Neural Network-Based Bird Detection Process

[1]Dr.G.V.V.Nagaraju, [2]M.Raga Namratha, [3]S.Nikhitha, [4]M.Kavyanjali, [5]G. Vivek Chandra

[12345]Vignan's Lara Institute of Technology and Science, Andhra Pradesh, India
[1]nagaraju006@gmail.com, [2]raganamratha123@gmail.com, [3]nikhithas997@gmail.com,
[4]muthyamkavyanjali03@gmail.com, [5]vivekgeddati1@gmail.com

Abstract: The occurrence of some bird species is decreasing, and even when they are discovered, it is difficult to classify and forecast which ones will be around in the future. Birds in diverse settings may, of course, appear from varying perspectives, sizes, colors, and forms. Efficacious monitoring systems and digitization are becoming global trends in the 21st century. Everyone has a cell phone these days, thus anybody may be able to snap a photo of the bird. The picture is gray scaled using a Convolutional Neural Network (CNN) technique. Then, the autograph is formed using the Pytorch model, which generates many comparison nodes. A score sheet is generated by comparing these various nodes with the testing dataset. The necessary bird species may be predicted when the score sheet is analyzed.

Keywords: Python, Convolutional Neural Network (CNN), dataset, greyscale format, and classification are some of the key terms.

1. INTRODUCTION

The current situation has elevated concerns about avian behavior and numbers. Birds play an important role in environmental organism detection. It is a significant and difficult difficulty to identify bird species just by their calls. We can also keep an eye on bird populations using a variety of techniques. The use of automated techniques for bird species identification allows for an efficient evaluation of the amount and variety of the birds that emerge in the area, since many birds move in response to changes in the environment.

The terms "artificial intelligence" and "machine learning" seemed like they were coming straight out of a science fiction novel. One of the easiest ways to use it is for image recognition. The way graphical data is organised and processed is being transformed by machine learning that is integrated into consumer websites and apps. The application of deep learning algorithms has greatly improved image recognition and identification. Developed to mimic the way the human brain works, this is a machine learning approach. This is how computers learn to identify objects in images.

Algorithms can interpret pictures and come up with pertinent tags and classifications by studying big databases and keeping an eye out for developing trends. The majority of the time, while classifying birds, we look at how they fit into a certain group. The closeness between the classifications makes bird classification much more difficult than category classification. For this reason, it is crucial to be able to identify which kind of bird a given photograph depicts. The process of bird species identification involves classifying images of birds according to their predicted characteristics.

1.1. Existing System

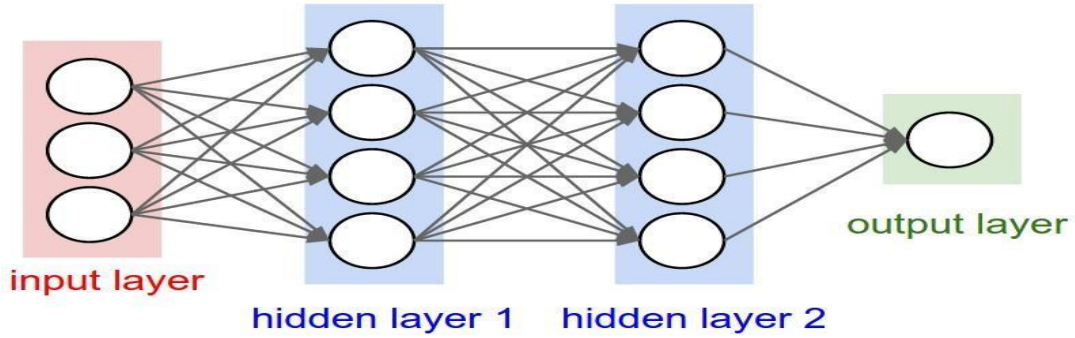
Bird identification has long been a challenge for ornithologists. They need to learn all there is to know about birds, including how they live, what they eat, where they live, and how the environment affects them. Ornithologists often use Linnaeus's five-part taxonomy to categorise birds: kingdom, phylum, class, order, family, and species.

