International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 4 ǁ 2022 ǁ PP. 10-14

SAS Security App: People Surveillance System

Aditya Sasmal
Department of Information Technology
Vidyalankar Institute of Technology (Wadala)
Mumbai, India

Shivam Vishwakarma
Department of Information Technology
Vidyalankar Institute of Technology (Wadala)
Mumbai, India

Sanya Gangwar
Department of Information Technology
Vidyalankar Institute of Technology (Wadala)
Mumbai, India

Prof. Dilip Motwani
Department of Information Technology
Vidyalankar Institute of Technology (Wadala)
Mumbai, India

Abstract
We present modified surveillance system using deep neural network and computer vision. A system that recognizes and locates people in a frame using object recognition and performs certain predefined tasks. People surveillance system is capable of people detection, keeping count, monitoring social distancing and many other operations. Main focus of the system is to recognize people in real time. It uses comparatively most efficient and accurate approach to detect a person in a pixelated image. The person detection model is capable of detecting and tracking the location of each person in the image or real time video. The tracking of person is required to add different operations to system like social distancing and other such operations. The approach we used for person detection is a neural-network based approach. Each layer of the network detects and studies different features of predefined training dataset, records values of pixels of those features and gives an estimation of how likely the object detected is a person. The network then discretizes the output space of bounding
boxes around the detected person. Many different operations can be added over this network. Our proposed model carries out person count and social distancing operations. This system can be easily deployed with standard surveillance system, modifying surveillance into an AI surveillance.
Keywords: Object detection, SSD, VGG16, python, social distancing, mask detection.

Date of Submission: 15-04-2022 Date of acceptance: 30-04-2022

I. INTRODUCTION

Today, technology is deviating towards AI rapidly and effectively. Every simple or supposedly difficult tasks are been carried out by AI. Controlling house lights to predicting weather, computer scientists are exploring capabilities of AI day by day. AI is capable of carrying out many calculations and estimations way effectively and rapidly than a human, though there are some tasks which humans can perform effectively, like detecting anddifferentiating objects, which AI is not at all good at. There are approaches been developed to make AI capable of such simple tasks for decades. The most popular and effective approach being worked on is Neural networks. Using neural networks, AI is able to recognize, distinguish and categorize objects with near to accurate estimation.

We aspire to use this new developed capability of AI into standard surveillance systems. Standard surveillance system can just record real time happenings, but is not capable of performing any operations to them. Using object recognition with surveillance, can increase its capabilities and also can be made to do different operations like keeping live count of people, monitoring social distancing and many more.

II. METHODOLOGY

There are numerous approaches developed for object detection using deep neural networks. KNN, RNN, YOLO and SSD [1][2]are some most efficient and near to accurate approaches preceding many more. YOLO and SSD provides object location tracking in real time with bounding boxes which is necessary for the proposed system to perform different operations [2][3]. YOLO (You only Look Once) and SSD (Single Shot Detection) approaches are similar but differ in architecture. SSD is more accurate and efficient as compared to YOLO according to studies. So, we use SSD approach for person detection task.

2.1 SSD (Single Shot Detection)

SSD is a multi-layer multiple box detection model. It detects an object in the frame and marks a bounding box across that object. The model works with multilevel neural networks, each level extracting different features of the object.SSD model is significantly faster in speed and highly accurate. SSD achieves high detection accuracy by using multiple boxes or filters with different sizes, and aspect ratio for object detection. It also applies these filters to multiple feature maps from the later stages of a network. This helps perform detection at multiple scale.

2.1.1 SSD Architecture

SSD model is divided into two components: A Convolutional Neural Network model as backbone and SSD Head [3][4]. A CNN model or base network extracts features from the input image. There are many CNN models like ResNet, Faster RCNN and VGG16 [5]. We use VGG16 as the base network.

