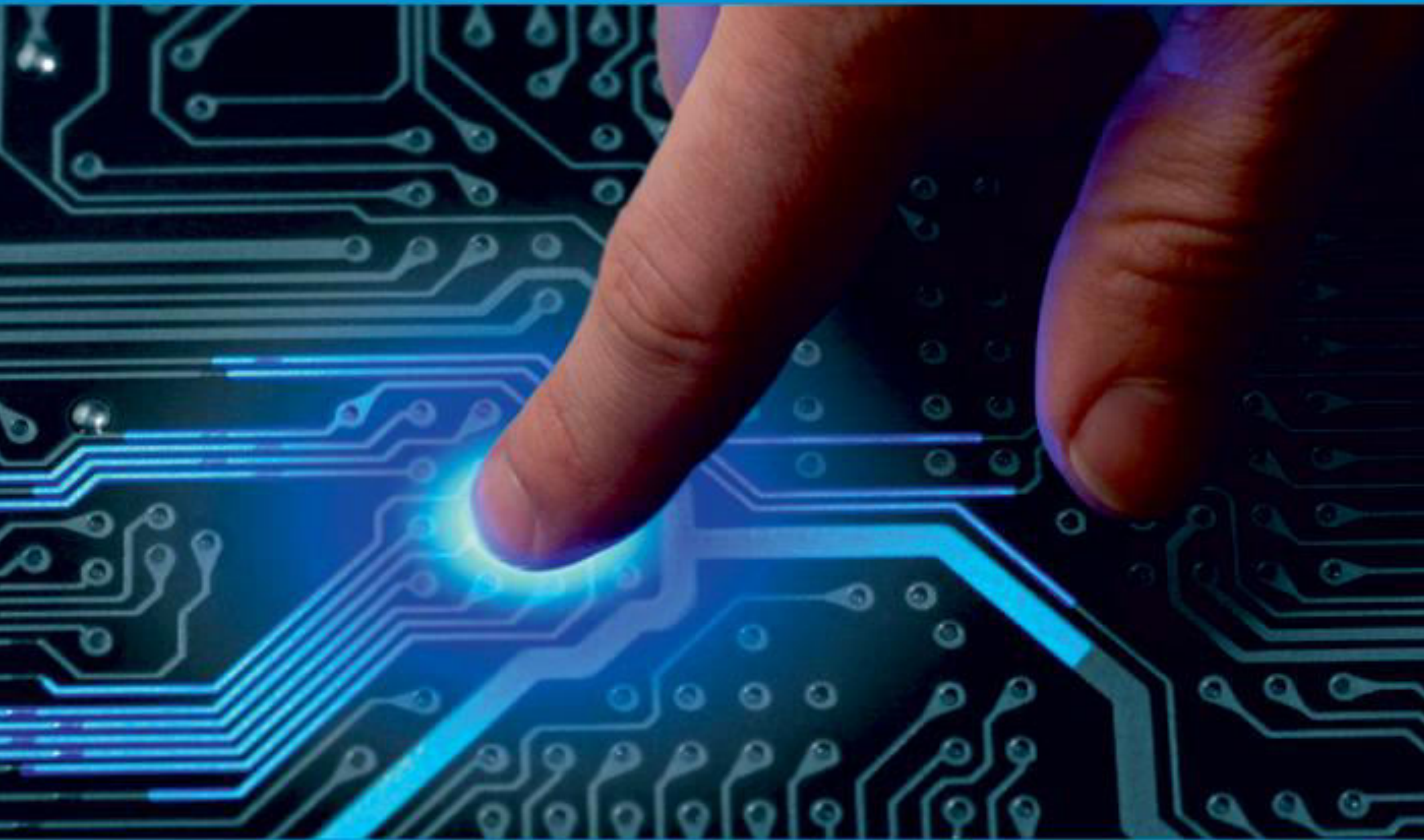
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 7, July 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107079 |

Smart Count: A Deep Learning Approach For People Counting System

M ANAND, V MALLESI

PG Student, Dept. of C.S.E., Bheema Institute of Technology & Science, Adoni, India

Assoc. Professor, Dept. of C.S.E., Bheema Institute of Technology & Science, Adoni, India

