National Conference on Advance in Science, Technology, Engineering and Management

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Real time crowd counting using deep learning

Sougandhika Narayan [1*],Anupa M.B [2*],Adithya S.R [3*],Chandrashekhar [4*],G Jaswanth [5*]
[1,2,3,4,5](Computer Science ,K.S School of engineering& management, Email: Sougandhika@kssem.edu.in)

****

Abstract:

Crowd counting has become indispensable in the era of smart cities and advanced surveillance for numerous applications, including public safety. This work provides a novel deep learning-based approach for real-time crowd counting, specifically using the Mobile Net SSD model. Our method uses deep learning skills to estimate crowd intake and outflow accurately in real time. Our approach is based on the Mobile Net SSD model, a powerful and lightweight convolutional neural network that performs identification tasks.

Keywords — **Crowd count, Mobile Net SSD, Public Safety**..

****

I. INTRODUCTION

In the contemporary world, the advent of smart cities and advanced surveillance systems has necessitated the development of efficient and accurate crowd counting mechanisms. These mechanisms are pivotal for a myriad of applications, including but not limited to, public safety, traffic management, and urban planning. The traditional methods of crowd counting, which often rely on manual counting or simple image processing techniques, are not only time-consuming and labor -intensive but also prone to errors, especially in scenarios with high crowd density. Moreover, these methods fail to provide real-time data, a feature that is increasingly becoming indispensable in today's fast paced world.In order to overcome all these obstacles, this project presents a revolutionary deep learning-based real-time crowd counting method that makes use of the Mobile Net SSD model.

By tracking the movement of detected individuals over time, our system can accurately estimate the inflow and outflow of crowds, providing valuable data for crowd dynamics analysis.

This real-time crowd counting system holds significant potential for deployment in smart city applications, contributing to safer and more efficient public spaces.

In order to overcome these obstacles, this project presents a revolutionary deep learning-based real-time crowd counting method that makes use of the Mobile Net SSD model. Crowd counting is a perfect application for deep learning, a branch of machine learning that has demonstrated impressive performance in pattern recognition tasks.

II RELATED WORK

People tracking and counting is a critical duty in many industries, including crowd control, retail, transit, and security. Numerous strategies, from more sophisticated ones based on deep learning to more conventional ones like video analysis and sensor networks, have been presented over the years for people counting and tracking. Among the oldest and most used methods for counting and monitoring individuals is video analysis. Video analysis techniques involve processing video data to detect and track people in real-time.

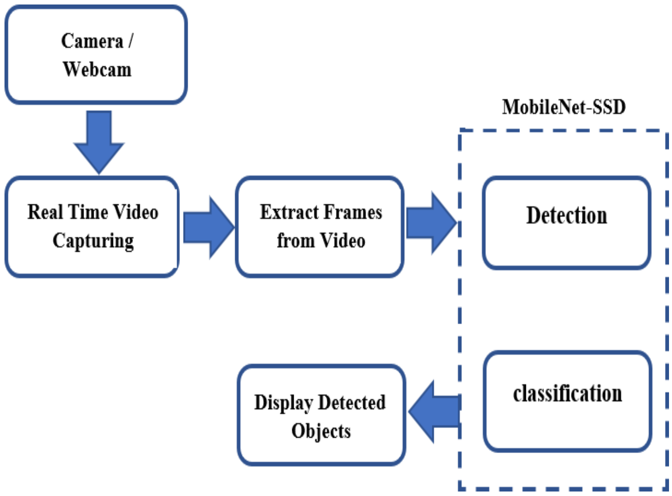
People can be detected and tracked using pixel-based approaches like contour tracking and blob analysis, which examine the movement of pixels over time. Although these methods can work well in straightforward situations, they are frequently hindered by problems like occlusions, shifting illumination, and clutter in the surrounding area. Many deep learning-based methods for tracking and counting people have been proposed by researchers in an effort to get around the drawbacks of conventional video analysis techniques. Neural networks are used in deep learning techniques to automatically extract features from massive volumes of data. These methods have demonstrated encouraging outcomes in a range of computer vision tasks, such as tracking, object identification, and recognition. Long Short-Term Memory (LSTM) networks, Recurrent Neural Networks (RNNs), and Convolutional Neural Networks (CNNs) are a few of the most widely utilized deep learning approaches for people tracking and counting. Convolutional Neural Nets (CNNs) are among the most widely used deep learning methods for object detection and recognition. CNNs have been effectively used for a number of computer vision tasks, such as tracking and object detection.

