Template based Animal face recognition system using Deep Learning

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 recognizes 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 wild animals in their prevalent environment is crucial. This project proposes to develop an algorithm for detecting animals in the wild.This proposed work develops an algorithm to detect the animals in wildlife. Since there are many different animals, manually identifying them can be an arduous task. This algorithm classifies animals 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.

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.

FACE RECOGNIZATION:

DIFFERENT APPROACHES OF FACE RECOGNITION:

There are two prevailing approaches for the face recognition problem: Geometric(feature based) and photometric(view based). As researchers' interest in face recognition grew, several different algorithms were developed.

There are two main approaches to recognition algorithms:

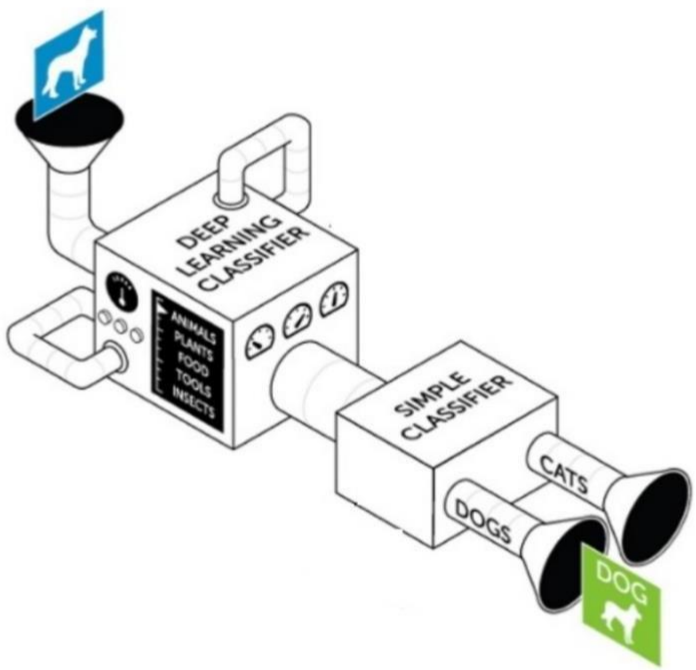
1. **Geometric:** It focus on geometrical relationship between facial landmarks, or in other words the spatial configuration of facial features. That means that the main geometrical features of the face like the eyes, nose and mouth are first located then faces are classified on the idea of various geometrical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

Fig:1.1

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 object s defined by a gradient map, which is made up of an array of surface normals (Zhao and Chellappa, 2006).

FACE DETECTION:

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 colour 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 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).



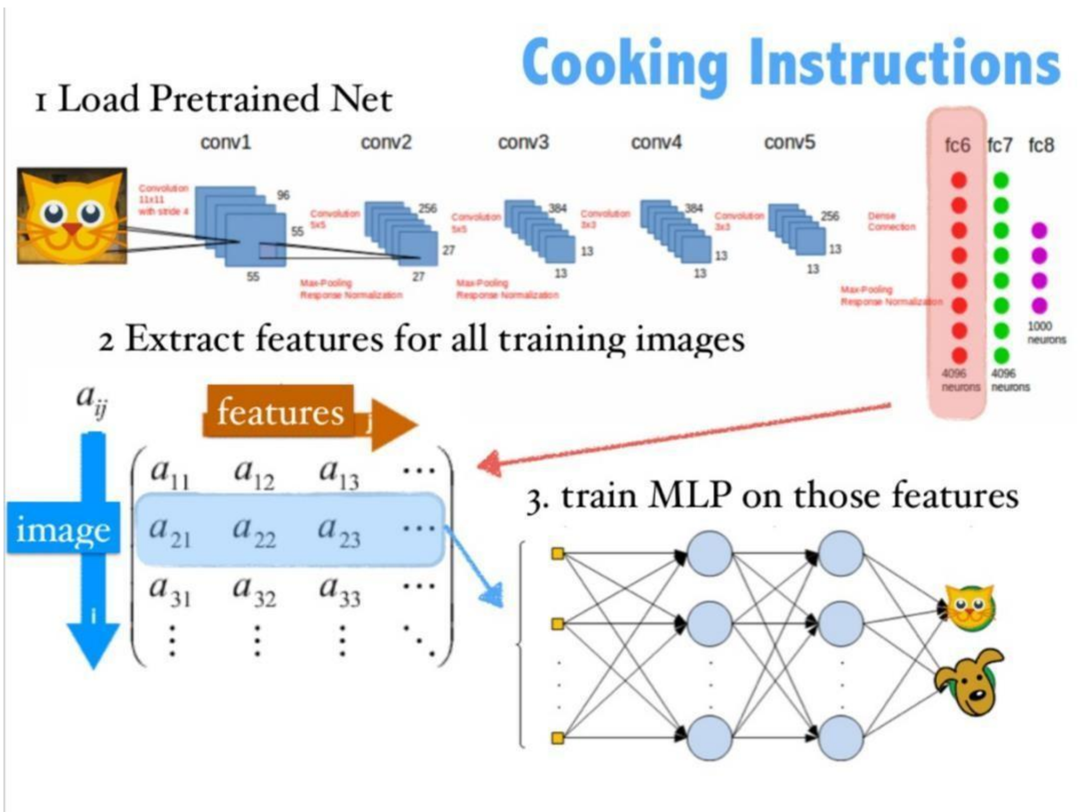
The face detection system is comprised into the following steps: -

