Human Detection and Counting with Faster R-CNN

B. Eswar Babu[1], Urella Ashwini[2],Attelli Pooja[2], Abhishek Patil [2]
[1]*Associate Professor,* [2]*UG Scholar*
Dept. of Information Technology, Vidya Jyothi Institute of Technology, Hyderabad, Telangana, India.eswar.banala@gmail.com

***Abstract*—** The Human Detection System represents a sophisticated computer vision application designed to accurately quantify individuals in high-density areas. This system addresses the necessity for real-time, reliable head detection to enumerate persons across diverse settings, including commercial centers, transit hubs, sports arenas, and mass gatherings. The development of this system is driven by the escalating need for automated and efficient crowd monitoring in the domains of security, retail analytics, and crowd management. A notable feature of the system is its real-time functionality, enabled by Faster R-CNN technology, which facilitates continuous and instantaneous population counts.

Keywords*—Computer Vision, Human Detection, Enumeration, Counting System, Crowd Management, Faster R-CNN*

I. INTRODUCTION

The rapid urbanization and increasing frequency of mass gatherings have necessitated advanced crowd monitoring solutions for safety, security, and operational efficiency. In response to this growing demand, the People Counting System has emerged as a cutting-edge computer vision application. This sophisticated system is designed to accurately quantify individuals in high-density areas, addressing the critical need for real-time, reliable head detection across diverse settings such as commercial centers, transit hubs, sports arenas, and large-scale events. [4,5]

The development of the Human Counting System is driven by the escalating requirements for automated and efficient crowd monitoring in various domains, including security, retail analytics, and crowd management. By leveraging state-of-the-art computer vision techniques, particularly the Faster R-CNN technology, the system offers real-time functionality that enables continuous and instantaneous population counts. This capability represents a significant advancement over traditional manual counting methods, providing stakeholders with timely and accurate data for informed decision-making. [1]

The Human Counting System's applications extend beyond mere numerical data collection. In the realm of security, it serves as a vital tool for identifying potential threats, managing crowd flow, and ensuring compliance with occupancy limits. For retail environments, the system offers invaluable insights into customer behavior, foot traffic patterns, and peak hours, enabling businesses to optimize staffing, layout, and marketing strategies. In transportation hubs, such as airports and train stations, the technology aids in queue management, resource allocation, and overall passenger experience enhancement. [2,3]

One of the key strengths of the People Counting System lies in its adaptability to various environmental conditions and crowd densities. Advanced algorithms allow for accurate detection and counting even in challenging scenarios, such as low-light conditions, partially occluded individuals, or rapidly moving crowds. This robustness ensures reliable performance across a wide range of real-world applications, from monitoring daily commuter flows to managing large-scale events and festivals. Moreover, the system's integration capabilities with other smart city technologies pave the way for comprehensive urban management solutions. By interfacing with traffic management systems, emergency response networks, and public transportation infrastructure, the People Counting System contributes to a holistic approach to urban planning and operations.

This integration facilitates proactive decision-making, enabling authorities to anticipate and respond to crowd-related challenges before they escalate. The ethical implications of widespread crowd monitoring technologies have not been overlooked in the development of the People Counting System. Stringent privacy measures are incorporated to ensure that individual identities are protected while still providing valuable aggregate data. This balance between functionality and privacy protection is crucial for public acceptance and regulatory compliance, especially in an era of increasing data privacy concerns. As urban environments become increasingly complex and populous, the People [6]

Counting System stands at the forefront of technological solutions aimed at enhancing public safety, optimizing resource allocation, and improving the overall management of crowded spaces. Its potential to transform urban living experiences through data-driven insights and real-time analytics positions it as a cornerstone technology in the smart cities of the future.

This introduction sets the stage for a comprehensive exploration of the system's architecture, methodologies, and potential applications in addressing the challenges of modern urban environments. The subsequent sections will delve into the technical intricacies of the Faster R-CNN algorithm, the system's implementation challenges, and case studies demonstrating its efficacy in various real-world scenarios. By examining these aspects in detail, we aim to provide a thorough understanding of the People Counting System's capabilities and its role in shaping safer, more efficient, and responsive urban spaces.[7,8]

II. RELATED WORK

Over the past decade, significant advancements have been made in the field of object detection. This review focuses specifically on human detection in thermal imagery, categorizing the approaches into traditional and deep learning-based methodologies.

