PERSON IDENTIFICATION AND TRACKING USING DEEP LEARNING

Mrs.SARANYA N
ASSISTANT PROFESSOR
ELECTRONICS AND COMMUNICATION ENGINEERING

Dr.KANTHIMATHI N
ASSISTANT PROFESSOR
ELECTRONICS AND COMMUNICATION ENGINEERING

DEEPIKA C
SOWMIYA A
SANTHIYA S
ELECTRONICS AND COMMUNICATION
BANNARI AMMAN INSTITUTE OF
SATHYAMANGALAM,ERODE,INDIA

ABSTRACT:

In the present scenario, digital data generation, data consumption becoming necessary due to advancement in technology. The human data processing becoming important in various types of applications like person authentication, verifications automatically by the machines. One of the applications is to identify the person automatically is by the machine. Numerous computer vision, machine and deep learning–based methods have been created in recent years. Majority of these methods are based on frontal view images/video sequences. The advancement of convolutional neural network reforms the way of object tracking.Person identification and tracking are critical tasks in many applications, such as security systems, video surveillance, and social media platforms. Deep learning has revolutionized the field of computer vision, enabling high accuracy in person identification and tracking. This paper provides an overview of person identification and tracking using deep learning techniques, such as Convolutional Neural Networks (CNNs). We discuss various deep learning architectures for person identification and tracking, including face recognition and body posture analysis. Finally, we suggest future directions for research in this field, such as multi-modal person identification and tracking and ethical considerations.

KEYWORD: *Detection, tracking, CNN, deep learning.*

I.INTRODUCTION:

It is essential to identify each individual person using multiple technologies because of the use of person identification in applications like airport verification, different units, digital transactions, and access to diverse resources. the information or limited area. The challenges of person identification have been studied for a while, however machine person recognition still falls short of human performance. Identification of the individual is complicated by a number of factors, including size, color, direction, and occlusion. Face recognition has recently become usable under constrained circumstances. To identify the person, a facial matching process is employed. In this case, pictures of faces are present in the face database. The unknown face image is compared to the face images in the face database. The usage of facial recognition technologies identify people, such as the subject having to be sufficiently close and facing the camera. Real-time face recognition applications are constrained by this process of face identification. As more and more video cameras are put in various locations, person recognition is becoming increasingly crucial in surveillance applications. Prior work involving the identification of people has only used facial recognition, and even then, the subject must appear in front of the camera with his face properly aligned. This method was quite time-consuming because the user had to personally present oneself in front of the camera each time to label himself as present in numerous regions. For processing, this generates a lot of video data. The difficulty of individual identification in surveillance footage is caused by to a number of problems, including lighting, illumination, scale, occlusion by other items, and person orientation. This essay explores the problem of person identification utilizing the person re-identification procedure. Today, models based on convolutional neural networks (CNNs) are incredibly successful, especially in the fields of computer vision, remote sensing, data categorization, pattern recognition, image processing, and smart surveillance analysis (specifically in object detection, tracking, and recognition). Analyzing the target trajectory of people in video sequences is done via person tracking. Since it covers fall detection in results for typical frontal view photos and video sequences, as well as unexpected event detection across a wide range of research fields, it is crucial. Elderly people, evaluating crowds or congestion, human-computer interface, robot navigation, and other things. As a changeable entity with variations in look, position, scale, and size, tracking people is a difficult undertaking. The accuracy of different tracking algorithms may be impacted by a variety of factors, including changes in lighting conditions, complex backgrounds, cluttered scenes, abrupt motion, shadows, the deformable nature of the person (scale and size variations), close human interaction, camera perspective view, and occlusion. Researchers have developed a number of person tracking techniques that have demonstrated good tracking.

II.LITERATURE SURVEY:

[1] DLD-APDT algorithm: This paper proposes a DLD-APDT algorithm was enhanced for effectual person detection and tracking on surveillance videos. The proposed DLD-APDT model initially performs frame conversion process where the input video is converted into a set of frames. And the proposed DLD-APDT model utilized Efficient Set model to detect persons and track them.

[2] "DeepSORT algorithm: This paper proposes a deep learning-based tracking algorithm called DeepSORT that combines a deep appearance feature extractor with a particle filtering framework. The proposed algorithm achieves state-of-the-art results on the MOT16 benchmark dataset.