1.2. Methodology

Deep learning is a branch of Machine learning based on a set of algorithms that attempt to model high level abstraction in data by using a deep graph with multiple processing layers,

composed of multiple linear and nonlinear transformation. Widely used algorithm of deep learning in image classification are convolutional neural network (CNN). Therefore, CNN allows classification based on Tenser Flow, Pytorch Models.

A large portion of this system's functionality is provided by the software, which classifies birds using the Python programming language, the Pytorch model, and a Raspberry Pi. The original picture is captured from a digital gadget and then transformed into a greyscale version. A large number of neurones were identified using deep learning algorithms. As the picture passes through a series of neural networks, these algorithms get a deeper understanding of it. The picture is classified for feature extraction using the following figure.



Above diagram shows the three layers of neural network.

1.3. Algorithm

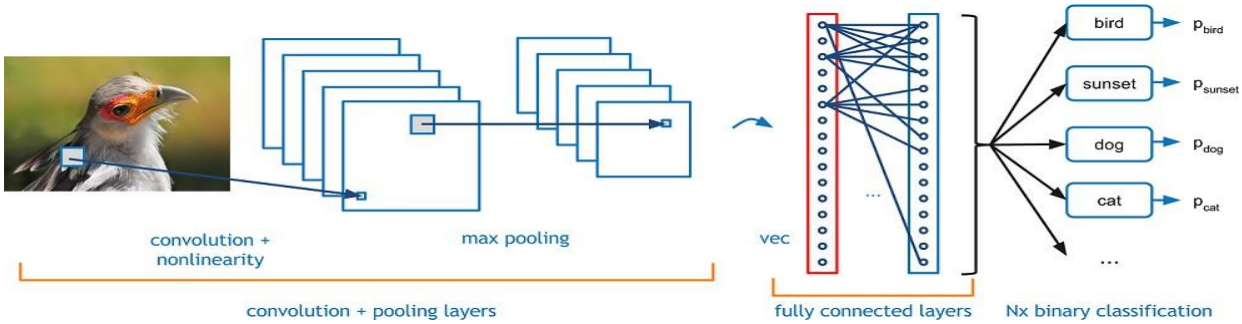
Due to the unknown nature of the input picture, a deep learning technique was used in its development. Convolutional neural networks (CNNs) are often used for image analysis. It has a number of hidden levels in addition to an input and an output layer. The neurones that make up each layer are totally coupled to each other and to the neurones that made up the layer before it. Predicting output is the job of the output layer. With an image as input, the convolutional layer generates a series of feature maps as output [2]. The convolutional layer maps one 3D volume to another, which is useful when the input picture has several channels, like a bird's beak, wings, eyes, or colour. The three dimensions studied here are height, breadth, and depth. In a CNN, there are two parts: Part one: feature extraction. The network detects features by doing a sequence of convolution and pooling operations.

Second, the classification part: a fully connected layer is used as a classifier and is fed the retrieved features.

Convolutional neural networks (CNNs) are structured with four layers: completely connected, activation, pooling, and convolutional. With a convolutional layer, you can extract a few key visual attributes from a picture. To save the crucial information while reducing the number of neurones from the preceding convolutional layer, pooling is used.

