e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:08/August-2022 Impact Factor- 6.752 www.irjmets.com

ANIMAL FACE RECOGNITION SYSTEM USING DEEP LEARNING

Waqas Qureshi*1, Dr. Ahmed Shaikh*2

*1Student, Department Of Computer Engineering, M.H.S.S.C.E, Mumbai, Maharashtra, India.

*2Professor, Department Of Computer Engineering, M.H.S.S.C.E, Mumbai, Maharashtra, India.

ABSTRACT

The face is one of the easiest ways to distinguish the identity of each animal. Face recognition technique is a personal identification system that uses personal characteristics to identify the Face. Face recognition procedure consists of two phases, namely face detection, where this process takes place very rapidly, except under conditions where the object is located at a short distance away, then the introduction, which recognize a face as individuals. Stage is then developed as a model for facial face recognition system which is one of the much-studied biometrics technology and developed by experts. Checking of wild animal in their prevalent environment is crucial This project proposes to develops an algorithm for detecting animals in the wild. This project creates an algorithm to identify animals in the outdoors. Given the diversity of creatures, manually identifying them can be a difficult undertaking. This algorithm classifies an animal based on their images so we can identify them more efficiently. Animal detection and classification may aid in prevention of animal accidents, tracking animals and can prevent theft. This task can be achieved by applying effective deep learning algorithms.

Keywords: Animal Detection And Classification, Deep Learning Algorithms, Neural Networks.

I. INTRODUCTION

A role of Face recognition is to identify an already detected object as a known or unknown face. Generally, the problem of face recognition is confused with the problem of face detection. Face Recognition on the other hand is process determining the "face" is someone known, or unknown, using for this purpose a database of faces in order to validate this input face.

Different Approaches of Face Recognition:

1. Geometric: It focus on geometrical relationship between facial landmarks, or in other words the spatial configuration of facial features. That means that the main geometrical features of the face like the eyes, nose and mouth are first located then faces are classified on the idea of various geometrical distances and angles between features.

Figure 1

2. Photometric Stereo: It used to recover the shape of an object from a number of images taken under varying lighting conditions. The shape of the retrieved objects defined by a gradient map, which is made up of an array of surface normals.

Face detection entails separating image windows into two groups; one containing faces (turning the background (clutter). It is difficult because, although commonalities exist between faces, they can differ considerably in terms of age, skin color and facial expression. Different lighting conditions, picture qualities, and geometries, as well as the possibility of partial occlusion and disguise, complicate the problem even more. An ideal face detector would therefore be able to easily detect the presence of any face under any set of lighting conditions, against any background. The face detection task can be divided down into two stages. The first stage

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:08/August-2022 Impact Factor- 6.752 www.irjmets.com

is a classification, that takes some arbitrary image as input and outputs a binary value of yes or no, indicating whether there are any faces present in the image. The second stage is the face localization task, which aims to take an image as input and output the location of any face or faces within that image as some bounding attributes (x, y, width, height).

