Crowd Analysis for Congestion Control Early Warning System on Foot Over Bridge

Narinder Singh Punn
Dept. of Information Technology
Indian Institute of Information Technology Allahabad
Prayagraj, India
pse2017002@iiita.ac.in

Sonali Agarwal
Dept. of Information Technology
Indian Institute of Information Technology Allahabad
Prayagraj, India
sonali@iiita.ac.in

***Abstract*—Crowds occur in a variety of situations like concerts, rallies, marathons, stadiums, railway stations, etc. Crowd analysis is essential from the point of view of safety and surveillance, abnormal behavior detection and thereby reducing the chance of a mishap. Generally, congestion in the crowd can lead to severe problems like a stampede. This congestion is due to increasing crowd count; thereby increasing the crowd density in regions and abnormal crowd motion. Most of the congestion control approaches follow a hardware-oriented approach. This paper proposes a software-oriented approach, Congestion Control Early Warning System (CCEWS), for congestion control with the help of object detection and object tracking technique. Object detection is performed by following the faster R-CNN architecture in which Google inception model is used as a pre-trained CNN model and with the help of proposed object tracking technique the crowd abnormality is analyzed. The proposed congestion control technique exhibits quite significant results on the proposed dataset made from the virtual simulation of FOB (foot over bridge) scenario.**

Keywords—Congestion control, Convolution, Crowd behavior, Neural network, Object detection, Object tracking, Stampede detection.

I. INTRODUCTION

With the increase in the sudden mass accidents over the year due to crowd abnormal behavior [1], crowd analysis plays a crucial role in the area of public safety. To survive with the chances of occurrence of any misfortune, crowd behavior analysis is necessary [1]. This analysis is mainly focused on abnormality detection, which can be defined based on crowd motion direction and people's action or activities. Based on the detected abnormality, the amount of congestion can be determined.

During huge gatherings like "Khumb Mela" in India, people gather from all over the world due to which places like railway station, becomes overcrowded which may lead to crowd disaster. Inspired from the incident of "2017 Mumbai stampede" [2], Congestion Control Early Warning System (CCEWS) is proposed on FOB. In the early-warning system, the aim is to generate the safety alarms for abnormal crowd behavior due to congestion which may result in some mishap. Abnormal behavior can be detected with the help of computer vision methodologies particularly object detection [3], object tracking, and object motion direction, a software-oriented approach.

This paper proposes a novel approach to generate the warning signals whenever there are high chances of occurrence of some mishap like a stampede on FOB. The task of object detection is achieved by using the faster variant of Region based Convolutional Neural Network also known by the acronym "Faster R-CNN" [4] architecture in which Google's "inception-resnet" architecture [5] is used as pre-trained CNN model and for object/people tracking centroid based algorithm is proposed. The object motion direction is decided by combining the features of both object detection and object tracking. Based on these tasks and some defined thresholds, congestion in the crowd is monitored and the chances of some accident are computed.

According to the Google survey paper [6], C. Szegedy et al. found that the faster R-CNN with Inception Resnet v2 model outperformed the other state-of-the-art methods like R-FCN [7] and SSD [8] in the object detection task covering both metrics speed and accuracy.

II. RELATED WORK

A lot of research is going on crowd management which involves continuous monitoring and tracking of the people. Most of the researched methods follow hardware oriented approach [9, 10, 11]. These systems are designed using smartphone GPS, integrated mobile with RFID (Radio-Frequency Identification) systems, wireless sensor networks, IoT (Internet of Things) devices and some fusion of technologies like RFID and smartwatch, RFID with ZigBee, etc.

