iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VII **Month of publication:** July 2020

DOI: http://doi.org/10.22214/ijraset.2020.7062

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VII July 2020- Available at www.ijraset.com

An Efficient Method for the Detection of Missing Person from Crowd

Akhil MP[1], Lavanya Anilkumar[2], Suraj Salim[3], Shahana Shaji[4], Linda Sebastian[5]
[1, 2, 3, 4, 5]*Department of Computer Science , College of Engineering Kidangoor*

Abstract: ***Numerous cases of missing are reported in Kerala. Our local department has the mechanism to watch every point on city. But it's not effectively helpful to seek out the missing person. It is found difficult to seek out an individual from crowd. During this paper, we present a mechanism to seek out the missing person or object from crowd by using opencv and image processing. The image of the person is subjected to image processing, which is taken as input for tracking the person or object from crowd using opencv. When the person or object is detected from video of crowd, an alert system is employed as a reminder. When an alert signal is produced, the authority can take necessary action within a short time. The objects are detected using one shot learning in real time scenarios. Main objective of the one shot learning is to detect various objects in real time video sequences. Once the network is trained by employing a single image of an individual, then the trained model can recognize that person in any angle. This real time analysis of the environment enables security and surveillance. This model can be implemented in different surveillance devices to detect attacks on many places like schools, government offices and hospitals where arms are completely restricted.***

Keywords: Alert system, Crowd, One-shot algorithm, Opencv, Person detection, Real time, Video sequence

I. INTRODUCTION

Different mechanisms are used by our department of local government to monitor each purpose on town. But it's not effectively useful to find the missing person from crowd. We have a tendency to gift a mechanism to find the missing person from crowd by victimisation one shot learning. Main objective of the one shot learning is to discover various objects in real time video sequences. Once the network is trained by employing a single image of someone, then the trained model will acknowledge that person in any angle. This real time analysis of the environment allows security and surveillance. This model can be implemented in different police investigation devices to discover attacks on several places like faculties, government offices and hospitals wherever arms are completely restricted.

II. PROPOSED METHOD

Detection of people from crowd may be a important component for a good range of applications including surveillance, group behaviour modelling and crowd disaster prevention. The efficient person detection in crowds, may be a highly challenging task during view variations and ranging density of individuals as well as the ambiguous appearance of body parts. Here, we are designing a product for detecting the missing person from crowd by giving an image of person as an input. By using the one shot algorithm, different features in different angles are extracted from the same single image provided. Then there is no need to keep many different images of the same person. And it is difficult to take so many images. After getting the extracted features, it is used for the comparison with each and every person in the crowd according to the direction, visibility, etc. When the identification is successful, an alert message is obtained that the person is identified. Otherwise, a message is obtained that person is not present in the crowd or identification failed. The alert message can be obtained through a web page.

A. Architecture

As a first step, an image of a person to be detected is inserted into the system. After that, insert the video from which the detection occurs. Both image and video are stored on system. Then the stored image and video are given to Siamese network. The input image is employed to coach the network. Then the same image is used for detection. The Siamese network consist of two sister networks. A bounding box will appear around the faces in video. The image to be detected is placed in the first sister network and each faces found in video are placed in the second sister network. Then the two faces get compared by using contrastive loss function, which calculates the similarity and dissimilarity of two faces. The similarity and dissimilarity are calculated by taking the Euclidean distance between key points in face. When the two faces are same, then a message will appear that person is matched along with person name and colour change of message from red to green. When the two faces are not matched, a message will appear that person is not matched and the colour of message remains as red. The architecture of the system can be viewed in fig 1.

©IJRASET: All Rights are Reserved

1) *The Dataset:* The data set consists of image that is input into the system for identification.
2) *Data Loading:* The architecture consist of an input pair. The image of a person to be detected is stored in the system folder. Another image in the image pair is face from crowd
3) *Training the Siamese Network:* The training process of a Siamese network is as follows:

a) Pass the primary image of the image pair through the network.
b) Pass another image in the image pair through the network.
c) Calculate the loss using the outputs from 1 and 2.
d) Back propagate the loss to calculate the gradients.
e) Update the weights using an optimiser.

