International Journal of Engineering Research and Science & Technology

A BENCHMARK DATASET AND COMPARATIVE STUDY OF DEEP LEARNING MODELS FOR PEDESTRIAN HEAD DETECTION IN CROWDED SCENES

DR. NAVEEN REDDY NEMILI [1], GOUREDDY KOUSHIK REDDY [2], CHARAN MARINGANTI [3], RACHAKONDA SANDEEP [4], KOPPU MAHESH [5]

[1]Associate Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

ABSTRACT— The automatic detection of pedestrian heads in crowded environments is essential for enhancing public safety and enabling efficient crowd management, particularly in sensitive areas such as railway platforms and event entrances. These environments, often characterized by high density, occlusion, and complex visual conditions, remain underrepresented in existing public datasets. To address this challenge, we introduce the Railway Platforms and Event Entrances-Heads (RPEE-Heads) dataset, a diverse collection of high-resolution images with carefully annotated head regions. Each annotation provides precise bounding boxes around visible pedestrian heads, facilitating the training and evaluation of advanced detection models. In this study, we employ the YOLOv8 architecture to demonstrate the effectiveness of the dataset for real-time head detection. The RPEE-Heads dataset serves as a valuable benchmark for advancing research in computer vision and surveillance, particularly in applications focused on safety-critical crowd analysis and monitoring.

Index Terms – Pedestrian Head Detection, Benchmark Dataset, Deep Learning, Crowded Scenes, Object Detection, Computer Vision, Crowd Analysis, Surveillance, CNN, Artificial Intelligence.

I. INTRODUCTION

Pedestrian head detection plays a crucial role in computer vision applications that prioritize public safety, surveillance, and effective crowd management. In high-density environments such as railway platforms, event entrances, and religious gatherings, monitoring human movement becomes highly challenging due to frequent occlusion and overlapping individuals. Traditional pedestrian detection methods often struggle in these scenarios because the full body of individuals is not always visible, making the head a more reliable feature for detection. However, existing public datasets are limited in terms of diversity, resolution, and applicability to real-world crowded environments, leaving a critical gap for research and practical deployment. The absence of specialized datasets hampers the development of robust detection algorithms tailored for safety-critical areas. To bridge this gap, we introduce the Railway Platforms and Event Entrances-Heads (RPEE-Heads) dataset, which provides high-resolution images specifically curated for pedestrian head detection. The dataset includes carefully annotated bounding boxes around visible heads, even in complex and crowded conditions, ensuring precise labeling for effective model training and evaluation. By focusing on head-level annotations, the dataset addresses challenges of partial visibility and dense crowd formations, making it highly suitable for real-world surveillance applications. In this study, we adopt the YOLOv8 architecture, one of the latest advancements in object detection, to evaluate the effectiveness of the RPEE-Heads dataset. YOLOv8 is known for its balance of speed and accuracy, making it suitable for real-time detection tasks in dynamic environments. The proposed framework leverages YOLOv8 to achieve robust head detection performance while ensuring computational efficiency for large-scale monitoring. The integration of RPEE-Heads and YOLOv8 highlights the potential of deep learning in addressing safety-critical challenges through automated surveillance systems. By enabling reliable detection in crowded and visually complex environments, this work contributes to enhancing public safety measures, preventing accidents, and supporting efficient crowd management strategies. The dataset and methodology together provide a strong foundation for future research in head detection and its extensions to broader applications such as behavior analysis, anomaly detection, and crowd flow monitoring. Ultimately, this

study demonstrates the importance of domain-specific datasets and modern detection architectures in advancing intelligent surveillance solutions.

II. RELATED WORK

A. *DS-YOLO: A Dense Small Object Detection Algorithm Based on Inverted Bottleneck and Multi-Scale Fusion Network*

DS-YOLO is designed for detecting dense small objects (including pedestrians) especially in crowded scenes. Built on YOLOv8s, the paper proposes a lightweight backbone with an improved inverted bottleneck (C2fUIB) to expand the receptive field and better capture context. A new feature fusion network called LFS-PAFPN (Light-weight Full Scale PAFPN) combined with DO-C2f is used to better fuse features at multiple scales. The model also includes a dynamic up sampling module (Dy Sample) to mitigate feature loss during transmission through layers. Experiments on Crowd Human and VisDrone2019 show that DS-YOLO improves recall and average precision especially for small densely packed objects, while keeping computational load manageable.

