute of Technology
Mumbai, India
e-mail: parinita192@gmail.com

Siddhant Bandiwadekar
Dept. of Computer Engineering
Rajiv Gandhi Institute of Technology
Mumbai, India
e-mail: siddhant.bandiwadekar@gmail.com

Prachi Chandanshive
Dept. of Computer Engineering
Rajiv Gandhi Institute of Technology
Mumbai, India
e-mail: prachic1210@gmail.com

Aakanksha Chaudhari
Dept. of Computer Engineering
Rajiv Gandhi Institute of Technology
Mumbai, India
e-mail: aakanksha.chaudhari@gmail.com

Mrs. Sonali Jadhav
Dept. of Computer Engineering
Rajiv Gandhi Institute of Technology
Mumbai, India
e-mail: khade.sonali@gmail.com

***Abstract*—**Having accurate, detailed, and upto-date information about wildlife location and behavior across broad geographic areas would revolutionize our ability to study, conserve, and manage species and ecosystems. Currently, such data are mostly gathered manually at great expense, and thus are sparsely and infrequently collected. Here we investigate the ability to automatically collect such data, which could transform many fields of biology, ecology, and zoology into big data sciences. In areas like an airportor the agricultural areas placed near the forest many animals destroy the crops or even attack on people therefore there is a need of system which detects the animal presence and gives warning about that in the view of safety purpose. In this project, we demonstrate that such data can be automatically extracted by deep neural networks (deep learning), which is a cutting-edge type of artificial intelligence.
Thus,the aim is to train neural networks that automatically identifiesanimals.

Keywords-*Deep learning; Back-Propagation algorithm;Inception model*

*****

I. Introduction

Monitoring biodiversity especially the effects of climate and land use change on wild populations is a critical challenge for our society. Having an updated knowledge about wildlife behaviour would impact our ability to study and manage species and our ecosystem. Researches regarding animals in image processing have been an important field to numerous applications. Many algorithms and methods have been developed by human being in order to have a better understanding on animal behaviour. Identifying animal attributes, analyzing their behavior in the pictures remains an expensive time consuming manual task performed by various researchers. Thus, we demonstrate that such detection of animal can be done by deep convolution neural network.

II. Related Work

Wherever Authorsin [8], [5] designed a system, which uses web cameras which are to be placed in the detecting areas from where the animal may cross their boundary. The videos are sent to the processing unit and then uses image mining algorithm, which identifies the change in set reference background. If there is a change in the newly acquired image,then authors applied content-based retrieval algorithm (CBIR) to identify the animal. The proposed method in [8] based on CBIR algorithm suffers from many issues like unsatisfactory querying performance-CBIR systems uses distance functions to calculate the dissimilarity between a search image and database images resulting in low-quality recovery results. This approach is veryslow and response times in the range of minutes maytake place if size of database is too large. To find the accurate location of fishes in the marine, researchers [9] aimed a technique using LIDAR (light detection and ranging). Some of the above-specified methods have been discussed in [10] also.

ISSN: 2321-8169
194 - 197

Researchers in [11] tried to discover an animal's presence in the scene (image) affecting the power spectrum of the image. This method of animal detection was also considered not appropriatesince quicker results with this approach would involve the massive amountof image processing in a short period.

III. METHODOLOGY

Our classifier comprises of two stages, training and testing. In the training stage, a set of images are provided as visual examples. In the testing stage, a newly captured image i.e. the test image is given as input to the classifier. With the help of the knowledge gained from training, the test image is accordingly classified into the most favorable class.

A. Receiving the input image

In the proposed system, an image is captured using the camera connected to the system. This test image is fed as input which is then converted into a binary pattern. A set of previously labelled images are present in the dataset whose features are matched with those of the test image, in order to determine the species of the animal present.

B. Feature Extraction

The test image which is received as input can be transformed into a reduced set of features. The selected features may contain the significant information from the input data, due to which the desired task can be performed by using this reduced amount of data instead of the initial unaltered data. Human crafted features are a form of fixed features directly extracted from images. Divergent from these, deep neural networks recognize features from images, and determine multiple levels of representation, with higher-level features depicting more abstract characteristics of the data.

C. Classifying the species present in an image

For the task of classification of species, the corresponding output layer generates the probabilities of the animal detected in the image belonging to one of the possible classes. Even though providing such a result would save a huge amount of human effort that will be needed in recognizing the correct species, the testing of this hypothesis will require human knowledge.

