Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 05, May-2019

Real Time Video Analytics for Object Detection and Face Identification using Deep Learning

Shrikant Jagannath Patro*
Vellore Institute of Technology,
Chennai, India

Prof. Nisha V M
Vellore Institute of Technology, Chennai
Chennai, India

***Abstract*-Video analytics is growing field in the** machine **learning and deep learning domain. This paper proposed the model that is capable of performing video analytics at large scale and faster pace and generated the appropriate inference on time. It include the detail of the algorithm for the automation in video analytics personalized cameras, security and Surveillance system using deep learning. This include much optimized algorithm for identification of the faces. The second module is consisting of object identification using deep learning and libraries that are capable of identifying 3 million object i.e. COCOAPI. Video Summarization is a technique introduced to increase the speed of investigation. The Module produced much optimized summary of video.**

keyword – Video Analytics , Object Detection , Face Identification , Video Summarization , Deep Learning.

I. INTRODUCTION

The Video analytics is growing field in analytics domain. There are lot of video resources available in libraries but due to lack of the processing infrastructure for the video they were not used effectively. Since rise of the deep learning with exceptional power to cater the need of the video analytics had made change in the life of every human being. Initially deep learning were introduced to reduce the cost and to increase the speed of the monotonous work. This had helped industries like automated vehicle manufacturing plant to increase the production of automobiles, food and beverage industry to food quality testing and packaging. Now, deep learning had started making impact on the social life of the human being. This paper primary focus is to use the deep learning for solving the social problem like identification of the thief on real time bases based on the CCTV Footage, Catch the person Red handed while committing crime in public place. Also, Woman's Safety must be ensured and can be protected well if deep learning is used for the identification of the misbehaving person at public place.

The security concern for the public and private place is the primary requirement for the owner of the shop and institutions. They spent lot of money for protecting their property with the help of the hired security agency. This can be saved by the institutions and land by automating the task for face identification and unknown baggage identification or any object caught in the camera frame. This automation for the object identification is made possible due to the introduction of the deep learning technology in the field of security and surveillance system that are developing and performing smarter operation comparatively.

This Paper introduces the usage of the libraries like dlib facial recognizer, ResNet-34 is facial recognition network. The network is trained on the 3-million images based on the labeled images of the wild (LFW). This network gives the accuracy of 99.8%. The imutils package is subordinate package along with dlib facial recognizer and ResNet-34 facial recognition network. It is specifically designed for the basic image processing task like translation, sorting contours, detecting edges, resizing, and display matplotlib images. This over all model obtained by this is used to train the network on the supplied image file and find the person face boxed in the rectangular box along with the name in the test set. The image detection technique being used is 'hog' or 'CNN' [2]. The Automated object identification is achieved with the Tensorflow object detection API. It is an open source built on the top of the Tensorflow. The object identification achieved is variable from various living to non-living creature. To achieve the automation in identification of the object the required libraries are OS, matplotlib, scipy, Numpy, Pandas, PIL, Tensorflow, skimage.transform, keras, Yolo_utils, amount of data within least amount of time. Raw feed can be processed in same place while integration and deeper analytics can be performed on cloud. The basic image processing can be achieved in low power gateway devices. The power of deep learning in crowd counting application using edge computing and basic image processing can be achieved in near real time on low powered gateway devices. Increasing the performance of the edge detection algorithm, will resolve the issue in crowd counting [6]. To achieve scalability for processing the video on demand. There are different factor that causes video analytics job to be done. The raw data processing is achieved using well engineered approach for the task distribution and processing. High scalability in this case can be achieved by distributing the component on demand. The wide utility of on-demand video analytics in found in forensic application where the video analytics is on demand, for regular inspection of law enforcement [5].