[3] Deep learning-based person detection algorithm: This paper proposes a deep learning-based person detection algorithm that uses a combination of deep convolutional neural networks and a sliding window approach. The proposed algorithm achieves high accuracy on several benchmark datasets.

[4] Deep learning-based multi-object tracking algorithm: This paper proposes a deep learning-based multi-object tracking algorithm that uses a quadruplet convolutional neural network to learn appearance and motion features. The proposed algorithm achieves state-of-the-art results on several benchmark datasets.

[5] Deep learning-based online multi-object tracking algorithm: This paper proposes a deep learning-based online multi-object tracking algorithm that uses a CNN-based single object tracker with a spatial-temporal attention mechanism. The proposed algorithm achieves state-of-the-art results on several benchmark datasets.

III.PERSON IDENTIFICATION:

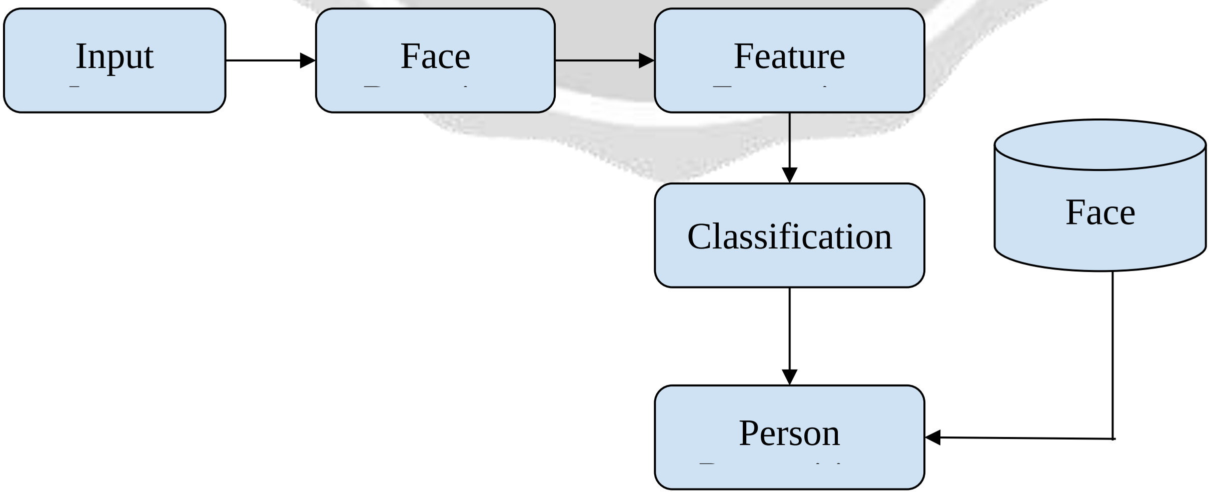


Fig.1. Person Identification

Without the need for faces, person identification attempts to identify people in pictures or videos taken by various cameras. It has numerous significant real-world applications, including person search, security monitoring, and no-checkout businesses. It is a significant problem in the computer vision community. However because of a number of elements, including varying lighting, shifting points of view, distortion in the human stance, and occlusion, this task is particularly difficult. Conventional methods typically concentrate on creating features by hand and/or learning distance measurements for matching to address these issues. Feature extraction and metric learning can be coupled in one framework using convolutional neural networks (CNNs). Deep learning is a technique for identifying people by analyzing and recognising important elements of a person's face or other biometric data. These characteristics could include things like eye color, hairline, and face structure. Large quantities of facial picture data can be used to train deep learning algorithms, which then enable them to detect and match people with high levels of accuracy. Applications for this technology range from security systems to photo tagging software. The potential for this technology to be abused for spying or other intrusive purposes has raised questions about privacy and ethical use. When utilizing deep learning for person identification, it is crucial to take these concerns into account and put in place the proper precautions.

