Real-Time Human Detection and Counting System Using Deep Learning

N. Siri Balakrishna Reddy
Department Of Artificial Intelligence And Machine Learning, Vardhaman College Of Engineering.
Shamshabad, Telangana, IndiaEmail: siribalakrishnareddy21aiml@student@vardhaman.org

Jayanth Rohan
Department Of Artificial Intelligence And Machine Learning, Vardhaman College Of Engineering.
Shamshabad, Telangana, India
jayanthrohanganna21aiml@student.vardhaman.org

H. Nimmagada Balaji
Department Of Artificial Intelligence And Machine Learning, Vardhaman College Of Engineering.
Shamshabad, Telangana, India
nimmagaddabalaji21aiml@student.vardhaman.org

G. Maktedar Pratham Singh
Department Of Artificial Intelligence And Machine Learning, Vardhaman College Of Engineering.
Shamshabad, Telangana, India
maktedarprathamsingh21aiml@student@vardhaman.org

Mr. Gagandeep Arora
Department Of Artificial Intelligence And Machine Learning, Vardhaman College Of Engineering.
Shamshabad, Telangana, India
gagandeeparora@vardhaman.org

Mr. Anand Kumar Biyani
Department of Artificial Intelligence And Machine Learning, Vardhaman College of Engineering
Shamshabad, India
Anandkumarbiyani@vardhaman.org

***Abstract*—Targeting the current Covid 19 pandemic situation, this paper identifies the need of crowd management. Thus, it proposes an effective and efficient real-time human detection and counting solution specifically for shopping malls by producing a system with graphical user interface and management functionalities. Besides, it comprehensively reviews and compares the existing techniques and similar systems to select the ideal solution for this scenario. Specifically, advanced deep learning computer vision techniques are decided by using YOLOv3 for detecting and classifying the human objects with DeepSORT tracking algorithm to track each detected human object and perform counting using intrusion line judgment. Additionally, it converts the pretrained YOLOv3 into TensorFlow format for better and faster real-time computation using graphical processing unit instead of using central processing unit as the traditional target machine. The experimental results have proven this implementation combination to be 91.07% accurate and real-time capable with testing videos from the internet to simulate the shopping mall entrance scenario.**

Keywords—detection, tracking, counting, deep learning, computer vision, Covid 19

Introduction

Concerts, political speeches, rallies, marathons, and stadiums are all examples of circumstances where crowds occur. Crowd counting, also known as density estimate, assists in crowd management for safety and surveillance purposes, such as law enforcement officer deployment and the detection of unusual behavior. It may also be used to determine the number of commuters, which is crucial for the construction of public transportation infrastructure. It may also be used to assess the political significance of demonstrations or protests, as different estimations are frequently given for the same event. And, because counting through turnstiles or by humans is not always practicable or practical, To estimate the number of individuals in dense crowds, use computer vision-based techniques.

The booming event industry has faced its biggest disruption due to the challenges of Covid 19, causing many events to be cancelled, postponed, relocated, and transformed into virtual events (Congrex, 2020). However, the demand of crowd management does not reduce as it is still significant to prevent virus spreads by controlling the crowd density in a specific environment. In this case, object detection is the ideal solution for crowd management in public areas like malls, shops, restaurants, parks, subway station, and more. Unlike the complicated and traditional manual crowd management which requires planning, risk assessment, and communication, object detection is far more efficient and effective without involving any human or manual operations.

I. Related Work

In a congested atmosphere, counting individuals has become a need for their safety. This count is essential in retail since it helps you to see when the business is busiest and plan accordingly. When a firm or government wants to know how crowded an area is at any given time, crowd counts is essential for infrastructure planning. It's also used to avoid congestion and create an escape route. With malls and subways 4 becoming increasingly popular, it's vital to estimate how many people will visit. There are a number of automatic detection techniques available. Although methods have been found to address this problem, achieving sufficient accuracy and counting efficiently remains a difficulty.