Tracking people and movement of the people is primary task in computer vision. The goal of the Tracking people and Surveillance system, object tracking is segmentation of a region of interest from video science. This will keep track of motion, positioning and occlusion. Object detection and classification of the preceding steps for the motion video surveillance. The object detection include living creature like person, animal, birds and motion tracking of the object is achieved by the spatial and temporal changes during the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 05, May-2019

video sequencing , including its presence , position , size , shape of the object. It is having application in robot vision, traffic monitoring, video in painting and monitoring [7]. The live environment of the video analytics the image data will be generated form the several devices, so there is higher chance of obtaining video in different format. In the integrated platform where the video analytics is performed at large scale we need to deal with such diverse data. The challenge is that video analytics platform must be able to perform multiple video analytics, ensure scalability, portability, adaptability are complete ingredients of the video analytics platform. This paper discuss about the SIGMA video analytics platform, specially designed to support the video from different variant of the sources and vendor. It is platform neutral analytics where more than one platform that perform analysis on video are bought under common platform to infer fruitful results out of it [8]. The Video data that is generated on daily bases had all time high. Now, In order to perform the analysis on them it is important for them to store and perform operation only on the significant portion of the image. This achieved with the Image indexing technique where the more relevant part is stored and processed while other part is discarded. This is having huge impact on the processing of the video. It is used to retrieve in time most significant information. It introduces the existing indexing technique, their implementation .They have explained the function of their in the proposed framework. It is great combination of the diverse field, in such a way that it is itself create a new learning domain for future and having powerful impact to bring dynamic change in the society [10].

The current generation of video analysis require lot of improvement, this is required to be made intelligent. This will bring this to main stream usage. This is having wide range of application like robotics, health care system and human computer interaction. This is achieved by Human action recognition (HAR). This will perform the self-analysis of the video feed and ongoing event and understand the behavior of the person. The real- time threat detection can be achieved using HAR. This paper will review different technique and method for the HAR. Actions are distinguished by the motion pattern executed by single person walking, running and hand waving etc. The method like bag of feature, multiple instance markov model, MRF Method are evaluated with experiments on these action [16].

The number of people getting educated through the Massive Online Courses are increasing in number every year. The developing countries like India and China have large population and having formal education for every person is difficult. Also, well verse with the new knowledge. So, the online courses play vital role in nurturing the development of the human skills. But, it is not enough to check whether how far person taking online courses had understand and the things explained in video. The facial analytics can be introduced during the online course to check the expression of the person taking course using that we can predict how far person is taking positively. The differentiation between interesting and not interesting topic can be achieved using this context. This result of the Facial Analytics can be generated and stored in real time for the analysis and inference can be generated [15].

II.METHODOLOGY:

Module 1: Face Identification

This Module is responsible for the face Identification. This task of face identification is divided into part, First generating encoding vector for the Training image dataset and then accept the Test image from the Test set Identify the face and display the label along the faces. The First part is consisting of dlib_facail_recognition libraries. This is responsible to generate the 128-d vector the image. This process is repeated for all the images and finally we work with only 128-d dimension vector per image to identify the faces uniquely. Once the network is sufficiently trained under the Training, We will begin with next part were the Test image from the Test set is Preprocessed. The Test image is converted into 128-dimension vector. Now, this 128-d vector of the test image is compared with the 128-d vector of the images belonging to training set. The nearest images is then identified by calculating the difference between the test image and the original image. Then the label of the nearest image is assigned to the original image.

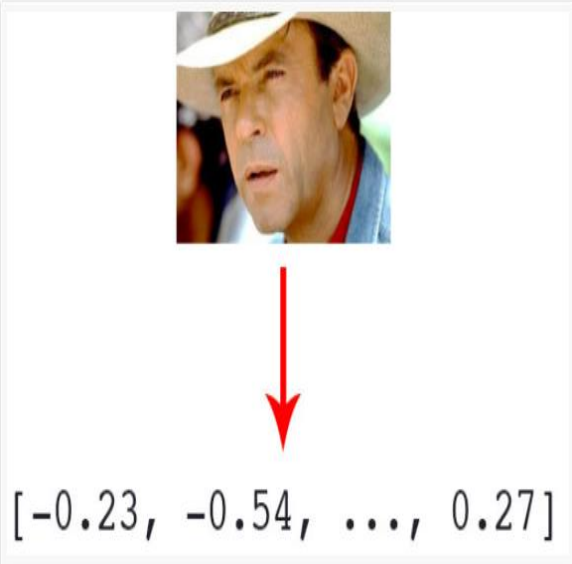


Fig 1: 128-d encoding feature vector for the images.