IV.PERSON TRACKING:

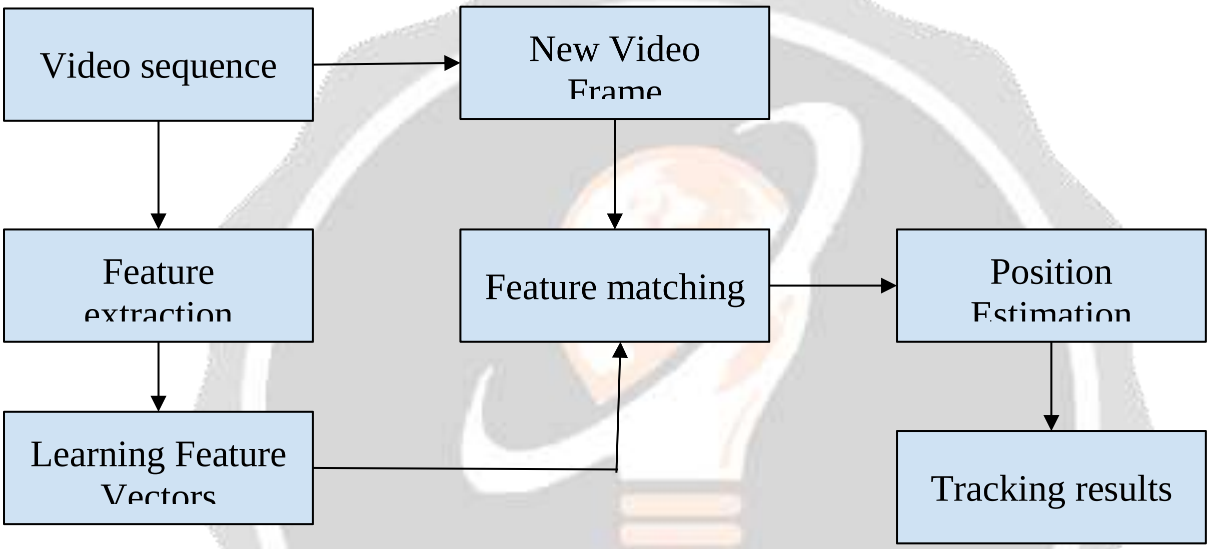


Fig.2. Person Tracking

In many surveillance or monitoring applications, one or more cameras view several people that move in an environment. Multi-person tracking amounts to using the videos from these cameras to determine who is where at all times. Abstractly, a multi-person tracker takes as its input a set of observations. At the lowest level, these are detections computed on each video frame by a person detection algorithm, and consist of a bounding polygon that encloses the person, together with an image position, a time stamp, an estimate of the person’s velocity, and an appearance descriptor. For the sake of efficiency, intermediate stages within the tracker may form higher-level observations by grouping detections. The tracker then partitions the input observations into sets, with the intent that each set corresponds to one person and vice versa. The literature calls these sets identities or when the linear time ordering of detections in a set needs emphasis trajectory. Two major pairs of conflicting challenges make multi-person tracking hard. Observations are ambiguous in that different people that look alike may be confused with each other. Conversely, changing,lighting, viewpoint, and other circumstances may cause variance of appearance for a given person, which may not be recognized to be the same in different observations. In the other pair of challenges, person occlusions whether caused by limited field of view, visual obstacles between camera and person, or algorithm failure generate gaps in the input observations that make tracking harder. Conversely, overactive person detectors may generate spurious observations that confuse the tracker. All of these challenges already show up in the convenient nutshell of the single-camera case, on which this paper is focused.

V.WORKING OF DEEP LEARNING:

Deep learning models belong to the category methods for supervised machine learning, which approaches, which identify hidden patterns in datasets by comparing them to reference datasets. It is a field that relies on studying algorithms to learn and advance on its own. Neural networks are used in deep learning because they created to mimic how humans think and learn, whereas machine learning uses simpler principles. Signals go between nodes and assign matching weights in an artificial neural network. A node with a higher weight will have a greater impact on the nodes in the layer below it. The weighted inputs are combined to create an output in the final layer. Multi-layered deep learning models are used. Features can be extracted by employing the neuron's built-in structure. Machines are trained utilising deep learning algorithms by learning from data sets.