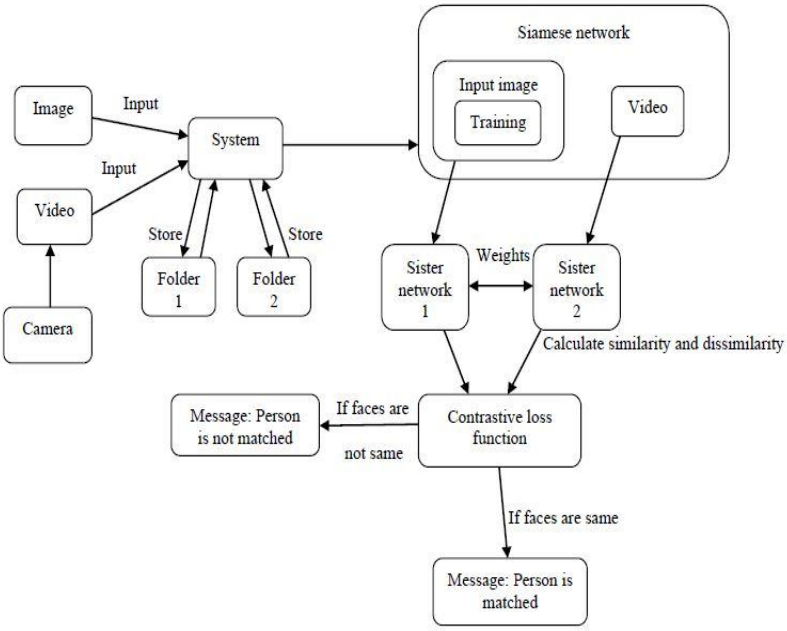


Fig 1. Architecture of proposed system

B. Modules of the System

The system is divided into five modules. They are:

1) *Camera Module:* To capture the video streams effectively. And to input an image into system.
2) *Database Module:* To store the input image and video stream for the purpose of detection.
3) *Neural Network Module:* To identify a person from a crowd in fast and accurate manner. Siamese network is used for this purpose. It has a special architecture, which consist of two sister networks used for comparing the inserted image with each and every faces in the crowd. One shot classification is the methodology adopted for this purpose. The input image is used for training the network. The steps for training the network are:

a) Pass the primary image of the image pair through the network.
b) Pass another image in the image pair through the network.
c) Calculate the loss using the outputs from 1 and 2.
d) Back propagate the loss to calculate the gradients.
e) Update the weights using an optimizer.

4) *Comparison Module:* For calculating the difference between face points using contrast loss function. And this value of difference decides the similarity and dissimilarity. Therefore, the comparison can be done by using the calculated difference.
5) Alert Module: A message should be arrived in both cases via webpage.

a) *Case 1:* When the person is matched successfully.
b) *Case 2:* When the person is not matched.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VII July 2020- Available at www.ijraset.com

III.ONE SHOT CLASSIFICATION

[7]In a one shot classification, we require just one training example for every class. Hence the name. One shot Classification models, on the opposite hand, requires that we've just one training example of every class we would like to predict on. The model remains trained on several instances, but they only have to be within the similar domain as our training example. A nice example would be face recognition. We will train a one shot classification model with a dataset that contains various angles, lighting, etc. of a couple of people. Then if we would like to recognize if an individual X is in a picture , we take one single photo of that person, then ask the model if that person is in the that image(note, the model wasn't trained using any pictures of person X). As humans, we will recognize an individual by his/her face by just meeting them once, and it's desirable by computers because repeatedly data is at a minimum.

A. Siamese Networks

[7]Siamese networks are a special type of neural network architecture. rather than a model learning to classify its inputs, the neural networks learn to differentiate between two inputs. It learns the similarity between them. Siamese network may be a pair of network or a single network. Because of the weights are equal for both networks, we use one model and feed it with two images at the same time. After that we calculate the loss value using both the pictures, then back propagate. this protects tons of memory at absolute no hit on other metrics (like accuracy).

B. Architecture

[7] A Siamese network consists of two identical neural networks, each taking one among the 2 input images. The last layers of the 2 networks are then fed to a contrastive loss function, which calculates the similarity between the 2 images. We have made an illustration to help explain this architecture fig 2. There are two sister networks, which are identical neural networks, with the precise same weights. Each image within the image pair is fed to at least one of those networks. The networks are optimised employing a contrastive loss function (we will get to the exact function).