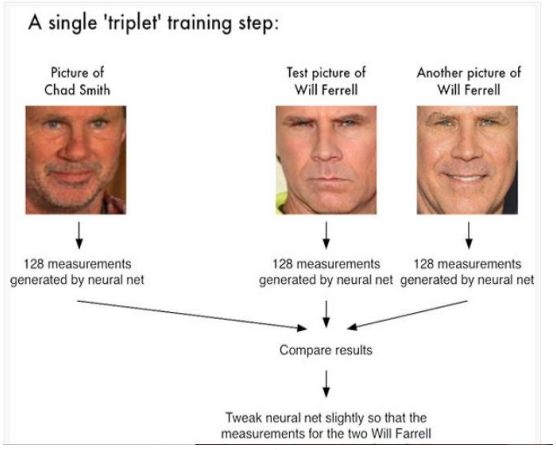


Fig 2: triplet of the training set

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 05, May-2019

Module 2: Object Detection

This Module is responsible for the unique identification of object in the image. The philosophy behind this method is divide and conquer. In division phase, the entire object is divided into M * N Grid size. For each grid, vector 'Y' will be generated. This Vector is consisting of the 8 Element. The element are listed as follows pc, bx, by, bh, bw, c0, c1, c2. Here c0, c1 and c2 are the classes of the object to be identified. The pc bit determine the presence or the absence of the images. The bx and by determine the x and y coordinate of the center of the bounding box. bh and bw will determine the height ratio and width ratio of the bounding box with respect to the grid respectively.

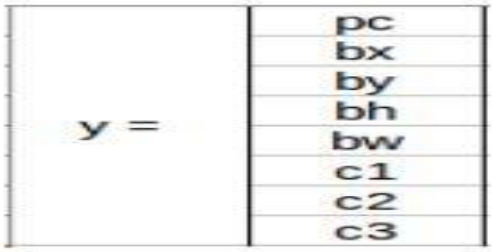


Fig 3: Output Vector

Once this vector is generated for all we can obtain the image along with the bounding boxes. Finally, images will be appearing with the bounding boxes. The summarized video is generated using the moviepy and pysrt packages in python. Moviepy library include the editor function used to edit the clip.

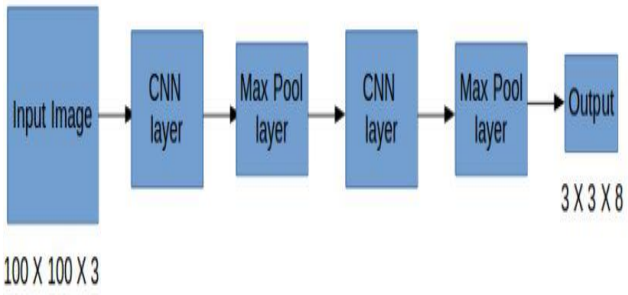


Fig 4: Convolution operation on images

1. Intersection of Union and Non-Max Compression

This is where Intersection over Union comes into the picture. It calculates the intersection over union of the actual bounding box and the predicted bonding box. Consider the actual and predicted bounding boxes for a car as shown below:

Fig 5(d) Bounding Box

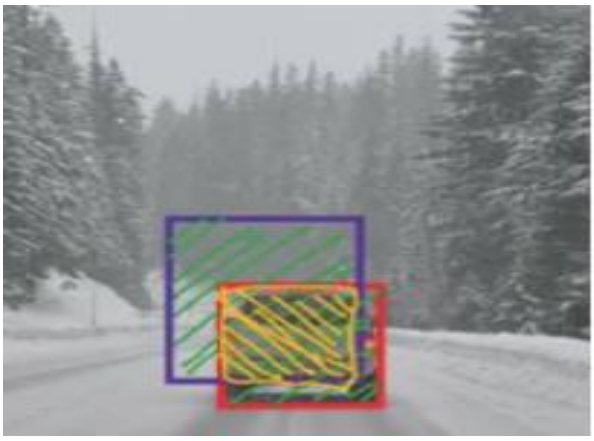


Fig 5(a): Intersection of the actual and predicted bounding box

Fig 5(b): overlapping boxes

IoU = Area of the intersection / Area of the union, i.e.