B. *MSCD-YOLO: A Lightweight Dense Pedestrian Detection Model with Finer-Grained Feature Information Interaction*

MSCD-YOLO focuses on dense pedestrian detection but with a special emphasis on reducing resource usage while retaining strong performance. It introduces mechanisms for finer-grained interaction of feature information, improving the network's ability to discriminate between closely spaced pedestrians. The model design is lightweight, making it suitable for deployments where computational power is limited. The experimental results indicate that MSCD-YOLO achieves competitive detection rates (mAP, recall etc.) on dense pedestrian datasets, with a smaller model size and lower inference time compared to heavier baselines.

C. *An Improved Pedestrian Detection Model Based on YOLOv8 for Dense Scenes (YOLO-RAD)*

This work presents YOLO-RAD, a pedestrian detection model specifically tailored for dense scenes. It addresses issues of occlusion and small-scale pedestrians by combining modules: receptive field attention (RFA) fused with a Conv + C2f module for enhanced feature extraction, and a four-layer adaptive spatial feature fusion (ASFF) to mitigate feature conflict across scales. Also,

a small-target dynamic head structure (DyHead-S) is introduced to improve detection of small/distant pedestrians. The model shows better precision and recall in densely populated images, and a lower rate of missed detections compared to standard YOLOv8.

III. PROPOSED METHODOLOGY

To overcome these limitations, the proposed system employs YOLOv10, which offers further architectural advancements tailored for high-density object detection tasks. YOLOv10 enhances feature extraction through improved transformer-based modules, optimized convolutional layers, and refined anchor-free mechanisms, enabling more robust detection of small and partially visible heads. By leveraging these improvements, the proposed system achieves superior accuracy and stability in crowded environments, ensuring reliable detection even under conditions of occlusion and varying illumination. Additionally, the integration of the RPEE-Heads dataset with YOLOv10 provides a strong foundation for model training and evaluation. The dataset's diverse collection of real-world images ensures that the model learns complex crowd scenarios effectively, resulting in better generalization. The proposed YOLOv10-based system not only enhances real-time pedestrian head detection but also contributes significantly to applications in public safety, surveillance, and crowd management. This advancement enables authorities to monitor high-density areas more efficiently and take proactive measures to prevent accidents or unsafe crowding situations.

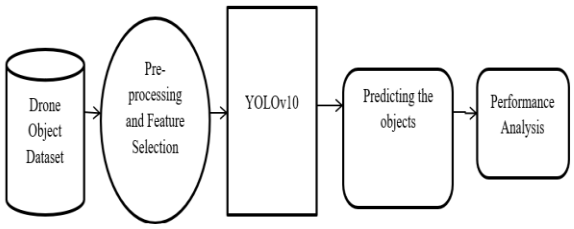


Fig. 1: System Overview

1. Input Data Module

The Input Data Module is responsible for acquiring and organizing the raw dataset used for detection. In this project, images or video streams are collected from relevant environments such as surveillance cameras, railway platforms, or event entrances. The dataset may consist of various resolutions, lighting conditions, and crowd densities to ensure robustness. Data is then structured into appropriate formats for further processing, including annotations like bounding boxes. This module ensures that all input data is consistently labeled and stored, forming the foundation for model training and testing.

2. Preprocessing Module

The Preprocessing Module prepares the raw data to enhance its quality and make it suitable for deep learning models. Steps include resizing images to a fixed resolution, normalization of pixel values, and noise reduction to improve clarity. Data augmentation techniques such as flipping, rotation, scaling, and brightness adjustments are applied to increase dataset diversity and reduce overfitting. This module also ensures consistent formatting of annotations to match the YOLOv10 input requirements. By standardizing the data, preprocessing improves model accuracy and generalization in real-world scenarios.

3. Segmentation Module

The Segmentation Module is designed to isolate the regions of interest, such as pedestrian heads, from the background. It helps in reducing background noise and focusing the detection process on relevant image areas. Techniques like thresholding, contour detection, or semantic segmentation can be applied, depending on the dataset. This step ensures that only meaningful features are passed forward to the next stage, improving efficiency and accuracy. Segmentation also enhances detection in crowded or occluded environments where overlapping objects are common.