ABSTRACT: Population growth has accelerated recently. Due to the population's linear increase, many individuals now frequent public spaces. Thus, this technique will give the number of people in a specific area's malls, supermarkets, etc. Some businesses exclusively depend on the timing and schedule of their customers. Therefore, by creating a system for counting people, our work satisfies the problem and offers a solution. Therefore, a centroid tracker and the single shot detector (SSD) mobile net are suggested. This model links six convolutional layers after the VGG16 base network for classification, replacing the base network with a Mobile Net network for improved feature extraction. In order to calculate the center of a bounding box, the centroid tracking algorithm uses bounding box coordinates from an object detector SSD. The model runs the dataset with training and testing data, as illustrated in fig. 4, and when the centroid is calculated, it will provide each person an ID. False positive rates (FPR) are limited to 0.08%, while the maximum true positive rate (TPR) is 95.03%. 1218 negative picture samples and 2416 positive image samples altogether. Next, the given data set is tested to determine the correctness of the model, which yields a 96.64% accuracy rate. Sometimes methods don't yield highly accurate results because of things like high traffic counting, people's clothing, and shadow problems.

KEYWORDS: people counting, mobile net, SSD, centroid tracker, CNN.

I. INTRODUCTION

The major goal of this approach is to raise the level of business through a people counting system. It serves as the basis for many common vision-based applications, including those used in intelligent transportation systems, retail establishments, shopping centers, and supermarkets. [1][2][10]. In this project, an overhead perspective is used to count the number of participants. Centroid tracker and mobile network are two distinct types of algorithms used for counting, with centroid tracker serving as an object tracking method and mobile network as an object detecting technique. If someone establishes the boundaries, they will send an alarm message any time someone crosses them.

In the past, people were found and counted using first-generation devices that used infrared sensors. The next to be introduced are the third-generation devices, which combine computer vision and video computing, which is based on image processing, for improved accuracy results. After that, the second-generation devices come with thermal image sensors for detecting and counting people. There are various approaches used in computer vision and image processing for counting people, including the overhead view people detection and counting system, Real-Time People Counting, Face-Detection, and strategies for detecting the head and shoulders [1]. By applying these algorithms, You only look once (YOLO) [5], Single Shot Detection (SSD), and RGB-D (RGB plus depth) There are multiple cameras used to capture the image of the crowd of people using the MCNN multi-camera convolution neural network [11]. It synchronizes the frames from the many input cameras and outputs an image of the crowd. For each convolution layer of a CNN, MCNN functions independently. In a SCNN single-camera convolution neural network, a single camera is used to synchronize the frames. Similar to a convolution neural network, the SCNN operates. The researchers have a very tough time using object detection because there are so many different approaches and procedures used for person detection. These methods use an overhead perspective to detect the individual from a different angle.

II. LITERATURE REVIEW

Many image and video processing techniques have been developed in the past few years for the purpose of detecting individuals. In computer vision and video computing generation, some examples are the Haar Cascade, Convolutional Neural Networks (CNN), Single-shot detector (SSD), and You Only Look Once (YOLO) [7][5]. Due to

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107079 |

each grid having only two bounding boxes, the YOLO algorithm has a limited recall and will not adequately detect items that are close to it.

An infrared beam line parallel to the ground acts as a first-generation infrared sensor, counting each time a person or item passes by and breaks the beam [6]. The second generation of technology is the thermal sensor. wherein the sensor picks up and records infrared light that the human body emits. Then, the video computing and computer vision algorithms are developed for detection purposes. There are numerous techniques available that are highly accurate for detection [9].

The author displays an overhead real-time RGB-D system view. In this strategy, the model first identifies the head-shoulder of the passing person and extracts the head-and-shoulder features before making predictions using a commercially available depth camera that is mounted in the primary entrance gate [11]. The primary objective of the foregoing discussion is often overhead view person detection using handmade feature-based algorithms [12].

III. METHODOLOGY

Both the SSD Mobile Net and centroid tracking algorithms were employed in this research, along with the deep learning object detector for people counting, OpenCV for general computer vision/image processing tasks, and the OpenCV. After that, correlation filters are implemented using the dlib package. Although dlib library was easier to work with for this project, OpenCV is used here. A deep learning network has been used to explore this SSD model for persons detection. Additionally, the SSD model uses the Mobile Net as its base network. This model was trained using a train data set and a test data set, which were both used to determine the model's correctness. Images, classes, and scaling the pixels by a NumPy array are all part of the model data. Once the model is trained, SSD predicts each class using six layers, as seen in Fig. 3. The model is then prepared to identify humans in an overhead perspective.

