COMPUTATIONAL INTELLIGENCE AND MACHINE LEARNING

Computational Intelligence and Machine Learning
Vol-3 Issue-2, April 2022, PP.9-14
ISSN : 2582-7464 (Online)

A Research Perceptive on Deep Learning Framework for Pedestrian Detection in a Crowd

Shaamili.R [1], Ruhan Bevi. A [2]

[1] Research Scholar, Department of ECE, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India.
[2] Associate Professor, Department of ECE, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India.

*Corresponding Author: [1] sr6582@srmist.edu.in, [2] ruhanb@srmist.edu.in

Abstract **- In populated cities, we often find crowded events like political meetings, religious festivals, music concerts, and events in shopping malls, which have more safety issues. Smart surveillance systems are used in big cities to keep crowds safe and make crowd security less complicated and more accurate. However, the surveillance systems proposed for a crowd are monitored by human agents, which are inefficient, error-prone, and overwhelming. Even with deep learning-based feature engineering in crowds, many variants of crowd analysis still lack attention and are technically unaddressed. Considering this scenario, the smart system requires the most advanced techniques to monitor the security of the crowd. Crowd analysis is commonly divided into crowd statics and behavior analysis. This paper explores more about crowd behaviour analysis, pedestrian and group detection which describes the movements that are noticed in the crowd image. Subsequently, the issues of the current methodology of pedestrian detection, datasets, and evaluation criteria are analyzed.**

Keywords: Crowd Analysis, Pedestrian and group detection, deep learning, Crowd IoT analysis, Human Activity Recognition.

1. Introduction

Mass events occur with a major crowd of human societies, which are highly probable in sports events, public places, and political meetings. The is an increase in population for these events that attract an ever-rising number of people[1]. The high population in the cities leads to multiple crowd situations, which need more ever-rising. Cities are establishing intelligence systems based on video cameras which humans have monitored over the last decades, and smart cities use the technology to amend the wellness of urban people. Computer vision techniques for crowd analysis become more and more popular as a result. To better understand how people and crowds behave, several research has been conducted in crowded settings. The two primary research facets of crowd analysis are crowd information, which provides the gathering count, and community behavior analysis, which examines how people behave in crowds. Fig 1 shows the basic framework of crowd detection.

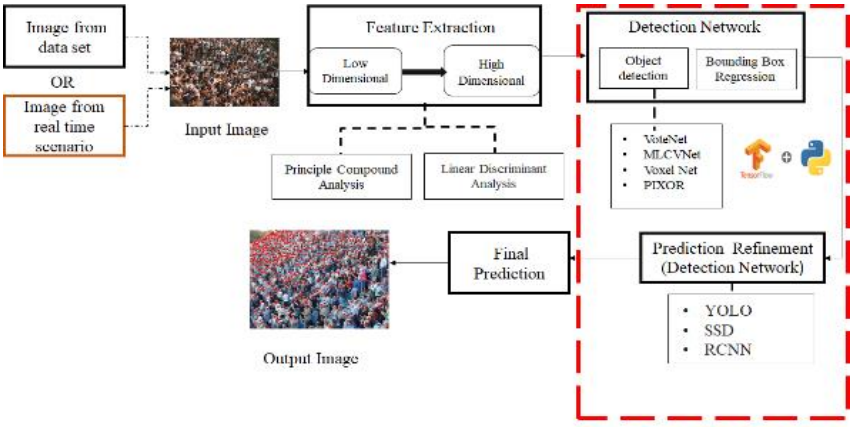

Fig1: Basic Crowd detection framework

Computational Intelligence and Machine Learning
Vol-3 Issue-2, April 2022, PP.9-14
ISSN : 2582-7464 (Online)

Real-time pedestrian identification is a significant field of study within the ideas of deep learning and computer vision, and it has become a significant issue in recent years. There are numerous potential applications for pedestrian detection, particularly in the field of surveillance. Machines incorporating vision-based intelligence systems can see objects according to computer vision. When deep learning algorithms are put under pressure, the deep learning models offer suitable improvements. Detecting, analyzing, and detecting public violations in streams of surveillance camera data are currently used to tackle deep learning approaches. The number of cameras in crowded places grows every year in terms of the aspect of security. The cameras are used to capture after the incident. The development and more systematic use of object detection, especially for pedestrian detection, helps as a pre-processing step for uninterrupted analysis of the video footage. Several algorithms have been developed, which focus on improvements in both accuracy and speed are analyzed and studied. The main properties of designing the detection algorithms are the evaluation speed and the accuracy. The detection accuracy is dependent on the convolution of the images.

