International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 Impact Factor: 8.176 ISSN: 2582-3930

IDENTIFICATION OF BIRD SPECIES USING CONVOLUTION NUERAL NETWORKS- vGG16.

Satvik Srivastav
Department of Computer Science and Engineering
SRM Institute of Science and Technology
Chennai, India
ss5465@srmist.edu.in.

Dr. N. Snehalatha
Department of Computer Science andEngineering
SRM Institute of Science and Technology
Chennai, India
@srmist.edu.in

***Abstract*— There are more than 9000 different kinds of birds in our world. Some bird species are hard to find, and if they are found, it's hard to tell what will happen next. To solve this problem, we have an easy and effective way to tell what kind of bird it is based on how it looks. Also, it's easier to understand how people can recognize birds from their pictures than from their sounds. So, we've used something called Convolutional Neural Networks (CNN). CNNs are a strong mix of machine learning techniques that have been shown to work well in image processing. In this paper, a CNN system for classifying bird species is shown. It is trained and tested on the BIRDS 500 SPECIES- IMAGE CLASSIFICATION dataset. By making this dataset and using the algorithm for comparing similarity, this system has been shown to work well in the real world. Using this method, anyone can easily figure out the name of the bird they want to know.**

Keywords— Bird species, Machine Learning, Convolutional Neural Networks

I. INTRODUCTION

Today, bird behavior and population patterns have become a major concern. Due to their fast response to ecological changes, birds assist us in identifying many forms of life on Earth. However, organizing and collecting information about bird species needs a great deal of human effort and is an incredibly expensive process. In such a circumstance, a strong framework is necessary that provides large-scale data preparation concerning birds and serves as a significant tool for scientists, legislative offices, etc. In this approach, bird species identification evidence plays an important role in determining which category a particular image of birds belongs to. Identifying bird species involves predicting which group each bird species belongs to given an image.

In this study, rather than examining the problem of recognizing an excessive number of distinct categories, the issue of recognizing an excessive number of classes within one category, that of birds, is explored. An additional difficulty in classifying birds is the high degree of resemblance between different groups. However, birds are not hard objects; they can bend and twist in different ways, so there is a great deal of variance even within groups. The majority of bird classifications done before have dealt with either a small number of classes or have relied on vocalizations.

II. LITERATURE SURVEY

Till date, many researchers have proposed several methods to solve this problem of bird species prediction. These methods are discussed below in this section.

The CNN method and deep residual neural networks were presented by John Martinsson et al. (2017)[1] to identify a picture in two different ways, namely, based on the extraction of features and the categorization of signals. They carried out an experimental investigation on datasets that contained a variety of photographs. However, their research did not take into account the background species. Larger volumes of training data are necessary in order to identify the background species, which may not be available.

The novel approach proposed by Andreia Marini, Jacques Facon, et al. (2013)[5] is based on color features extracted from unconstrained images and uses a color segmentation algorithm to try to remove distracting background elements and to delimit candidate regions where the bird may be present in the image. Processing at an aggregate level was used to reduce the histogram intervals to a fixed number of bins. Results from experiments conducted using the CUB-200 dataset show that this approach improves accuracy.

It was proposed by Juha Niemi, Juha T. Tanttu, and colleagues (2018) [2] to use convolutional neural networks that have been trained using deep learning techniques. Images can be categorized with the help of these networks. It also presented a technique for data augmentation, in which pictures would be changed and flipped to the desired hue. Using both the radar's characteristics and the image classifier's predictions, a definitive identification may be made.

A piece of software was developed by Madhuri A.Tayal, Atharva Magrulkar, and their colleagues (2018)[4] to assist in the process of identifying birds. This program needs nothing more than an image of the bird as input in order to correctly identify the species of the bird. On the other hand, it gives back the right identification as output. In order to accomplish the detection, we make use of MATLAB in conjunction with transfer learning.

