Animate Object Detection Using Deep Learning

Aishwarya Patil, Pooja Maske, Ashvini Maske, Nitin ILhe
Student, Student, Student, Student
Computer Engineering,
AISSMS Institute Of Information Technology, PUNE, INDIA

Abstract ***The face is one of the easiest ways to distinguish the individual identity of each other. Face recognition is a personal identification system that uses personal characteristics of a person to identify the person's identity. Human face recognition procedure basically consists of two phases, namely face detection, where this process takes place very rapidly in humans and animals, except under conditions where the object is located at a short distance away, the next is the introduction, which recognize a face as individuals. Stage is then replicated and developed as a model for facial image recognition (face recognition) is one of the much-studied biometrics technology and developed by experts. There are two kinds of methods that are currently popular in developed face recognition pattern namely, Eigenface method and Fisherface method. Facial image recognition Eigenface method is based on the reduction of face-dimensional space using Principal Component Analysis (PCA) for facial features. The main purpose of the use of PCA on face recognition using Eigen faces was formed (face space) by finding the eigenvector corresponding to the largest eigenvalue of the face image. The area of this project face detection system with face recognition is Image processing. The software requirements for this project is Net- Beans/Eclipse Java software.In face recognition systems, variables such as direction of light, facial expression and reflection are making difficult to identify. Thus, in recent Years, Convolutional Neural Network (CNN) models, which are deep learning models as an alternative to traditional feature extraction and artificial neural network methods, have begun to be developed. Data mining is used to analyze human activities or machine learning techniques, these are infer properties such as the gender or age of the people. Different algorithms are implemented on static and nonstatic conditions. Static conditions include stable and uniform background, identical poses, similar illumination, neutral frontal face Non static conditions include position, partial occlusion and facial hair, which makes recognition process a complex problem. The main stages for face recognition include face detection, feature representation and classifications. In this work we present face detection techniques, methods used, their performance, limitations and proposed a new technique for Face Detection based on Viola and Jones algorithm and principal component analysis.***

IndexTerms - face detection,Gender Detection, Classification, Eigen face, PCA,java,Data Analytics, Machine learning, Object detection, CNN (Convolution Neural Network), Facial Expression Detection, deep learning

1 INTRODUCTION

Object detection is one of the most fundamental yet challenging problems in computer vision community. The human brain is a very powerful machine. We see multiple images every second and process them without realizing how the processing is done. But**,** that is not the case with machines. The first step in image processing is to understand how to represent an image so that the machine can read it? In simple terms every image is an arrangement of dots **(**pixels**)** arranged in a special order**.** If you change the order or color of a pixel, the image would change as well. To understand an image it's extremely important for a network to understand how the pixels are arranged. What we did for image processing that we were trying to extract features from an image by using **a special arrangement** of images.

Deep convolutional neural networks (CNNs) have dominated many tasks in computer vision. For object detection, region based CNN detection methods are the main paradigm. It is such a rapidly developing area that three generations of region –based CNN detection models, from the R-CNN, to the fast R-CNN, and finally the Faster R-CNN, used with increasingly better accuracy and faster processing speed.

Classification:

Image classification is one of the most challenging issues, for computerized machine, so to solve this issue we propose this work. Which includes classification of objects into animals and the Human beings and Human beings are again classified into male and female. Using the Convolution Neural Network Model.

Facial Expression:

For facial expression we use SVM classifier baseline model and develop a CNN to classify these emotions we use VGG-16 and ResNet50 art architecture.FACS (Facial Action Coding System) is a system that classifies the human facial movements by their appearance on the face using the action unit (AU). An AU is one of 46 atomic elements of visible facial movements.

Gender Detection:

The classic task of face gender recognition has recently attracted new attention, mostly due to the availability of large sets of images collected in the wild. Applications are readily found in many areas, for example, in the analysis of gender bias in news media content. The emphasis of this new phase of research is on avoiding images collected under controlled conditions (e.g. in the background, pose or illumination), and focusing efforts on the more challenging case of natural images.

2. SURVEY DETAILS

SAMIKSHA AGRAWAL AND PALLAVI KHATRI [1]

Facial Expression Detection techniques: Based on Viola and Jones algorithm and principle component analysis. In thus using Viola and Jones algorithm and principle analysis component which consist of Eigen face approach. This system can recognize animate object's expression, but this system is more time consuming and detection rate is low. All over this technique used in this work to detect the human facial expression and recognize them on the basis of accuracy and computational time.