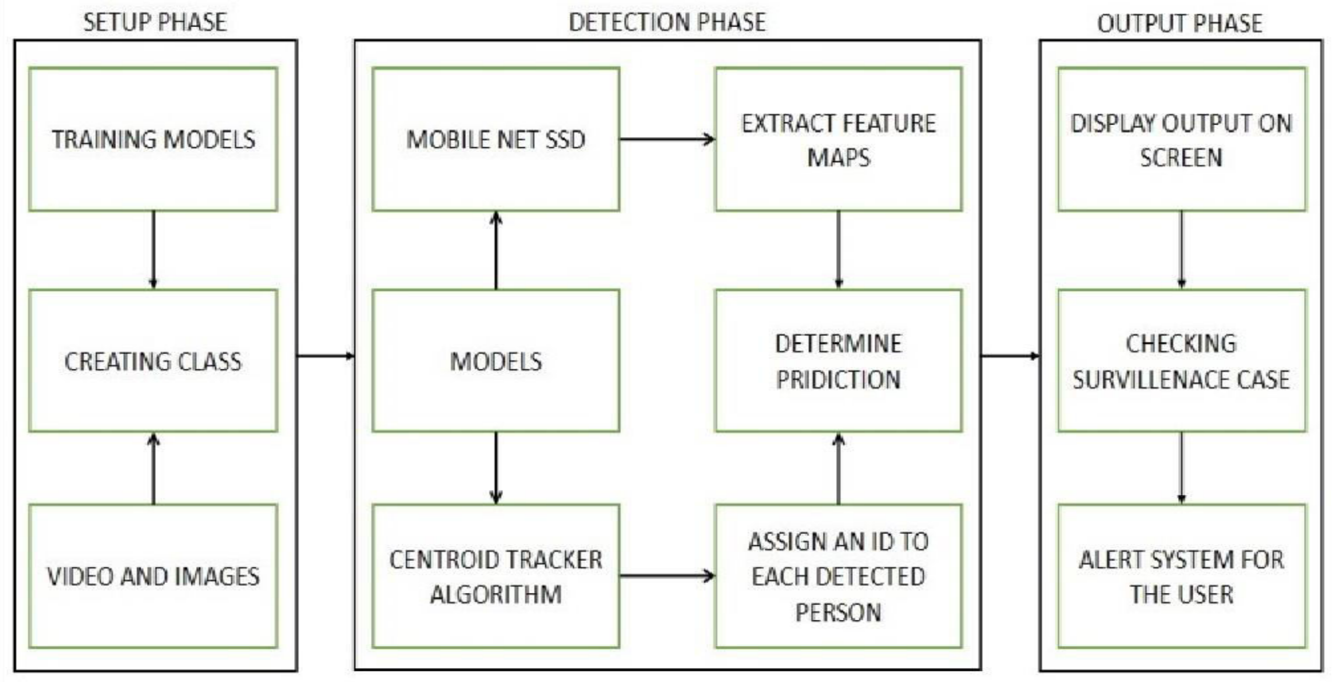


Fig. 1.*Block diagram of working model*

The model is composed of three parts, including setup, detection, and output. The mobile net model is trained using the below-mentioned Inria data set during the setup phase. Every time an object enters the field of view of the camera, a class is created for that object. Class then enters the detecting phase. The mobile net SSD recognizes an object or person in the form of a bounding box during the detection phase and extracts an object's feature. A unique ID is given to each person by the centroid tracker, which also tracks the individual. The model makes a forecast if the discovered individual is real. The prediction then proceeds on to the output phase, where it shows the overall number of predictions that were made. Whenever the number of persons counting surpasses the limit, a notice is issued to the staff if a limit has been put on the number.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107079 |

A. Data sets:

The dataset is the set of information used to train the algorithm and test the model's precision. The datasets used in this model are taken for training from this website: http://pascal.inrialpes.fr/data/human/. The model in this system has been flawlessly trained using train data and tested using test data. Pixels, pictures, and scaling the pixels via NumPy array are all part of the model data.