The research of Marcelo T.Lopes, Lucas L.Gioppo, et al. (2011)[6] entered on the automatic identification of bird species from their audio recorded song. Here, the scientists applied signal processing and machine learning techniques to the avian species identification problem utilizing the MARSYAS feature set. Included were the results of a battery of studies run on a database containing the songs of seventy-five different species of birds; just 12 of these proved problematic in terms of their overall performance.

Peter Jancovic and Munevver Kokuer et al. (2012)[7] studied acoustic modeling for bird species recognition from audio field recordings. Hybrid deep neural network hidden Markov model (DNN-HMM). The models identified, detected, and recognized various bird species vocalizing in a recording. This paper scored 98.7% identification and 97.3% recognition accuracy.

Mario Lasseck et al. (2013)[8] showed how to use deep convolutional neural networks and data augmentation to identify bird species based on their sounds. The author of this paper used the Xeno-Canto collection of recordings of birdsongs.

Li Jian, Zhang Lei et al. (2014)[3] proposed an effective picture feature-based automatic bird species identification. Using standard picture database and similarity algorithm.

III. DATASET

Dataset Source: Kaggle

Structured/Unstructured data: Structured Data in CSVformat.

Dataset description:

This study made use of the BIRDS 500 SPECIES- IMAGE CLASSIFICATION dataset, which included 500 species of birds, 11,788 pictures, annotated viewable parts, binary attributes, and bounding boxes surrounding the birds. Among these 500 items, we will be evaluating the following:

IV. METHODOLOGY

The algorithm of a convolutional neural network is a multilayer perceptron, which is a specially designed network for the recognition of two-dimensional picture data. The structure is composed of four layers: the input, the convolution, the sample, and the output. It is possible for a deep network's convolution layer and sample layer to have many instances of the same type of computation. In contrast to the Boltzmann machine, convolutional neural network (CNN) techniques only require each neuron to sense the local portion of the image rather than the entire image. Each neuron contributes the same convolution kernels to the deconvolution image, and all of the neuron's parameters are identical.

In order to extract features and minimize the size of the training parameters, the central era of CNN is the local receptive field, sharing of weights, subsampling by using time or space. The CNN algorithm's strength lies in its ability to intuitively learn from the training data rather than relying on explicit feature extraction. Due to sharing the same neuron weights on the feature mapping's surface, the network can train in parallel and with less overhead. Using a subsample structure based on the time stability, size, and deformation displacement of the data. The input data and network structure may be a perfect fit. When applied to the processing of images, it offers certain distinct benefits. Below are the stages of a Convolutional Neural Network:

Convolution Layer: A convolutional layer is the fundamental building component upon which a CNN is built. A collection of self-sufficient feature detectors are contained within the convolution layer. Every single Feature map undergoes an own convolving process using the photos.

Pooling Layer: The pooling layer has a feature that gradually shrinks the spatial dimension of the illustration in order to cut down on the vast number of different parameters and computations that are performed in the network. The pooling layer performs its tasks separately on each individual function map. The following methods are utilized when conducting pooling:

Maximum Use of Pooling

Weighted Average Pooling

Sum Pooling

Fully Connected Layer: Neurons that are part of the layer that is fully connected are connected to all of the activations that are part of the layer below them. The procedure continues in the same manner as the ANN model, with the output gained from max pooling being converted to a one-dimensional array and serving as the input layer. The following taxa were taken into consideration under this suggested system:

"Linear" in "multiple linear regression" refers to the model's linear parameters. Each parameter multiplies an x-variable, and the regression function is a sum of these terms. Each x-variable might be a predictor or a transformation (such as the square of a predictor variable or two predictor variables multiplied together). Allowing non-linear transformation of predictor variables allows the multiple linear regression model to reflect non-linear responses and predictors.

(1) VGG16-

ConvNets are a type of artificial neural network that are able to perform convolutions. There is an input layer, an output layer, and multiple hidden layers in a convolutional neural network. The VGG16 CNN is widely regarded as one of the most effective computer vision models available today. In order to improve upon state-of-the-art setups, the model's developers conducted an evaluation of the networks and improved the depth utilizing an architecture with very small (3 3) convolution filters. Their efforts resulted in a model with 16–19 weight layers and 138 tunable parameters.