Detectors experience several forms of occlusions, and real-time crowd scenes are frequently congested. It is still difficult to create a high-precision pedestrian and group detector that will work on systems with quick response times and reasonable computational resources. The system developed must react fast since individuals usually outpace the camera's field of vision. Therefore, an efficient deep framework with multi-modal image fusion is vital to prevent false alarms.

This study extends to the surveillance systems[2] offered for pedestrian detection in crowds because there is a possibility of a high-risk factor in accidents when pedestrians walk on footpaths and intersections in a crowded area. Therefore, the safety systems should focus on the accuracy and effectiveness of pedestrian tracking, localization, and potential obstacles. The rest of the survey extends with related work, methodology of crowd analysis, human behavior analysis, and pedestrian behavior concluded with the above-related datasets, annotators, and evaluation metrics.

2. Methods of Crowd Analysis

Over the last few decades, a great number of works on crowd analysis and pedestrian detection have been analyzed and discussed. This section follows the literature review for providing an overview of crowd analysis that include (i)Overall crowd analysis, (ii)Anomaly Detection, (iii)pedestrian and group detection, (iv) Human Action Recognition, (v) IoT-based crowd analysis

A. Crowd Analysis

A spatiotemporal feature [3] of crowd motion is used as a set of priors for the individual tracking in the video of densely populated environments. sequences from the film that show pedestrians moving at different speeds and directions both physically inside each frame and chronologically over the entire scenario Reviews of crowd analysis taxonomies are made in the following subsections.

B. Anomaly Detection in crowd

The existing crowd anomaly detection models have high complexity due to the incompetence of traditional deep learning methods to extract the time-related features and the absence of training of images. Yan Hu [4] created an enhanced spatial-temporal convolution to address this problem. This improved convolution uses an aggregation channel feature model to perform picture monitoring and selects aberrant behaviors in the image with low-level features. Sonkar [5] compared the ViBe and CNN algorithms to identify anomalous behaviors in the photos. Due to their computing flexibility, statistical approaches are extensively utilized in video frame computation designs, and real-time abnormal event detection typically makes use of fast algorithms with minimum computational costs. [6]. However, these assumptions cause some anomalous event detection to degrade, making it impossible for them to deliver real-time outputs. In addition, some strategies for reducing computing complexity result in a more minor video sequence, which can negatively affect predictive modeling and even the overall appearance of people in the crowd, leading to poor monitoring or higher recognition error [7,8].

C. Pedestrian and Group Detection

Recognizing pedestrians is a challenging problem in real-world situations that video surveillance software frequently observes. Detectors must deal with a range of occlusions in the usually crowded scenes. Deep Learning models frequently deliver fantastic results, whereas conventional techniques frequently fail to recognize pedestrians in challenging environments. The major goal of the research is to develop a robust, fast, and highly accurate pedestrian detector which is an efficient system to be improved with stringent processing power constraints.

Computational Intelligence and Machine Learning
Vol-3 Issue-2, April 2022, PP.9-14
ISSN : 2582-7464 (Online)