The smartphone GPS system proposed by M Mohandes to track and identify the location of the pilgrims [12] is dependent on server availability to share UID, latitude, longitude and time stamp, thereby position obtained might be inaccurate if the server is unavailable. In RFID based approach, an RFID tag is provided to each object. But doing so in crowd management is not feasible, and because of this, an improved method for object detection and tracking was proposed by Yuanyuan Fan and Qingzhong Liang [13]. Though they have used one RFID reader, two RFID tag arrays and one processing unit, the total system setup, and maintenance overhead is high. S. Liu et al. proposed the framework for computing the chances of occurrence of crowd stampede with the help of live footage from the surveillance cameras and utilizing the potential of computer vision and image processing methodologies [14]. But this framework needs improvement from the perspective of real-time implementation to produce more accurate stampede detection results. Pathan et al. used a background subtraction approach to separate the moving objects from the background; thereby reducing the computation time for estimating the pedestrian density [15]. This approach shows a good result for low-density crowd gathering but not suitable for regions where pedestrian density is high.

So, in this paper a software-oriented approach is designed for CCEWS by utilizing the real-time surveillance footage and compute the amount of congestion in the crowd and generate the corresponding alert.

978-1-7281-3591-5/19/$31.00 ©2019 IEEE

III. Proposed Architecture

In the proposed system of congestion control on FOB the crowd behavior is analyzed with the help of three tasks; object detection, object tracking and object motion direction, where the object is a human head.

Figure 1 shows the design of the proposed approach for an early-warning system to detect high congestion and generate safety alarms for the execution of safety plans.