IoU = Area of yellow box / Area of green box

If IoU is greater than 0.5, we can say that the prediction is good enough. 0.5 is an arbitrary threshold we have taken here, but it can be changed according to your specific problem. Intuitively, the more you increase the threshold, the better the predictions become.

One of the most common problems with object detection algorithms is that rather than detecting an object just once, they might detect it multiple times. Consider the below image:

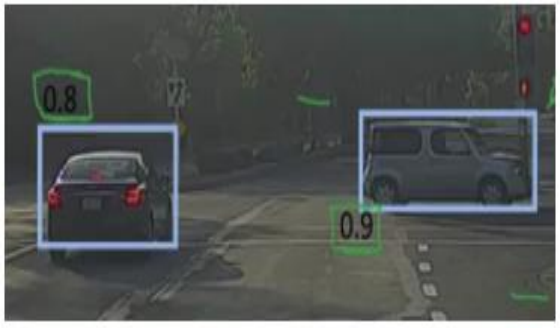


Fig 5(c): Original boxes

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 05, May-2019

Module 1: Face Identification
Algorithm**:**

Step 1**:** Import library imutils, face_recognition, argparse, pickle, cv2 and OS.

Step 2: Create arparse object specify the parameter expected dataset, encoding file path, detection method are either 'hog' or 'CNN'. This parameter are passed while running encoding python file.

Step 3: Two vector is created to store mapping of the name and 128-d encoding of the file as known encoding and known names.

Step 4**:** Traverse through the images stored in the folder, load them using OS and extract person name from image path and color property of the image is in BGR after loading change it to RGB.

Step 5: Identify the faces from the image using face_recognition.face_locations. The bounding boxes is converted into 128-d encoding of the images.

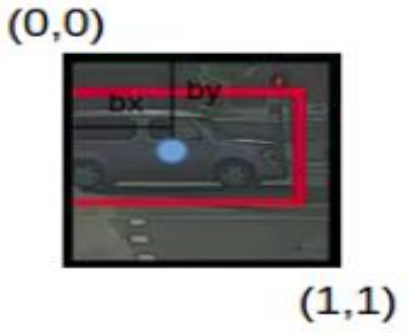
Step 6: Dump the encoding into the file for each image using (pickle.dumps[data]).

Step 7: Execute the face_rcognition program by providing encoding.pickle file, image file and detection method as 'hog' or 'CNN'**.**

Step 8: Load the pickled encodings and face names from disk. Also, load the test image and convert it into rgb color.

Step 9: Proceed to detect all the faces in the images and identify the list of names for each faces that is detected

Step 10: Note: Face identification also work on video either collected from the disk or through webcam real-time.



Module 2: Object Detection

Step 1: Accept the input image and divide that image into M * N parts. Where M and N is variable from the 3 to 19 and M=N. So, this will produce grid of images where images are divided into M rows and N Columns.

Step 2: for each grid, the vector will be produced with 8 distinct element in the vector listed as pc, bx, by, bh, and bw, c0, c1, c2. Here, pc = 0 / 1 (0 – for no object and 1 for object exist), bx and by are the x and y coordinate for the center of the bounding box. bw and bh are the ratio of the corresponding width of the box to the width of the grid. Also, bh is the ratio of the height of the bounding boxes to the height of the grid.

Step 3: The vector mentioned in step is repeated for each of the grid. So, we have same number of vector as of number of grid.

Step 4: The Grids in which the object is existing we will generate the bounding box for those object.

EXPERIMENT

The Experiment begin with the dataset folder consisting of the subfolder with the name of the person to which the images within the folder belongs to. It may contain one or many images for the person associated with that sub-folder. This folder structure is acting as the Training set for the Face Identification Module.