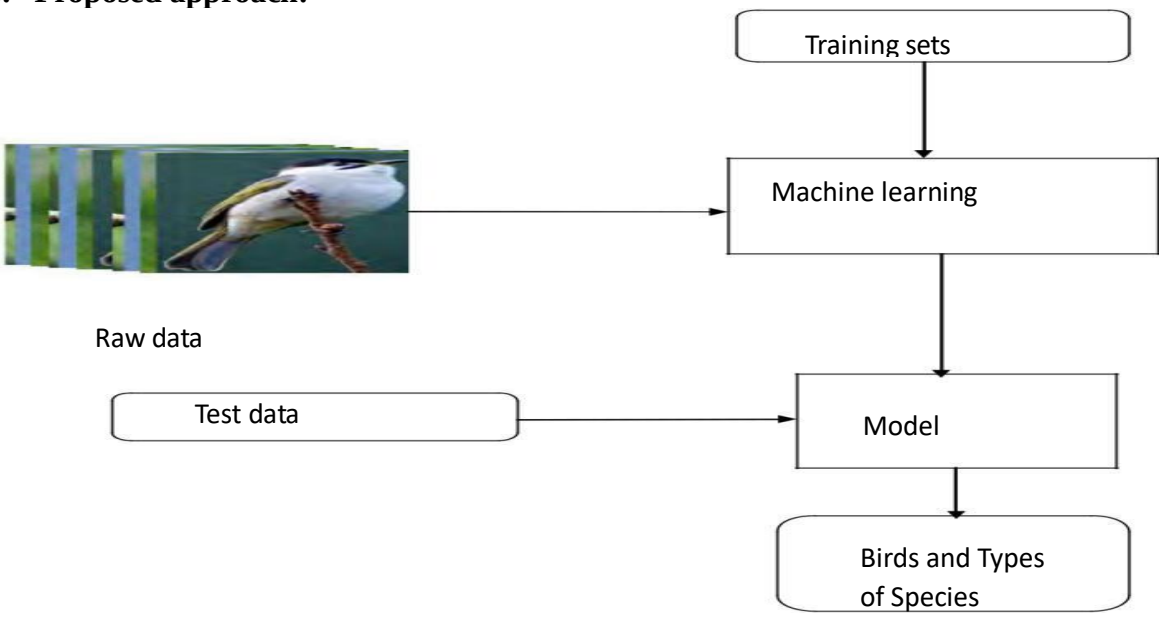
An activation layer's function condenses data into a range before passing them on. In a completely linked state

layer links each neurone in one layer to every neurone in the layer below it. The increased precision is a result of CNN's thorough neurone classification.



There are two main approaches to picture categorization in machine learning.
1.Greyscale
2. By making use of the RGB codes.
The majority of the time, all the data is transformed into greyscale. In a greyscale algorithm, the computer will give each pixel a value depending on the pixel's actual value. In order to categorise the data, the computer will execute an operation on an array that contains all the pixel values.
Libraries:
One of the most important libraries utilised by this system is Pytorch. Developed by Facebook's AI Research unit (FAIR), Pytorch is an open-source software library. Useful for tasks like computer vision and NLP, this open-source machine learning library is based on the Torch library. Python forms the core component of Pytorch. A modified version of the BSD license is used to release Pytorch.
Pytorch has two main functions:
1. Tensor computation (e.g., NumPy) with GPU acceleration for performance.
2. Constructing deep neural networks using an autodiff method that is based on tape.
Dataset:
A dataset is a group of related data sets. Tabular data is organised in a way that makes sense when viewed via a database: each row represents a record from the set in question, and each column represents a different variable.

3. Proposed approach:



3.1.Block diagram Explanation:

1. Pure Data: Unstructured data is what it's called. It does not allow us to infer any useful information. It stands in for a single piece of table data that is implicitly organized.
2. Exercise regimen: For the purpose of training the model to identify certain feature parameters and carry out a co-relational job, the training dataset contains raw data samples.
3. In-depth education CNN: It's a module that uses CNN to anticipate the most categorized categories for input photos and extracts unique bird traits. In order to identify birds, the CNN model used a stack of convolutional layers, which included an input layer, two FC layers, and an output layer.

4: Test Results: The classifier parameters and the network model's actual prediction performance may be evaluated using the test dataset. The trained prediction model will be used to categorise fresh input photos once features have been retrieved from the raw data.