III. PROPOSED SYSTEMS

Using the Mobile Net-SSD architecture, the Proposed System will quickly and effectively detect objects in real time. Using OpenCV 3.4,

we will write a Python script that uses a deep neural network and to recognize objects. The way the system operates is as follows: Real-time video from a camera or webcam will be used for input. The simplified Mobile Net Architecture utilizes depth-wise separable to create lightweight deep neural networks, use convolutions. The input video was split up into frames and sent to the layers of the mobile net.

The difference between the amount of pixel intensity under the bright region and the amount of pixel intensity under the dark area determines the value of each feature. These components are computed using all of the image's available sizes and areas. There may be few pertinent aspects in an image that can be utilized to identify the object and many irrelevant ones. The Mobile Net layers' task is to convert the input image's pixels into highlights that provide information about the image's contents. The Mobile Net-SSD model is then used to ascertain the bounding boxes and associated class (label) of objects.



A DESCRIPTION OF MOBILE NET SSD

The item detection version of Mobile Net SSD computes the item elegance and output bounding field from the input image. Using Mobile Ne t as a spine, this Single Shot Detector (S SD) item detection version may produce fast item detection that is optimized for mobile devices. Figure 1: Mobile Net SSD Layered Architecture depicts the architecture. This model offers quick and precise object detection and is tailored for mobile devices. Mobile Net and Single Shot Detector (SSD) are the two primary parts of the Mobile Net SSD architecture.

A lightweight deep neural network architecture called Mobile Net was created for mobile devices with constrained computing power. Computational assets. The foundation of it consists on depth-wise separable convolutions, which split a standard convolution into two parts: a depth-wise convolution and a po int-wise function. Whereas the point-wise convolution applies a 1x1 filter to each output channel of the depth-wise convolution, the depth-wise convolution applies a single filter to each input channel. By using this method, the number of parameters in the network is drastically decreased, leading to a more compact and effective model. Within the Mobile Net SSD design, Mobile Net serves as a feature extractor.

Several convolutional layers process the input image, applying depth-wise separable convolutions at each layer. A feature map, containing high-level features helpful for object detection, is the result of the final convolutional layer. The architecture's SSD component uses the feature map generated by Mobile Net to perform object detection. With the SSD technique, the item class and location are predicted in a single forward pass of the network, making it a single-shot detector.

The SSD method involves creating a grid of cells out of the input image, and each cell's job is to forecast the presence of objects within it. For each grid cell, the SSD predicts a set of bounding boxes, each with a different aspect ratio and size. These bounding boxes are defined as offsets from the coordinates of the grid cell.

The SSD also predicts the confidence score for each bounding box, which represents the likelihood that the object is present in that box. The overlap between the ground truth bounding box and the predicted bounding box is used to determine the confidence score. Additionally, the SSD predicts the class of each bounding box's item.

Based on a set of learnt features unique to each object class, the class prediction is made. These features are calculated using a collection of convolutional layers unique to each class from the feature map produced by Mobile Net.