C. Contrastive loss function

[7]The objective of the Siamese architecture isn't to classify input images, but to differentiate between them. So, a classification loss function (such as cross entropy) wouldn't be the simplest fit. Instead, this architecture is best suited to use a contrastive function. Moreover, this function just evaluates how well the network is distinguishing a given pair of images mathematically, it is defined as:

(1-y)1/2(Dw)2 + (y)1/2max(0,m-Dw)2 (1)

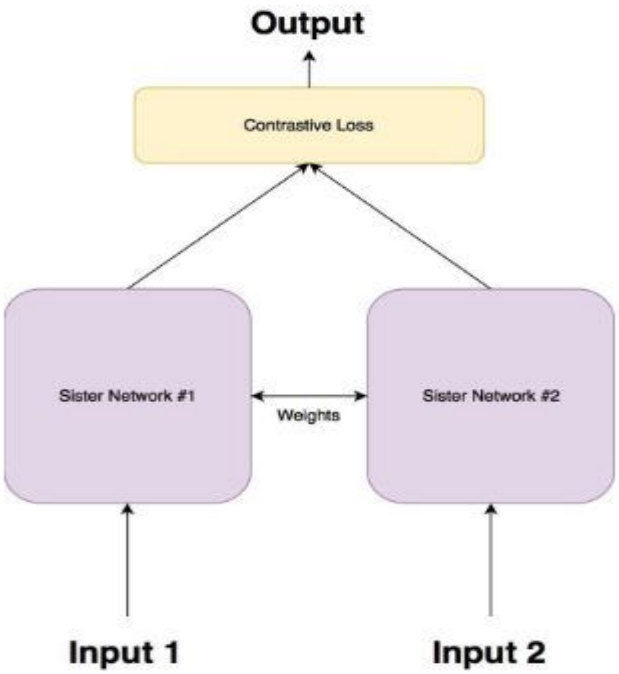


Fig 2. Architecture of Siamese network

©IJRASET: All Rights are Reserved

IV. TECHNOLOGIES USED

1) *Visual Studio Code Editor:* Visual Studio Code may be a free source-code editor made by Microsoft for Windows, Linux and macOS. It has features for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded Git.
2) *Tensor Flow:* Tensor flow is an open-source software library used for machine learning applications such as neural networks.
3) *Neural Network Using one shot Algorithm:* One-shot learning is a classification task where one, or a few, examples are used to classify many new examples in the future. This characterizes tasks seen in the field of face recognition, such as face identification and face verification, where people must be classified correctly with different facial expressions, lighting conditions, accessories, and hairstyles given one or a couple of template photos. Modern face recognition systems approach the problem of one shot learning via face recognition by learning a rich low dimensional feature representation, called a face embedding that can be calculated for faces easily and compared for verification and identification tasks. The training of Siamese networks with contrastive loss functions resulted in better performance.

a) How Siamese Neural Networks are used for one shot Image Recognition: generally, we have a tendency to learn image representations via a supervised metric-based approach with Siamese neural networks, then we use the network options for one-shot learning with none grooming. To develop a model for one-shot image classification, we have a tendency to aim to find out a neural network which will discriminate between the class-identity of image pairs, that is that the common place verification task for image recognition. we have a tendency to theorize that networks that act at verification ought to generalize to one-shot classification. The verification model learns to spot input pairs in line with the chance that they belong to identical category or completely different categories. This model will then be usually value new pictures, precisely one per different category, throughout a try wise manner against the take a look at image. The pairing with the best score in line with the verification network is then awarded the best chance for the one-shot task. If the options learned by the verification model are enough to substantiate or deny the identity of characters from a collection of alphabets, then they got to be enough for alternative alphabets, as long because the model has been exposed to a range of alphabets to encourage variance amongst the learned options.

4) *Flask:* Flask is a web framework which means that the flask provides you with tools, libraries and technologies that allow you to build a web application. This web application can be some sites, a blog, a wiki or go as big as a web-based calendar application or a commercial website.

V. RELEVANCE

1) *Security:* As a part of security, we can implement this method in places where unauthorized people are restricted. When an unauthorized person enters, the authority will get an alert message. Since the network is trained by using the images of authorized people.
2) *Surveillance:* As a part of surveillance, we can implement this method in crucial areas like ATM etc. We can observe the activities of people through surveillance.

VI. APPLICATIONS

A. Early threat detection
B. ID verification
C. Prison visitor system
D. Banking using ATM
E. Physical access control of buildings, areas etc
F. Voting system
G. Computer Security
H. Automation

VII. EXPERIMENTAL ANALYSIS AND RESULTS

This section presents a case study about detection of missing person from crowd. It serves as an illustration of our proposed idea. Numerous cases of missing are reported in Kerala. Our police department has the mechanism to monitor every point on city. But, it is not effectively helpful to find the missing person. It is found difficult to find a person from crowd. We implemented this idea as two parts:

©IJRASET: All Rights are Reserved

1) Real time detection of a person from group picture/photo.
2) Real time detection of a person from video streams.