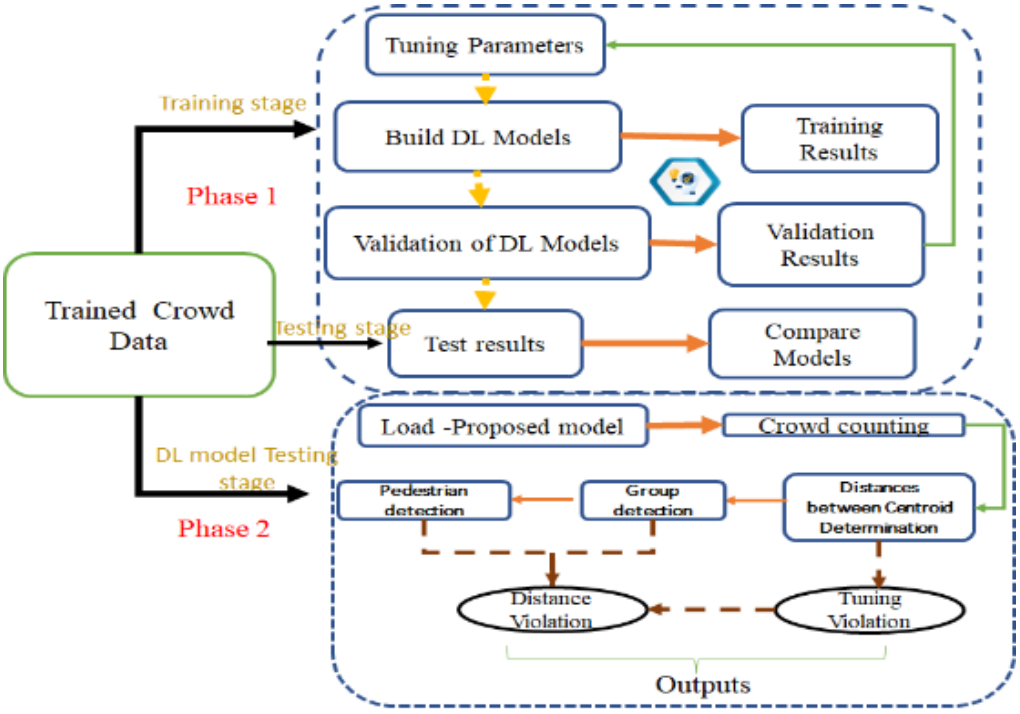

Fig 2: General Deep learning framework for pedestrian and group detection

Several studies employed conventional techniques to extract features and nourish those data into machine-learning algorithms. This is done to identify pedestrians and groups before the emergence of deep learning-based approaches for crowd analysis. The study in this field includes two different detection frameworks and uses hand-crafted approaches for identifying pedestrians and groups. Despite being less recent, it is nonetheless highly regarded. A framework with a two-stage detection framework is divided into different parts: on the image, a series of region suggestion boxes are given before object identification methods are used. One approach is RFCN [9], which provides information of location to the pooling layer, enhances location sensitivity, and enhances processing results for location-sensitive pedestrian identification problems. Adding FCN leads to more network parameters and feature sharing compared to Faster RCNN, as well as a smaller overall network size. As a result, the network's performance is improved while the amount of repetition is decreased. How quickly you're moving. When performing mask prediction tasks, the Mask RCNN follows the pooling layer with a convolutional layer. This structure can perform tasks like segmenting and recognizing pedestrians and separating them from the background. The output may also be used to identify bodily motions in people. The SAF RCNN [10] enhances general object detection, but general object detection improvement is constrained because object size changes are more frequent in pedestrian identification. Area proposal and classification are the two components of the two-stage pedestrian detection framework [11]. By offering innovative preselection box generation and feature extraction approaches, improving the prediction component, or both, researchers might boost the detection effect. The overall structure is highly complex than a single stage framework but it offers more accuracy.

Ge[12] demonstrated that cluster analysis can find small groups of people moving together and that video may be used to construct paths from automated pedestrian detection and tracking. As far as we know, the first study demonstrates that the outcomes of agglomerative clustering are statistically consistent with judgments of one-to-one human in a crowd. As a discipline like a computer vision develops, its influence on other fields is one way to assess the significance of research in that area. The results suggest that automated monitoring may quantitatively characterize actual crowds more quickly and accurately than human observation, offering up a new avenue for empirical research on social behavior.

D. Human Activity Recognition

Human activity recognition is extremely important in various sectors, including human-computer interface, robotics, everyday monitoring, ecological observation, and video surveillance systems. This activity recognition task may be effectively completed by using several datasets such as Sports-1M, UCF-101, and HMDB-51 and training them. Convolutional Neural Network [13] model implementation for image identification using OpenCV aids in the model's effective operation. The use of several datasets in the action recognition model has made it simple to classify activity according to whether it is normal or abnormal and suspicious. The server delivers an alert to the authorities on the occurrence

Computational Intelligence and Machine Learning
Vol-3 Issue-2, April 2022, PP.9-14
ISSN : 2582-7464 (Online)

of abnormal behavior occurring in real-time by the specified nature. Many hazardous actions can be prevented with their negative effects reduced due to the adoption of this concept.