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 use 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.



Review of Literature:

Face detection is a computer technology that determines the location and size of animal face in random (digital) image. The facial features are detected and any other objects in digital image such as trees, buildings and bodies etc. are ignored from. It can be cosidered as a specific ‘case of object-class detection, where the job is to locate and finding the sizes of all objects in an image that belong to a particular class. Face detection, can be regarded as a more general ‘case of face localization. In face localization, the aim is to find the locations and sizes of a known number of faces (usually one). Basically, there are two approaches to detect facial part in the given image i.e., feature base and image base approach. Feature base approach attempts to extract features of the image and match it against the knowledge of the face features. While image-base approach seeks the possible match between training and testing images.

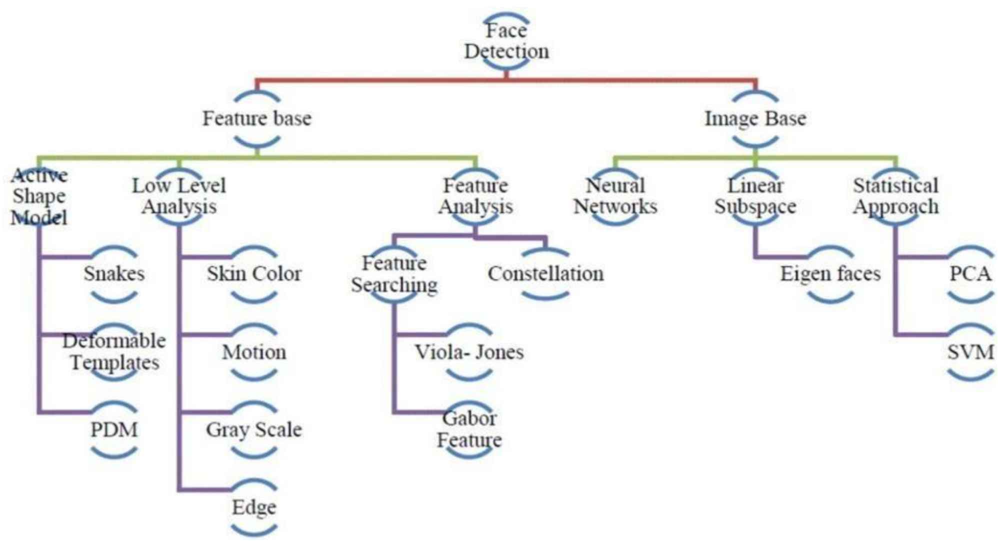


Fig3.1:Approaches Of Face Detection

FEATURE BASE APPROCH

Active Shape Model.