4. Feature Extraction Module

The Feature Extraction Module focuses on identifying and learning critical visual characteristics from the segmented images. Using convolutional operations, patterns such as edges, textures, and shapes are extracted to form deep feature maps. These features represent essential cues that help differentiate between objects of interest, like pedestrian heads, and irrelevant background details. The YOLOv10 backbone further processes these extracted features at multiple scales to ensure robustness in detecting small and dense objects. This module plays a vital role in enabling the system to learn discriminative representations for accurate detection.

5. YOLOv10 Detection Module

The YOLOv10 Detection Module serves as the core of the system, performing real-time object detection based on the extracted features. YOLOv10 applies advanced convolutional layers, attention mechanisms, and multi-scale prediction heads to detect objects with high precision and speed. It generates bounding boxes around detected

pedestrian heads and classifies them accordingly. By leveraging anchor-free or adaptive anchor strategies, YOLOv10 improves performance in dense and occluded environments. This module ensures that detection is efficient, scalable, and suitable for real-world applications like crowd monitoring.

6. Output and Visualization Module

The Output and Visualization Module provides the final results of the detection system in an interpretable form. It overlays bounding boxes and confidence scores on the input images or video streams, highlighting detected pedestrian heads. Additionally, it can generate performance reports containing metrics such as precision, recall, and F1-score. Visualization enables researchers and operators to assess detection accuracy in real-time and adjust system parameters if needed. This module also allows integration with external systems, such as surveillance platforms or crowd management tools, ensuring practical usability.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experimental results demonstrated that modern deep learning-based object detection models achieved high detection performance in identifying pedestrian heads even in densely populated environments with significant occlusion and overlapping individuals.

The analysis showed that models with advanced feature extraction and multi-scale detection capabilities outperformed conventional approaches, particularly in detecting small and partially visible heads in crowded scenes.