IV. IMPLEMENTED SYSTEM

In machine learning, a convolutional neural network (CNN, or ConvNet) is a class of deep, feed-forward artificial neural networks that has successfully been applied to analyzing visual imagery. [1] CNN compares any image piece by piece and the pieces that it looks for in an image while detection are called as features. CNN gets trained by finding approximate feature matches in the same position in two different images. Every neuron in CNN will be connected to small region of neurons below it, this would allow handling fewer amounts of weights and number of neurons required will also be less

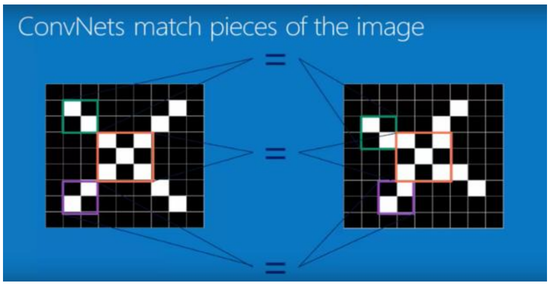

Figure 1. Principle of Convolutional Neural Network

A. Convolution Layer

The Convolution layer is the core building block of a Convolutional Network that does most of the computational heavy lifting. [2] It preserves spatial relationship between pixels thereby extracting and learning features out of them. It captures the local dependencies in the original image. The image is represented as a matrix and a filter, which is also a matrix is used to obtain the convolved feature map or activation map by sliding the feature matrix over the image matrix. We can perform numerous operations just by changing values in the filter matrix.

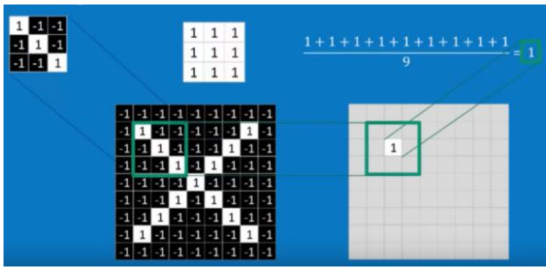

Figure 2. Convolving the filter with image matrix to obtain final value in feature map

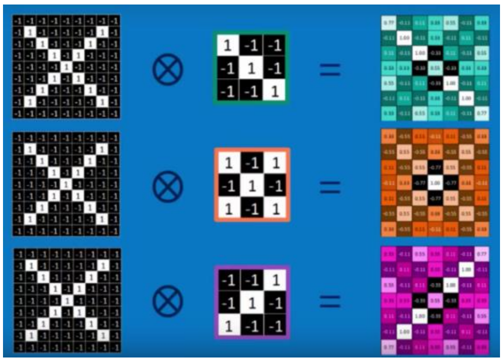

Figure 3. Convolving each filter with all image matrices to obtain various feature maps

B. Rectified Linear Unit Layer

It is called as rectified linear unit(ReLU), which is an activation function that activates a node if an input is above a certain quantity(threshold); while if the input is 0 then the output will be 0. But, if the input is above certain threshold it has a linear relationship with the dependent variable.

Figure 4. ReLU Operation to obtain rectified feature maps

It replaces all the negative values in the feature map by zero and generates a rectified feature map as shown in Fig. 4. ReLU introduces non-linearity in the ConvNet since most of the image data are non-linear in nature in the real world.

C. Pooling Layer

In this layer, we reduce the dimensionality of the feature map to get smaller or shrunk maps that would reduce the parameters and computations. Pooling can be Max, Average or Sum Pooling from the rectified and downsized feature map. Number of filters in convolution layer is same as the number of output maps from pooling as shown in Fig. 3. It also makes the network invariant to small transformations, distortions and translations in the input image.