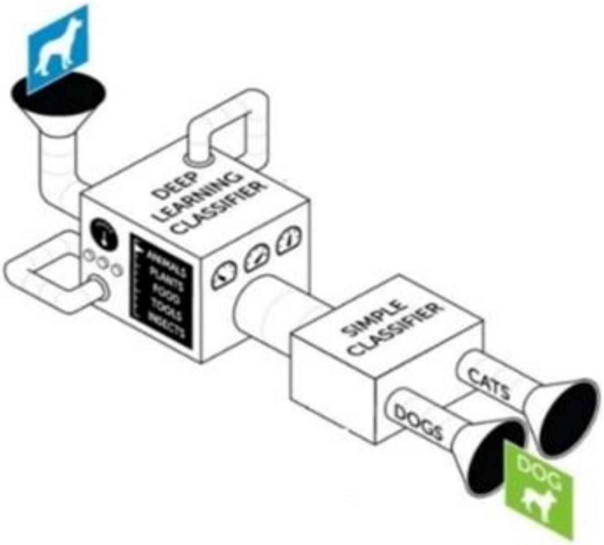

Figure 2

Approaches of Face Detection

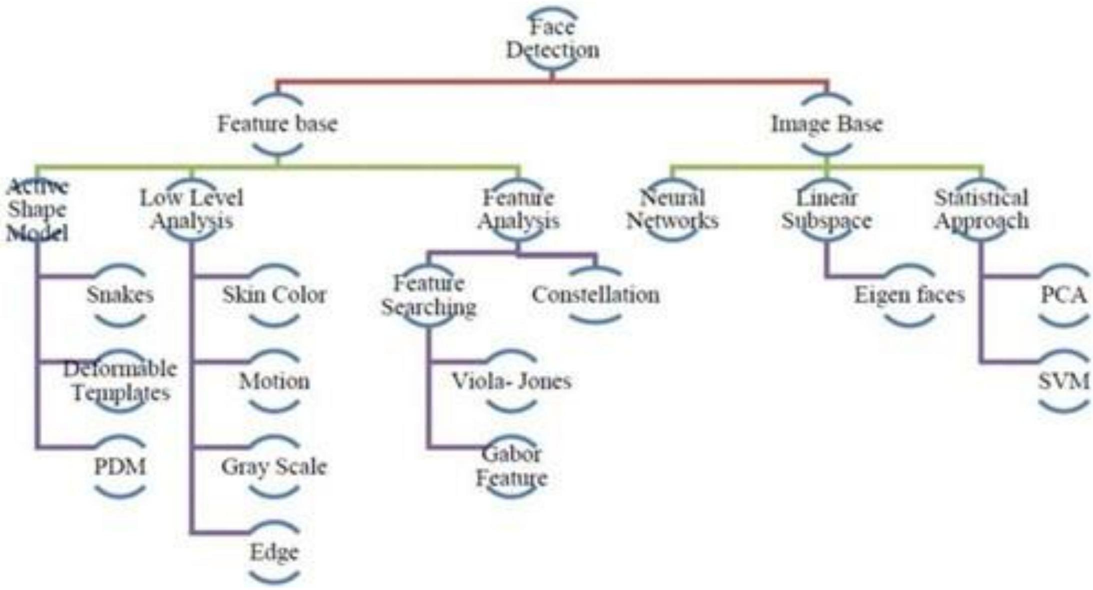

Figure 3

Feature Base Approach

Active Shape Model: Active shape models concentrate on complex non-rigid features such as actual physical and higher-level appearance of features. It means that Active Shape Models (ASMs) are designed to automatically locating a landmark point that describe the shape of any statistically modelled object in an image. When of facial features like eyes, lips, nose, mouth and eyebrows. The training stage of an ASM entails the development of a statistical facial model from a training set of images with manually annotated landmarks. ASMs is divided into three groups i.e., snakes, PDM, Deformable templates.

Deformable Templates: Deformable templates were introduced by Yuille et al. to take into account a priority of facial features and to better the performance of snakes. It is difficult to Locate a facial feature boundary because the local evidence of facial edges is difficult to organize into a sensible global entity using standardized contours. The low contrast in brightness around some of these features also complicates the edge detection process. Yuille et al. took the concept of snakes a step further by adding global knowledge of the eye to improve the reliability of the extraction process. Approaches based on deformable models are built to solve this problem. local valley, edge, peak, and brightness are used to determine deformation. Aside from face boundary, extracting salient features (eyes, nose, mouth and eyebrows) is a great challenge in face recognition. E = Ev + Ee + Ep + Ei + Einternal, where: Ev, Ee, Ep, Ei, Einternal are external energy result from valley, edges, peak and image brightness and internal energy. [1]

Point Distribution Model (PDM): Prior to the development of ASMs, researchers developed statistical methods of shape independent of computerized image processing. The idea is that once shapes are represented as vectors, they can be subjected to standard statistical methods just like any other multivariate object. These models learn allowable constellations of shape points from training examples and construct a Point Distribution Model using principal components. These have been used in variety of ways, for example for categorizing Iron Age broaches. Distribution of ideal Points Models can only deform in ways that are characteristic of the object. Cootes and his colleagues were looking for models which do exactly that so if a beard, say, covers the chin, the shape model can \override the img" to approximate the location of the chin under the beard. It was therefore natural to adopt Point Distribution Models. This combinations of ideas from image processing and statistical shape modelling led to the Active Shape Model. The first parametric statistical shape model for image analysis based on principal components of inter- landmark distances was introduced by Cootes and Taylor. On this approach, Cootes, Taylor, and their colleagues, then published a series of papers which cumulated in what we call the classical Active Shape Model. [2][3]

Low Level Analysis: Based on low level features such as color, intensity, edges, motion etc.

Skin Color Base: Color is a crucial feature of a faces. There are several advantages of using skin- color as a feature for tracking a face . Color processing is much quicker than processing other facial features. Color is orientation invariant under certain illumination conditions. This property makes motion estimation even easier because only a translation model is needed for motion estimation. Tracking faces using color as a feature has several issues such that the color representation of a face obtained by a camera is affected by many factors (ambient light, object movement).There are Majorly three different face detection algorithms are available based on RGB, YCbCr, and HSI color space models. In the implementation of the algorithms there are three major steps viz.