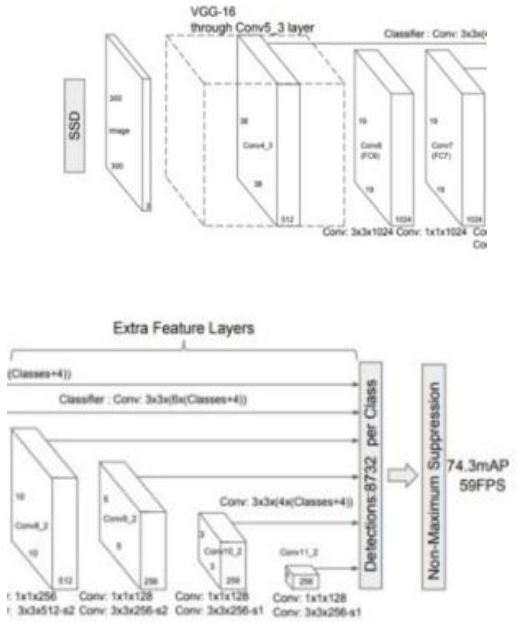

Figure 1. SSD Architecture

VGG16 is a convolutional neural network model. The basic idea is to extract features from the input image. The model uses convolutional layers with filters of size 3x3 and stride1 with same padding [6]. All the max-pooling layers have a filter size of 2x2 with stride 2. The width and height of the reduces by half with each pooling layer and the number of filters doubles with each set of convolutional layers. All the hidden layers are equipped with rectification non-linearity. The width of convolutional layer is small, starting from 64 in the first layer and then increasing by a factor of 2 after each max pooling, until it reaches 512 [7].

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**

Figure 2.VGG 16 Configuration

The SSD head is added layers of convolution to the VGG16 backbone. The outputs of this architecture are interpreted as the bounding boxes and object detection in the spatial location of the final layer activations [3].

2.2 Person Counter

The SSD model is a trained and tested just for detecting people in an image. This model detects people in an image with 97% accuracy. Whether there are only few scattered people or a crowd, this model is capable of detecting them with near to accurate estimation. We use this accuracy to count the number of people in an image or in real time.

We have designed a counter function to keep the count of number of people in a frame. Each bounding box located across each person is counted by the system and the count is displayed in real time. We can store the data of each day or each hour or even each second. We store hourly data of the count in the database. The system can estimate the count of people at any point in a day using the hourly data count.

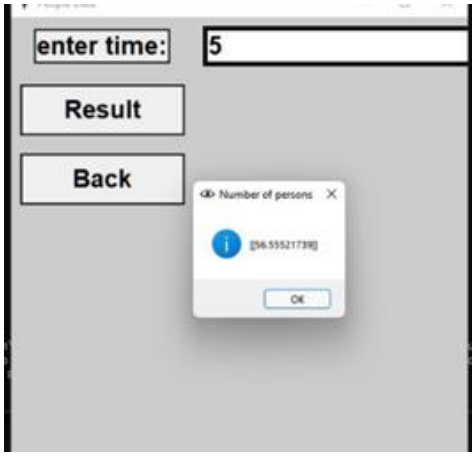

Figure 3. Person Count

2.3 Social Distancing Monitoring

The system recognizes a person in the image and plots a bounding box of its size across it. This bounding box locates the person in the image. The location of the bounding box i.e., co-ordinates of the box in xy plane is extracted. Using the xy co-ordinates, we calculate the centroid of the box [8]. Centroid of every bounding box in the frame is calculated in real time. Then we calculate Euclidian distance between every centroid in the frame. All of this is done by centroid tracking algorithm [9].

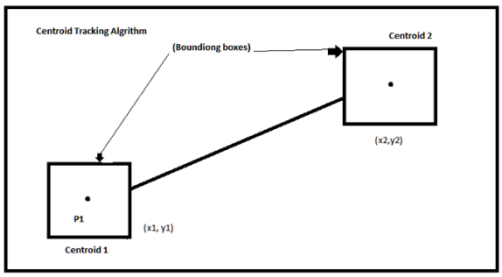

Figure 4. Centroid Tracking

In real world, social distancing is maintained by people being atleast one meter apart from each other. Likewise, in the computer system, we set a threshold based on the angle and size of the frame. If the calculated Euclidian distance is less than set threshold between any two or more people, their bounding boxes turn red indicating social distancing is violated. And the boxes remain or turn green, if the social distancing is maintained. This system when deployed with surveillance can monitor social distancing in real time with high speed and accuracy.