Fig 2. *Trained the model from the above Datasets.*

B. SSD with mobile net

A well-known object detection algorithm is single shot detector (SSD). It is superior because it just requires one shot to detect humans, as opposed to two shots required by RPN-based proposals like R-CNN. The base network and detection network are the two categories into which the deep neural networks' network is divided. For the detection layer, the base networks offer a better collection of extract features. Additionally, it features six convolutional layers for object detection and categorization following the basic network.

Because the Mobile Net network contains numerous weight settings, it is used in the work reported here instead of the VGG16 base network. As a result, the model grows large, which also results in a lengthy inference period. The mobile network is thirty times smaller and faster than VGG16, and for better extracting features. After the mobile network processes the input from the detection network to produce the output that contains the detected people coordinate that contains the midpoint and detected bounding boxes, which have many multiple boxes so that non-maximum suppression chooses the best bounding box for the object, these two algorithms have similar accuracy.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107079 |

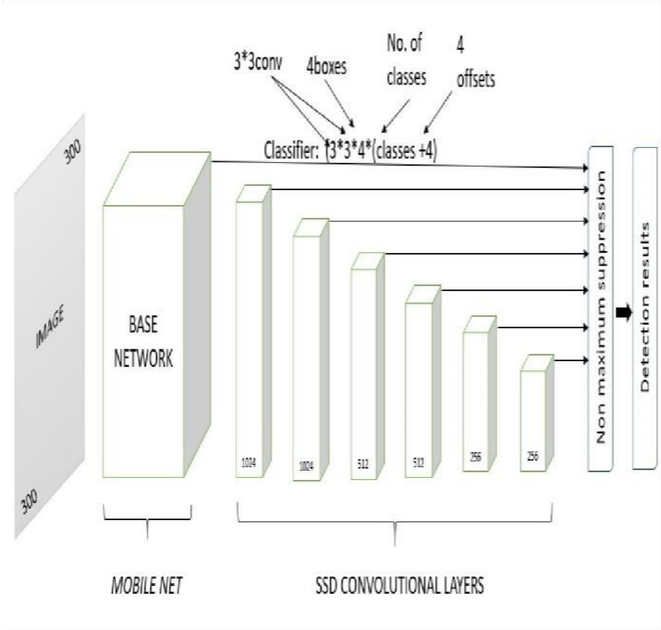


Fig. 3. *SSD architecture with Mobile Net.*

C. Centroid tracker algorithm

The centroid tracking algorithm is an object tracking method. In centroid tracking algorithm, it is a combination of object tracking and object detection. For implementing, bounding box coordinates is used from an object detector SSD and applies them to calculate the center of a bounding box. After the centroid is calculated, it will assign an id to each person which is detected by the SSD algorithm. If a new person appears on a screen whose id did not match the previous person's id, a new id will be assigned to that person. The bounding box itself provides an object detector and object tracking. Object detectors such as SVM and R-CNN, as well as correlation filters, are used for object tracking.

IV. Experimental And Results

In this section, the results of people counter system and accuracy of our model was depicted. The model is trained on pc Intel core i5 with 8GB ram. The model is trained and observed the different parameters such as true positive rate (TPR), false-positive rate (FPR), true negative rate (TNR), false-negative rate (FNR), and also recall, precision, is shown in Table 1. After that, the evaluation is calculated for the model by using true positive, true negative and total samples images. The model gives a good result as compared to other, previous model. The model parameter carried a maximum TPR of 95.03% which is greater than the previous model. And the FPR is also less than the previous model which is 0.08 and the previous model FPR is 0.2%. In this data, positive image sample of 2416 and negative image sample of 1218 were considered. The accuracy of our model is achieved by this given formula:

Table 1: *Performance results of trained model.*

S. NO	RESULT	PERCANTAGE
1	TPR	95.03%
2	FPR	0.08%
3	TNR	99.83%
4	FNR	9.85%
5	PRECISION	0.999
6	RECALL	0.950
7	ACCURACY	96.64%

```
Positive Image Samples: 2416
Negative Image Samples: 1218
Reading Images
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
libpng warning: iCCP: known incorrect sRGB profile
True Positives: 2296 False Positives: 2
True Negatives: 1216 False Negatives: 120
Precision: 0.999129677980853 Recall: 0.9503311258278145
F1 Score: 0.9741196436147647
```

people was counted with the help of centroid for assigning

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107079 |

This model is successfully detected and the number of
the id and SSD for object detection. The accuracy of our
TPR= TP / Positive model is 96.64% which is tested for the above-given dataset.
FPR = FP/ Positive The performance of the people counting system detects the **TNR = TP/Negative** people which are shown in Fig.5.
FNR = FN/Negative
Fig. 5. *Detected and counted people on a different video*
Precision =TP/(TP+FP) *shown below.*
Recall = TP / (TP+FN)
Accuracy = TP+TN/Total

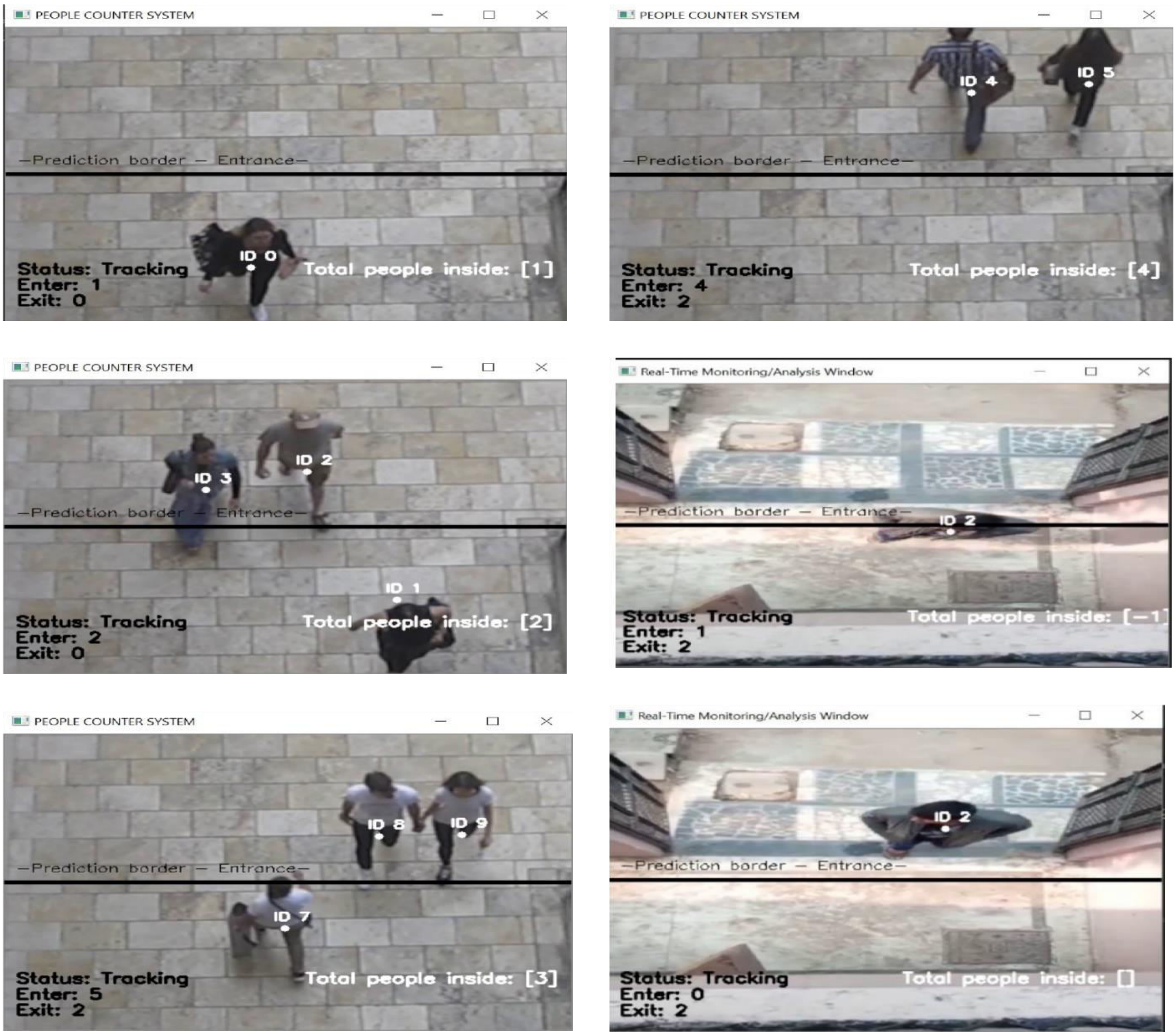


Fig. 4. Tested model output result

V. CONCLUSION AND FUTURE WORK

This study attempts to concentrate on the People Counting system that the deep learning SSD model investigates. It was suggested that the single shot detector (SSD) might offer superior outcomes to the YOLO paradigm. When