Fig 6: structure of input Dataset

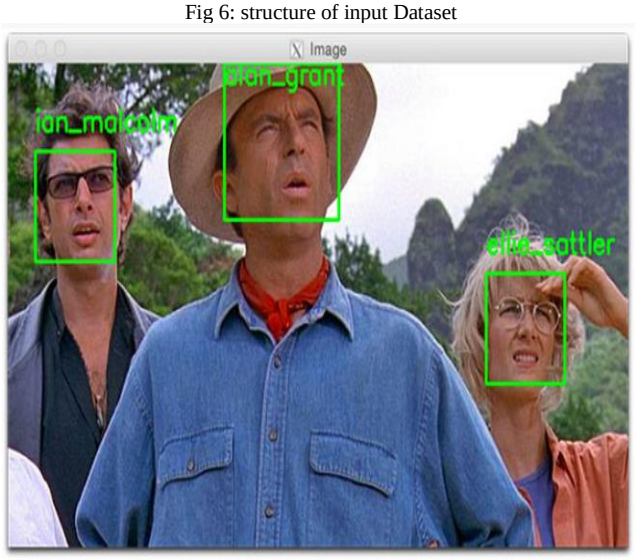


Fig 7: output of the face identification operation

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 05, May-2019

The Object identification module is consisting of the input image. This input image is operated in following manner as represented by fig 7.

Fig 8(a): output of single grid

y =	1
	0.4
	0.3
	0.9
	0.5
	0
	1
	0

Fig 8(b): Vector bounding boxes

RESULT:

The First module produce the bounding box around the face of the person. The bounding is generated labelled around the face and name of the person is written along with the bounding boxes. The second module perform identification of the object in current frame using single short learning(YOLO version 3). Third module produce the summarization of video. Time frame is used to compress the video. Fig 7 and 8(a) and 8(b) represent final output snipt.

CONCLUSION:

CCTV Security and surveillance is very promising industry in future. To enhance the prooductivity in video analytics, this paper produced the novel approach of identification of persion from live video strraming. Also, identification of the suspecious object is achived using Yolo version 3. To reduce the investigation time , the complete video footage analysis is more time consuming task. The summarized video will help to obtain overview of the video footage. So, basic evidence can be collected without wasting much time.

REFERENCE:

[1] Ganesh Ananthanarayan, Paramvir Bahl, Peter Bodik, Krishna Chintalapudl, Matthai Philipose, Lenin Ravindranath, and Sudipta Sinha, Microsoft Research "Real – Time video Analytics : The Killer App for the Edge computing" , IEEE computer society , 2017

[2] https://www.pyimagesearch.com/2018/06/18/face-recognition-with-opencv-python-and-deep-learning/

[3] https://www.analyticsvidhya.com/blog/2018/12/practical-guide-object-detection-yolo-framewor-python/

[4] Pankaj Mendki , Senior Principle Engineer, Member of R & D, Talentica Software Pvt. Ltd, Pune India "Docker container based analytics at IOT edge – video analytics use case" , IEEE 2018.

[5] George Matthew ,"Architectural Considerations for Highly Scalable Computing to Support On-demand Video Analytics", IEEE international conference on Big data analytics , 2017.

[6] Camille Ballas , Mark Marsden, Dian Zhang, Noel E. O'Connor, and Suzzane Liitle, Insight Center for data analytics Dublin City University, Dublin, Ireland "Performance of video processing at the edge for crowd –monitoring application", IEEE, 2018.

[7] Hetal K. Chavda , Prof. Maulik Dhamecha , V.V.P Engineering College Rajkot, Gujarat, India "Moving Object tracking using PTZ Camera in video Surveillance System" , International Conference on Energy , Communication , Data Analytics and Soft Computing(ICEDS-2017)

[8] George Mathew, Lincolan Laboratory, Massachusetts Institute of Technology "The Challenges and Solution for the building and integrated Video analytics Platform", 2017 International Conference on Information Reuse and Integration.