The experiment setup for detection from both group photo and video are same. The image of the person to be detected is treated as dataset, and it is used for training the Siamese neural network. The implementation has been done by dividing the project into five modules as Camera module, Database module, Neural network module, Comparison module, Alert module.

A. Real time Detection From Group Photo

We considered a pretrained Siamese network model for the detection purpose. Hence numerous datasets are not required for training. Since the network is pretrained, only the last layer of network is trained using input image. As a web based user interface is used, image of a person to be detected and group picture/photo from which person to be detected are provided as input via the webpage. We can see the simple user interface depicted in fig 3. When the input image is provided to the network, it is stored into a system folder named as pictures, which acts as a database for this purpose. Siamese network has a special type of architecture such that the input image and each image from group picture are compared without taking much time. It find outs the similarity and dissimilarity of two faces by taking the Euclidean distance of all possible key points and converting into 128 dimensional form using the contrast loss function. Then the nplinalgnorm method in the numpy function differentiates the two people. The accuracy of detection can be improved by adjusting the value of function. When we press the upload image button on webpage, there is box in which we can enter the person name, depicted in fig 4. And the system folder will opens and we can choose the image of person to be detected from the picture folder. It can be viewed on fig 6. Then there is another option to start recognition. When clicked on it we can give a group picture/photo in front of camera. The Siamese network will take all possible key points of a face to find the similarity and dissimilarity. After some processing time, we can see the detection along with the accuracy rate and person name. Each person in image is bounded by a bounding box in blue color. If the person is matched with input image, then the color of alert message generated along with accuracy rate and person name will be green. It can be viewed on fig 7. As a part of experiment, we took an image of Christiano Ronaldo and a group photo in which he is present. A simple user interface is used for the experimentation.