- Classify the skin region in the color space
- Apply threshold to mask the skin region
- Draw bounding box to extract the face image

Crowley and Coutaz proposed the most basic skin color algorithms for detecting skin pixels. The perceived color of human varies in relation to direction to the illumination. The pixels for skin area can be detected using a normalized color histogram, and can be normalized for changes in intensity on dividing by luminance. Converted (R, G, B) vector is converted into an (r, g) vector of normalized color which provides a quick means of skin detection. This algorithm fails when there are some additional skin regions such as legs, arms, etc. Cahi and Ngan suggested a skin-color classification algorithm with YCbCr color space. According to the research pixels belonging to skin region having similar Cb and Cr values So that the thresholds are (Cr1, Cr2) and (Cb1, Cb2), a pixel is considered to have skin tone if the values are (Cr, Cb) fall within the thresholds. The skin color distribution determines the face portion in the color image. This algorithm is also having the restriction that the image should be having only face as the skin region. K.jeldson and K.ender defined a color predicate in HIS color space to separate skin regions from background. Skin color classification in HSI color space is the similar as YCbCr color space but here the responsible values are hue(H) and saturation (S). As like to above the threshold is chosen as (H1, S1) and(H2,S2), and a pixel is classified to have skin tone if the values (H, S) fall within the threshold and this distribution results the localized face image. Similar to previous two algorithm this algorithm is also having the same limitation. [4][5][6]

Motion Base: When use of video sequence is available, motion information can be used to find moving objects. By simply thresholding cumulative framed disparities, moving silhouettes such as faces and body parts can be extracted. Aside from face regions, facial feature scan be located by frame differences

Gray Scale Base: Grayscale detail on a face may also be treated as essential features.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:08/August-2022 Impact Factor- 6.752 www.irjmets.com

Facial features such as the eyebrows, pupils, and lips appear darker than the rest of the facial regions .several recent feature extraction algorithms search for local gray minima within segmented facial regions. In these algorithms, the input images are first improved by contrast-stretching and gray-scale morphological routines to improve the quality of local dark patches and thereby make identification easier. The extraction of dark patches is accomplished by low-level gray-scale thresholding Based method and consist three stages. Yang and huang proposed a new method that is, faces gray scale behavior in pyramid (mosaic) images. This system employs a hierarchical Face location consist three levels. Higher two level ae focused on mosaic images at different resolution. In the lower level, edge detection method is proposed. Further, this algorithm gives good response in complex background where size of the face is unknown. [7]

Edge Base: Edge base Face detection was introduced by Sakai et al. This work was based on analyzing line drawings of the faces from photos in order to locate facial features. Later, Craw et al. suggested a hierarchical framework for tracing a human head outline based on Sakai et al works. Following that, several researchers in this field performed outstanding work. Method proposed by Anila and Devarajan was very simple and quick. They proposed frame work that consist three steps. Initially the images are improved by applying median filter for noise removal and histogram equalization for contrast adjustment. In the second step, the edge image is constructed from the improved image using a sobel operator. Then, based on the edges, a novel edge tracking algorithm is used to extract the sub windows from the enhanced image. Further, they used Back propagation Neural Network (BPN) algorithm to determine the sub-window as either face or non-face. [8]

Feature Analysis: These algorithms seek structural features that exist even when the pose, viewpoint, or lighting conditions differ, and then use these to locate faces. These approaches are primarily intended for face localization.

Viola Jones Method: Paul Viola and Michael Jones proposed object detection approach which minimizes computation time while achieving high detection accuracy. Paul Viola and Michael Jones presented a fast and robust method for face detection which is 15 times faster than any technique available at the time of release with 95 percent accuracy at about 17 fps. The technique is based on the use of simple Haar-like features that are easily evaluated using a new image representation. Based on the idea of an Integral Image it produces a large set of features and uses the boosting algorithm AdaBoost to minimize the over complete set and the introduction of a degenerative tree of the boosted classifiers provides for robust and quick interferences. The detector is used in a scanning fashion on gray- scale images, and the scanned window can also be scaled, as well as the features evaluated. [9]

Gabor Feature Method: Sharif et al presented an Elastic Bunch Graph Map (EBGM) algorithm that successfully implements face detection using Gabor filters. This system uses 40 different Gabor filters on an image. As a result of which 40 photographs with different angles and orientations are obtained. Then, maximum intensity points in each filtered image are determined and mark them as fiducial points. The system reduces these points based on the distance between them. The next step is calculating the distances between the reduced points using distance formula.[10] Finally, the distances are link to the database. If match found, it indicates that the faces in the image have been detected. Equation of Gabor filter is shown below

$$\psi_{u,v}(z) = \frac{\| k_{u,v} \|^2}{\sigma^2} e^{\left(\frac{\| k_{u,v} \|^2 \| z \|^2}{2\sigma^2} \right)} \left[e^{i\vec{k}_{u,v} z} - e^{-\frac{\sigma^2}{2}} \right]$$

Where

$$\phi_u = \frac{u\pi}{8}, \quad \phi_u \in [0, \pi) \text{ gives the frequency.}$$

Figure 4

Constellation Method: All methods discussed so far can track faces but still some issues such as locating faces of different poses in complex background is extremely difficult. To a alleviate this challenge , investigator arranged a group of facial features in face-like constellations using more robust modelling approaches such as statistical analysis. Burl et al. suggested various forms of face constellations. They established the application of statistical shape theory to features identified by a multi scale Gaussian derivative filter. In a paradigm focused on image feature analysis, Huang et al. also use a Gaussian filter for pre-processing. [11]