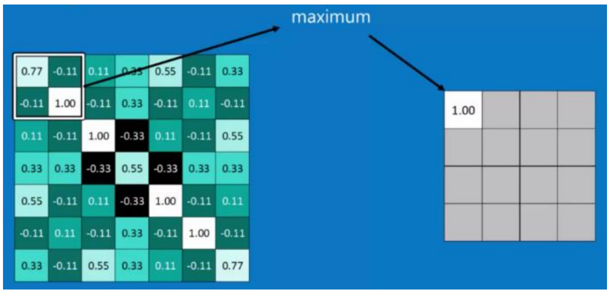

Figure 5. Max pooling technique

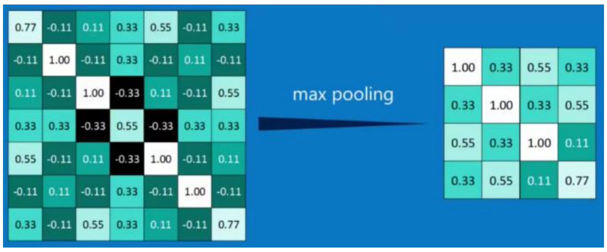

Figure 6. Max pooling applied to rectified feature maps

D. Fully Connected Layers

This is the final layer where the actual classification occurs where we take our downsized or shrinked images obtained after processing through convolution, ReLU and pooling layer and put them into single list or a vector.

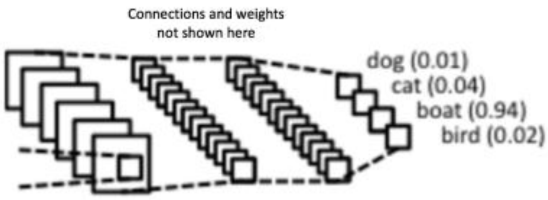

Figure 7. Fully Connected Layer

It is a traditional Multi-layer perceptron uses a softmax activation function. Convolution and pooling layers generate high-level features. The purpose of the fully connected layers is to use these features to classify the data into various classes based on the dataset.

E. Training of CNN Using Back-Propagation

Initialize all filters and parameters with random variables. Take input images for training, go through the forward propagation and find output probability for each class.

Calculate total error by the formula:

$$\text{Total Error} = \sum \tfrac{1}{2} (\text{target probability} - \text{output probability})^2 \quad (1)$$

Calculate gradients of the error with respect to the weights and use gradient descent to update the filter values and parameters to minimize the output error.

Repeat above steps for all images in the training set.

F. Inception Model

Very deep convolutional networks have been central to the largest advances in image recognition performance in recent years. One example is the Inception architecture that has been shown to achieve very good performance at relatively low computational cost.

The Inception deep convolutional architecture was introduced as GoogLeNet in (Szegedy et al. 2015a), here named Inception-v1. Later the Inception architecture was refined in various ways, first by the introduction of batch normalization (Ioffe and Szegedy 2015) (Inception-v2). Later by additional factorization ideas in the third iteration which is referred to as Inception-v3.

We can train a model from scratch to its best performance on a desktop with 8 NVIDIA Tesla K40s in about 2 weeks. In order to make research progress faster, we are additionally supplying a new version of a pre-trained Inception-v3 model that is ready to be fine-tuned or adapted to a new task.

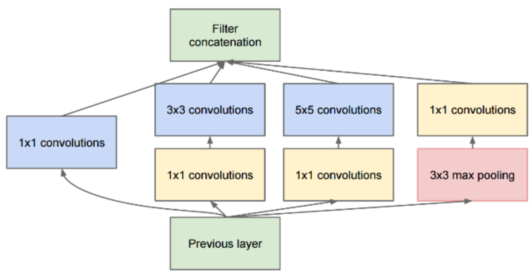

Figure 8. Inception Model Architecture

V. Simulation Results

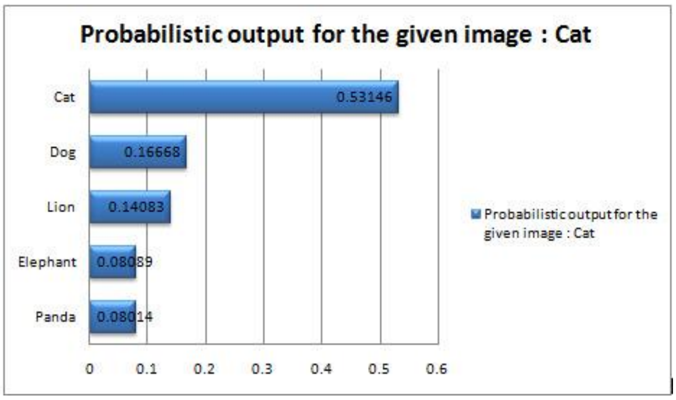

Figure 9. Probabilistic output for a sample image

The obtained results after experimentation have been shown in the figure 8. The dataset through which features gotextracted helped in creating various classes. For a test image to be correctly classified the output showed variousclasses of animals. Each class showed the probabilistic value of whether the image of the animal belongs to thatparticular class of animal. The class with the maximum probabilistic value thus showed that the features of the animalresemble to that specific class on large scale and hence the image will be classified to the class with large probabilisticvalue.