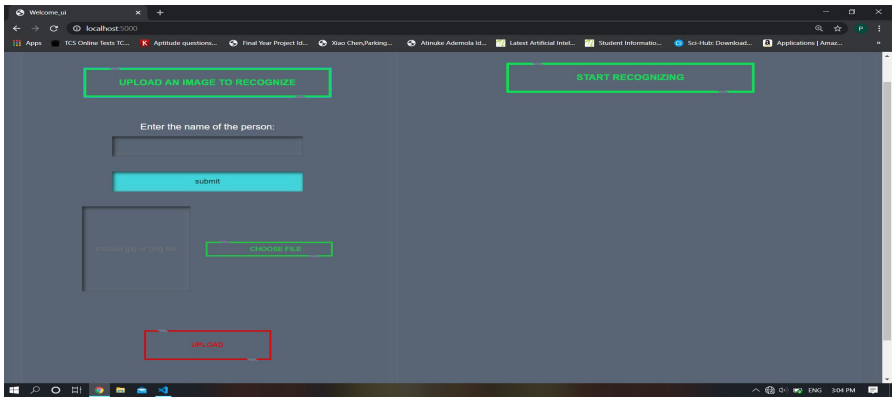


Fig 3. A simple user interface

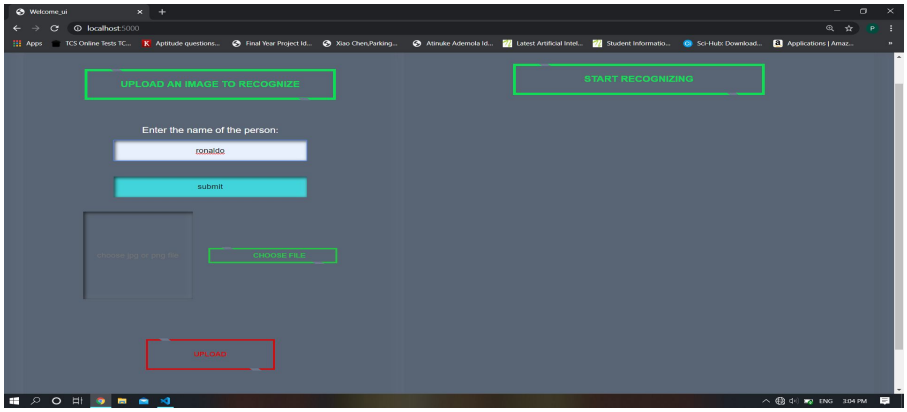


Fig 4. Insertion of person name

©IJRASET: All Rights are Reserved

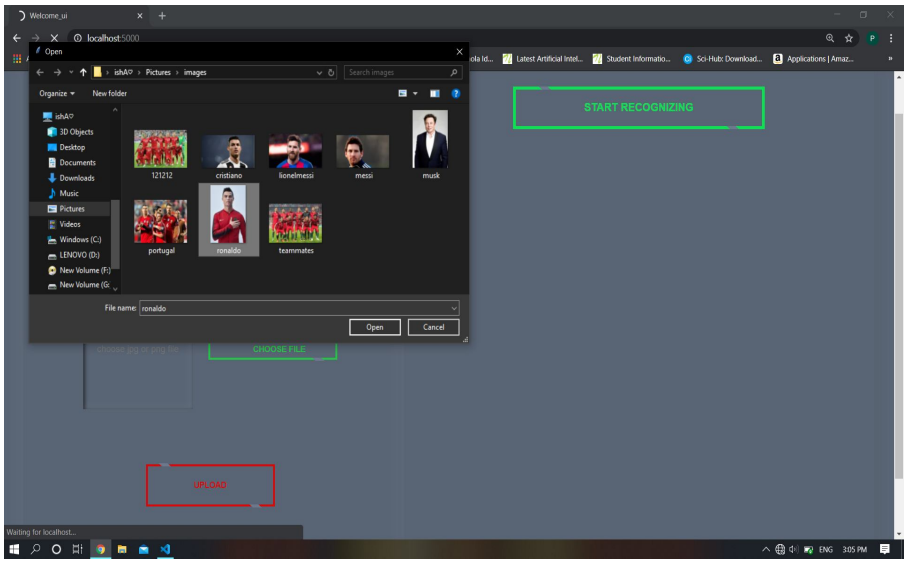


Fig 5. Uploading an image

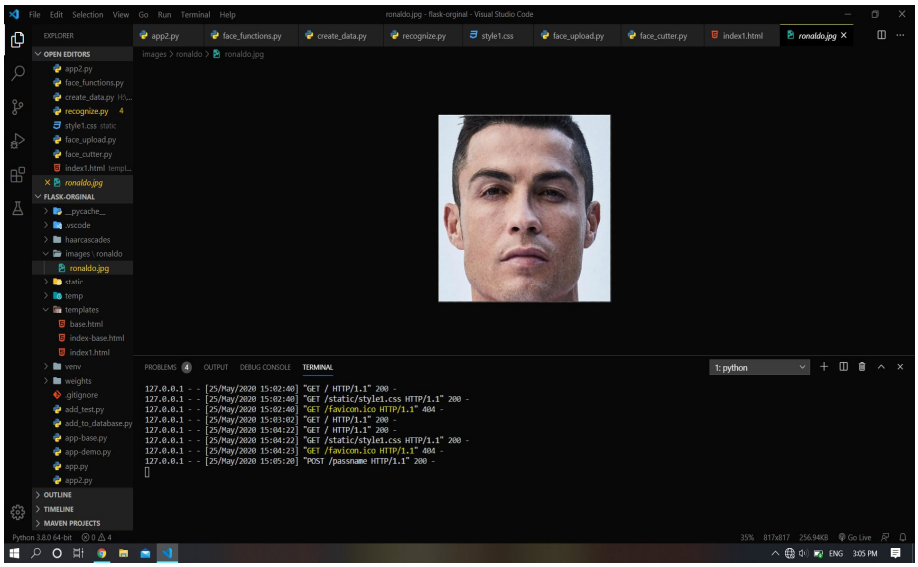


Fig 6. View of inserted image

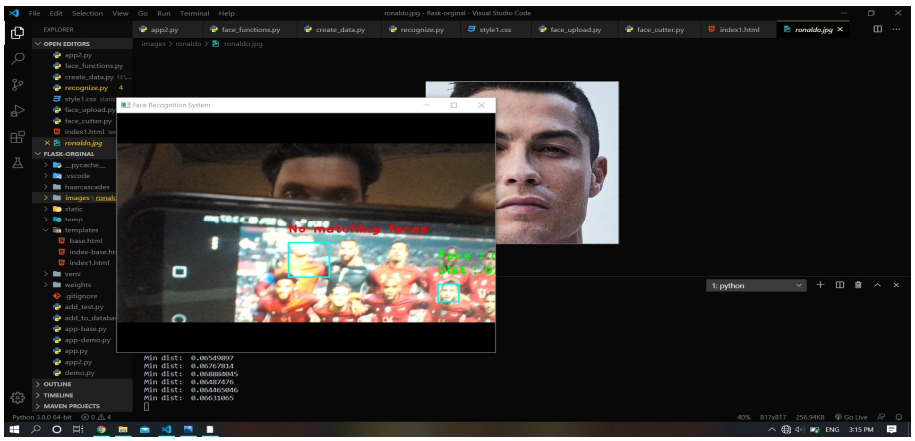


Fig 7. Detection from group photo

©IJRASET: All Rights are Reserved

B. Real time Detection from Video Streams

We considered a pretrained Siamese network model for the detection purpose. Hence numerous datasets are not required for training. Since the network is pretrained, only the last layer of network is trained using input image. The detection of the person from video streams is similar to detection from group picture. Instead of group picture, a real time video is provided to the system. It can be viewed on fig.8. When pressing the input image option, it moves to the system folder and from the picture folder, we can choose the image of person to be detected fig.5. Hence the image is entered into the network. When we click on start recognition option, we can provide the video from which detection occurs. The provided video is also stored in the system. Siamese network has a special type of architecture such that the input image and each image from group picture are compared without taking much time. It find outs the similarity and dissimilarity of two faces by taking the Euclidean distance of all possible key points and converting into 128 dimensional form using the contrast loss function. Then the nplinalgnorm methods in the numpy function differentiate the two people. After some processing time, we can see bounding boxes will appear around the faces of each and every person in the crowd. Then the input image of person to be detected and faces in each bounding boxes get compared by Siamese network. When matching becomes successful on any face, then the alert message generated along with accuracy rate and person name will be in green color. It can be viewed on fig.9.