Image Base Approach

Neural Networks: In many pattern recognitions issues, such as OCR, object recognition, and autonomous robot driving, neural networks are gaining a lot of attention. Since face detection is a two- class pattern recognition challenge, several neural network algorithms have been proposed. The advantage of using neural networks for face detection is the ability to train a machine to capture the complex class conditional density of face patterns. However, one disadvantage is that in order to achieve exceptional efficiency, the network architecture must be extensively calibrated (number of layers, number of nodes, learning rates, etc.). Agui et al. suggested the most hierarchical neural network in the early days. The first stage consists of two parallel sub networks, the inputs of which are filtered intensity values from an original image. The subnetwork outputs and derived function values are used as inputs to the second stage network. The presence of a face in the input region is indicated by an output at the second level. Propp and Samal developed one of the first neural networks for facial recognition. Their network is made up of four layers, each of 1,024 input units, 256 in the first hidden layer, eight units in the second hidden layer, and two output units. Feraud and Bernier proposed a detection approach based on auto associative neural networks. The concept is based on, which demonstrates that an auto associative network with five layers can perform nonlinear principal component analysis. One auto associative network detects faces in the frontal view, while another detects faces turned up to 60 degrees to the left and right of the frontal view. Lin et al. then proposed a face recognition method based on probabilistic decision making. [12][13][14]

Linear Subspace:

Eigen Faces Method: Kohonen demonstrated the use of eigen vectors in face recognition early on, using a simple neural network to perform face recognition for aligned and normalized face images. Photos of Kirby and Sirovich were proposed by Kirby and Sirovich. Faces can be encoded linearly using a small number of basis images. The concept was allegedly first suggested by Pearson in 1901, and then by HOTELLING in 1933. Provided a training set of n by m pixels. The mean square error between the projection of the training images onto this subspace and the original images is reduced when images are interpreted as a vector of size m X n, basis vectors spanning an ideal subspace are calculated. Eigen pictures are given to the set of optimal basis vectors since they are simply the eigen vectors of the covariance matrix computed from the vectorized face images in the training set. Experiments with a range of 100 images show that a (91x50) pixel face image can be effectively encoded using just 50 Eigen images. [15][16]

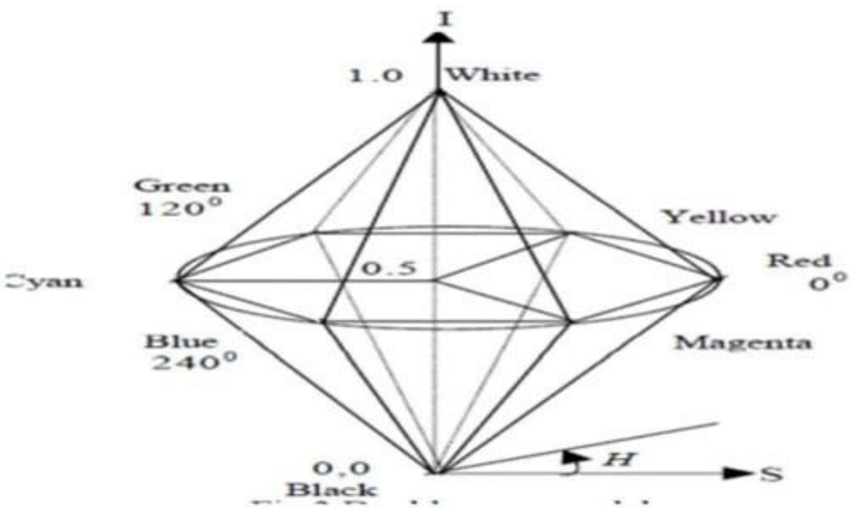

Figure 5

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:08/August-2022 Impact Factor- 6.752 www.irjmets.com

Statistical Approach:

Support Vector Machine(SVM): Osuna et al. originally presented SVMs for face detection. SVMs was used to train polynomial function, neural networks, or radial basis function (RBF) classifiers as new paradigm. SVMs works on principle of n induction, called structural risk minimization, which aims to minimize an upper bound on the expected generalization error. An SVM classifier is a linear classifier in which the separating hyper plane is selected to minimize the expected classification error of the unseen test patterns. In Osuna et al. create an effective method for training an SVM for large scale problems, and applied it to face detection. On basis of two test sets of 10,000,000 test patterns of (19 x19) pixels, their system has slightly lower error rates and runs approximately 30 times faster than the system by Sung and Poggio. SVMs can also use to detect faces and pedestrians in the wavelet domain.[17]

II. METHODOLOGY

Face Detection System comprises of the following steps:

1. Pre-Processing: To reduce the variability in the faces, the images are first processed before they are fed into the network. All positive examples i.e., the face images are produced by cropping images with frontal faces to include only the front view. After that, All the cropped images are then corrected for lighting through standard algorithms.
2. Classification: To classify the images as faces or non-faces Neural networks are implemented by training on these examples. We have used both our implementation of the neural network and the MATLAB neural network toolbox for this task. Different network configurations are tested with to optimize the results.
3. Localization: The trained neural network is then used to search for faces in an image and if they found localize them in a bounding box. Different Feature of Face on which the work has done on: - Position Scale Orientation Illumination

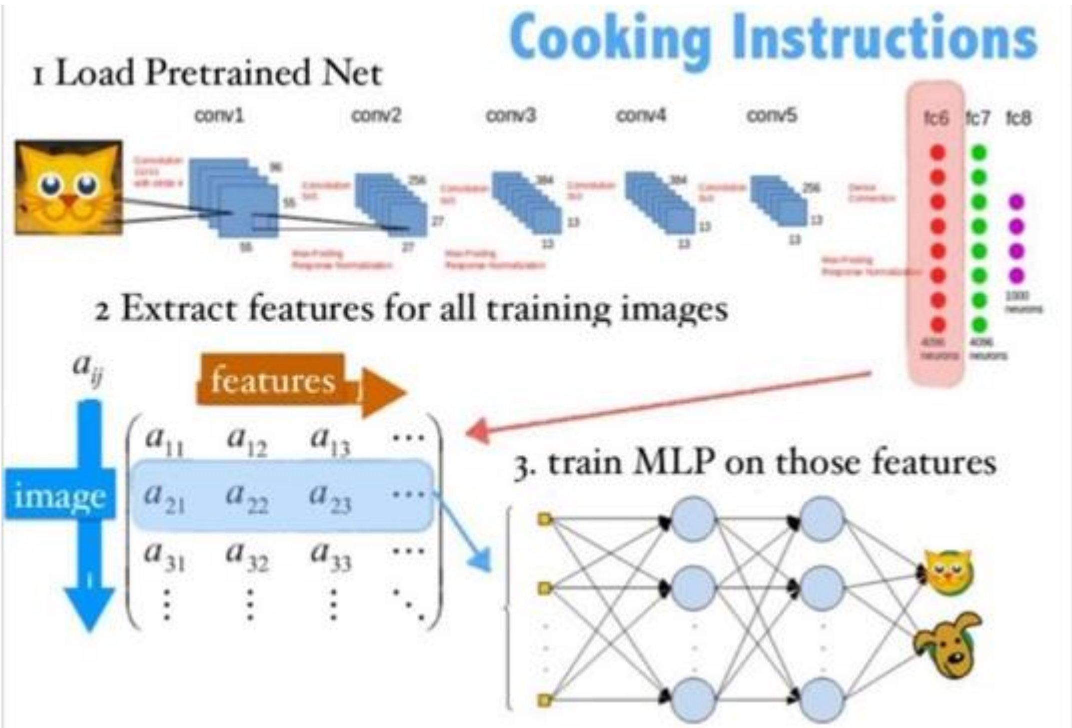

Figure 6

III. MODELING AND ANALYSIS

1. CNN image classifications take an input image,[Input image will have the raw pixel values of color channels Red(R), Green (G), Blue (B).] Computers sees an input image as array of pixels and it depends on the resolution of an image. On basis of image resolution, it will see h x w x d(h = Height, w = Width, d = Dimension).

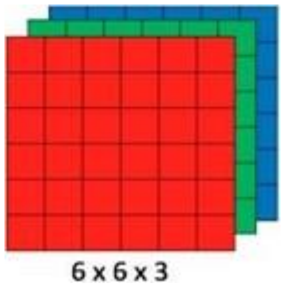

Figure 7

2. The next layer is called, Convolution layer. Convolution is the first layer which extract features from an input image. Convolution maintains the relationship between pixels by utilizing image features using small squares of input data. It is a mathematical operation This layer will calculate the output of neurons that are connected to local regions to the input layer. Each calculates a point product between their weights and a small area connected to the input layer.

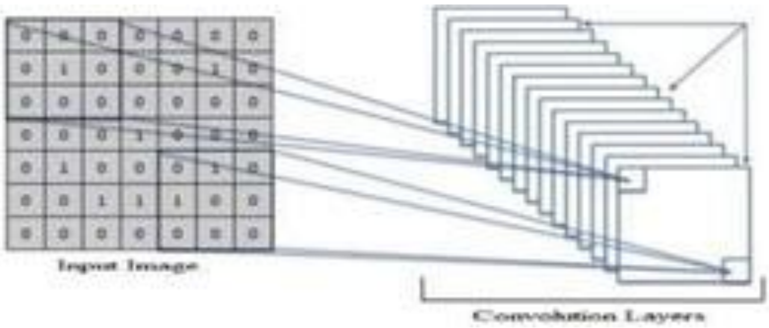

Figure 8

3. Third layer is called Rectified Linear activation function(ReLU) layer, its function in a neural network is in charge of turning the node's summed weighted input into the activation of the node or output for that input. which will apply a unit wise activation function. It maintains the volume of the input.