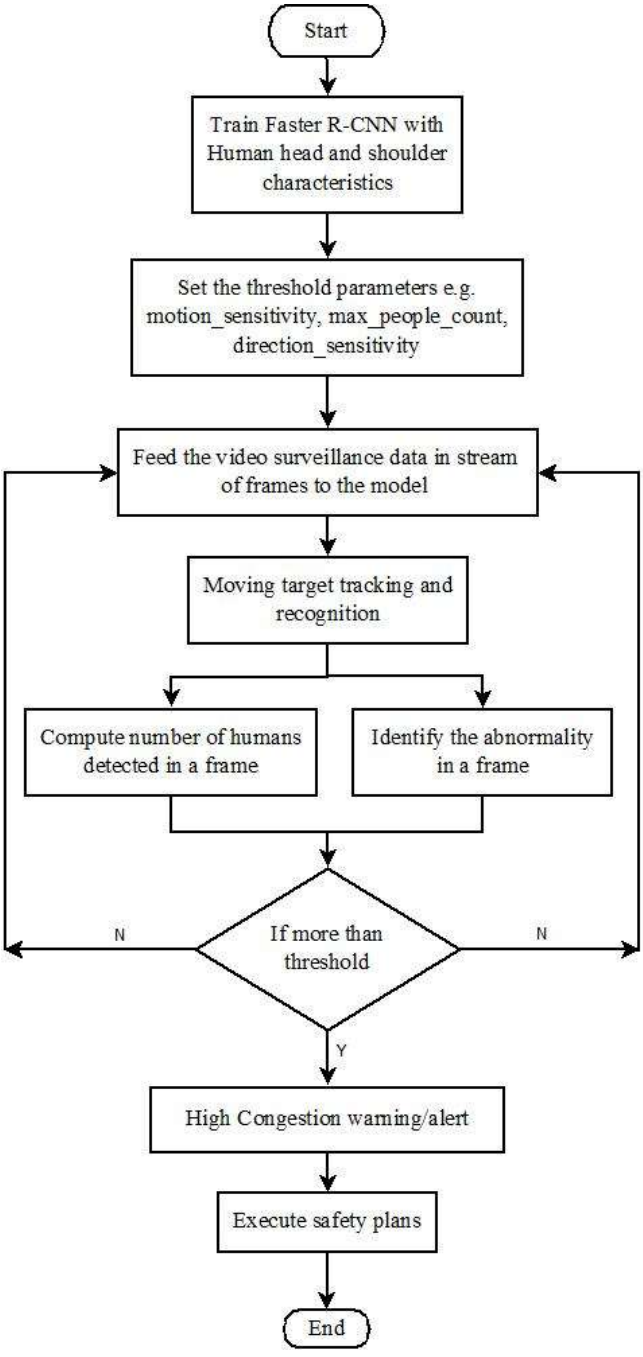

Fig. 1. System Design for Congestion Control

A. Object Detection

Faster R-CNN architecture [4] as shown in figure 2, helps to accomplish the task of object/head detection by deploying the input frame on a pre-trained CNN model, such as *Inception* architecture of *GoogLeNet* [5]. Then, *Region Proposal Network* (RPN) is used to detect the regions that might contain the objects in the feature map by generating the object proposal score and bounding boxes. Then the *Region of Interest* (RoI) pooling layer is used to extract the feature maps according to the regions proposed by RPN and the feature maps output from CNN. Finally, the output feature map is then used for classification and in fine-tuning of the bounding boxes via fully connected layers.

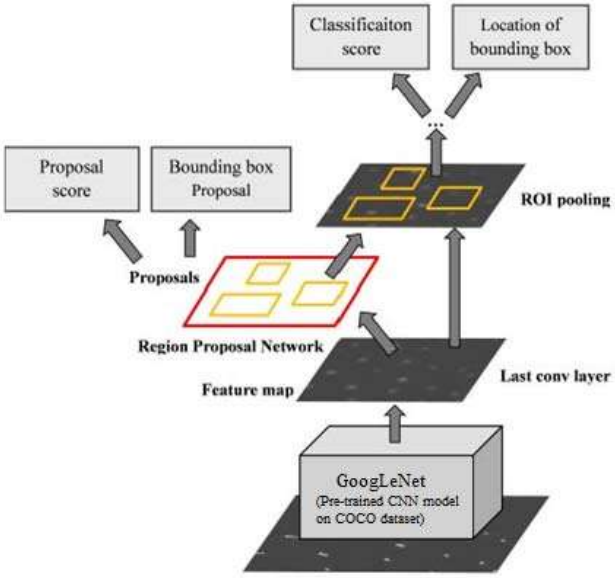

Fig. 2. Faster R-CNN architecture with GoogLeNet as pre-trained CNN model

B. Object Tracking in Crowd

This paper proposes a novel method to track the object. The idea is that the movement of a particular object in the consecutive frame is gradual, i.e. the distance traveled is less. The procedure to track an object (figure 3) is as follows.

- Compute the centroids of the bounding boxes of the objects in the consecutive frames.
- Centroids, which are closer to each other in the consecutive frames, belong to the same object. The Euclidean distance is used to measure closeness between the centroids of the two frames.
- If closeness property is not satisfied, then that object is treated as a novel object.

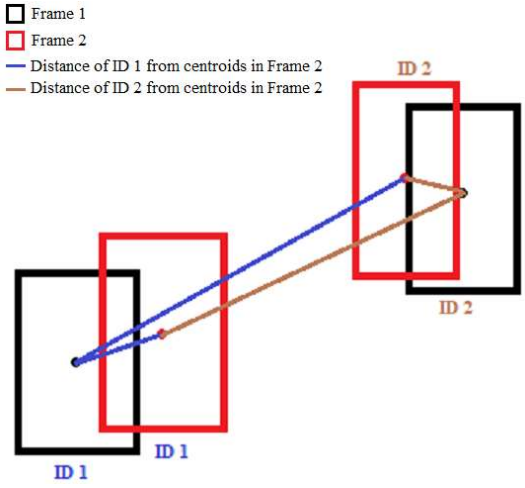

Fig. 3. Object tracking approach

For instance, consider two detected objects as shown in figure 3 (denoted by black color bounding box) with the tag as ID 1 and ID 2. Suppose they moved at a certain distance and their new position is represented by a red color bounding box. Now compute the distance of the centroid of ID 1 with the new centroids. The centroid which is closer to ID 1 is provided an ID 1 tag. Similarly for ID 2. The new tag is given to the object that has no such closer centroid.

IV. Implementation

The proposed system is implemented using TensorFlow [16] and is visualized with the help of Tensorboard.

Consider two flags, good_flag: Indicating the normal situation of the crowd and *bad_flag*: Indicating the abnormal situation of the crowd, leading to high congestion.