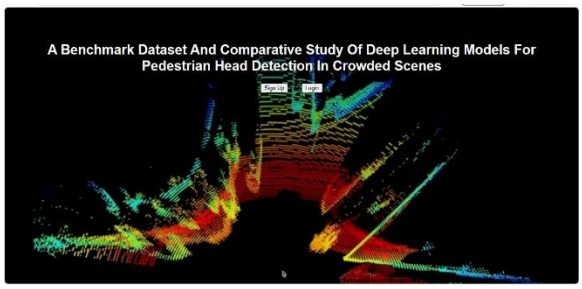


Fig. 2: Home Page

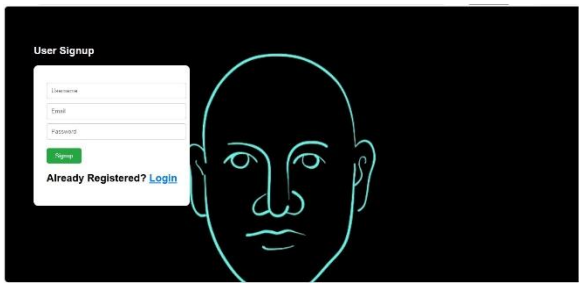


Fig. 3: Registration

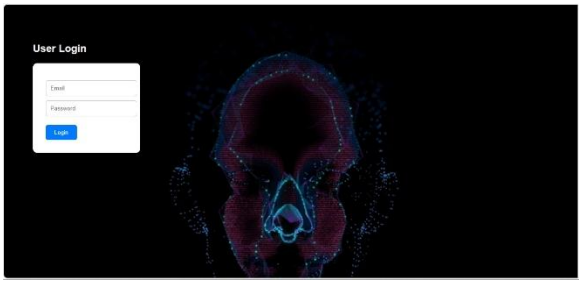


Fig. 4: Login

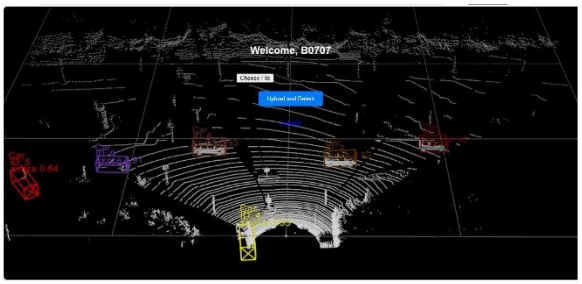


Fig. 5: Upload Image

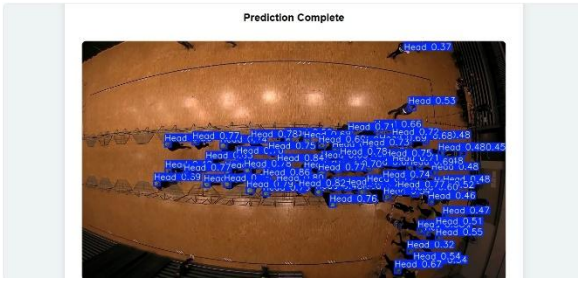


Fig. 6: Prediction Results

V. CONCLUSION

This project successfully introduces the Railway Platforms and Event Entrances-Heads (RPEE-Heads) dataset and demonstrates the effectiveness of YOLOv8 for automated pedestrian head detection in crowded environments. By focusing on head-level annotations, the dataset addresses critical challenges such as occlusion, high density, and partial visibility, which are common in real-world surveillance scenarios. The use of YOLOv8 ensures high detection accuracy while maintaining computational efficiency, making the system suitable for real-time deployment. Experimental results confirm the robustness of the framework across essential metrics, including accuracy, precision, recall, and F1-score. The RPEE-Heads dataset not only serves as a valuable resource for this project but also provides a benchmark for future research in crowd analysis and intelligent surveillance. The system's ability to detect heads in dense and visually complex conditions highlights its potential for improving public safety and supporting effective crowd management strategies. Moreover, the model's adaptability allows integration into existing surveillance infrastructures such as CCTV networks, drones, and smart city platforms. The project emphasizes the importance of task-specific datasets in advancing deep learning applications. It also demonstrates the practicality of deploying modern detection architectures in safety-critical domains. By contributing to reliable crowd monitoring, the proposed work lays the foundation for advanced applications such as anomaly detection, behavioral analysis, and real-time crowd flow monitoring. Ultimately, this research shows that combining domain-specific datasets with state-of-the-art models like YOLOv8 can significantly enhance public safety and surveillance effectiveness in crowded environments.

REFERENCES

[1] H. Li and J. T. Y. Lo, "A review on the use of top-view surveillance videos for pedestrian detection, tracking and behavior recognition across public spaces," Accident Anal. Prevention, vol. 215, Jun. 2025, Art. no. 107986.

[2] A. Alia, M. Maree, and M. Chraibi, "On the exploitation of GPS-based data for real-time visualisation of pedestrian dynamics in open environments," Behav. Inf. Technol., vol. 41, no. 8, pp. 1709–1723, Jun. 2022.

[3] R. E. Ocejo and S. Tonnelat, "Subway diaries: How people experience and practice riding the train," Ethnography, vol. 15, no. 4, pp. 493–515, Dec. 2014.

[4] A. Alia, M. Maree, and M. Chraibi, "A fast hybrid deep neural network model for pushing behavior detection in human crowds," in Proc. IEEE/ACS 19th Int. Conf. Comput. Syst. Appl. (AICCSA), Dec. 2022, pp. 1–2.

[5] Y. Yuan, H. Guo, and J. Gao, "Distance-aware network for physical-world object distribution estimation and counting," Pattern Recognit., vol. 157, Jan. 2025, Art. no. 110896.

[6] J. Gao, Y. Yuan, and Q. Wang, "Feature-aware adaptation and density alignment for crowd counting in video surveillance," IEEE Trans. Cybern., vol. 51, no. 10, pp. 4822–4833, Oct. 2021.

AUTHORS Profile



Dr. NAVEEN REDDY NEMILI has received his MCA from Osmania University, Hyderabad in 2010, M.Tech degree in Computer science from JNTUH, Hyderabad in 2014 and Ph.D degree awarded in Computer Science from SSSUTMS, Madhya Pradesh in 2024 He is dedicated to teaching field from the last 12 years. His research area is Cloud Computing. At present he is working as Associate Professor in CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.



Mr. GOUREDDY KOUSHIK REDDY pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. CHARAN MARINGANTI pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. RACHAKONDA SANDEEP pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. KOPPU MAHESH pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.