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107079 |

compared to SSD, the YOLO model has a lower recall and more localization error. The VGG16 convolutional neural network was swapped out in this model. The primary goal is to replace the VGG16 convolutional neural network model, which is 30 times bigger than mobile net SSD in terms of size. Because our model needs improved extraction features and non-maximum suppression choices, the SSD cannot be compacted.

This model was created to help interpret experimental findings, and it provides good accuracy, with a TPR of 95.06% and an FPR of 0.08%. By using a different data set and training it in a different location, the output can be further improved. Due to factors like clothing, shadows, high traffic counting, etc., this method does not always provide results with great precision.

In the future, various crowd detection, traffic counting, and vehicle detection algorithms can be used in conjunction with top-down view datasets using the same algorithm. The main drawback of SSD is the lack of deep layers in neural networks, which prevents them from producing accurate predictions for small objects. As a result, the SSD technique is incompatible with the detection of small objects. In further work, our model can be trained using a huge training dataset or a sophisticated dataset augmentation.

REFERENCES

[1] Sun, S., N. Akhtar, H. Song, C. Zhang, J. Li, & A. Mian. Real-Time People Counting in Cluttered Scenes with Depth Sensors: Benchmark Data and Method. 20(10), 3599–3612, IEEE Transactions on Intelligent Transportation Systems.

[2] Jing Liu, Jing Liu, and Ming Zhang (2013). a real-time people counting technique based on detection and tracking. Chinese Automation Congress of 2013.

[3] Vishaka The authors are Gayathri, D., Shree, S., Jain, and Sornalakshmi (2018). Real-time Human Tracking and Identification from Surveillance Videos System. Engineering & Technology International, 7(3.12), 244.

[4] Yi-Li Kuo, Da-Jinn Wang, Chao-Ho Chen, and Tsong-Yi Chen. (2010). a face-based system for counting people. Fourth International Genetic and Evolutionary Computing Conference, 2010.

[5] Adnan, A., I. Ahmed, and M. Ahmad (2019). YOLO Overhead View Person Detection. UEMCON is the 2019 IEEE 10th Annual Conference on Ubiquitous Computing, Electronics, and Mobile Communications.

[6] Goncalves do Patrocinio, Z. K., Guimaraes, S. J. F., and Roldao Reis, V. H. (2020). Capsule networks for dense crowd counting. The IWSSIP 2020 International Conference on Systems, Signals, and Image Processing.

[7] "Bench mark data and method for real-time people counting in cluttered scenes using depth sensors," arXiv preprint arXiv, by H. Song, S. Sun, N. Akhtar, C. Zhang, J. Li, and A. Mian.

[8] Cho, S. I., et al. (2018) Kang, S.-J. Customer movement analysis using a real-time people counting system. 1 IEEE Access.

[9] According to Wang, Z., Cai, G., Zheng, & Fang (2018). Estimating bus crowds using a shallow convolutional neural network. 2018 SNSP (Sensor Networks and Signal Processing) International Conference.

[10] S. Liu, S. Zhai, C. Li, and J. Tang. Crowd counting method that works well using CNN-based statistical features. International Conference on Smart Cities 2017, or ISC2.

[11] H. Fu, H. Ma, and H. Xiao (2012). Based on RGB-D data, precise crowd counts in real time. IEEE's 19th International Conference on Image Processing was held in 2012. doi:10.1109/icip.2012.6467452.

[12] Cho, Sung Ho; Choi, Jeong Woo; Quan, Xuanjun (2017). IR-UWB Radar Sensor-Based Bidirectional Passing People Counting System. Journal of the IEEE Internet of Things, (), 1-1.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com