Active shape models concentrate on complex non-rigid features such as actual physical and higher-level appearance of features. Means that Active Shape Models **(ASMs)** are designed to automatically locating a landmark points 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 the 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 builted 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. $E = E_v + E_e + E_p + E_i + E_{internal}$,where: E_v, $E_{e,}$ E_p, E_i, $E_{internal}$ are external energies resulting from valley, edges, peak and image brightness and internal energy.

PDM (Point Distribution Model): 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 a 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 a 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.

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. Under certain lighting conditions, color is orientation invariant. This property makes motion estimation even more 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,etc).There are Majorly three different face detection algorithms are available based onRGB, YCbCr, and HSI color space models. In the implementation of the algorithms there are three major steps viz.

i. Classify the skin region in the color space

ii. Apply threshold to mask the skin region and

iii. Draw bounding box to extract the face image.

Crowley and Coutaz proposed the most basuc 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 region such as legs, arms, etc. Cahi and Ngan [27] 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 colour 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 are 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.

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. 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 an 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.

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 work's. 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.

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.

Feature Searching 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 [39] 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.

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.Finally, the distances are link to thd 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

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

Fig3.2: Gabor Feature Method

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. The Image-Based Approach

Image Base Approach

Neural Network

In many pattern recognition 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. [43] 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 [44]. 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 [45, 46, 47] proposed a detection approach based on auto associative neural networks. The concept is based on [48], 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.

Linear SubSpace Method

Eigenfaces method:

Kohonen demonstrated the use of eigenvectors 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 eigenvectors 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.

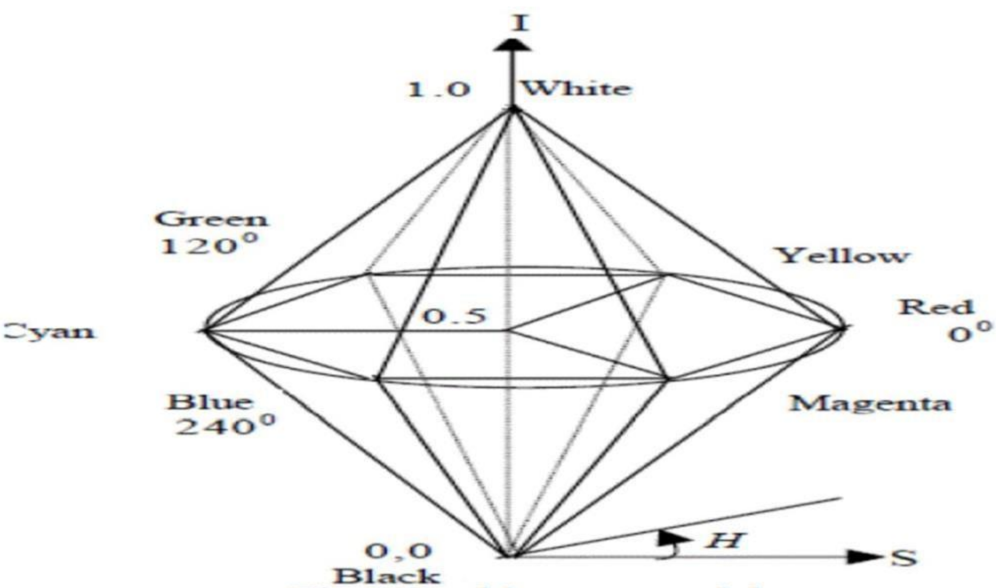


Fig 3.3: Eigen faces method

Statistical Approach

Support Vector Machine (SVM)

SVMs were first introduced by Osuna et al. 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 used to detect faces and pedestrians in the wavelet domain.

Existing System:

The System basically consists of parts:

- Creating a CNN model.
- Training the model on the dataset.
- Evaluating the model.
- Creating a custom classifier
- Making a prediction of the image.

Drawbacks of existing system:

- Less effective in low light condition
- Facial recognition system can tell the difference between identical twins.
- Less effective if the angle of the face is not properly inclined towards camera.
- Less accurate in low resolution image.