Ma et al. employed a conventional technique involving blob extraction and classification for pedestrian detection and tracking. Their method utilized region-wise gradient and geometric constraints filtering for blob extraction, while HoG and DCT features were combined with SVM for blob classification. [9]

An efficient framework for pedestrian detection and tracking in thermal images was introduced by Lahouli et al. Their approach incorporated saliency maps with contrast enhancement for region extraction, and discrete Chebychev moments (DCM) served as features for SVM classification. Younsi et al. implemented GMM for moving object extraction, employing a similarity function based on shape, appearance, spatial, and temporal characteristics for detection purposes. [10]

The extraction of hot spots using maximally stable extremal regions (MSER) was performed by Teutsch et al., who then applied discrete cosine transform (DCT) and a random naive Bayes classifier for classification. [11]

For pedestrian detection in far infrared images, Biswas et al. employed local steering kernel (LSK) as low-level feature descriptors. Oluyide et al. developed a method for candidate generation and ROI extraction in IR surveillance videos, utilizing histogram specification and partitioning techniques for pedestrian detection. [12]

Zhang et al. presented an infrared-based video surveillance system that enhanced data resolution prior to applying a Faster R-CNN approach. A substantial and diverse thermal dataset (AAU PD T) was created by Huda et al., incorporating variations in capture time, weather conditions, camera distance, body and background temperatures, and shadows. They utilized the YOLOv3 detector for human detection tasks.[13]

Chen and Shin conducted pedestrian detection in infrared (IR) images utilizing an attention-guided encoder–decoder convolutional neural network (AED-CNN), which generates multi-scale features. Tumas et al. developed a 16-bit thermal pedestrian dataset (named ZUT), captured during severe weather conditions, and employed YOLOv3 to perform pedestrian detection. [14]

Huda et al. analyzed the impact of utilizing a pre-processed thermal dataset with the YOLOv3 object detector. The AAU data were enhanced using histogram stretching, and the performance was compared with the data in its original form.[15]

The optimal performance was obtained for the AAU data in its original form without the application of pre-processing techniques. Cioppa et al. proposed a novel system to detect and enumerate players in a football field, utilizing a network trained in a student–teacher distillation approach with custom augmentation and motion information.[16,17]

Haider et al. proposed a fully convolutional regression network to perform human detection in thermal images. This network was designed to map the human heat signature in the thermal image to spatial density maps. Additionally, various recent approaches found in the literature include the cascaded parsing network (CP-HOI) for multistage structured human object interaction recognition and differentiable multi-granularity human representation learning, which can be adopted for these tasks due to their superior performance in similar vision tasks.[18]

The literature review reveals that limited research has been conducted to perform small target detection specifically on aerial thermal images. As numerous object detection algorithms exist for generic object detection tasks, researchers may encounter ambiguity in selecting an algorithm for small target detection.

The proposed research conducts a performance analysis of Faster R-CNN and SSD algorithms to detect human targets with small dimensions in aerial view thermal images. Furthermore, an attempt is made to fine-tune these algorithms to enhance detection performance.[19,20]

III. METHODOLOGY

Faster R-CNN is an advanced object detection algorithm that combines two key components: the Region Proposal Network (RPN) and Fast R-CNN. This innovative approach represents a significant advancement in computer vision and object detection.

The RPN, a deep fully convolutional network, generates high-quality region proposals, which are potential areas in an image where objects might be located. This component efficiently scans the entire image and identifies regions of interest with high objectness scores, thereby reducing the computational burden of processing the entire image uniformly.[21]

These proposals are subsequently fed into the Fast R-CNN detector for final object classification and bounding box refinement. Fast R-CNN builds upon its predecessor, R-CNN, by introducing several key improvements, including the utilization of a single convolutional neural network to process the entire image and the implementation of a region of interest (RoI) pooling layer. [22]

This unified network architecture enables end-to-end training and significantly enhances both the accuracy and speed of object detection compared to its predecessors. A notable advantage of Faster R-CNN is its ability to share convolutional features between the RPN and Fast R-CNN stages. This feature sharing not only reduces computational overhead but also enhances the overall performance of the network. By leveraging a common set of convolutional layers, the model efficiently extracts meaningful features from the input image, which are then utilized by both the RPN for proposal generation and Fast R-CNN for object classification and localization.

In the context of transport system flow analysis, Faster R-CNN proves to be an invaluable tool for developing vision-based solutions. By leveraging its ability to detect and classify multiple object types simultaneously, such as cars, buses, trucks, motorcycles, bicycles, and pedestrians, the system provides comprehensive insights into traffic patterns and urban mobility.