Fig 3: Human Activity Recognition Framework

E. IoT-based crowd analysis

The use of overhead view video sequences to conduct people detection and counting is shown using an IoT-based crowd monitoring system [14]. The SSD-Mobilenetv2 detection model is examined for detecting purposes. Because people's appearances from the above perspective differ significantly in visibility, shapes, sizes, body articulations, and postures, the detection model was trained using the standard frontal view data set. Eventually, during the phase of transfer learning, the additional training is done using an overhead data set. The recently learned features are combination of existing trained model in which two virtual lines are utilized to count the number of people after detection. The experimental findings describe the efficiency and determined the learning-based crowd surveillance system.

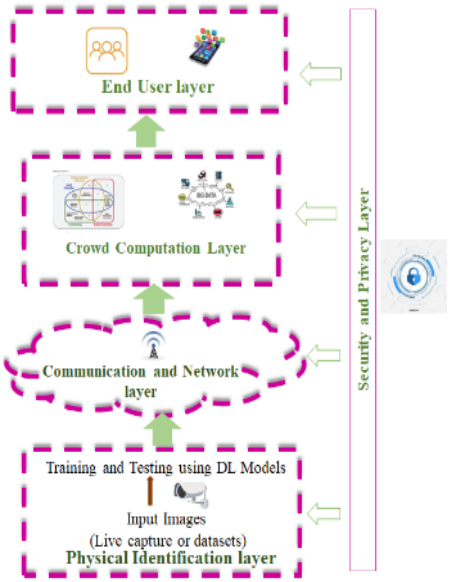

Fig 4 IoT Architecture for Crowd Analysis

3. Data Sources

Crowd Human dataset: CrowdHuman[15] is a dataset that may be used to compare detectors in crowd settings. The CrowdHuman dataset is extensive, well-annotated, and diverse. For training, validation, and testing, CrowdHuman has 15000, 4370, and 5000 photos, respectively. The dataset contains 470K human examples from the train and validation subsets, with 23 people per image and various types of occlusions.

KITTI dataset: Kitti[16] is a dataset created with help of autonomous driving platform.The tasks, such as stereo, optical flow, visual odometry, are listed in the comprehensive benchmark. In addition, to the object detection dataset the monocular pictures and bounding boxes, are included. This dataset tagged 7481 training photos with 3D bounding boxes

Computational Intelligence and Machine Learning
Vol-3 Issue-2, April 2022, PP.9-14
ISSN : 2582-7464 (Online)

UCF crowd datasets[17]: The goal of the Static Floor Field is to capture the scene's beautiful and consistent features. Favorite places, such as prominent paths typically chosen by the audience as it flows around the scene, and preferred departure spots are among these qualities. For human action detection UCF101,UCF50 and UCF11 are used .

Pedestrian and Group Detection datasets

COCO datasets: The Microsoft Common Objects in Context dataset [18] is large-scale object detection, segmentation, key-point detection, and captioning dataset with 328K images. Bounding boxes and per-instance segmentation masks with 80 item types are included in the dataset as annotations for object detection. More than 200,000 photos and 250,000 human instances have been classified with key points. Per-pixel segmentation masks containing 91 categories, such as grass, wall, and sky, are used in the stuff-based image segmentation.

Behave Dataset[19]: The dataset consists of eight individuals interacting with twenty items in five different naturalistic environments. With four Kinect RGB-D cameras, a total of 321 video sequences were captured. Human and object masks, as well as segmented point clouds, are included in each frame. Each picture is coupled with 3D SMPL and object mesh registration in-camera coordinates. For each sequence, the camera takes a different stance. Reconstructions of the 20 objects using textured scanning.

4. Challenges and Future Direction

The multiscale challenges and occlusion issues listed above are the primary problems impacting pedestrian detection. The multiscale issue among them calls for it to be possible to measure the size of pedestrians correctly; simultaneous detection places more demands on the system. on the network for feature extraction. The occlusion issue demands specific pedestrian detection parts and suggests additional restrictions and improvements in the recognition algorithm. These concerns directly enhance pedestrian effects and sophisticated scene detection, a crucial strategy to enhance pedestrian detectors. The hardware requirements are frequently considerable, even though the large distribution network has significantly improved.