Concerts, political speeches, rallies, marathons, and stadiums are all examples of circumstances where crowds occur. Crowd counting, also known as density estimate, assists in crowd management for safety and surveillance purposes, such as law enforcement officer deployment and the detection of unusual behavior. It may also be used to determine the number of commuters, which is crucial for the construction of public transportation infrastructure. It may also be used to assess the political significance of demonstrations or protests, as different estimations are frequently given for the same event. And, because counting through turnstiles or by humans is not always practicable or practical, To estimate the number of individuals in dense crowds, use computer vision-based techniques.

The study's main goal is to use head detection to count the number of persons present. It aids in the correct evaluation of crowd count statistics . The system is linked to a simple graphical interface that allows users to complete jobs, keep track of persons by taking action against overcrowding, and aid with evacuation routes.

Research Methodology

In recent years, the need for efficient and accurate human detection and counting systems has surged, driven by applications in security, crowd management, retail analytics, and more. Leveraging the advancements in deep learning, particularly in computer vision, presents a promising approach to address these needs. This essay outlines the research methodology for developing a real-time human detection and counting system using deep learning, focusing on the key steps from problem definition to deployment.s.

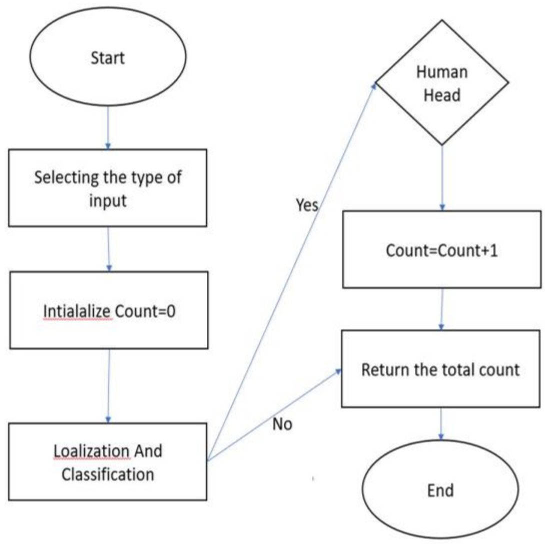
The primary problem addressed in this research is the real-time detection and counting of humans in various environments. The objective is to build a system that is not only accurate in detecting and counting individuals but also fast enough to operate in real-time scenarios. This involves overcoming challenges such as varying lighting conditions, occlusions, and different crowd densities.

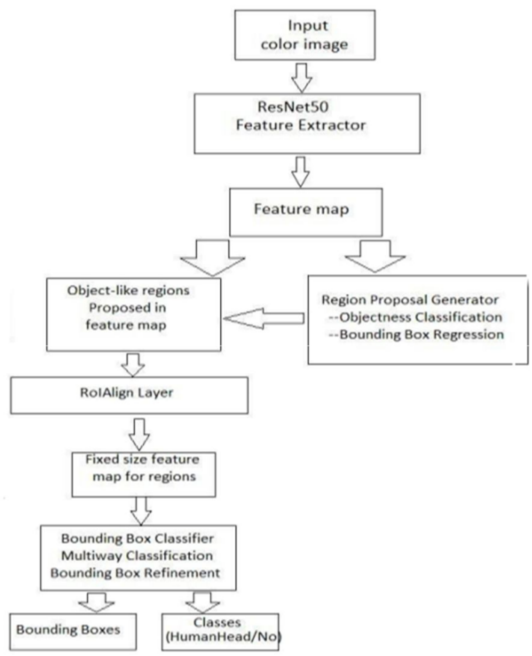
Literature Review

A comprehensive literature review is essential to understand existing methods and identify areas for improvement. Current approaches often rely on traditional image processing techniques or machine learning algorithms that lack robustness and scalability. Recent advancements in deep learning, particularly the development of Convolutional Neural Networks (CNNs), have significantly improved object detection capabilities. Models like YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector), and Faster R-CNN have demonstrated high accuracy and efficiency in various detection tasks. By reviewing these methods, we can identify their strengths and limitations, guiding the development of a more robust system.