IMPLEMENTATION DETAILS:

A. Animal detection using template matching algorithm:

In this paper, reviews on different object detection algorithms were proposed. Considering efficiency, proposed system has low false positive rate and false negative rate.

Template Matching

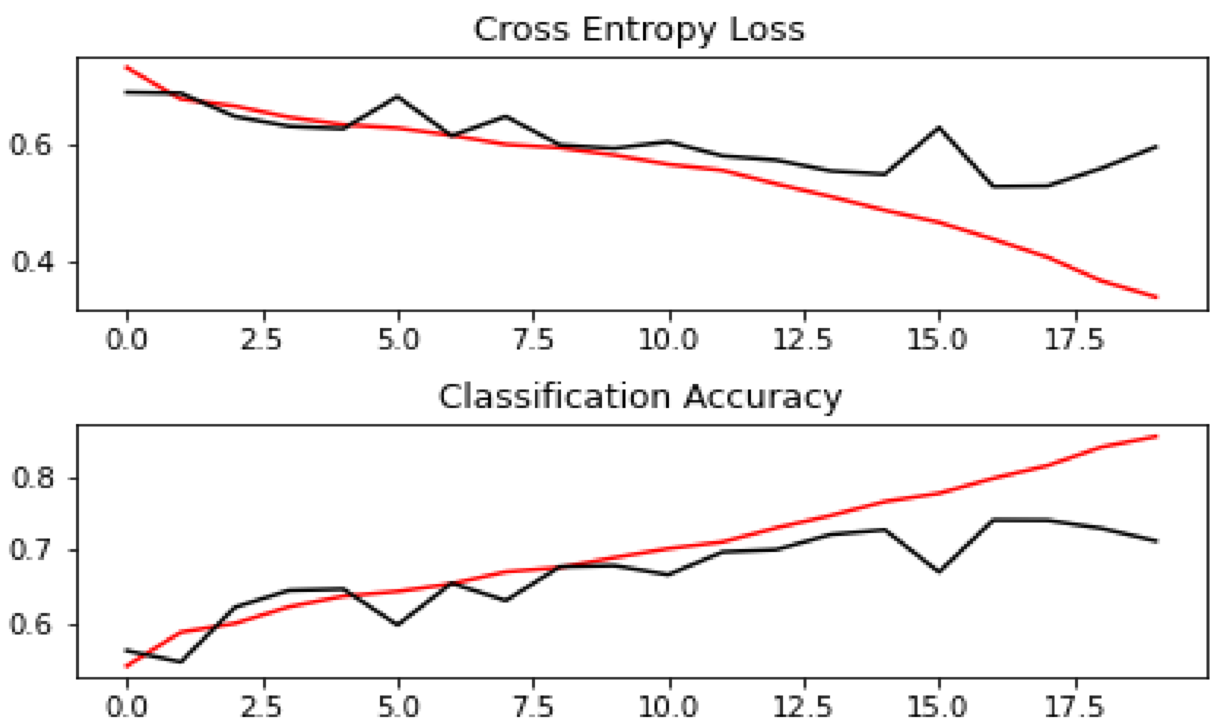
Template matching is a method for identifying small parts of an image which should match the template image. Normalized cross correlation is introduced to perform template matching. In signal processing, cross-correlation is a measure of alikeness between two waveforms as a component of time-slack applied to one of the waveforms.
That is also known as the sliding point product or/and sliding inner-product. Most commonly, template matching is used for searching a long- duration signal for an identified feature. For applications involving image processing techniques to find the brightness of an image, the

template can differ due to lighting and exposure surroundings, so the images can first be normalized. This is usually done at each step by subtracting the mean and separating by the standard deviation. In this paper we have discussed the feature based template matching technique using NCC.

B. *Identifying, Counting, and Describing Wild Animals with images with Deep Learning*

In this research, the location and activities of animals in the wild is known earlier using deep learning. This paper analysis the ability to involuntarily and accurately gather image data, also a motion sensor is also present for collecting the movements of wildlife. Although, extracting data from these images remains expensive, sustained, physical task. It’s observable that such information can be automatically extracted by using deep learning.