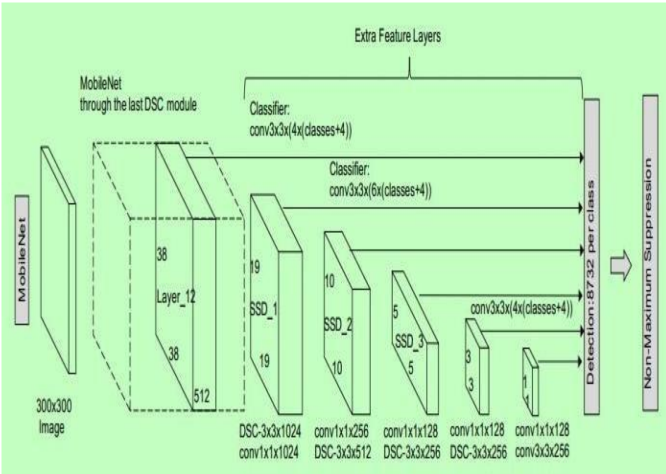


Fig.1 Mobile Net SSD layered Architecture

B. DESCRIPTION OF CENTROID TRACKER

In order to calculate the Euclidean distance between the centroids of the most recent detections and the earlier ones, the Centroid Tracker method keeps track of the most recent centroids for each tracked object. A straightforward object tracking algorithm called Centroid Tracker tracks an object's motion by using its centroid. For monitoring objects in real-time video feeds, it is especially helpful. The approach uses a distance-based matching algorithm to associate centroids in successive frames of the video stream.

The context of object detection, the Centroid Tracker is typically used to track the motion of objects that is detected by a detection algorithm such as Mobile Net SSD. Once objects are detected in the video frame, the Centroid Tracker uses the coordinates of the bounding boxes around the objects to calculate their centroids. It then associates these centroids with objects detected in subsequent frames, based on the distance from the centroids in the previous frame. The advantages of the Centroid Tracker algorithm is computationally efficient and can be implemented in real-time. It is also able to handle cases where objects move in and out of the frame, and where objects are partially occluded by other objects in the frame.

IV. DESIGN ARCHITECTURE AND CONSIDERATIONS

Based on standard design assessment we have considered few designs requirement of the design, they are

a. DESIGN CONSIDERATIONS:

1. Operating system: Windows /Linux
2. Programming Language: Python 3.0
3. Packages:Numpy,argparse,imutils,dlib,open cv-python,scipy, cmake, schedule
4. Model : Mobile Net SSD
5. Code implementation: Jupyter Notebook

b. DESIGN SCOPE

This project focuses on developing a real-time crowd counting system using the Mobile Net SSD deep learning model. We'll collect and process video data to train our model to count people in various crowd scenarios. The system will analyze video frames in real-time, track individual movements, and estimate crowd inflow and outflow. We aim to integrate all this system with the existing surveillance infrastructure. The performance of the system will be evaluated using appropriate metrics. We'll also ensure that the system respects privacy norms and is scalable and robust.

c. DESIGN HIGHLIGHTS

The design feature includes the solution highlights like:

1. Efficient Model
2. Real-Time Analysis
3. Robust and Scalable
4. Integration
5. Performance Evaluation

V. IMPLEMENTATION

1. Data Collection: Collect video data representing various crowd scenarios. This could be from public datasets or your own data collection efforts.
2. Data Preprocessing: Preprocess the data to make it suitable for training the model. This could involve resizing the frames, normalizing the pixel values, and augmenting the data to increase its diversity.
3. Model Training: Train the Mobile Net SSD model on the pre processed data. You would need to define a suitable loss function and optimization algorithm for this task. The model should be trained to detect and count individuals in the video frames.
4. Tracking Algorithm: Develop an algorithm to tracking the movement of detected individuals over time. This could be based on simple motion tracking techniques or more advanced methods like Kalman filtering or optical flow.

5. Inflow and Outflow Estimation: Use the tracking data to estimate the inflow and outflow of the crowd. This could involve counting the number of individuals moving in and out of a defined region in the video frame.

6. System Integration: Integrate the crowd counting system with the existing surveillance infrastructure. This could involve developing APIs or other interfaces to communicate with the surveillance system.

7. Performance Evaluation: Evaluating the performance of the system using appropriate metrics. This could involve testing the system on a separate test dataset and calculating metrics like the Mean Absolute Error (MAE) and Mean Squared Error (MSE).

8. Privacy Measures: Implement measures to ensure the privacy of individuals in the video frames. This could also involve techniques like anonymization or differential privacy.