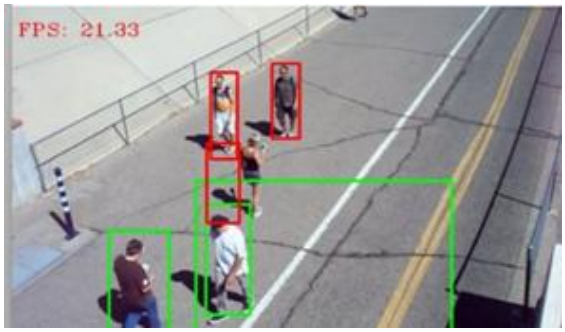

Figure 5. Social Distancing Monitoring

2.4 Mask Detection

The system is trained to detect mask as an entity. The system locates a person's face with bounding box and search for inferences of mask entity over inferences of person's face. If the inferences collide to certain point, the mask is detected on the face and the bounding box on the face of person is green, and if the mask is not detected, the bounding box turns red [10].

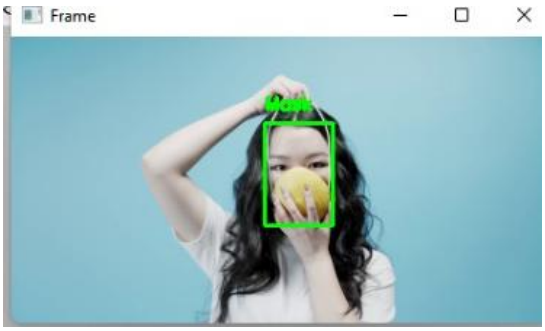

Figure 6. Mask Detection

III. SCOPE OF IMPROVEMENT

SSD model is near to accurate model. The accuracy can be improved. When two people are in very close premises, sometimes they are detected into one bounding box. To overcome this, the model can be trained with more such datasets.

The mask detection does not work with person faced sideways. This can be improved with more training of the system with different datasets.

The data is stored in hourly basis in the database due to shortage of storage. Real time data can be stored using cloud databases. Dynamic database can be helpful for real time storage and operations.

IV. CONCLUSION

This paper introduces an AI surveillance system. Apart from usual surveillance, this system is capable of recognizing people in the frame and tracking them with bounding boxes in real time. For the person detection and tracking, SSD is comparatively the most efficient and near to accurate model. Using this model, the system is able to track people in real time and perform different operations. Person counting is one of the operations, it counts people in real time, whether it be a crowd or few scattered people. Social distancing being necessary in the pandemic, this system is capable of monitoring social distancing via surveillance. Another regulation in the pandemic was wearing masks, which is also detected by this system, if the person in the frame is wearing mask or not. The proposed surveillance system can be used in malls, schools, societies, streets and events, in order to make sure that pandemic regulations are being followed or not.

REFERENCES

[1]. Uijlings, J.R., van de Sande, K.E., Gevers, T., Smeulders, A.W.: Selective search for object recognition. IJCV (2013)
[2]. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: CVPR. (2016)
[3]. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg: Single Shot MultiBox Detector. arXiv (2016)
[4]. Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam: MobileNets Efficient Convolutional Neural Networks for Mobile Vision Application. arXiv (2017)
[5]. Yuhong Li, Xiaofan Zhang, Deming Chen: CSRNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes. CVPR (2018)
[6]. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation.CVPR. (2015)
[7]. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Object detectors emerge in deep scene cnns. ICLR. (2015)
[8]. Maya Aghaei, Matteo Bustreo, Yiming Wang, Gianluca Bailo, Pietro Morerio, Alessio Del Bue: Single Image HumanProxemics Estimation for Visual Social Distancing. arXiv (2020)
[9]. Qiuyu Zhu, Pengju Zhang, Xin Ye: A New Loss Function for CNN Classifier Based on Pre-defined Evenly-Distributed Class Centroids. arXiv (2019)
[10]. Haiwei Wu, Lin Zhang, Lin Yang, Xuyang Wang, Jun-Jie Wang, Dong Zhang, Ming Li: Mask Detection and Breath Monitoring from Speech: on Data Augmentation, Feature Representation and Modeling. arXiv (2020)