Xushen Han, Dajiang Zhou, Shihao Wang, and Shinji Kimura[2]

An FPGA-Based Memory-Efficient Reconfigurable Processor for forward and backward propagation of CNN. Which consist CNN-MERP that incorporates an efficient memory hierarchy that significantly reduces bandwidth from multiple optimization including on\off chip data allocation, flow optimization on and data reuse. It has advantages as the computational component of these architectures are extensible i.e.

higher performance can be achieved by leveraging parallelism and having limitation that it can't utilize multiple FPGA's to accelerate CNN.It needs to enhance the computational density in hardware to attain higher speed of acceleration for CNN's.

Xiaofeng Ning, Wen Zhu, Shifeng Che[3]

propose recognition, object detection, segmentation of white background photos based on deep learning. Due in advance of deep learning uses the white background, photos object detection, object segmentation. If using the variety of convolutional networks. Labeling of objects tough. Complex background image can be conducted ware application range in this method used to obtain an end to end deep learning solution for pixel level semantic image segmentation problem.

Luis Tobias, AurelinDucournau Francois, Rousseau, Gregoire Mercier, Ronan Fablet [4]

Convolutional Neural Network for object recognition on mobile devices using different methodology machine learning, deep learning. Object detection and the conventional networks. In this case study the object detection on mobile devices based on the object detection. The most important function of this is to develop iterative interfaces on mobile platform based on object recognition it is domain specific real time application. In this the highly accurate domain specific object recognition but it depends on the large input.

Zipeng Deng, Lin Lei, Hao Sun [5]An enhanced deep Convolutional Neural Network for densely packed object detection in remote sensing Images. This uses object detection, remote sensing image and densely packed. The most important feature of it is a remote sensing application. It is available publicly with same address with high revolution remote sensing and high capacity. Limitation of it is the small object over large dataset extract good objects and dataset demonstrate effectiveness.

Chungkeun Lee, H. Jin Kim,Kyeong Won Oh[6]

Comparison of faster R-CNN models" proposed object detection using human body, pattern recognition, vision system methodologies, for image classification. The computation time for faster R-CNN is hundreds milliseconds per image. The precision cabinet models are ResNet>VGG16>Google>caffeNet but the computation time of those models is reverse order of followed orders. The computation time of faster R-CNN is hundreds milliseconds per image which is acceptable.

Huaizu Jiang, Erik Learned-Miller[7]

Gender Classification by Deep Learning on Millions of Weakly Labelled Images" proposes /focuses on the sub problem of gender recognition. It uses the hardware such as Graphics Processing Units (GPUs) to train the networks and the creation of very large sets of images obtained in uncontrolled conditions that are labelled by Gender. In this study, the performance ever reported on the LFW test sets having an accuracy from 97.31% to 98.90%. A deeper CNN out-performed a shallower one with the same training data that using a larger bounding for the face region can improve performance.

3. SYSTEM DISCRIPTION

In this paper, we proposed a new method which detects animate objects and classifies them as a in the two groups as an animal and the human beings. To develop Animate Object Detection Using Deep Learning system which classifies Human Beings and Animals and Identify the Facial expression, gender and age of human beings. In our proposed system, the Classification of gender, emotions and age is easy to identify, also useful for classification of domestic and wild animals.

As shown in following fig (a), System takes the input in the form of image ,then this input image is processed and that image is compared with our datasets which are trained by us .If this image is present in our datasets it gives the proper output and shows the classifications else window message will be display on the screen. For this classification we use the R-CNN algorithm and Viola-Jones Algorithm for detecting the facial expressions.