V. RESULTS AND DISCUSSIONS

The outcomes of the Centroid-Tracking and Mobile Net-SSD algorithms are covered in this section. A video serves as this system's primary input. The goal of this research project was to create a people counter and location tracker that would allow for real-time crowd monitoring. People counting and tracking are essential components in security and traffic monitoring applications, where the system was built for use in video surveillance environments. The system is a popular choice for real-time security applications because of its capacity to track objects via background subtraction and detect them in real-time video. The region that can be monitored depends on the camera's range, and businesses like diamond stores, banks, and the military can all profit substantially from this technology.

The process of detecting people in crowded scenarios involves feeding input images as shown in Fig 2 to our model. However, the input image needs to be preprocessed to meet the requirements of our model. The frame is resized to a maximum width of 500 pixels and changed from BGR to RGB to accomplish this. Our centroid-based object tracker then receives these bounding box coordinates as input and uses them to determine the centroids of the objects it has observed. We can precisely count the number of people in a crowded situation by following these centroids.

To precisely identify and count people in congested situations, this procedure entails image preprocessing, object tracking using centroid-based techniques, and object detection using a deep neural network.

Fig.2 people moving in open space

The dnn module is then used to convert the preprocessed image to a blob. This blob is then passed to our model to generate predictions, which is a list containing seven floating values. At the first index of this list, we have the class ID of the detected object, and at the second index, we have the confidence or probability of the detection.

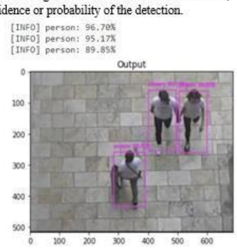


Fig.3 Detection of people in open space along with bounding box coordinates

The Hungarian algorithm is used by the program to match the current collection of detected centroids with a set of previously detected item centroids.

VI. CONCLUSIONS

Systems for counting and monitoring employees can offer insightful information about how a company is run, facilitating the creation and application of performance-enhancing plans. As people enter and leave a physical location, these systems can count them, keep track of them, and provide unique reports for each. Business executives who want to comprehend visitor behavior, such as when they enter, how they navigate the area, and when they leave, will find this data to be extremely helpful. Furthermore, individuals counting Technology has the ability to display visitor behavior, such as where they spend their time and which products are most popular. In the retail industry, people counting systems are able to pinpoint the exact number of visits and track the percentage of those visitors who make purchases. Understanding the effectiveness of marketing strategies in drawing in visitors and turning them into customers requires knowledge of this information. People counting equipment at shopping malls can assist executives in understanding the most desirable regions of the mall and how traffic patterns vary over time.

People counting systems can also be helpful to management of exhibitions and events, enabling them to better allocate staff during the event's peak hours. Furthermore, people counting systems can offer event totals, which are crucial for assessing the event's success and organizing similar events in the future. Systems for tracking and counting people need to have the ability to identify and recognize objects. These technologies can be used to identify any kind of object in a physical space and are crucial for real-time applications. People counting and tracking systems can reliably follow the motions of individuals in a video and count the number of people entering and leaving in real time by integrating object detection and recognition with centroid-based object tracking.

REFERENCES

[1] Harshal Honmote, Pranav Katta ,Shreyas Gadekar ,Madhavi Kulkarni “Really Time Object Detection and Recognition using Mobile Net-SSD with OpenCV”International Journal of Engineering Research & Technology (IJERT) ,2022.

[2] Ujwala Bhangale, Suchitra Patil, Vaibhav Vishwanath, Parth Thakker, Amey Bansode ,Devesh Navandhar,“Near Real-Time Crowd Counting using Deep learning Approach”,2021,In Science direct Third International Conference on Computing and theNetworkCommunications(CoCoNet’19),pages770-779

[3]Xiaowen Shi , Xin Li, Caili Wu, Shuchen Kong , Jing Yang , Liang He, “A real time deep network on crowd counting”,2021 IEEE International conference on Acoustics, speech and signal Processing.

[4] Ilyas, Naveed and Shahzad, Ahsan and Kim, Kiseon,2020”Convolutional-neural network-based image crowd counting: review, categorization, analysis, and performance evaluation,” *Sensors.*

[5] D. Zhang, X. Du, and L. Zhang, “A Review on people counting and crowd density estimation”.