This multi-class detection capability is particularly crucial in complex urban environments where various modes of transportation coexist and interact. The application of Faster R-CNN in traffic analysis enables researchers and urban planners to gather rich, detailed data on vehicle counts, traffic density, and movement patterns.

This information can be utilized to identify bottlenecks, optimize traffic signal timings, and design more efficient road layouts. Moreover, the algorithm's ability to detect and track pedestrians and cyclists allows for a more holistic approach to urban mobility planning, ensuring that the needs of all road users are considered. Furthermore, Faster R-CNN's robustness to variations in object scale, orientation, and partial occlusion makes it well-suited for real-world traffic scenarios. It can accurately detect vehicles and pedestrians even in crowded urban scenes or under challenging lighting conditions, providing reliable data for analysis and decision-making.[23,24]

The high speed and accuracy of Faster R-CNN also make it suitable for real-time traffic monitoring applications. By processing video feeds from traffic cameras in real-time, the system can provide up-to-the-minute information on traffic conditions, enabling rapid response to incidents or congestion.

This capability is particularly valuable for intelligent transportation systems and smart city initiatives, where timely and accurate data is crucial for effective management of urban infrastructure. In addition to its applications in traffic analysis, Faster R-CNN has demonstrated promise in related fields such as autonomous driving and advanced driver assistance systems (ADAS). Its ability to quickly and accurately detect objects in a vehicle's surroundings is essential for safe navigation and collision avoidance in self-driving cars.

The application of Faster R-CNN in the domain of transport system flow analysis demonstrates its versatility and potential for addressing real-world problems beyond traditional computer vision tasks. As urban populations continue to grow and the demand for efficient transportation systems increases, the role of advanced object detection algorithms like Faster R-CNN in shaping smarter, more sustainable cities is likely to become increasingly significant. By providing accurate, real-time data on urban mobility patterns, these technologies empower decision-makers to develop data-driven solutions that enhance the quality of life for city residents.

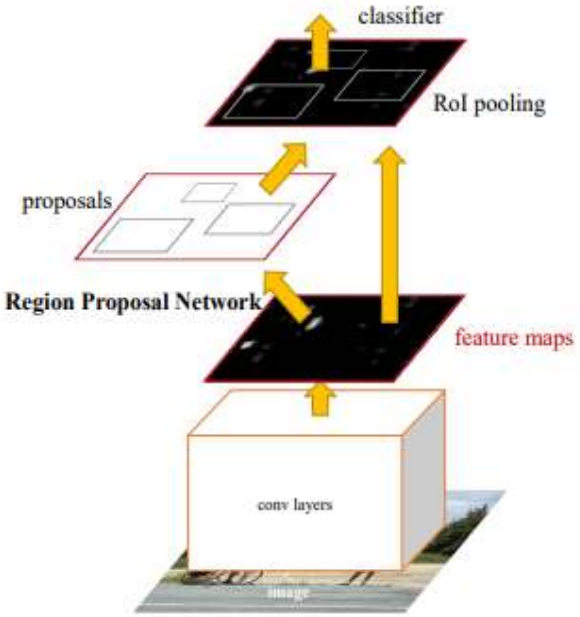

Fig 1: *Region Proposal Network*

IV. IMPLEMENTATION

The text describes a sophisticated head detection system that leverages the Faster R-CNN (Region-based Convolutional Neural Network) model, a state-of-the-art deep learning approach for object detection. This system is designed to accurately identify and count human heads in images or video streams. Key components and functionalities of the system include:

1. Model Implementation: The Faster R-CNN model is loaded using PyTorch's torchvision module, which provides pre-trained models and utilities for computer vision tasks. This implementation allows for efficient and accurate object detection.

2. Core Function: The get_prediction function serves as the central component of the system. It takes two primary inputs: the path to the image file and a confidence threshold. This function orchestrates the entire detection process, from image preprocessing to final prediction output.

3. Image Preprocessing: Before feeding the image into the model, several preprocessing steps are applied. These include converting the image to a tensor format, resizing it to a standard dimension, and normalizing the pixel values. These steps ensure that the input is compatible with the model's requirements and help improve detection accuracy.

4. Prediction Generation: The Faster R-CNN model processes the preprocessed image and generates predictions for various objects within the image. These predictions include bounding box coordinates (indicating the position of detected

objects) and confidence scores (representing the model's certainty about each detection).