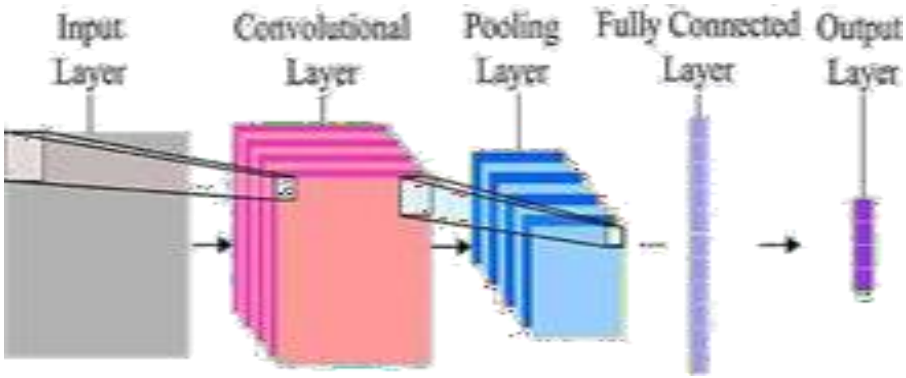
5: Extracting Features: First and foremost, our main objective is to extract features from the raw input photos. These features will provide descriptive and useful information for fine-grained object detection. Nevertheless, feature extraction will be difficult due to intra- and semantic-class variability. We will first learn which features in the model were directly mapped to which picture sections by extracting features in appropriate places for each component of the image and then using that knowledge.

6. Predictive model: If you submit a picture of a bird, the suggested model can identify it. Predicting and differentiating between photos of different birds is the task of the suggested system.

4. Software Implémentation:

1. Pytorch with Python:
   Once in a while, a new python library comes up with revolutionary promise for deep learning. Among these libraries is PyTorch. It has been my experience that of all the deep learning packages, PyTorch is the most user-friendly and adaptable. Since we don't need to wait till the whole code is created before finding out whether it works or not, this fits well with the python programming technique. Running a small section of code and seeing its output in real-time is a breeze. PyTorch is a library for deep learning development created in Python. It offers versatility.

2. Vertex-based computation networks:
   Since PyTorch gives us a framework to operate inside, we don't need premade graphs with particular functionality.
   In order to construct computational graphs dynamically and even modify them while functioning. This is helpful in cases when the amount of memory needed to construct a neural network is uncertain.
3. CNN:
   One kind of deep neural network that finds widespread use in image analysis is the convolutional neural network (CNN). It has a number of hidden levels in addition to an input and an output layer. The neurones that make up each layer are totally coupled to each other and to the neurones that made up the layer before it. Predicting output is the job of the output layer. The input picture is processed by the convolutional layer, which then outputs a series of feature maps. The convolutional layer maps one 3D volume to another, which is useful when the input picture has several channels, like a bird's beak, wings, eyes, or colour. The three dimensions of volume are breadth, height, and depth. The approach that has been suggested for doing this is based on convolutional neural networks.

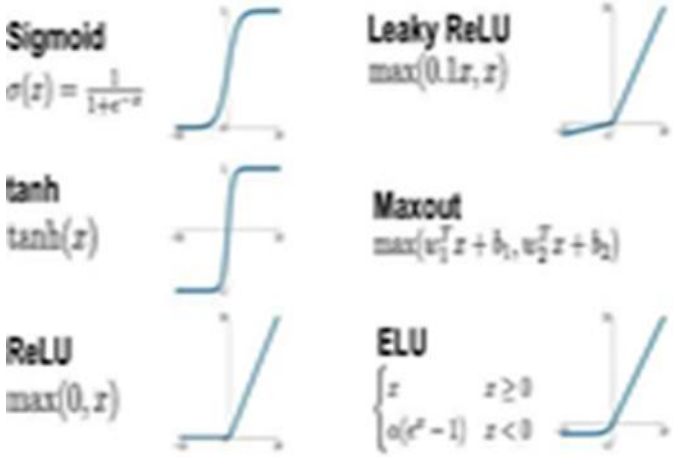


Basic architecture of CNN

Image pixel inputs with R, G, and B colour channels are sent to the convolutional layer via the input layer. Convolutional neural networks (CNNs) differ from basic neural networks in that their neurones are built in three dimensions, where depth denotes activation volumes, rather than two. The primary function of the convolutional layer is to retain the spatial relationships between input pixels while extracting visual information. By calculating the output of neurones, it alters the picture.