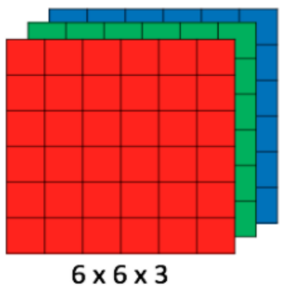
A Deep convolutional neural network is trained to recognize, count, and illustrate the behaviors of different breeds from multiple image Snapshot dataset. The network can automatically recognize animals with accuracy of 93.8%. More efficiently, if the system classifies only self- confident images, it automates animal identification to 99.3% of data. But still the performance remains likewise 96.6% accurate, saving more than 8.4 years of human labeling effort (over 17,000 hours) on multiple image dataset. This effectiveness highlights the importance of using deep neural networks to automate the extraction of data from images.The results of this study emphasize that this technology can enable inexpensive, significant, and real- time collections in the natural habitats of large numbers of animals.



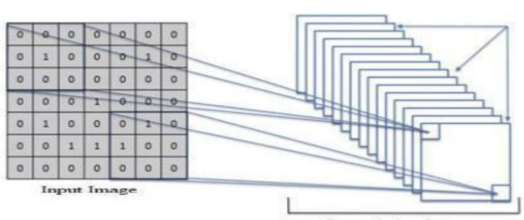
C. *Creating a CNN Model :*

A convolutional neural network (CNN) is a subset of artificial neural networks which uses perceptron, of a machine learning algorithm for supervised learning to analyze large amount of data. CNNs can be used for image processing and natural language processing (NLP) applications and any kind of cognitive tasks. A convolutional neural network (CNN) has an input layer, an output layer and many numbers of hidden layers. Few of these layers are convolved, using mathematical models to carry on results to succeeding layers.

- CNN image classifications takes 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).

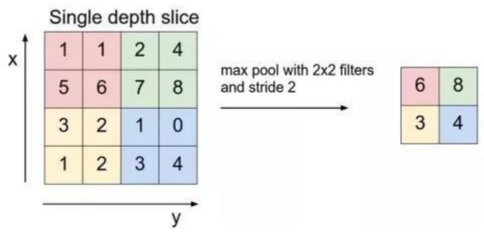


- The next layer is called, CONV layer. Convolution is the first layer which extract features from an input image. Convolution maintains the relationship between pixels by utilizing image features using small squares of input data. It is a mathematical operation This layer will calculate the output of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

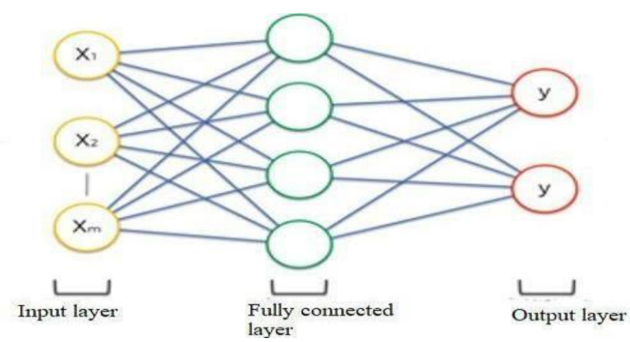


- Third layer is called RELU layer, which will apply a unit wise activation function. This makes the size of the volume unaffected.