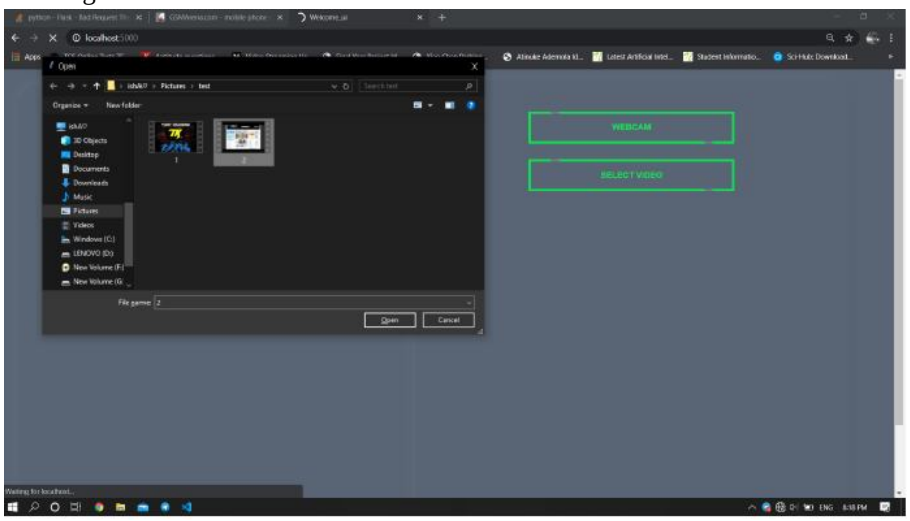


Fig 8. Insertion of video

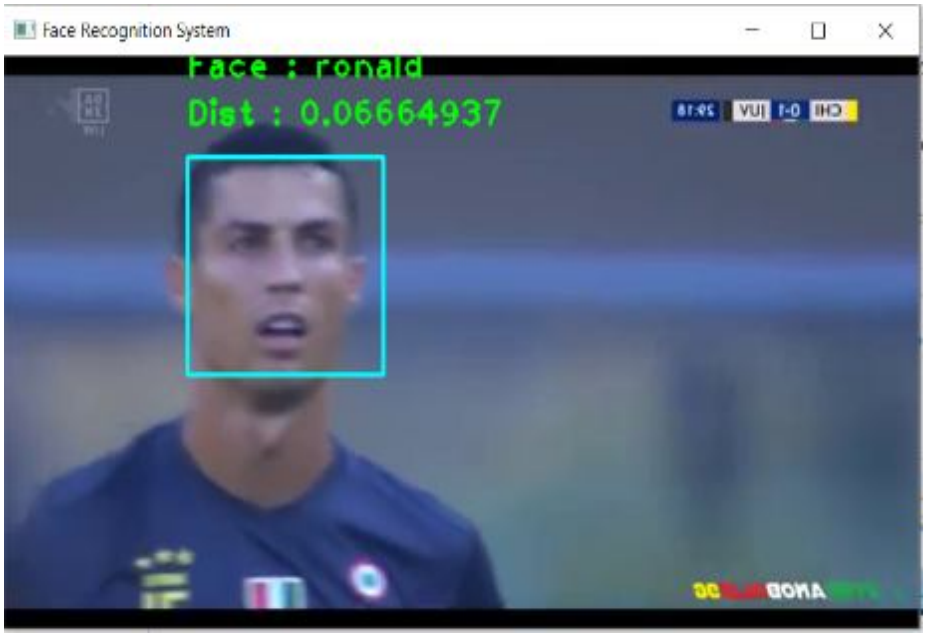


Fig 9. Detection from video stream

©IJRASET: All Rights are Reserved

VIII. RESULTS AND ANALYSIS

During the experimentation, we tried both contrastive loss function and triplet loss function for detection process. We noticed that triplet loss function is perfectly accurate by theory. While implementing, it seems that accuracy rate of triplet loss function is lesser than contrastive loss function. The distance loss of contrastive loss function is in the range of 0-1. And the distance loss of triplet loss function is in the range of 1-2. The comparison of distance loss of contrastive loss function and triplet loss function can be viewed in fig 10.

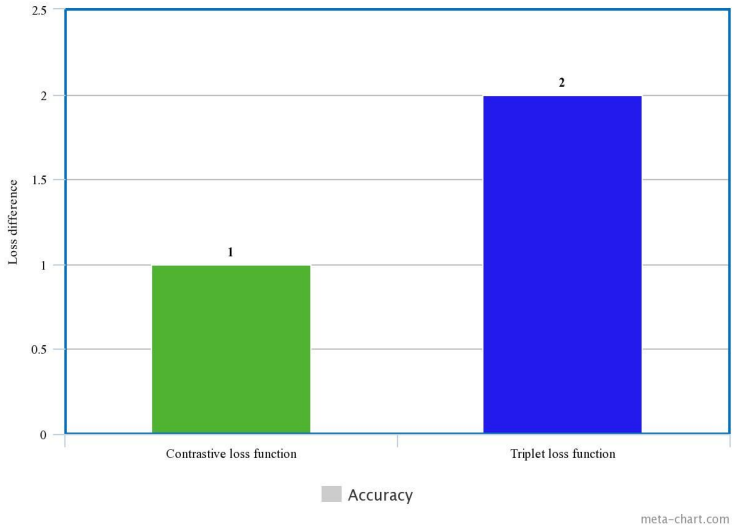


Fig 10. Comparison of distance loss between contrastive loss function and triplet loss function

IX. CONCLUSION

In this paper, the facial detection is done by implementing the one-shot algorithm using One shot learning. The image of the person is given to system as an input, which is used for detection of person from crowd. When the person or object is detected from video of crowd, an alert system is used as a reminder. When an alert signal is produced, the authority can take necessary action within a short time. And, if the input image and person in the video matches, the face is recognized.

Main objective of the one shot learning is to detect objects from real time videos. Once the network is trained by employing a single image of an individual, then the trained model can recognize that person in any angle. This real time analysis of the environment enables security and surveillance. This model can be implemented in different surveillance devices to detect attacks on many places like schools, government offices and hospitals where arms are completely restricted.