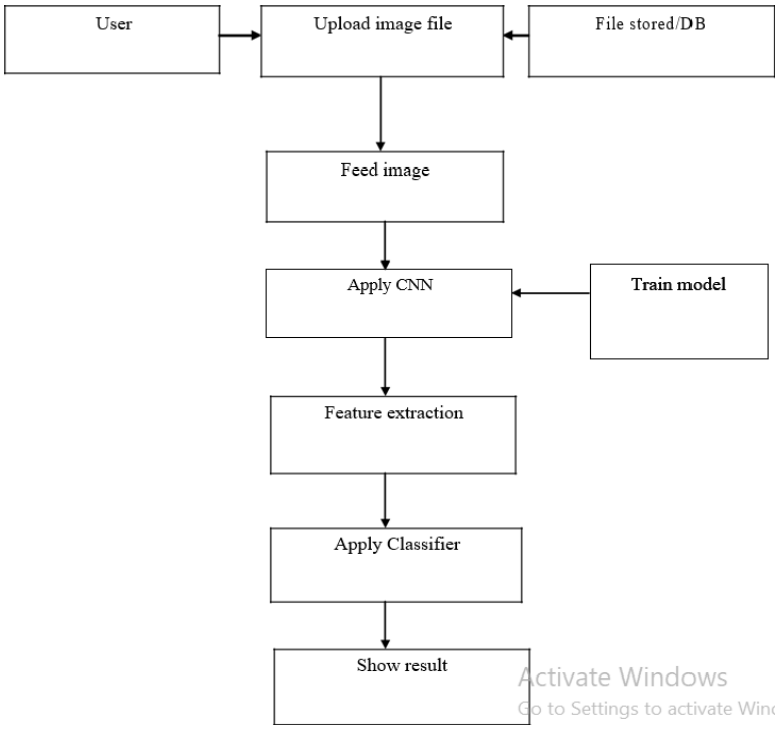
Converting pixel data into output volume or final class scores. The striding filter is used to perform this convolution function on the input picture. In convolutional neural networks (CNNs), each picture is essentially a matrix of pixel values; the filter in this context is a matrix by another name, kernel, and the idea is derived from image processing. Every place in the feature map has its filter strided by one pixel, and the procedure calculates computed by summing the results of the matrices' multiplication. The final feature map is computed by each convolutional layer. Additionally, activation functions are used to take input volume from earlier layers and neurone characteristics like bias and weights.

This layer is used in the ReLU layer to convert all negative pixel values to zero. Contrary to the non-linear nature of most real-world pictures, CNN operations follow a linear pattern. In order to train CNNs to handle non-linear data, the ReLU function is used to induce non-linearity.