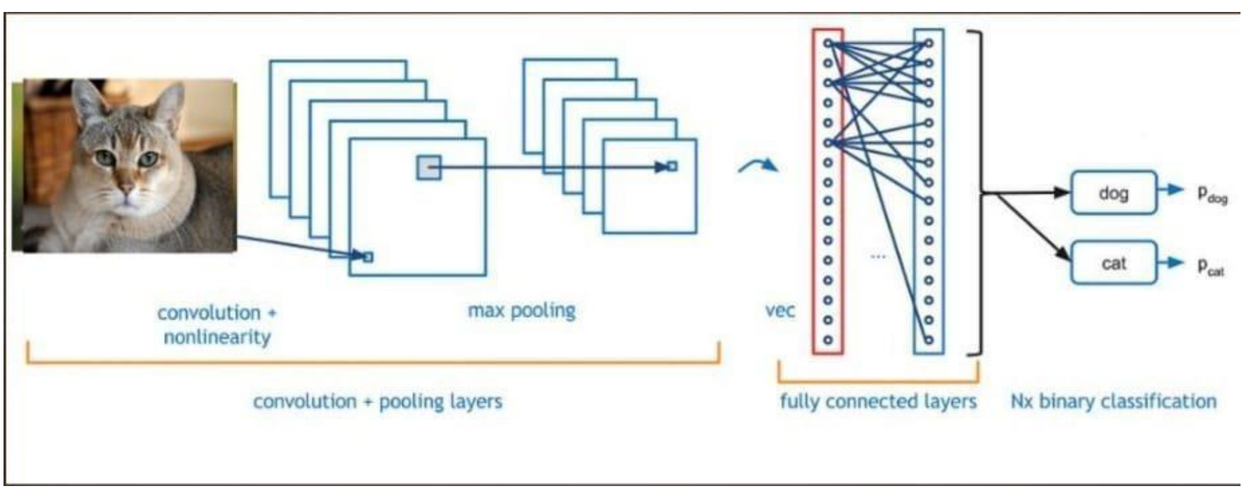
- The next layer is called the POOL layer, which performs down sampling operation along the spatial dimensions (width and height) resulting in size such as

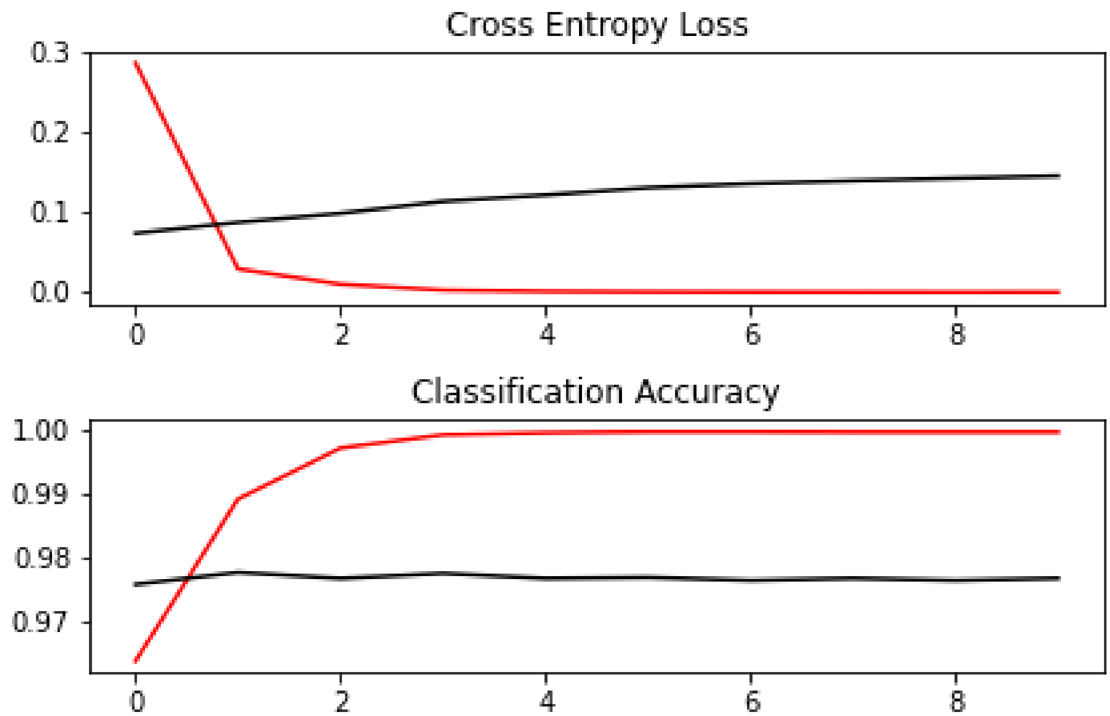


- Last layer is FC (i.e. fully-connected) layer; we flattened our matrix into vector and feed it into a fully connected layer like a neural network .this layer will calculate the class scores, resulting in volume of size. Like traditional Neural Networks, each neuron in this layer is connected to all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



- Finally we have an softmax or sigmoid to classify the outputs as cat,dog etc.