The foundation of any deep learning model is a well-curated dataset. For this project, datasets containing annotated images or videos with human figures are required. Publicly available datasets like COCO (Common Objects in Context) and PASCAL VOC are excellent starting points. Additionally, creating a custom dataset tailored to specific application needs can further enhance model performance. Data augmentation techniques, such as rotating, scaling, and flipping images, are applied to increase dataset diversity and improve model generalization.

Training the selected model involves several steps. First, setting up the training environment with the necessary hardware (such as GPUs) and software libraries (TensorFlow, PyTorch) is essential. Next, the model is trained using the annotated dataset, where it learns to identify and count humans in various scenarios. Key hyperparameters, including learning rate and batch size, are adjusted to optimize the training process.The loss function, typically a combination of classification loss (cross-entropy) and localization loss (mean squared error), guides





2.1. Other importance of human detection and counting

Human detection and counting with computer vision not only can solve crowd management problems but also help to study people behavior in a specific environment or scenario. In Velastin et al. (2020), the same approach is used to perform passenger traffic management for reducing delays in public transport systems (Mokayed et al., 2022). Especially in railway system where traveling at optimized and highest speed at their own track, passenger transit time is the only critical factor affecting the system's effectiveness and efficiency. For example, detecting and counting how many passengers can enter or exit the train in each transit. The challenge is related not only to human detection but also to other object detection like plate numbers, vehicles, and other objects that add value to the field of the intelligent transportation system (Velastin et al., 2020). Thus, an intelligent model is developed to determine the best dwell time, suitable physical conditions such as door width, gap between the train and platform, allocation of train interior, and train and platform floor material while ensuring passengers can be boarded and exited easily, quickly, and safely.

2.2. Challenges and difficulties in human detection and counting

In a comparative research paper (Raghavachari et al., 2015), the main difficulties in vision-based human object detection and counting applications are its accuracy in different scenarios and conditions. Different camera orientations will obtain distinctive angles of the human objects which require a different type of implementation to perform the detection and counting. Moreover, the accuracy of detection commonly changes with different people density as crowd size can be increased or decreased throughout the day. Occlusion happens when human objects are too concentrated with overlaying, causing difficulty to distinctly detect each object separately. In addition, lighting or weather conditions might also decrease the quality of images sent for detection such as blurred caused by the camera glare. Therefore, different types of algorithms or techniques should overcome these four common issues to be compatible and suitable for various experimental environments. Another concern in real-time human detection and counting is the high computational efficiency to execute multiple image frames at one time.

2.3. Computer vision-based approaches

Jalled and Voronkov (2016) perform human body and vehicle detections using image processing to perform background subtraction and obtain the foreground mask using corresponding shape features to detect human and vehicle objects. The initial result shows many false detections for human object due tocommonly shared shapes with other objects. Therefore, the face detected from human object is sent to the Haar classifier to further determine the human object using a series of sequential classifiers where the object must pass all the classifiers to obtain a true result. Based on the conducted experiment, Haar classifier only works for face detection while the performance drops on vehicle detection due to its characteristics. The tracking on face or human object is performed using simple constant facial features like color which does not change when the object rotates or moves. However, these limited tracking features can suffer through light and camera issues.