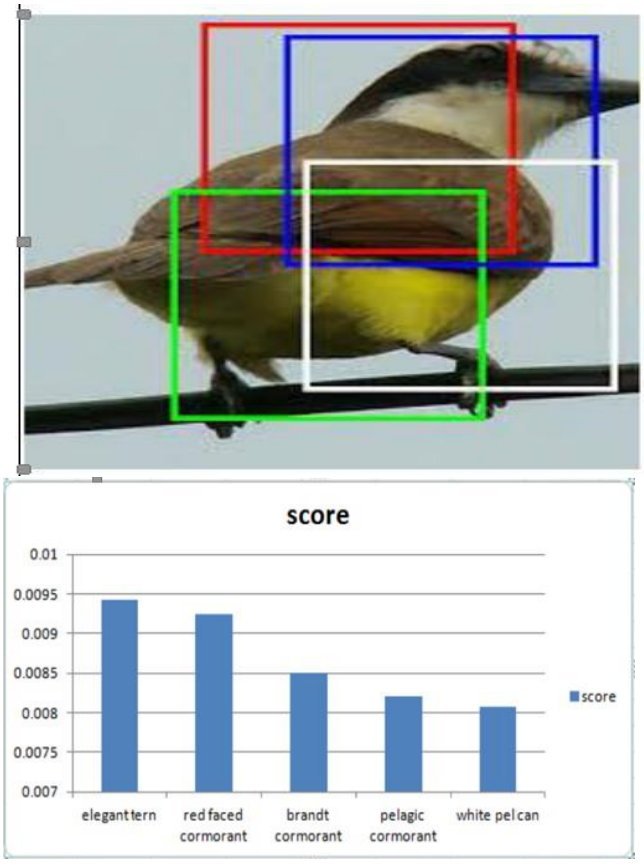


Even though the pooling function shrinks the feature map, it keeps the crucial spatial information. The pooling function manages the overfitting problem by reducing the network's parameters and computation. Max, sum, average, and many more kinds of pooling functions are available. In this study, the pooling layer will use the max form of pooling. For the max pooling process, you need to start by defining a spatial neighbourhood, such as [2x2], and then

add it to the corrected feature map, striding it by 2, until you find the region with the biggest pixel value.



5. SIMULATIONS/RESULTS: This is an example image that was found online. The picture is processed using a number of parameters shown in the figure using deep learning convolutional neural networks approach.



It has been noted from the data analysis that the accuracy is reduced when just one parameter is employed. However, when other factors are taken into account, such as the exquisite tern, red-faced cormorant, Brandt cormorant, and countless more, a combined approach might be used.

6. **Conclusion:**

The classification research looks at a way to identify bird species utilising a dataset for picture classification and a deep learning system. A user-friendly interface will be included into the system, allowing users to easily submit photos for identification purposes and get the appropriate results. Part detection and CNN feature extraction from many convolutional layers would form the basis of the proposed system's operation. The classifier will be provided these attributes so that it can classify the data. The algorithm will use the data as a starting point to improve its bird species prediction accuracy. In order to attain optimal efficiency, the system will run a number of trials on a dataset that contains several images.

REFERENCES:

1. XIE, Z., A. SINGH, J. UANG, K. S. NARAYAN and P.ABBEEL. Multimodal blending for high-accuracy instance recognition. In: 2013 IEEE/RSJ International Conference on Intel-ligent Robots and Systems. Tokyo: IEEE, 2013, pp. 2214–2221. ISBN 978-14673-6356-DOI: 10.1109/IROS.2013.6696666.

[2] EITEL, A., J. T. SPRINGENBERG, L. D. SPINELLO, M.RIEDMILLER and W. BUR-GARD. Multimodal Deep Learning for Ro-bust RGB-D Object Recognition. In: 2015IEEE/RSJ, International Conference on Intelligent Robots and Systems (IROS). Hamburg: IEEE, 2015, pp.681–687. ISBN 978-1-4799-9994-1.DOI: 10.1109/IROS.2015.7353446.

[3] RUSSAKOVSKY, O., J. DENG, H. SU,J. KRAUSE, S. SATHEESH, S. MA, Z. HUANG,A. KARPATHY, A. KHOSLA, M. BERNSTEIN,A. C. BERG and L. FEI-FEI. Image Net Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV). 2015, vol. 115, no. 3, pp. 211–252. ISSN 1573-1405. DOI: 10.1007/s11263-015-0816-y.

[4] KRIZHEVSKY, A., I. SUTSKEVER and G. E. HINTON.Image Net classification with deep convolutional neural networks. Annual Conference on Neural Information Processing Systems (NIPS). Harrah's Lake Tahoe: Curran Associates, 2012, pp. 1097– 1105. ISBN 978-1627480031

[5] Tóth, B.P. and Czeba, B., 2016, September. Convolutional Neural Networks for Large Scale, Bird Song Classification in Noisy Environment. In CLEF (Working Notes) (pp. 560-568).

[6] Avinash P, Venkateswarlu T, Anand D (2048) A detail study on biometrics with Matlab. Int J Eng Technol (UAE) 7(2.20 Special Issue 20), pp 243–249

[7] Srinivasu SVN, Venkateswarlu T, Avinash P (2018) A valuable role of digital payments in building smart cities using IoT technology. J Adv Res Dynamical Control Syst 10(2):1890–1896