FINAL MODEL ACCURACY GRAPH

Conclusion:

Thus this project uses Convolutional Neural Network (CNN) algorithm to detect wild animals. The algorithm classifies animals efficiently with a good accuracy and also displays the image of the detected animal 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.

References

M. Kass, A. Witkin, and D. Terzopoulos, “Snakes: Active Contour Models,” Proc. First IEEE Int’l Conf. Computer Vision, pp. 259-269, 1987. [1]

Y.H. Kwon and N. da Vitoria Lobo, “Face Detection Using Templates,” Proc. Int’l Conf. Pattern Recognition, pp. 764-767, 1994. [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]

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. [4]

R. Kjeldsen and J. Kender, “Finding Skin in Color Images,” Proc. Second Int’l Conf. Automatic Face and Gesture Recognition, pp. 312- 317, 1996. [5]

D. Saxe and R. Foulds, “Toward Robust Skin Identification in Video Images,” Proc. Second Int’l Conf. Automatic Face and Gesture Recognition, pp. 379-384, 1996. [6]

I. Craw, H. Ellis, and J. Lishman, “Automatic Extraction of Face Features,” Pattern Recognition Letters, vol. 5, pp. 183-187, 1987. [7]

I. Craw, D. Tock, and A. Bennett, “Finding Face Features,” Proc. Second European Conf. Computer Vision, pp. 92-96, 1992. [8]

H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. In IEEE Patt. Anal. Mach. Intell., volume 20, pages 22–38, 1998. [9]

Sahoolizadeh, Sarikhanimoghadam and Dehghani (2008) “Face Detection using Gabor Wavelets and Neural Networks”, World Academy of Science, Engineering and Technology, Vol. 45, pp552-554. [10]

Burl, M.C., Leung, T.K., and Perona, P.(1995). Face localization via shape statistics. Int. Workshop on Automatic Face and Gesture Recognition, Zurick. [11]

Yow, K.C. and Cipolla, R.(1997). Feature-based human face detection. ImageVision Comput., 15(9). [12]

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. [13]

S.-H. Lin, S.-Y. Kung, and L.-J. Lin, “Face Recognition/Detection by Probabilistic Decision- Based Neural Network,” IEEE Trans. Neural Networks, vol. 8, no. 1, pp. 114132, 1997. [14]

M. Propp and A. Samal, “Artificial Neural Network Architectures for Human Face Detection,” Intelligent Eng. Systems through Artificial Neural Networks, vol. 2, 1992. [15]

H. Rowley, S. Baluja, and T. Kanade, “Human Face Detection in Visual Scenes,” Advances in Neural Information Processing Systems 8, D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo, eds., pp. 875- 881, 1996. [16]

R.C. Gonzalez and P.A. Wintz, Digital Image Processing. Reading: Addison Wesley, 1987. [17]

K. Fukunaga, Introduction to Statistical Pattern Recognition. New York: Academic, 1972. [18]

I.T. Jolliffe, Principal Component Analysis. New York: SpringerVerlag, 1986. [19]

T. Kohonen, Self-Organization and Associative Memory. Springer 1989. [20]

M.M. Loe`ve, Probability Theory. Princeton, N.J.: Van Nostrand, 1955. [21]

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. [22]

E. Osuna, R. Freund, and F. Girosi, “Training Support Vector Machines: An Application to Face Detection,” Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 130-136, 1997. [23]

C. Papageorgiou, M. Oren, and T. Poggio, “A General Framework for Object Detection,” Proc. Sixth IEEE Int’l Conf. Computer Vision, pp. 555-562, 1998. [24]

K.-K. Sung and T. Poggio, “Example-Based Learning for ViewBased Human Face Detection,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 1, pp. 39-51, Jan. 1998. [25]