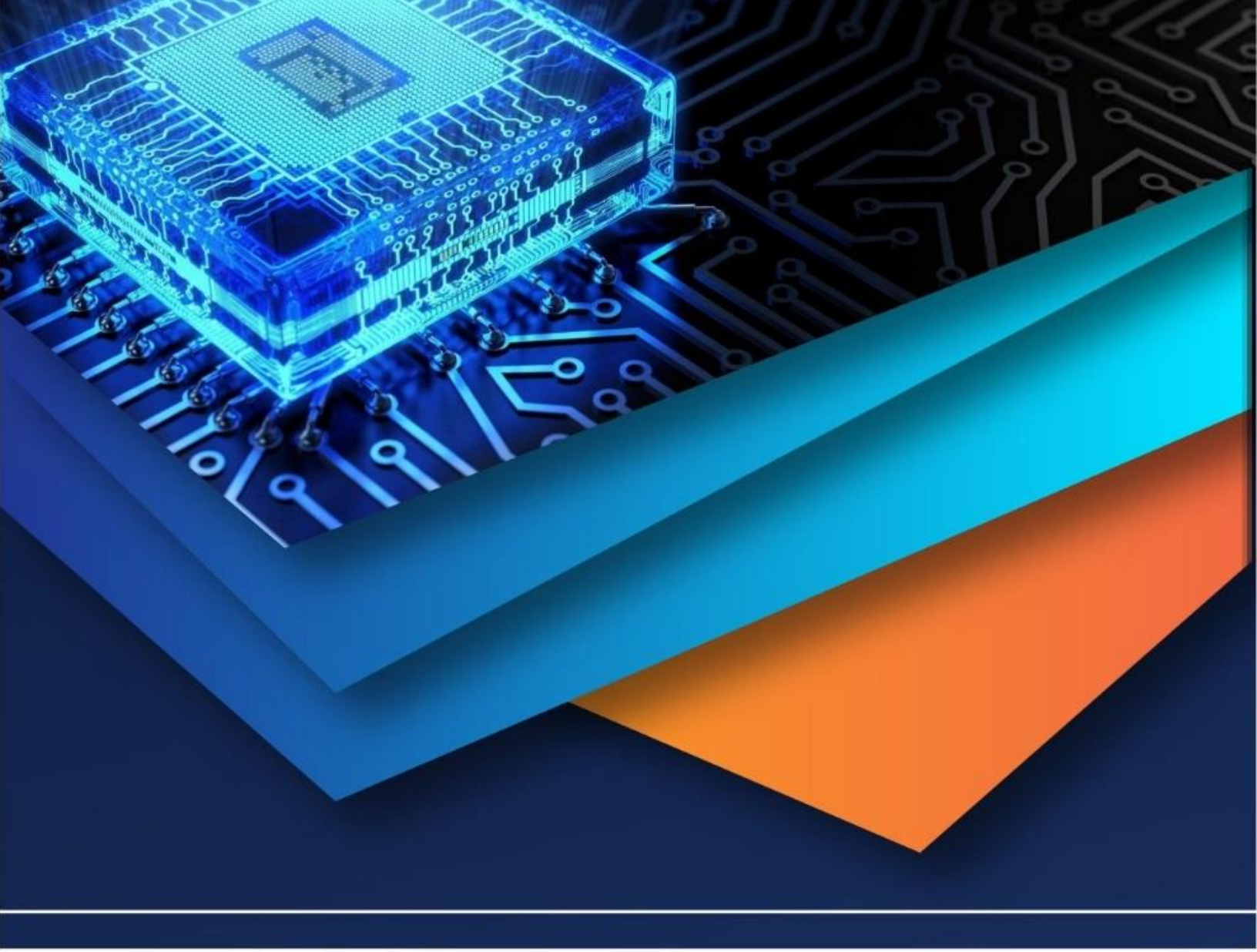
X. ACKNOWLEDGEMENT

The authors would like to thank all who were involved directly or indirectly in the data collection, experiment conduction, and analysis stage of the project.

REFERENCES

[1] Chandan G, Ayush Jain, Harsh Jain, Mohana," Real time object detection and tracking using deep learning and opencv", ICIRCA 2018.
[2] Witsarut Sriratana, Suppakorn Mukma, Narin Tammarugwattana, Kaset ,Sirisantisamrid, ""Application of the opencv-python for personal identifier statement", 978-1-5386-4956-5/18/2018 IEEE
[3] Kruti Goyal, Kartikey Agarwal, Rishi Kumar "Face detection and tracking", ICECA 2017
[4] Sachin Umesh Sharma,Dharmesh J.Shah "A Practical Animal Detection and Collision Avoidance System Using Computer Vision Technique" 2016 Electronic ISSN: 2169-3536
[5] A. Ali and M. A. Bayoumi. Towards real-time dpm object detector for driver assistance. In 2016 IEEE International Conference on Image Processing (ICIP), pages 3842–3846, Sept 2016.
[6] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). Comput. Vis. Image Underst., 110(3):346–359, June2008.
[7] https://hackernoon.com/one-shot-learning-with-siamese-networks-inpytorch-8ddaab10340e
[8] https://towardsdatascience.com/one-shot-learning-with-siamesenetworks-using-keras-17f34e75bb3d
[9] https://medium.com/@enoshshr/triplet-loss-and-siamese-neuralnetworks-5d363fdeba9b

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)