1. Train the faster R-CNN model for human head detection on the proposed dataset (discussed in Section V).
2. Do object detection and get coordinates of bounding boxes.
3. Find the area and centroid for each detected bounding box.
4. Maintain the previous areas and previous centroids to compare the boxes in the consecutive frames.
5. Track the objects by using proposed centroid based object tracking algorithm (discussed in Section 3.2).
6. Compare area of bounding boxes in two frames belonging to the same object to get the direction of motion of the crowd.
7. If the total number of people is greater than the number of people allowed in the frame
    - If the motion is in one direction (either towards the camera or away from the camera) then increment the good_flag by 1. (Increment is done by two because of this being excellent condition than abnormal).
    - If an equal number of people are moving in the opposite directions to each other, i.e. both towards and away from the camera then increment the bad_flag by 2. (Increment is done by two because of this being terrible condition than usual)
    - If an unequal number of people are moving in the opposite direction to each other, i.e. both towards and away from the camera then increment the bad_flag by 1.
    - If an object/person stays still for some time (computed by tracking the number of frames in which the person stays still) then also increment the bad_flag by 1.
    - This condition also follows the sensitivity parameter which ranges between 0 and 1; 1 being more sensitive. 1 means that if people are moving in one direction, then no one can come in the opposite direction, else increment the bad_flag by 1.
8. If in a frame the total number of people is less than or equal to the number of people allowed.
    - Then increment the good_flag by 2 (this is the most relaxed condition).
9. Do steps 5, 6, 7, 8, 9 for 20 frames. (Number of frames equal to 20 is not fixed can be changed as per requirement, but for accuracy concerns, it should be between 10 to 50).
10. Finally, if bad_flag > good_flag then generate the high congestion alert and store the particular frame in the output for analysis and execution of contingency plans.

V. Experiment

A. Dataset

Since the dataset was not available to satisfy the requirement of the objective of crowd analysis for congestion control on FOB, the new dataset was created via video shoot. This was done by virtually simulating the scenario of FOB on regular stairs at Indian Institute of Information Technology Allahabad (IIITA). Some of the frames of the video shoot are shown in figure 4.

Fig. 4. Sample video frames from the dataset

In this dataset, video (648 x 1152) is taken from four different viewpoints; people going up the stairs and people going down the stairs with two different camera positions, i.e. at the bottom and top of the stair. The congestion scenario is created by making motion as random as possible and label the corresponding node abnormal.

B. Training and Testing

Faster R-CNN model is trained and tested with the shot video frames. These video frames are manually annotated by the bounding boxes around the human head as shown in figure 5. The model is trained to learn the characteristic features of the human head even with the different orientations. The model was trained until there was consistently low loss, thereby it almost took 22,000 iterations for training.

Fig. 5. Sample annotated video frames from the dataset

C. Losses

Faster R-CNN is divided into modules; a deep convolutional neural network (output from RPN) and fast RCNN detector (final output). Each of these module exhibit two losses.

- *RPN Losses (module 1)*: Binary classification loss or objectness loss (based on object proposal score for being background or foreground) is shown in figure 6 (value: $8.2006e^{-3}$), and bounding box regression loss (based on bounding box coordinates proposal) is shown in figure 7 (value: 0.01053).
  Smoothing value is just for visualization of the graph, i.e. to make the curve as general as possible; not utilized in any computation. In the following graphs (figure 5,6) the dark orange colored curve is the generalized representation of the actual curve indicated by light orange color.

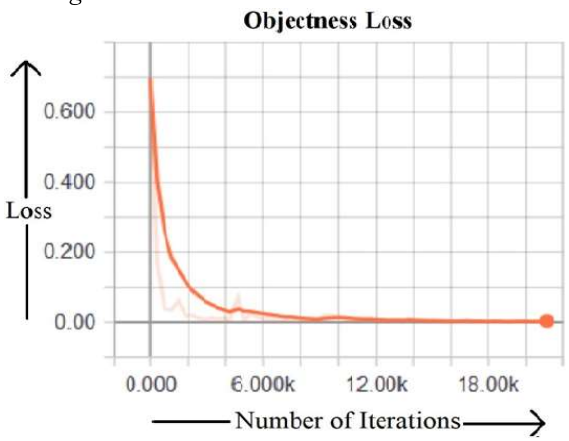

Fig. 6. Binary Classification loss over the number of iterations (smoothing value: 0.8)

- *Classification Losses (module 2):* Classification loss (based on some class labels) and bounding box localization loss (based on the fine-tuned bounding box coordinates).