5. Threshold-based Filtering: To enhance accuracy and reduce false positives, the system applies a confidence threshold to the predictions. Only detections with confidence scores above this threshold are considered valid. This step is crucial for accurately counting human heads while minimizing errors.

6. Graphical User Interface (GUI): The system features a user-friendly GUI that displays the results of the head detection process. It shows the processed images or video frames with the detected heads highlighted and provides a count of the identified heads. This visual feedback allows users to easily interpret the results.

7. Versatile Input Handling: The GUI offers options for processing both static images and video streams. Users can select image files or video sources through file selection dialogs, making the system adaptable to various use cases.

8. R-CNN Technology: The text provides context on R-CNN (Region-based Convolutional Neural Network) technology, describing it as a two-stage detection algorithm. This approach first proposes regions of interest and then classifies objects within those regions. R-CNN and its variants, including Faster R-CNN, have found applications in diverse fields such as autonomous driving, surveillance systems, and facial recognition.

9. Related Techniques: The text briefly mentions related object detection techniques like Fast R-CNN, suggesting that the system is built upon a foundation of evolving deep learning methodologies for computer vision tasks.

Fig 3 : Output2

Fig 4: Output3

Fig2 : Output1

Fig 5: Output4

This head detection system demonstrates the application of advanced machine learning techniques to solve practical problems in computer vision. Its ability to process both images and videos, coupled with a user-friendly interface, makes it suitable for various applications ranging from crowd monitoring to security and safety systems. The use of the Faster R-CNN model, known for its speed and accuracy, ensures that the system can perform real-time detection tasks efficiently.

V. CONCLUSION

The Human Detection System presents a significant advancement in crowd monitoring technology, offering real-time, accurate head detection for population quantification in various high-density environments. By leveraging Faster R-CNN technology, this system provides instantaneous and continuous crowd counts, addressing critical needs in security, retail analytics, and crowd management. As urban populations grow and public gatherings become more frequent, the importance of such sophisticated monitoring tools is likely to increase. The system's ability to operate across diverse settings positions it as a versatile solution for a wide range of applications, potentially revolutionizing how we approach crowd safety, space utilization, and event planning in densely populated areas.

REFERENCES

[1] Lee, D.; Lee, S.H.; Masoud, N.; Krishnan, M.S.; Li, V.C. Integrated digital twin and blockchain framework to support accountable information sharing in construction projects. Autom. Constr. 2021, 127, 103688.

[2] Zhao, Z.Q.; Zheng, P.; Xu, S.T.; Wu, X. Object detection with deep learning: A review. IEEE Trans. Neural Netw. Learn. Syst. 2019, 30, 3212–3232.

[3] Felzenszwalb, P.F.; Girshick, R.B.; McAllester, D.; Ramanan, D. Object Detection with Discriminatively Trained Part-Based Models. IEEE Trans. Pattern Anal. Mach. Intell. 2010, 32, 1627–1645

[4] Liu, L.; Ouyang, W.; Wang, X.; Fieguth, P.; Chen, J.; Liu, X.; Pietikäinen, M. Deep learning for generic object detection: A survey. Int. J. Comput. Vis. 2020, 128, 261–318

[5] Jiao, L.; Zhang, F.; Liu, F.; Yang, S.; Li, L.; Feng, Z.; Qu, R. A survey of deep learning-based object detection. IEEE Access 2019, 7, 128837–128868.

[6] Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587

[7] Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Washington, DC, USA, 7–13 December 2015; pp. 1440–1448.

[8] Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans. Pattern Anal. Mach. Intell. 2017, 39, 1137–1149.

[9] Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.

[10] He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 2015, 37, 1904–1916.

[11] He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.

[12] Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.

[13] Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. 2018. Available online: http://xxx.lanl.gov/abs/1804.02767

[14] Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv 2020, arXiv:2004.10934.

[15] Ramesh Cheripelli and A. N. K. Prasannanjaneyulu Generating Automatic Ground Truth by~Integrating Various Saliency Techniques,Proceedings of Second International Conference on Advances in Computer Engineering and Communication Systems.doi 10.1007/978-981-16-7389-4_35

[16] Fu, C.Y.; Liu, W.; Ranga, A.; Tyagi, A.; Berg, A.C. Dssd: Deconvolutional single shot detector. arXiv 2017, arXiv:1701.06659.

[17] Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. Centernet: Keypoint triplets for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 6569–6578.

[18] Law, H.; Deng, J. Cornernet: Detecting objects as paired keypoints. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 734–750.