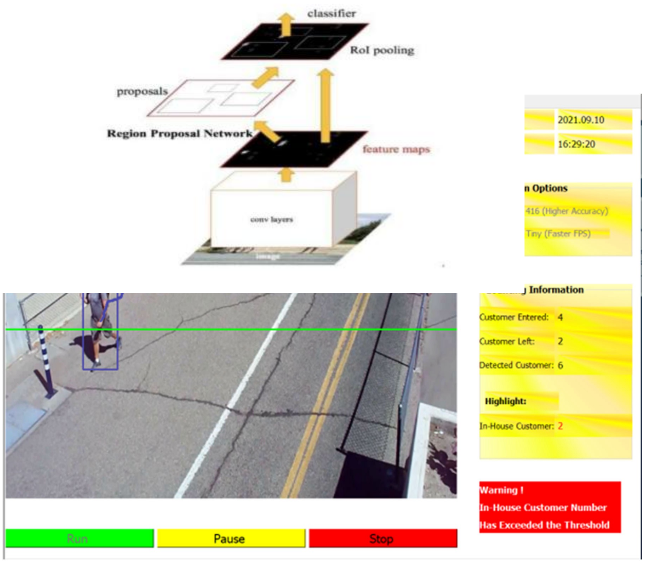
Al-Zaydi et al. (2018) proposed another image processing-based technique using regression model, low-level feature representation, and perspective normalization with frame-to-frame analysis. First, the Gaussian mixture model algorithm is used to remove unwanted noises and background. Then, the Gaussian process regression (GPR) model is trained with identified features and crowd size like foreground segment, texture, edge, and key point features which could be extracted using image processing through translation and rotation. Besides, prospective normalization helps to resolve the issue of people size changes from different camera angles by allocating different weights to the pixels at different locations. With these human-based features extracted and a trained GPR model, human counts in a particular frame are estimated with little processing. The result errors were measured using mean absolute error and mean squared error matrixes and resulted in an acceptable margin with slight difference between the estimated count and true count.

Alternatively, Al-Zaydi et al. (2016) have proposed another image processing method using multiplexer cascade model and multiple independent people detectors to enhance true positive detections. Different detectors are equipped with independent feature extractions, deep learning method, and human models. Furthermore, the confidence level of each detector is used to classify the detected window into four levels where only the highest level will be considered for counting. Different from normal cascade Haar classifier, the multiplexer cascade solution is more advanced and fuses the different confidence levels of each detector with three defined quality thresholds. When experimenting the two different detectors separately, Haar-like detector resulted in higher miss rate because of low confidence level while full-body detector resulted in much lesser miss rate. Ultimately, the paper has proven that applying deep learning and combining both detector obtains the best accuracy, highest true positive detections and lowest missing rate compared to using single people detector. However, this approach requires extensive cascading processes and might lack real-time processing capability although a pipeline is used.

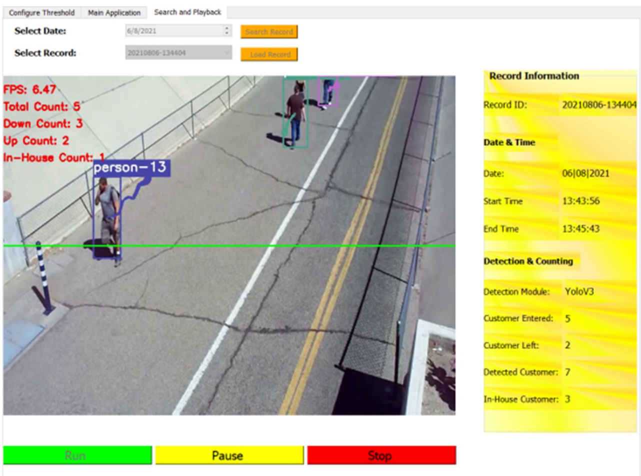
2.4. Similar system comparisons

This section performs comprehensive analysis and comparison between commercial systems that are already deployed in the market which are offering similar features as the proposed system in this paper. The four popular and reliable similar systems or solutions selected from different Information Technology companies are Prodco, Vivotek, Xovis, and Hikvision.