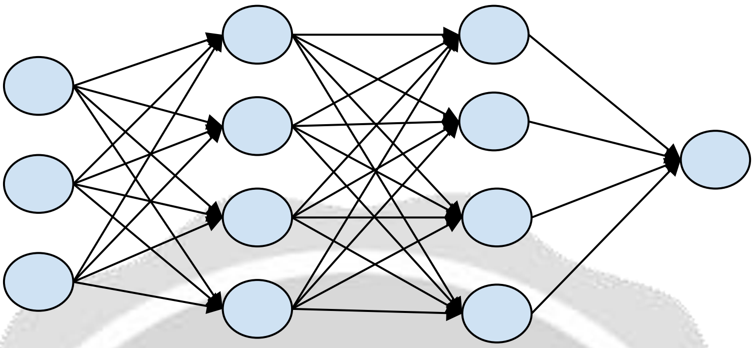


Fig.3. Working of Deep Learning

VI.FEATURE EXTRACTION OF CNN:

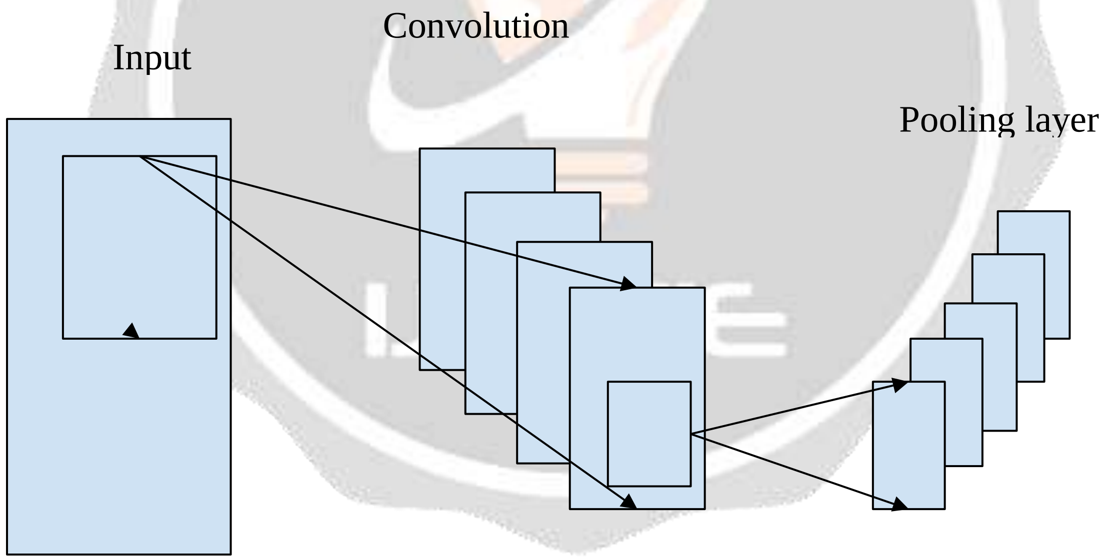


Fig.4.Feature Extraction of CNN

The neural network is often used for multiclass categorization in CNN's output layer. Instead of doing it by hand, CNN utilises a feature extractor in the training phase. The feature extractor used by CNN is made up of unique neural network types, the weights of which are determined during training. When CNN's neural network feature extraction is more in-depth (has more layers), it improves picture identification, but at the expense of the learning process complexity that had previously rendered CNN ineffective and unappreciated. A neural network called CNN extracts the features of the input images, while a different neural network categorises the features. The feature extraction network uses the input image as a starting point. The neural network uses the extracted feature signals for classification.The result is subsequently generated by the neural network categorization based on the image features. The convolution layer stacks and sets of pooling layers are part of the neural network for feature extraction. The convolution layer, as its name suggests, uses the convolution method to modify the image. It could be compared to a collection of digital filters. The One pixel is created by combining neighbouring ones by the layer of pooling. The image dimension is subsequently reduced by the pooling layer.

VII.BASIC ARCHITECTURE OF CNN:

A CNN architecture consists of two fundamental components.

- A convolution tool that, through a procedure known as feature extraction, separates and identifies the distinct characteristics of the image for study.
- A fully connected layer uses the output from the convolution process and predicts the class of the image based on the characteristics collected in earlier stages.
- This CNN feature extraction model seeks to minimise the quantity of features in a dataset. It generates new features that compile an initial set of features' existing features into a single new feature.

VIII.CONVOLUTIONAL LAYER:

The CNN is made up of three different kinds of layers: fully-connected (FC), pooling, and convolutional layers. A CNN architecture is created when these layers are stacked.

1. Convolutional Layer

The numerous features from the input pictures were first extracted using this layer. Convolution is a mathematical process that is carried out between the input image and a filter with MxM dimensions at this layer. The dot product between the filter and the elements of the input image is obtained by sliding the filter over the input image in respect to the filter's size (MxM). The feature map is what provides details about the image, such as its corners and edges. This feature map is later made accessible to other layers so they can pick up new features from the source image. The CNN convolution layer applies the convolution and then forward the result to the following layer.The spatial link between the pixels is preserved thanks to convolutional layers of CNN.