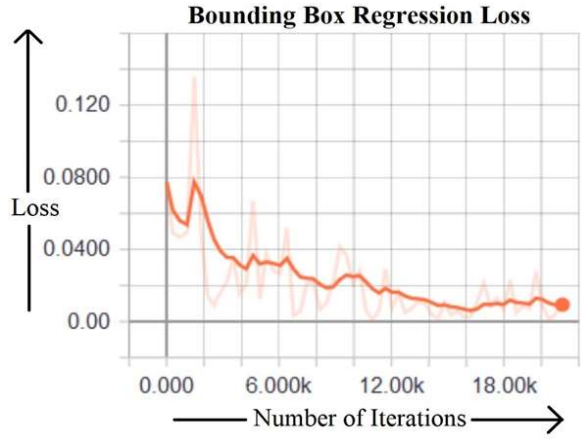

Fig. 7. Bounding Box regression loss over the number of iterations (smoothing value: 0.8)

Since only one type of class label was needed (head) and the objective of the second module of faster R-CNN is to classify the detected objects into multiple classes; thereby only the first module of faster R-CNN is needed to serve the purpose of human head detection. Thus, the overall loss (value: 0.06497) is computed as the weighted sum of the classification and the regression loss as shown in figure 8.

D. Threshold Parameters

The values are set pertaining to the experimental environment. These values are not fixed and may vary for a different environment. Table 1 shows the threshold parameters and respective values, which were kept during experimentation.

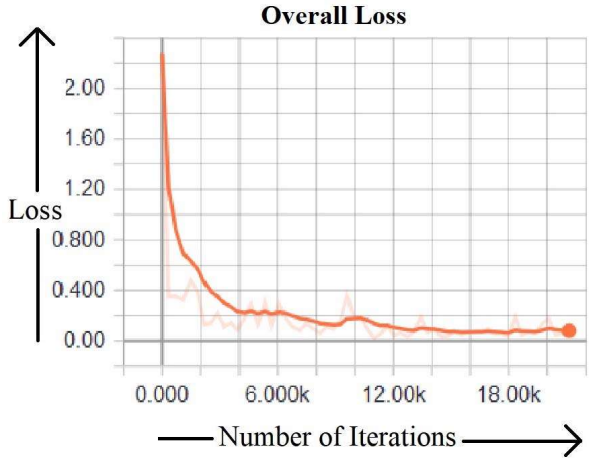

Fig. 8. Overall loss over the number of iterations (smoothing value: 0.8)

TABLE I. THRESHOLD PARAMETERS

Parameter Name <Value>	Description
area_diff_threshold <0.2>	Controls the sensitivity for an object's motion i.e. if the area difference between two consecutive frames is less than 0.2 then it is considered as a still object.
max_people_count <5>	A maximum number of objects allowed in a frame. The stampede risk is conducted if the number of people exceeds the max people count.
direction_sensitivity <0.8>	Ranges between 0 and 1, where 1 being maxed sensitivity. Controls the abnormality in the direction of motion of the object. For instance, consider its value to be 1. Now not even a person is allowed to move in the opposite direction to the crowd flow.
still_sensitivity <0.4>	Ranges between 0 and 1, where 0 is more sensitive. Track the number of frames in which still object count exceeds the fraction of total objects (still sensitivity * object_count) detected.

E. Output

The CCEWS output displays the people detected by drawing the bounding boxes over the head along with their count at the top. And if an alert is generated, an alert message is displayed beside head-count, and that frame is recorded for further analysis. Figure 9 shows the sample output in case of congestion alert. This alert signal was generated by making the model very sensitive (using the threshold values as discussed) towards congestion detection.