Tables 1 and 2 have comprehensively shown the comparisons of different criteria on the four popular similar systems in the existing market. All the systems have similar deployment environment like being ceiling mounted on a shopping mall entrance and use computer vision-based approach to handle huge crowds that require detection and counting on larger and faster scale. Similar to the approaches reviewed in research papers, these systems use image processing techniques or fast deep learning algorithms for preprocessing, human detection, and counting to reduce computational resources. Based on the research, almost all commercialized systems use GPU as the target machine due to the



Proposed system search and playback tab



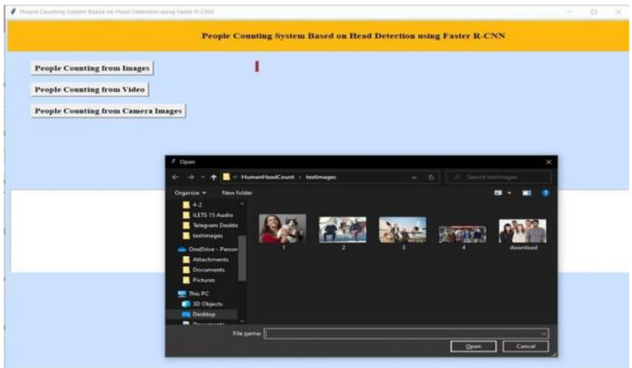
ALGORITHM

The piecemeal technique of the programme or its implementation is shown by the algorithm. In our planned system, we've got used quicker R-CNN Object Detection algorithmic program. This algorithmic program doesn't use any selective search. quicker R- CNN design consists of 2 networks. 1. Region Proposal Network 2. Object Detection Network CNN is the network's backbone, and it's shared by the Region Proposal Network and therefore the Object Detection .

I.CONCLUSION

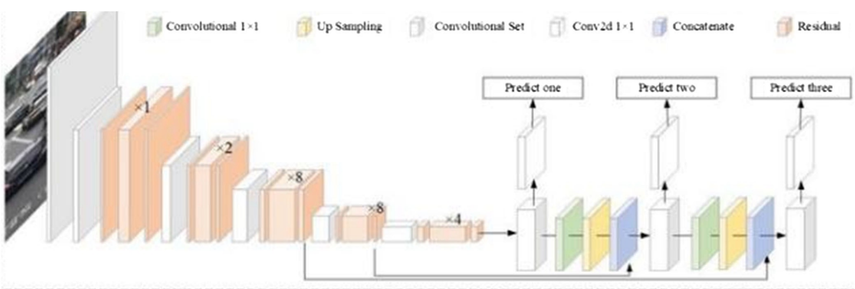
We suggested a system for counting individuals in crowded movies based on human head recognition. Its detection accuracy has increased while the time it takes to classify has dropped. We did experimental evaluations on the data set. The results showed significant progress. People counting systems are in high demand in the travel and transportation industry to improve

passenger management systems. Our proposed method can be used in a variety of nations to manage the personnel count in workplaces in compliance with new regulatory standards. Many businesses are turning to advanced people counting technologies to keep track of employee data and provide protection in the event of a COVID-19 emergency.



Picture Uploading

Output of selected picture



REFERENCES

[1] Al-Zaydi, Z., Vuksanovic, B., & Habeeb, I. (2018). Image processing based ambient context-aware people. International Journal of Machine Learning and Computing, 8, 268–273. https://doi. org/10.18178/ijmlc.2018.8.3.698

[2] Al-Zaydi, Z. N., Ndzi, D, & Sanders, D. A. (2016). Cascade method for image processing based people detection and counting. In Proceedings of 2016 international conference on image processing, production and computer science (pp. 30–36).

[3] Bernama. (2021, February 26). RM10,000 fine for SOP violations beginning March 11. Retrieved from newstraitstimes: https:// www.nst.com.my/news/nation/ 2021/02/669100/rm10000- fine-sop-violations-beginning-march-11

[4] Congrex. (2020, March 30). Disruption in the business events industry: rising to the challenges of Covid-19. Retrieved from congrex https://congrex.com/blog/ disruption-business- events-industry-challenges-covid-19/

[5] Jalled, F., & Voronkov, I. (2016). Object detection using image processing. In Computer vision and pattern recognition (pp. 1–6).