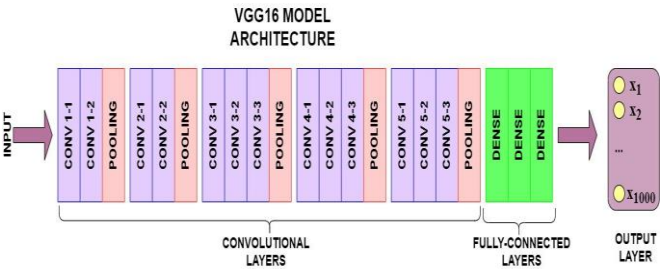
VGG16 is an object recognition and classification method that has a success rate of 92.7% when used to the classification of 1000 images belonging to 1000 distinct categories. It is one of the most well-known algorithms for picture classification since it is simple to implement. The images in the ImageNet collection all have RGB channels and have a fixed resolution of 224 by 224 pixels. So, the input that we have is a tensor with the values (224, 224, 3). This model processes the image that is provided as input and generates a vector containing 1000 values as the result. This vector represents the likelihood of assigning the appropriate class to the given category.

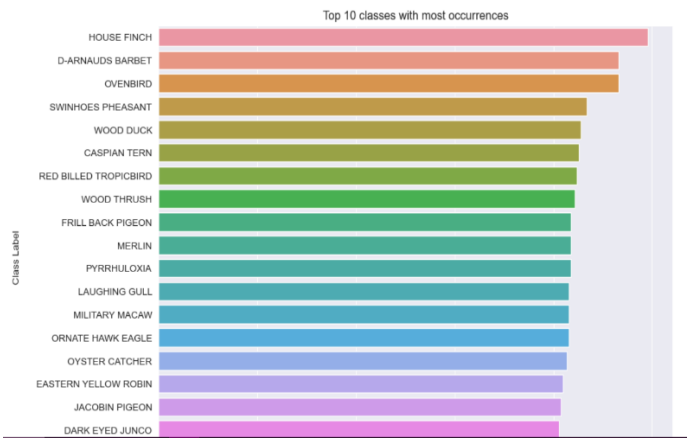
VGG Architecture:

An image with certain dimensions serves as the network's input (224, 224, 3). The initial two layers both include 64 channels with a filter size of 3 by 3, and they both use the same padding. Next, following a maximum pool layer of stride (2), two levels have convolution layers of 128 and filter size respectively (3, 3). After this, a max-pooling layer of stride (2, 2) will be applied, which is the same as the layer that came before it. Following that are two convolution layers with filter sizes of three and three and 256 filters respectively. Two sets of three convolution layers follow, and then a maximum pool layer. Each has the same padding and size (3, 3) in its 512 filters. The image is then passed to a stack that includes two convolution layers. Our filters in the convolution and max-pooling layers are only 3 by 3, whereas those in AlexNet and ZF-Net are 11 by 11 and 7 by 7. It uses the 1x1 pixel in some layers to change the number of input channels. This is accomplished in a few of the deeper levels. After each convolution layer, an additional one-pixel padding is added to protect the image's spatial details. This filler is always the same..

Following the stack of convolutional and max-pooling layers, we were left with a feature map with the dimensions (7, 7, 512). This output is then made into a feature vector with the dimensions (1, 25088), which we flatten. Following this, there are three fully connected layers: the first layer takes input from the most recent feature vector and outputs a (1, 4096) vector; the second layer also outputs a vector of size (1, 4096); however, the third layer outputs 1000 channels for the 1000 classes of the ILSVRC challenge; in other words, the third fully connected layer is used to implement the softmax function in order to classify the 1000 classes. The activation function of each of the buried layers is represented by the letter ReLU. Because it leads to faster learning and reduces the risk of difficulties with vanishing gradients, the ReLU algorithm is more computationally efficient than other similar algorithms.