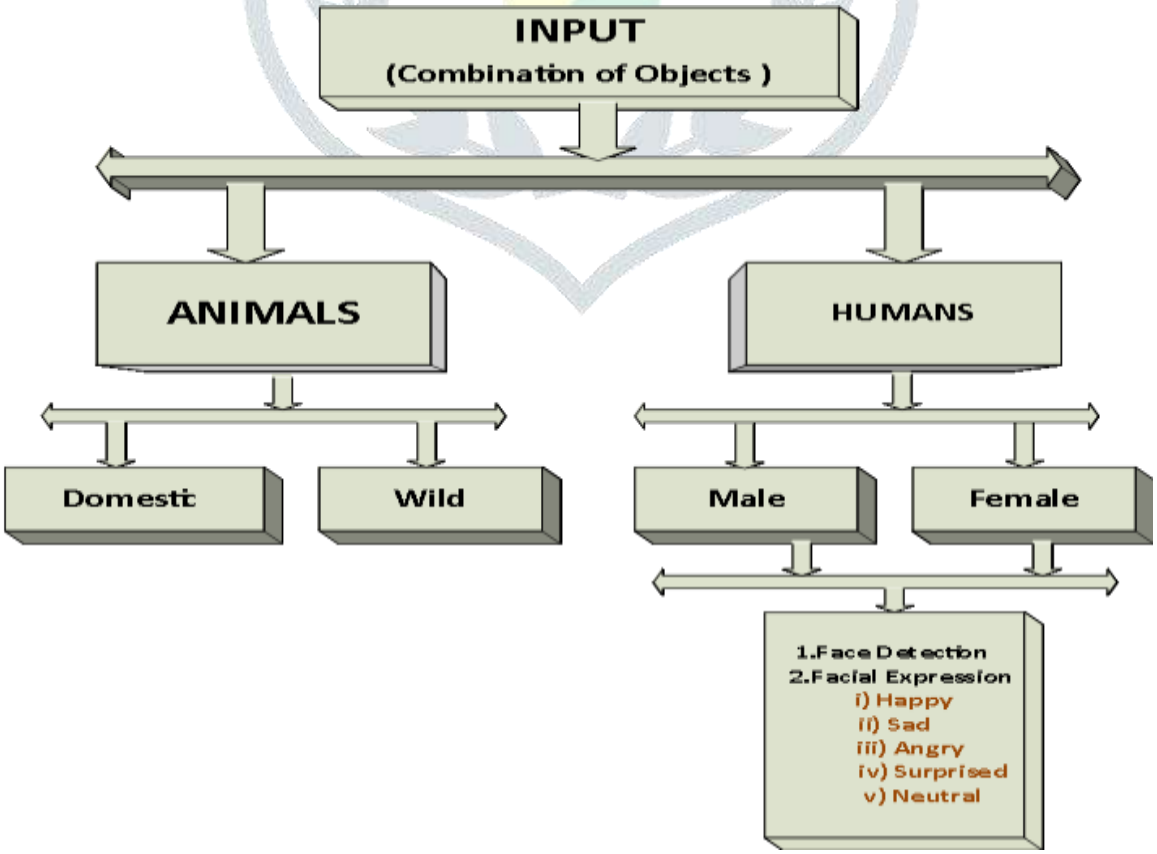

Fig 1: System Architecture

4. PROPOSED SYSTEM ARCHITECTURE

1. CLASSIFICATION: THE MOST WELL-KNOWN PROBLEM IN COMPUTER VISION. IT CONSISTS OF CLASSIFYING AN IMAGE INTO ONE OF MANY DIFFERENT CATEGORIES. ONE OF THE MOST POPULAR DATASETS USED IS IMAGENET, COMPOSED OF MILLIONS OF CLASSIFIED IMAGES.
2. Localization: Similar to classification, localization finds the location of a single object in the image.
3. Instance segmentation: In object detection not only find objects in an image, but find a pixel by pixel mask on each of the detected objects. We refer to this problem as an instance or object segmentation.
4. Object detection: Object detection is the problem of finding and classifying a variable number of objects on an image.
5. Face detection: The detection of face is done here.
6. Counting: One simple but often ignored use of object detection is counting. The ability to count people and Animals is a real world need that is broadly required for different types of systems using images.

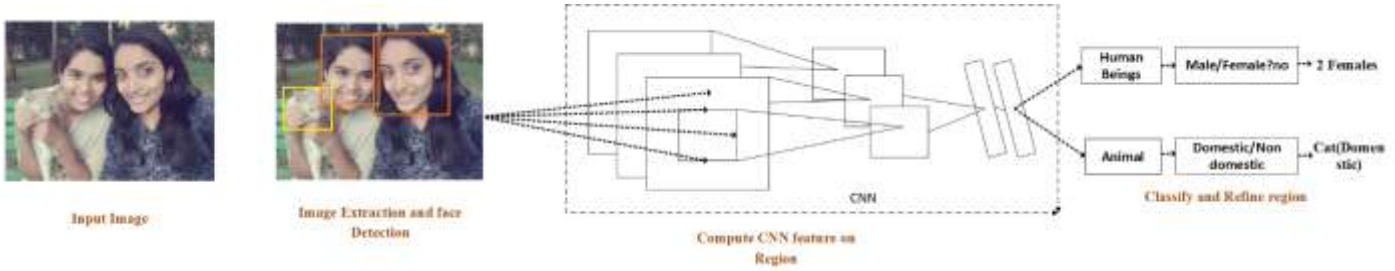

Fig. 2: Proposed architecture of System

4.1 Classical Approach:

Although there have been many different types of methods throughout the years, we want to focus on the two most popular ones.
The first one is the Viola-Jones framework proposed in 2001 by Paul Viola and Michael Jones in the paper Robust Real-time Object Detection. The approach is fast and relatively simple, so much that it's the algorithm implemented in point-and-shoot cameras which allows real-time face detection with little processing power.