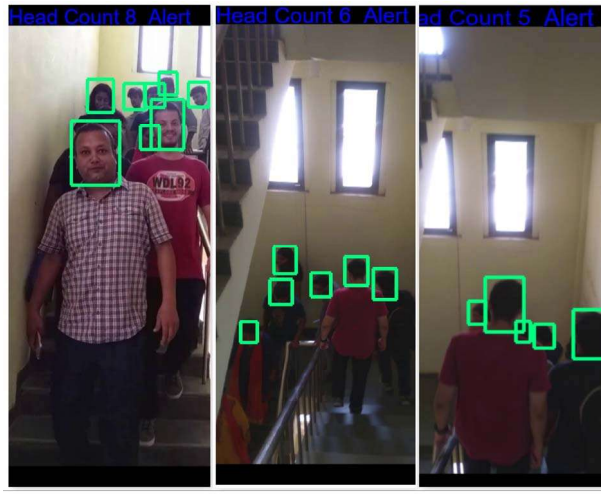

Fig. 9. Sample output

VI. Results And Discussion

The faster R-CNN model along-with google inception resnet v2 CNN model produced significant results by detecting the humans in the respective frames with an accuracy of 93.503% at the rate of 28 FPS. Figure 10 shows the detection accuracy and the output FPS comparison with the other state-of-the-art methods like R-FCN and SSD tested on the same proposed dataset and environment. The overall human detection accuracy and output FPS for Faster R-CNN is significantly better than other methods.

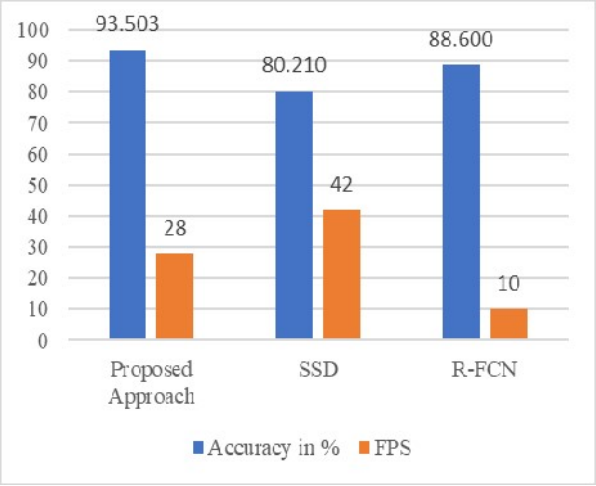

Fig. 10. Human detection accuracy and output FPS comparison

The performance of the proposed approach can be observed from the following confusion matrix as shown in figure 11. The overall accuracy, precision and recall for the CCEWS is 88.79 %, 88.52% and 89.46%.

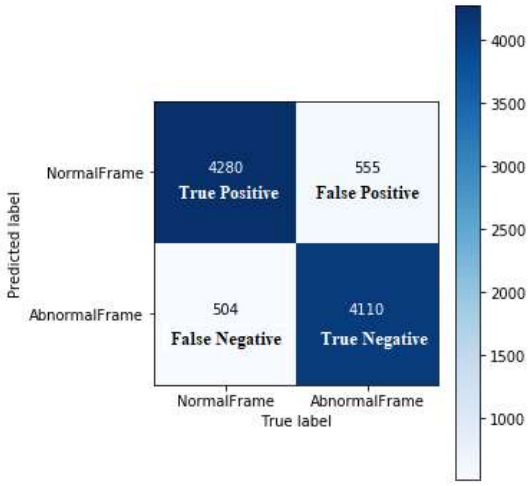

Fig. 11. Confusion matrix between true lable and predicted label for normal and abnormal frame

VII. Future Research Direction

Based on the accidents that can happen due to abnormal crowd behavior, there is a huge need for research in the area of crowd behavior management. The proposed approach is efficient enough to execute the contingency plans for the detected crowd congestion, but still, it can be improved further. The one challenging task which can be tackled is of occlusion; thereby this approach can be enhanced further to handle the cases where people are hidden behind some other person or object as per camera vision. The Frames per Second (FPS) of the output can be increased by following different and new architectures. The approach can be updated even to recognize people in the low-resolution frames in which the head is barely visible. Indeed, there are countless possibilities for improvements in the field of crowd behavior analysis. With the passage of time technology is evolving and new types of challenges are originating which are leading towards new corresponding research fields.