TABLE I. Accuracy after each iteration

No. of Iterations	Train Accuracy (%)	Validation Accuracy (%)	Final Test Accuracy (%)
1	53.0	37.0	59.0
2	80.0	64.0	69.0
3	53.0	37.0	59.0
4	82.0	79.0	82.3
5	92.8	82.0	83.9
6	97.0	83.0	82.3
7	66.8	91.8	89.5
8	75.0	72.0	86.2
9	85.0	65.0	93.0
10	97.0	90.0	95.2

VI. Conclusion

In this report, we first briefly explained our motivation of this project and showed some background materials. Then, we precisely illustrated our task of demonstrating that wild animals' detection can be done using deep convolutional neural networks. Features learned by Deep Neural Networks are hierarchical. With more hidden layers, the learned features become more high level and specific. In this paper, we tested the ability of state-of-the-art computer vision methodscalled deep neural networks to automatically identify animals. Automating animal identification can thus dramatically reduce the cost to extract informative and actionable information from wild habitats, potentially revolutionizing studies of animal behavior, ecosystem dynamics, and wildlife conservation.

VII. Future Scope

In the system proposed above, we have used a fixed number of training images of each species. Increasing the number of images for each species will result in increase in the prediction of the class to which the animal in the captured image belongs. Furthermore, a newly captured image can be added to the dataset and in the next iteration, the system will consider this image for training along with the previous images. Also, the number of features that can be extracted from a single image can be increased in order to make the result unambiguous. The challenge in some cases is how to train a model without access to a large number of labeled images. Here, transfer learning can help, wherein a deep neural network is trained on a huge, labeled dataset at first and then the knowledge learned is remodeled to classify a different dataset with fewer labeled images. Given the rapid pace of progress in this field, all of the performance measures should be considered just a preview of what is possible and it is likely that they will be substantially improved on each year for the foreseeable future. Therefore in future we want to extract regions for addressing the location of objects and extract other features as well to get better results.

References

[1] Leila Mansourian, Muhamad Taufik Abdullah, Lilli Nurliyana Abdullah and Azreen Azman ,Evaluating classification strategies in Bag of SIFT Feature method for Animal Recognition, Research Journal of Applied Sciences, Engineering and Technology, pg 1266-1272, August 2015

[2] Bang Liu, Yan Liu and Kai Zhou ,Image Classification for Dogs and Cats

[3] Kaggle DogVCat Competation: http://www.kaggle.com/c/dogs-vs-cats

[4] Griffin, G., Holub, A., Perona, P.: Caltech-256 object category dataset (2007)

[5] Mohammed Sadegh Norouzzadeh, Anh Nguyen, Margaret Kosmala, Ali Swanson, MeredithPalmer, CraigPacker, and JeffClune1, Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning , ArXiv , 1703.05830v5, November 2017

[6] Chattopadhyay P, Vedantam R, Selvaraju RR, Batra D, Parikh D (2017) Counting everyday objects in everyday scenes, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1604.03505v3 [cs.CV], 9 May 2017.

[7] Krizhevsky, A., Sutskever, I., & Hinton, G. Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems 25 , pg. 1106-1114, 2012.

[8] Shaikh S, Jadhav M, Nehe N and Verma U, 2015, Automatic animal detection and warning system, International Journal of Advance Foundation and Research in Computer, vol. 2, pg: 405-410, 2015.

[9] Mitra V, Wang C and Edwards G, "Neural network for LIDAR detection of fish", Proceedings of the International Joint Conference on Neural Networks, pg. 1001-1006, 2003.

[10] Sharma S and Shah D, Real-Time automatic obstacle detection and alert system for driver assistance on Indian roads, Indonesian Journal of Electrical Engineering and Computer Science, vol.1, pg:635-646, 2016

[11] Ragheb H, Velastin S, Remagnino P and Ellis T, "Human Action Recognition using Robust Power Spectrum Features", Proceedings of the 15th International Conference on Image Processing, (ICIPO'08), San Diego, CA, pg.753-756, 2008