4.2 Deep learning approach:

OverFeat: One of the first advances in using deep learning for object detection was OverFeat from NYU published in 2013. They proposed a multi-scale sliding window algorithm using Convolutional Neural Networks (CNNs).
R-CNN: After OverFeat, Regions with CNN features or R-CNN from Ross Girshick, et al. at the UC Berkeley was published which boasted an almost 50% improvement on the object detection challenge. What they proposed was a three stage approach: as shown in figure above.
Extract possible objects using a region proposal method.
Extract features from each region using a CNN.
Classify each region with SVMs.
TensorFlow: Creating accurate machine learning models capable of localizing and identifying multiple objects in a single image remains a core challenge in computer vision. The TensorFlow Object Detection API is an open source framework built on top of TensorFlow that makes it easy to construct, train and deploy object detection models.
For object detection and classification we have trained the following datasets:
1. PASCAL VOC [11]
2. Caltech 256[12]
3. ImageNet [13]
4. Coco [14]
5. CASIA [15]

For reorganization of images we trained the above datasets in TensorFlow. Following are some steps to train the datasets.
Tensorflow Object Detection API uses the TFRecord file format, so at the end we need to convert our dataset to this file format
Training can be either done locally or on the cloud (AWS, Google Cloud etc.). If we have GPU (at least more than 2 GB) at home then we can do it locally otherwise we would recommend to go with the cloud. Here we do it locally.
1.Create your own PASCAL VOC format dataset
2.Generate TFRecords from it
3.Configure a pipeline
4.Visualize
With the dataset prepared, we create the corresponding label maps. Then we generate the TFRecord file format. The instructions should be self-explanatory to cover this segment. Sample configs can be found in *object detection/samples/configs*. It is recommended that during the training we should start the evaluation job. So we then monitor the process of the training and evaluation jobs by running Tensorboard on our local machine.

5. RESULT ANALYSIS:

Our system gives 98% accuracy for face detection of human and animals .90% accuracy for gender detection. And we classify between the human beings and animals and then it also classifies between males and females from human beings. We can also classify between animals such as wild animals and domestic animals. We have studied on the R-CNN algorithm for object detection which is faster than CNN.

6. CONCLUSION AND FUTURE SCOPE:

In this paper, a new aspect has been provided by using scene viola Johns and Watson algorithm for classification of animate object and the gender detection algorithms. This proposed scheme gives near to 98% results which makes it better than previous techniques used for animate object detection and gender classification. Though it is a bit expensive and time consuming because of the use of different detection algorithm along with training algorithms which makes it robust. The result provided by this technique is directly related to the properties of the input provided. Having its applications ubiquitously, it can be applied anywhere, where one wants to recognize the face and gender of

person also for the animate object detection. Example Malls, where it for checking for how many Animals and the human are visited. Also used for the Passport for photo recognizing of animals

7. REFERENCES

1.Samiksha Agrawal and Pallavi Khatri "Facial Expression Detection Techniques: Based on Viola and Jones algorithm and Principal Component Analysis" 2015 Fifth International Conference on Advanced Computing and Communication Technologies

2.Xushen Han, Dajiang Zhou, Shihao Wang, and Shinji Kimura "CNN-MERP: An FPGA-Based Memory-Efficient Reconfigurable Processor for Forward and Backward Propagation of Convolutional Neural Networks"2016 IEEE 34th International Conference on Computer Design (ICCD)

3.Zipeng Deng, Lin Lei, Hao Sun "an enhanced deep Convolutional Neural Network for densely packed object detection in remote sensing Images"2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP).

4.Chungkeun Lee, H. Jin Kim,Kyeong Won Oh" Comparison of faster R-CNN models"2016 16th International Conference on Control Automation and Systems (ICCAS).

5.Huaizu Jiang, Erik Learned-Miller "Face detection with the faster R-CNN"2017 12th IEEE International Conference on Automatic Face and Gesture Recognition.

6.Sen Jia, Thomas Lansdall-Welfare and Nello Cristianini " Gender Classification by Deep Learning on Millions of Weakly Labelled Images" 2016 IEEE 16th International Conference on Data mining workshops.

7.A.V. Anusha, J. K. Jayasree, Anusree Bhaskar, R.P.Aneesh "Facial expression recognition and gender classification using facial pathes" in 2016 International Conference on Communication Systems and Networks(ComNet)

8.Ma. Christina D. Fernandez, Kristina Joyce E. Gob,Aubrey Rose M. Leonidas, Ron Jason J. Ravara, Argel A. Bandala, Elmer P. Dadios "Simultaneous Face Detection and Recognition using Viola-Jones Algorithm Artificial Neural Networks for Identity Verification" in 2014 IEEE 10 SYMPOSIUM.

9.http://host.robots.ox.ac.uk/pascal/VOC/

10.https://en.wikipedia.org/wiki/Caltech_101

11.https://en.wikipedia.org/wiki/ImageNet

12.https://github.com/nightrome/cocostuff

13.http://www.cbsr.ia.ac.cn/english/CASIA-WebFace-Database.html