4. The next layer is called the Pooling layer, it reduces the dimensions of the feature maps and summarizes the feature presented in a region. It helps in making the model more robust to variance in size of input image.

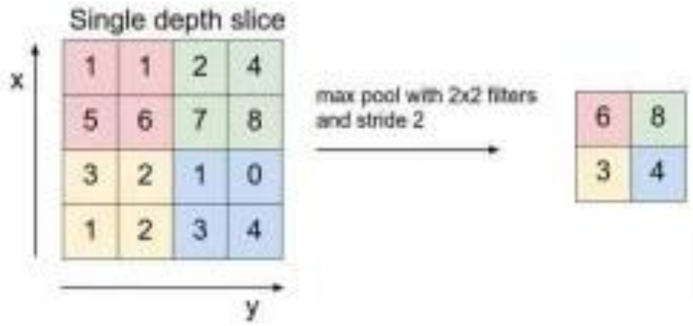

Figure 9

5. Last layer is Fully Connected Layer (FC): We flattened our matrix into vector and feed it into a fully connected layer like a neural network .this layer will calculate the class scores, resulting in volume of size. Like traditional Neural Networks, each neuron in this layer is connected to all the weights in the previous set.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:08/August-2022 Impact Factor- 6.752 www.irjmets.com

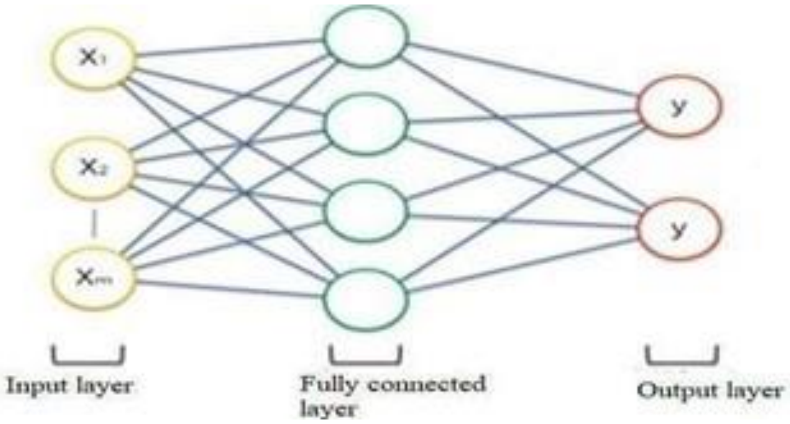

Figure 10

6. Finally, we have a softmax or sigmoid to calculate the probabilities of the classes and to classify the outputs as cat, dog, etc.

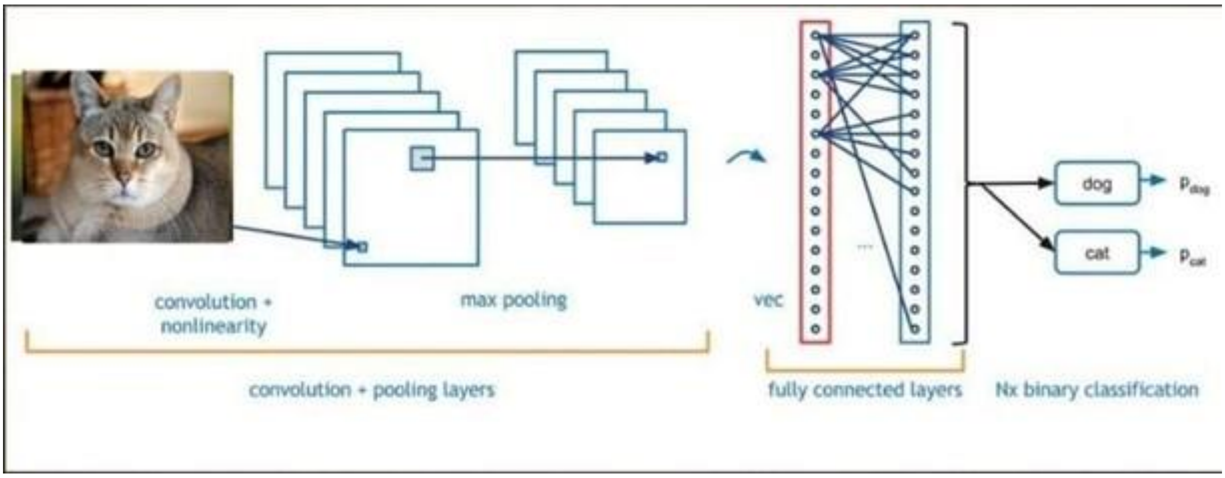

Figure 11

IV. RESULTS AND DISCUSSION

We have successfully implemented a system that can predict the animals from the input image. To evaluate the performance of the model on the dataset we made the use of Cross Entropy loss and Classification accuracy of both the test dataset and the train dataset. Initially we got an accuracy of 71.22%.