VGG16 is much better than the models that competed in the ILSVRC-2012 and ILSVRC-2013 events. Also, the VGG16 result is in competition with GoogleNet, which won the classification task with a 6.7% error rate, and it does a lot better than Clarifai, which won the ILSVRC-2013 competition. It got about 11.7% without the external training data and 11.2% with it. In terms of single-net performance, the VGGNet-16 model gets the best result with about 7.0% test error, which is about 0.9% better than a single GoogleNet.





V. RESULTS AND CONCLUSION

The primary objective in building the identification website was to raise people's levels of awareness regarding bird-watching as well as bird species and how they can be identified, with a particular focus on birds that may be found in India. In addition to this, it satisfies the demand for streamlining the process of bird identification, which consequently makes bird watching much less difficult. Convolutional Neural Networks are the form of technology that was implemented in the experimental setup (CNN). Image recognition is accomplished through the extraction of features. The method that was utilized was adequate for the extraction of features and the classification of images.

The purpose of this paper is to provide a synopsis of our project. The primary objective of the project is to determine the species of the bird based on an image that is provided by the user as input. CNN was chosen because of its suitability for the implementation of complex algorithms as well as its high level of numerical precision and accuracy. In addition to that, it has a scientific and broad function. We were able to attain an accuracy of between 85% and 90%. We believe that the scope of this project exceeds quite a bit beyond what is necessary to accomplish the aim. This idea can be used in camera traps for the purpose of wildlife research and monitoring; the goal is to keep a record of the behavior of any animal as well as the movement of wildlife within a certain environment..

VI. REFERECES

[1] Elias Sprengel, Martin Jaggi, Yannic Kilcher, and Thomas Hofmann. Audio Based Bird Species Identification using Deep Learning Techniques. 2016.

[2] Yoonchang Han and Kyogu Lee. Acoustic scene classification using convolutional neural network and multiple- width frequency-delta data augmentation. 14(80:1-11, 2016.

[3] Kurt, A.; Oktay, A.B. Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. Expert Syst. Appl.2010, 37, 7986–7992.

[4] Corani, G. Air quality prediction in Milan: Feed-forward neural networks, pruned neural networks and lazy learning. Ecol. Model. 2005, 185, 513–529.

[5] Stefan Kahl, Thomas Wilhelm-Stein, Hussein Hussein, Holger Klinck, Danny Kowerko, Marc Fitter, and Maximilian Eibl Large-Scale Bird Sound Classification using Convolutional Neural Networks.

[6] Ni, X.Y.; Huang, H.; Du, W.P. Relevance analysis and short-term prediction of PM 2.5 concentrations in Beijing based on multi-source data. Atmos. Environ. 2017, 150,146–161.

[7] Mayer, H. Air pollution in cities. Atmos. Environ. 1999, 33, 4029–4037.

[8] Sankar, Ganesh, S., Arulmozhivarman, P., &Tatavarti, R. (2018). Air quality index forecasting using artificial neural networks-a case study on Delhi. International Journal of Environment and Waste Management, 22(1- 4), 4-23. https://doi.org/10.1504/IJEWM.2018.094105.

[9] M. Dong, D. Yang, Y. Kuang, D. He, S. Erdal, and D. Kenski, “PM 2.5 concentration prediction using hidden semi-markov model-based times series data mining,” ExpertSyst. Appl., vol. 36, no. 5, pp. 9046–9055, Jul. 2009.

[10] S. Thomas and R. B. Jacko, “Model for forecasting expressway pm2.5 concentration – application of regression and neural network models.” Journal of the Air & Waste Management Association, vol. 57, no. 4, pp. 480–488, 2007.

[11] R. Bardeli, D. Wolff, F. Kurth, M. Koch, K.-H. Tauchert, and K.-H. Frommolt, “Detecting bird songs in a complex acoustic environment and application to bioacoustic monitoring,” Patt Recog Letters, vol. Vol.31, pp. Pp. 1524-1534, 2010

[12] https://machinelearningmastery.com/autoregressionmodels-time-series-forecasting-python

[13] https://machinelearningmastery.com/arima-fortime-series-forecasting-with-python/

[14] https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/

[15] https://www.kaggle.com/c/air-pollution-prediction