2. Pooling Layer

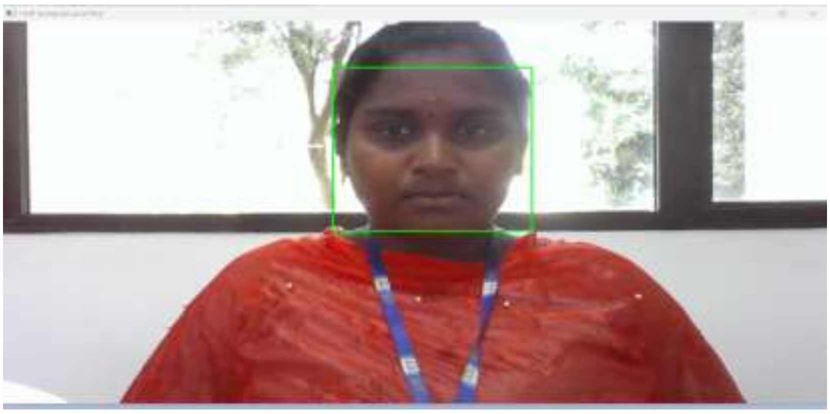
After a convolutional layer, a pooling layer is frequently applied. The primary objective of this layer is to reduce the size of the convolved feature map in order to reduce computing costs. Using fewer links between layers and independently modifying each feature map, this is accomplished. Several pooling operations exist, depending on the technique used. It is essentially a summary of the features that a convolution layer produced. The feature map is where MaxPooling gets the bulk of its components. With average pooling, the average of the elements in a picture part of a certain size is calculated. Sum Pooling calculates the components' cumulative sums inside the given section. Usually, the FC Layer and the Convolutional Layer are connected by the Pooling Layer..This CNN model generalises the features extracted by the convolution layer, and helps the networks to recognise the features independently. With the help of this, the computations are also reduced in a network.

3. Fully Connected Layer

The Fully Connected (FC) layer, which also has weights and biases, is used to connect the neurons between two layers. These layers are often positioned before the output layer and make up the last few layers of a CNN design. This flattens the input image from the layers beneath and provides it to the FC layer. The typical operations on mathematical functions are then performed on the flattened vector through a few more FC levels. At this moment, the classification process begins to take place. Two layers are connected because two fully connected layers perform better than one connected layer. The degree of human oversight is reduced by these CNN levels.

IX.RESULT:

PERSON IDENTIFICATION :

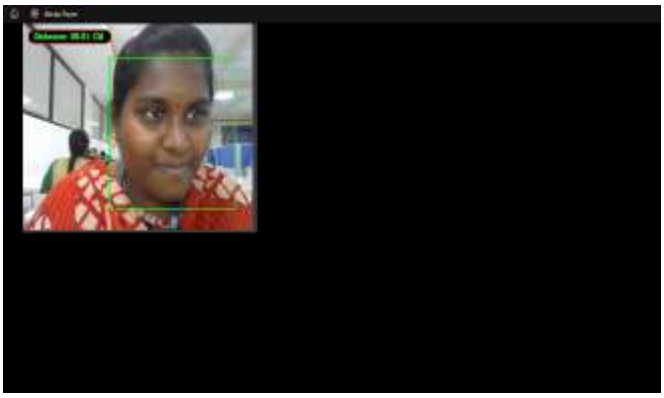


PERSON TRACKING :

PERSON DISTANCE:

Below 30:

Above 30:



DISCUSSION:

Person tracking and identification play a crucial role in a variety of applications, including security, surveillance, and HCI. It has been demonstrated that deep learning methods work well for handling these jobs. The process of person identification is identifying a person using their biometric characteristics, such as their face or fingerprint. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Siamese Networks are examples of deep learning models that have been utilised for face, voice, and gait detection, respectively. These models may learn discriminative features that are resistant to changes in light, position, or occlusion since they are trained on huge datasets of labelled biometric data. Finding and tracing a person's path across a video series is known as person tracking. Region-based Convolutional Neural Network is one example of a deep learning model (R-CNN), for object detection and tracking, as well as CNN. Based on a person's look, motion, and surrounding circumstances, these models may identify and follow them. They can deal with occlusions, different poses, and crowded backdrops. But tracking and identifying people is not without difficulties. The privacy of the data and ethical issues are two of the major difficulties. It is important to carefully analyse the ethical and legal implications of using biometric data for person identification in relation to concerns of privacy, consent, and fairness. The resilience of deep learning models to adversarial assaults and spoofing attempts is another issue. A mask or a false photo, for instance, can be used to trick face recognition models. In conclusion, deep learning-based person recognition and tracking have produced positive outcomes in a number of applications. However, it is important to address the challenges related to data privacy and security, and to ensure the fairness and robustness of the models.

CONCLUSION:

In person identification and tracking using a webcam in deep learning without a dataset can be a challenging task as it requires developing a deep learning model without the use of a large pre-existing dataset. This approach can be useful in scenarios where collecting a large labeled dataset is not feasible or where the specific application requires a highly personalized model.we can identify and track individuals with high accuracy in real-time. However, the use of these technologies raises ethical and privacy concerns, particularly with regards to surveillance and individual privacy. Therefore, it is important to establish clear regulations and guidelines to ensure that these technologies are used ethically and responsibly. Despite these challenges, person identification and tracking using a webcam in deep learning without a dataset has the potential to enable new applications and improve the performance of existing ones. Future work may focus on exploring novel techniques for developing deep learning models without the use of a dataset, such as using generative models or transfer learning. Furthermore, the research in this field is still ongoing, and there is much room for improvement in terms of accuracy, speed, and robustness. Future work may focus on developing more efficient and effective deep learning models, incorporating additional modalities such as audio and depth information, and exploring novel applications. In summary, person identification and tracking using a webcam in deep learning without a dataset can be a valuable approach in certain scenarios, but requires careful consideration of the limitations and potential trade-offs.

REFERENCES :

1. Bolívar Chacua, Iván García, Paul Rosero, Luis Suárez, Iván Ramírez,Zhima Simbaña, Marco Pusda “People Identification through Facial Recognition using Deep Learning”(2019).

2. Dimitrios Meimetis, Ioannis Daramouskas, Isidoros Perikos “Real-time multiple object tracking using deep learning methods”(2021).
3. Dendorfer P et al MOT20 “a benchmark for multi object tracking in crowded scenes” (2020).
4. Emami P, Pardalos PM, Elefteriadou L, Ranka S “Machine learning methods for data association in multi-object tracking” (2020).
5. Karunasekera H, Wang H, Zhang H “Multiple object tracking with attention to appearance, structure, motion and size” (2019).
6. Khan G, Tariq Z, Khan MUG. “Multi-person tracking based on faster R-CNN and deep appearance features” (2019).
7. Liu G, Liu S, Muhammad K, Sangaiah AK, Doctor F “Object tracking in varying lighting conditions for fog based intelligent surveillance of public spaces” (2018).
8. Markus Marks, Qiuhan Jin, Oliver Sturman, Lukas von Ziegler “Deep-learning-based identification, tracking, pose estimation and behavior classification of interacting primates and mice in complex environments” (2022).
9. Misbah Ahmad, Imran Ahmed, Fakhri Alam Khan, Fawad Qayum, Hanan Aljuaid “Convolutional neural network–based person tracking using overhead views” (2020).
10. Ramar Ahila Priyadarshini, Selvaraj Arivazhagan , Madakannu Arun “A deep learning approach for person identification using ear biometrics”(2020).
11. S. Sivachandiran, K. Jagan Mohan , G. Mohammed Nazer “Deep Learning driven automated person detection and tracking model on surveillance videos” (2022).
12. Sun S, Akhtar N, Song X, Song H, Mian A, Shah M “Simultaneous detection and tracking with motion modeling for multiple object tracking” (2020).
13. Voigtlaender P, Krause M, Osep A, Luiten J, Sekar BBG, Geiger A, Leibe B MOTS: “multi-object tracking and segmentation” (2019).
14. Wojke N, Bewley A, Paulus D “Simple online and real-time tracking with a deep association metric” (2017).
15. Y. Al Atrash “Detecting and Counting People Faces in Images Using Deep Learning Approach” (2019).
16. Yingkun Xu, Xiaolong Zhou, Shengyong Chen “Deep learning for multiple object tracking” (2019).