[6] javaTpoint. (2021, August 18). Advantage and Disadvantage of TensorFlow. Retrieved from javatpoint: https://www. javatpoint.com/advantage-and-disadvantage-of-tensorflow

[7] Li, B., Zhang, J., Zhang, Z., & Xu, Y. (2014). A people counting method based on head detection and tracking. In 2014 International conference on smart computing (pp. 136–141).

[8] Liu, G., & Sun, Y. (2012). An in-vehicle system for pedestrian detection. In 2012 11th International symposium on distributed computing and applications to business, engineering & science, vol. 1(1109) (pp. 328–331).

[9] Mao, Q.-C., Sun, H.-M., Liu, Y.-B., & Jia, R.-S. (2019). Mini-YOLOv3: Real-time object detector for embedded applications. IEEE Access, 7, 1–9. https://doi.org/10.1109/ ACCESS.2019.2941547

[10] Mokayed, H., Meng, L. K., Woon, H. H., & Sin, N. H. (2014). Car plate detection engine based on conventional edge detection technique. In The international conference on computer graphics, multimedia and image processing (CGMIP 2014). The Society of Digital Information and Wireless Communication.

[11] Mokayed, H., Shivakumara, P., Woon, H. H., Kankanhalli, M., Lu, T., & Pal, U. (2021). A new DCT-PCM method for license plate number detection in drone images. Pattern Recognition Letters, 148, 45–53. https:// doi.org/10.1016/j. patrec.2021.05.002

[12] Nikouei, S. Y., Chen, Y., Song, S., Xu, R., Choi, B.-Y., & Faughnan, S. Y. (2018). Real-time human detection as an edge service enable by a lighweight CNN. In 2018 IEEE international conference on edge computing (pp. 125–129).

[13] Piciarelli, C., & Foresti, G. L. (2011). Surveillance-oriented event detection in video streams. In Video event detection (pp. 32–41).

[14] Raghavachari, C., Aparna, V., Chithira, S., & Balasubramanian, V. (2015). A comparative study of vision based human detection techniques in people counting applications. In Second international symposium on computer vision and the internet (pp. 461–469).

[15] Redmon, J. (2021, August 18). yolo. Retrieved from pjreddie: https:// pjreddie.com/darknet/yolo/

[16] Reis, J. V. (2014). Image descriptors for counting people with uncalibrated cameras. MIEEC 2013/2014-Master in Electrical and Computers Engineering, University of Porto (pp. 1–73).

[17] Sumit, S. S., Watada, J., Roy, A., & Rambli, D. R. A. (2020). In object detection deep learning methods, YOLO shows supremum to Mask R-CNN. In The 2nd Joint International Conference on Emerging Computing Technology and Sports (JICETS 2019) (pp. 1–8).

[18] Times, T. S. (2020, May 8). Malaysia issues list of malls, hypermarkets seen as potential Covid-19 hot spots using new AI system. Retrieved from straitstimes: https://www. straitstimes.com/asia/se-asia/malaysia-lists-names-of-malls- hypermarkets-seen-as-covid-hotspots-using-new-ai-system

[19] Velastin, S. A., Fernández, R., Espinosa, J. E., & Bay, A. (2020). Detecting, tracking and counting people getting. Sensors, 20, 1–20. https://doi.org/10.3390/s20216251

[20] How to Cite: Mokayed, H., Quan, T. Z., Alkhaled, L., & Sivakumar, V. (2022). Real-Time Human Detection and Counting System Using Deep Learning Computer Vision

Techniques. Artificial Intelligence and Applications https://doi.org/10.47852/bonviewAIA2202391

[21] Dan Kong, Douglas Gray, and Hai Tao. "A viewpoint invariant ap- proach for crowd counting". In: 18th International Conference on Pattern Recognition (ICPR'06). Vol. 3. IEEE. 2006, pp. 1187–1190.