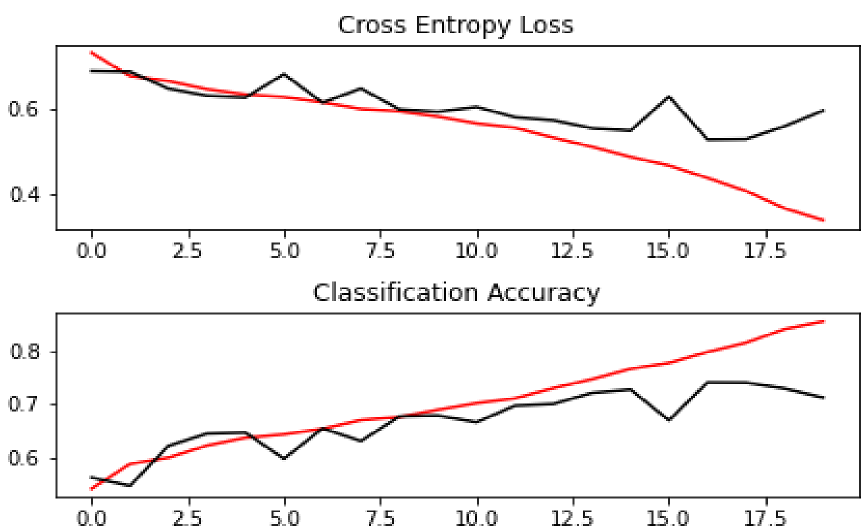

Figure 12

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:08/August-2022 Impact Factor- 6.752 www.irjmets.com

After tuning some parameters and introducing more complexity, making use of 3-block VGG model with data augmentation, to the model we were able to get the accuracy up to 86.63%.

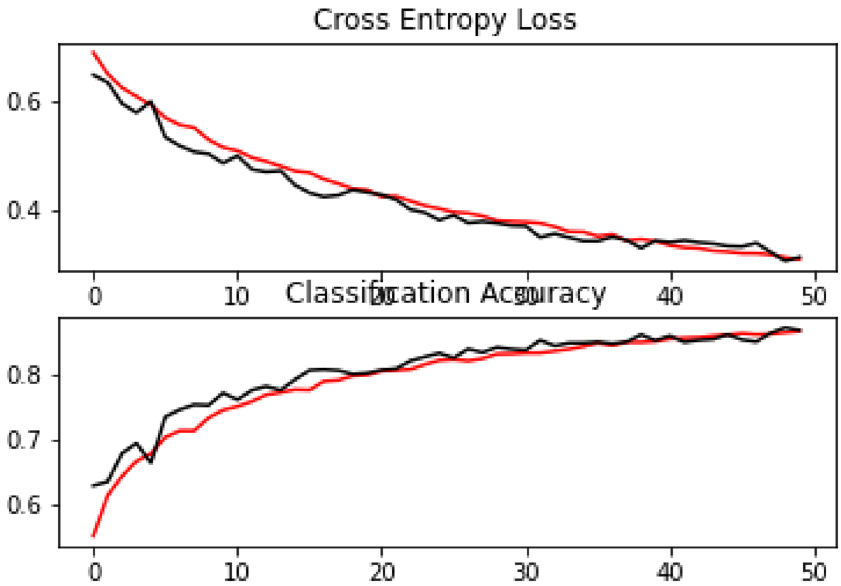

Figure 13

By making the use of VGG-16, a convolution neural network with a depth of 16 layers, the accuracy we got was 97.67%.

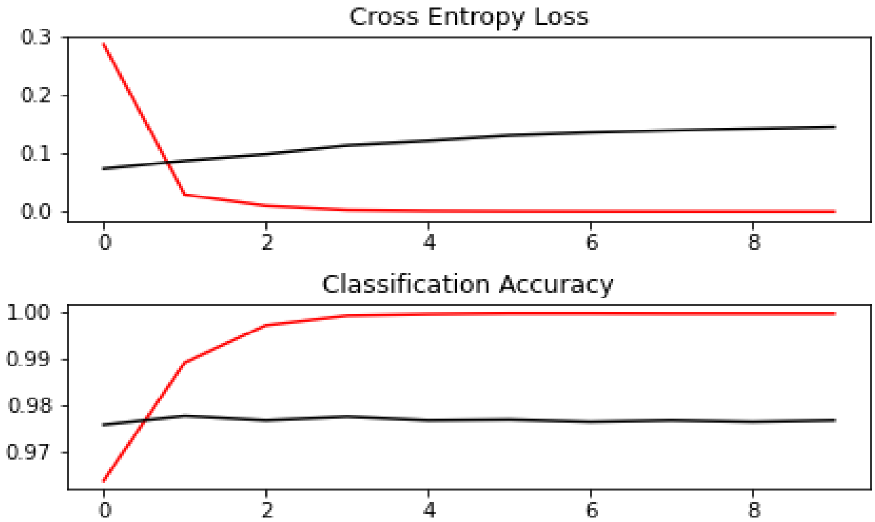

Figure 14

Table 1. Comparison of models

SN.	Model Type	Depth	Accuracy
1	Model-A	1	71.220%
2	Model-B	2	75.472%
3	Model-C	3	78.471%
4	Model-D	3	86.625%
5	Model-E	16	97.668%