[9] Jennifer Rasch, Jonathan Pfaff, Michael Schafer, Heiko Schwarz, Martin Winken, Mischa Siekmann, Detlev Marpe, Thomas Wiegand, Video coding and Analytics department , Fraunhofer Institute of Telecommunication Heinrich Hertz institute, of Berlin ," A Single Diffusion filter for video coding", IEEE, 2018.

[10] Abderrahmane EZ-ZAHOUT and Jawed OUBAHA , Higher National School of IT/ENSIAS College of Engineering , Mohammed V University – Rabat , Morocco , " A Framework for Big data Analytics in Secure network of Video surveillance System based on image indexation" IEEE, 2018

[11] Jennifer Rasch, Jonathan Pfaff, Michael Schafer, Heiko Schwarz, Martin Winken, Mischa Siekmann, Detlev Marpe, Thomas Wiegand, Video coding and Analytics department , Fraunhofer Institute of Telecommunication Heinrich Hertz institute, of Berlin ,"A Signal Adaptive diffusion filter for video coding using directional and Total Variation ", IEEE, 2018.

[12] Gayathri Venugopal, Philipp Merkle, Detlev Marpe and Tomas Wiegand Video Coding and analytics department ,Fraunhofer Heinrich Hertz Institute(HHI), Berlin , Germany , "Fast Template Matching for intra prediction" , IEEE, 2017

[13] Robert Skupin , Yago Sanchez, Dimitri Podborski, , Cornelius Hellge, Thomas Schierl , Video Coding and Analytics Department , Berlin, Germany , " HEVC Tile based Streaming to Head Mounted Displays", 2017 14th IEEE Annual Consumer Communications and Network Conference (CCNC).

[14] Xingxing Zhang, Zenfeng Zhu, Yao Zhao, Senior Member, IEEE and Dongxia Chang, "Learning a General Assignment Model for Video Analytics" , IEEE Transaction on Circuits and Systems for Video Technology, VOL. PP . NO. PP, June 2017.

[15] Vincent Tam, Mimansha Gupta , Department of Electrical and Electronic Engineering , The University of the Hongkong, Hongkong, " Facilitating and Open learning through Facial Analytics and Video Streaming", 2017 IEEE 17th International Conference on Advanced Learning Technology.

[16] Rashmi S R, Shubha Bhat, Sushmitha V C , Computer Vision Lab , Computer Science and Engineering , Dayananda Sagar College of Engineering, Bangalore , India , " Evaluation for the Human Action Recognition Technique intended for Video Analytics " , IEEE, 2017

[17] Dimtri Podborski, Yago Sanchez, Robert Skupin, Cornelius Hellge, Thomas Schierl , Video Coding and analytics department, Fraunhofer Heinrich-Hertz-institute , Berlin, Germany ,"Tile Based Panoramic streaming using shifted IDR Representations", Proceedings of the IEEE international Conference on MultiMedia and expo (ICME) 2017.

[18] Tung Nguyen and Detlev Marpe , Department of Video coding and analytics, Fraunhofer Institute of Telecommunications – Heinrich Hertz Institute Einteinufer , Berlin , Germany , "Future Video Coding Technologies : A Performance Evaluation of AVI , JEM , VP9, and HM", IEEE 2018.

[19] Soumen Kenrar , Niranjan Kumar Mandal, Department of Electrical Engineering UEM university, Calcutta , India, "Approximation of Bandwidth for the Interactive Operation in video on Demand System" , International Conference on I-SMAC , 2017.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 05, May-2019

[20] Iveel Jargalsaikahan, Sizanne Little and Noel E 'O'Connor ,Insight center of Data Analytics, Dublin City University , Ireland, "Action Localization in Video using a Graph- based Feature Representation", IEEE , 2017.
[21] Anchal Kathuria , Dr. S.N. Panda, Chitkara university , India. "Video Capturing and streaming over Ad-HOC Network", 2017 International Conference on Big Data Analytics and computational Intelligence (ICBDACI).
[22] Rasmika Nawratne , Tharindu Bandragoda , Achini Adikari, Damminda Alahakoon, Daswin De Silva , Research center for Data Analytics and Cognition , La Trobe University, Victoria , Australia