VIII. Conclusion

This paper proposes the novel CCEWS architecture to analyze the crowd behavior and generate the warning/alert signals to execute the contingency plans in order to control the congestion and prevent from mishap. The system follows three tasks of object detection, object tracking, and object motion direction. Each of this task is achieved by following the modified faster R-CNN architecture (because of the requirement of one class classification only first module of faster R-CNN, RPN is required), centroid-based algorithm and analysis of the output of these two tasks for abnormality detection concerning the crowd motion respectively. And through testing, it was found that alert signals were generated for every corresponding frame for which there was a higher chance of occurrence of mishap like stampede due to congestion in crowd motion.

Acknowledgment

We are very grateful for our institute “Indian Institute of Information Technology Allahabad”, India. We would like to thank our supervisors and guides who helped us in conducting this research. Our Institute provided us with all the necessary resources such as powerful computer systems

and allocated BDA labs our analysis. Authors are also indebted to CPS Programme, DST vide Reference No.244 for their financial support to carry out this research.

REFERENCES

[1] K. Rohit, K. Mistree and J. Lavji, "A review on abnormal crowd behavior detection," *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, Coimbatore, 2017, pp. 1-3.

[2] https://en.wikipedia.org/wiki/2017_Mumbai_stampede, as on 29th Oct., 2018.

[3] H. Jiang and E. Learned-Miller, "Face Detection with the Faster R-CNN," *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, Washington, DC, 2017, pp. 650-657.

[4] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017.

[5] C. Szegedy, S. Ioffe, and V. Vanhoucke. Inception-v4, inception-resnet and the impact of residual connections on learning. In ICLR Workshop, 2016.

[6] Huang, Jonathan, et al. "Speed/accuracy trade-offs for modern convolutional object detectors." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[7] Dai, Jifeng, et al. "R-fcn: Object detection via region-based fully convolutional networks." Advances in neural information processing systems. 2016.

[8] Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016.

[9] R. O. Mitchell, H. Rashid, F. Dawood, and A. AlKhalidi, "Hajj crowd management and navigation system: People tracking and location based services via integrated mobile and RFID systems," *2013 International Conference on Computer Applications Technology (ICCAT)*, Sousse, 2013, pp. 1-7.

[10] N. Farooqi, "Intelligent safety management system for crowds using sensors," *2017 12th International Conference for Internet Technology and Secured Transactions (ICITST)*, Cambridge, 2017, pp. 144-147.

[11] S. Vidyasagaran, S. R. Devi, A. Varma, A. Rajesh and H. Charan, "A low cost IoT based crowd management system for public transport," *2017 International Conference on Inventive Computing and Informatics (ICICI)*, Coimbatore, 2017, pp. 222-225.

[12] Mohandes M "Pilgrim tracking and identification using the mobile phone" *Consumer Electronics (ISCE)*, 2011 IEEE 15th International Symposium, 2011.

[13] Y. Fan and Q. Liang, "An Improved Method for Detection of the Pedestrian Flow Based on RFID," 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), Guangzhou, 2017, pp. 69-72.

[14] Shangnan Liu, Qiang Cheng, Zhenjiang Zhu, Hao Zhang "Analysis and Design of Public Places Crowd Stampede Early-Warning Simulating System." International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration , IEEE , 2016.

[15] S. Pathan, A. Al-Hamadi, and B. Michaelis, "Crowd behavior detection by statistical modeling of motion patterns," in Soft Computing and Pattern Recognition (SoCPaR), 2010 International Conference of, pp. 81–86, 2010.

[16] M. Adabi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015.