Model-A – Initial model

Model-B – 2 Block model

Model-C – 3 Block model

Model-D – 3 Block model (with data augmentation)

Model-E – VGG-16

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:08/August-2022 Impact Factor- 6.752 www.irjmets.com

V. CONCLUSION

Thus, this project successfully utilizes Convolutional Neural Network (CNN) algorithm to detect wild animals. So main objective of the proposed work is to benchmark different techniques used to detect the faces of the various animals and to build a unified model which classifies two different species of animals. The algorithm classifies animals efficiently with a good accuracy and also display the image of the detected animal is for a better result so that it can be used for other purposes such as detecting wild animals entering into human areas and to prevent wildlife poaching and even human animal-conflict. The accuracy of the prediction of the previous systems were low and it was a necessity to increase it before implementing in our model. Now making use of the extensive dataset which was available online open source and applying some pre-processing techniques, we can make an apt model which can be used for detection purpose

ACKNOWLEDGEMENTS

I would like to thank Dr. Mohammed Ahmed Shaikh for guiding me throughout the project.

VI. REFERENCES

[1] A. Yuille, P. Hallinan, and D. Cohen, “Feature Extraction from Faces Using Deformable Templates,” Int’l J. Computer Vision, vol. 8, no. 2, pp. 99-111, 1992.
[2] T.F. Cootes and C.J. Taylor, “Locating Faces Using Statistical Feature Detectors,” Proc. Second Int’l Conf. Automatic Face and Gesture Recognition, pp. 204-209, 1996.
[3] M. Kirby and L. Sirovich, “Application of the Karhunen-Loe`ve Procedure for the Characterization of Human Faces,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 1, pp. 103-108, Jan. 1990.
[4] D. Chai and K.N. Ngan, “Locating Facial Region of a Head-and Shoulders Color Image,” Proc. Third Int’l Conf. Automatic Face and Gesture Recognition, pp. 124-129, 1998.
[5] J.L. Crowley and J.M. Bedrune, “Integration and Control of Reactive Visual Processes,” Proc. Third European Conf. Computer Vision, vol. 2, pp. 47-58, 1994.
[6] J.L. Crowley and F. Berard, “Multi-Modal Tracking of Faces for Video Communications,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 640-645, 1997.
[7] H.P. Graf, T. Chen, E. Petajan, and E. Cosatto, “Locating Faces and Facial Parts,” Proc. First Int’l Workshop Automatic Face and Gesture Recognition, pp. 41-46, 1995.
[8] I. Craw, H. Ellis, and J. Lishman, “Automatic Extraction of Face Features,” Pattern Recognition Letters, vol. 5, pp. 183-187, 1987.
[9] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. In IEEE Patt. Anal. Mach. Intell., volume 20, pages 22–38, 1998.
[10] Sahoolizadeh, Sarikhanimoghadam and Dehghani (2008) “Face Detection using Gabor Wavelets and Neural Networks”, World Academy of Science, Engineering and Technology, Vol. 45, pp552- 554.
[11] Burl, M.C., Leung, T.K., and Perona, P.(1995). Face localization via shape statistics. Int. Workshop on Automatic Face and Gesture Recognition, Zurick.
[12] T. Agui, Y. Kokubo, H. Nagashashi, and T. Nagao, “Extraction of Face Recognition from Monochromatic Photographs Using Neural Networks,” Proc. Second Int’l Conf. Automation, Robotics, and Computer Vision, vol. 1, pp. 18.8.1-18.8.5, 1992.
[13] R. Fe´raud, “PCA, Neural Networks and Estimation for Face Detection,” Face Recognition: From Theory to Applications, H. Wechsler, P.J. Phillips, V. Bruce, F. Fogelman-Soulie, and T.S. Huang, eds., vol. 163, pp. 424-432, 1998.
[14] R. Fe´raud and O. Bernier, “Ensemble and Modular Approaches for Face Detection: A Comparison,” Advances in Neural Information Processing Systems 10, M.I. Jordan, M.J. Kearns, and S.A. Solla, eds., pp. 472-478, MIT Press, 1998.
[15] R.C. Gonzalez and P.A. Wintz, Digital Image Processing. Reading: Addison Wesley, 1987.
[16] K. Fukunaga, Introduction to Statistical Pattern Recognition. New York: Academic, 1972.
[17] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio, “Pedestrian Detection Using Wavelet Templates,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 193-199, 1997.