5. Conclusion

Smart city development is mainly concerned with intelligent surveillance systems. The deployment of these devices necessitates the creation of a framework that can accurately scan video surveillance settings. In addition, crowd analysis-related approaches are in high demand since video surveillance is frequently used in public places. This overview is captured in a review paper by considering both parent fields and recent trends within the area. In this review study on crowd analysis, we discussed current developments in the industry. We looked at earlier reviews of crowd analysis throughout this work. We observed recent works just on branches and many sub-branches of crowd analysis, pedestrian and group detection, and human activity recognition

REFERENCES

[1] Bendali-Braham, Mounir, Jonathan Weber, Germain Forestier, Lhassane Idoumghar, and Pierre-Alain Muller. "Recent trends in crowd analysis: A review." Machine Learning with Applications 4 (2021): 100023

[2] Ang, Kenneth Li Minn, Jasmine Kah Phooi Seng, and Ericmoore Ngharamike. "Towards crowdsourcing internet of things (crowd-IoT): Architectures, security, and applications." Future Internet 14, no. 2 (2022): 49.

[3] Kratz, Louis, and Ko Nishino. "Tracking pedestrians using local Spatio-temporal motion patterns in extremely crowded scenes." IEEE transactions on pattern analysis and machine intelligence 34, no. 5 (2011): 987-1002.

[4] Hu, Yan. "Design and implementation of abnormal behavior detection based on deep intelligent analysis algorithms in massive video surveillance." Journal of Grid Computing 18, no. 2 (2020): 227-237.

[5] Sonkar, Riddhi, Sadhana Rathod, Renuka Jadhav, and Deepali Patil. "CROWD ABNORMAL BEHAVIOUR DETECTION USING DEEP LEARNING." In ITM Web of Conferences, vol. 32, p. 03040. EDP Sciences, 2020.

[6] Aldissi B, Ammar H (2020) Real-time frequency- based detection of a panic behavior in human crowds. Multimed Tools Appl 79(33):24851–24871

[7] Kh R, Ghezelbash MR, Haddadnia J, Delbari A (2012) An intelligent surveillance system for falling elderly detection based on video sequences. 19th Iranian Conference of Biomedical Engineering (ICBME), Tehran, Iran Dec (pp. 20-21)

[8] Qasim, T. and Bhatti, N., 2019. A low dimensional descriptor for detection of anomalies in crowd videos. Mathematics and Computers in Simulation, 166, pp.245-252

[9] Dai, J., Li, Y., He, K. and Sun, J., 2016. R-fcn: Object detection via region-based fully convolutional networks. Advances in neural information processing systems, 29.

[10] Li, Jianan, Xiaodan Liang, ShengMei Shen, Tingfa Xu, Jiashi Feng, and Shuicheng Yan. "Scale-aware fast R-CNN for pedestrian detection." IEEE transactions on Multimedia 20, no. 4 (2017): 985-996.

[11] Tian, Di, Yi Han, Biyao Wang, Tian Guan, and Wei Wei. "A review of intelligent driving pedestrian detection based on deep learning." Computational intelligence and neuroscience 2021 (2021).

[12] Ge, Weina, Robert T. Collins, and R. Barry Ruback. "Vision-based analysis of small groups in pedestrian crowds." IEEE transactions on pattern analysis and machine intelligence 34, no. 5 (2012): 1003-1016.

[13] Patil, Swati, Harshal Lonkar, Ashutosh Supekar, Pranav Khandebharad, and Mayur Limbore. "Human Activity Recognition using Machine Learning."

[14] Ahmed, Imran, Misbah Ahmad, Awais Ahmad, and Gwanggil Jeon. "IoT-based crowd monitoring system: Using SSD with transfer learning." Computers & Electrical Engineering 93 (2021): 107226.

[15] [Online] Available: https://www.crowdhuman.org

[16] [Online]Available: https://www.cvlibs.net/datasets/kitti/

[17] [Online]Available: https://www.crcv.ucf.edu/data/tracking.php

[18] [Online] Available: https://cocodataset.org/#download

[19] [Online] Available: https://virtualhumans.mpiinf.mpg.de/behave/license.html