WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Real-Time Pedestrian Detection System Using Deep Learning

Mahanthi Hakshika, Annepu Sirisha, Dr. Kotti Jayasri

CSE-AI&DS, GMR Institute of Technology, Rajam, India.

ABSTRACT:

Ensuring accurate and real-time pedestrian detection is a critical challenge in the development of intelligent transportation systems and urban safety management. Factors such as occlusions, varying lighting conditions, diverse pedestrian postures, and low-resolution or multi-spectral images can significantly degrade detection performance, increasing the risk of accidents. Analysis on the effectiveness of deep learning techniques, particularly Convolutional Neural Networks integrated with YOLO and OpenCV, highlights improvements in detection accuracy and processing speed across benchmark datasets. Advanced computer vision methods, image preprocessing, and model optimization reduce errors in real-world scenarios, enhancing the safety of both pedestrians and autonomous vehicles. Continuous development and real-world validation are needed to ensure robust and reliable detection in dynamic urban environments.

Keywords: Pedestrian detection, autonomous vehicles (AVs), deep learning (DL), Convolutional Neural Networks (CNNs), Open Computer Vision, You Only Look Once, real-time decision-making, Urban Traffic Optimization, Accident Reduction.

Introduction:

Pedestrian detection is a critical aspect of modern computer vision, playing a vital role in urban planning, traffic optimization, and autonomous driving. With the rapid increase in population and road accidents, ensuring pedestrian safety has become a central concern for smart transportation systems. However, accurate detection remains challenging due to occlusions, varying lighting conditions, low-resolution images, and complex urban scenarios. This study focuses on addressing these challenges by leveraging deep learning techniques, particularly Convolutional Neural Networks (CNNs), in combination with OpenCV for robust pedestrian detection. By incorporating the YOLO (You Only Look Once) framework, the proposed system aims to achieve real-time detection while maintaining high accuracy across diverse environments. OpenCV enhances the pipeline through efficient image pre-processing, feature extraction, and frame-by-frame analysis, enabling continuous tracking of pedestrians in dynamic scenes. The integration of YOLO and OpenCV allows the system to distinguish pedestrians from complex backgrounds, handle crowded and multi-spectral scenarios, and maintain performance under challenging conditions. Benchmark datasets are utilized to evaluate the model's effectiveness, demonstrating its ability to manage occlusions, varying illumination, and different pedestrian postures. The ultimate goal is to develop a scalable and reliable pedestrian detection framework that can support traffic surveillance, autonomous vehicle operation, and smart city applications, contributing to safer urban environments and improved traffic management.

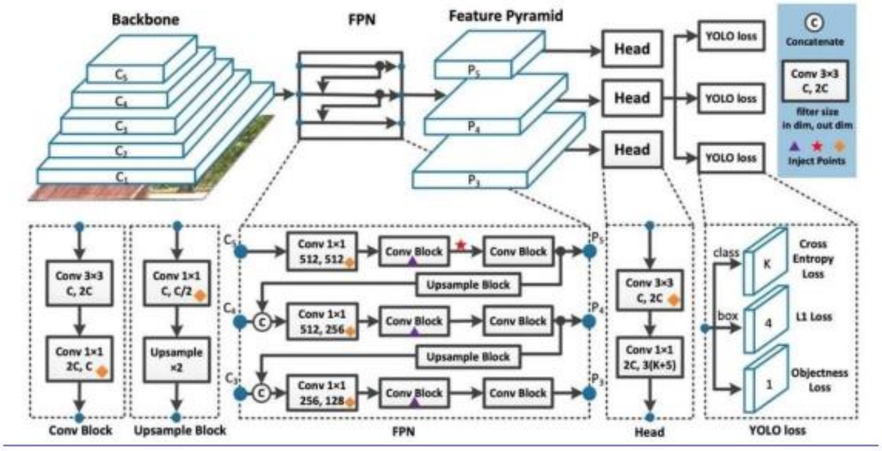
Literature Survey:

The paper [1] proposes the CBODL-RPD model, combining YOLOv4 with SqueezeNet and Colliding Bodies Optimization for hyperparameter tuning, to achieve robust pedestrian detection in images and video sequences. The authors [2] review deep learning-based pedestrian detection approaches, highlighting challenges such as occlusion, low-quality images, and multispectral data, and compare CNN, Faster R-CNN, MobileNet-SSD, and YOLO models for autonomous vehicle applications. The research [3] introduces a fusion framework combining object detection networks and semantic segmentation networks to improve pedestrian detection accuracy, particularly for partially occluded or low-light scenarios. The survey [4] traces the evolution from handcrafted features to deep CNN-based methods, discussing multi-scale, part-based, anchor-free, and multi-task models while emphasizing multi-spectral detection and real-time performance. The PF-YOLOv4 algorithm [5] enhances small-scale pedestrian detection by integrating soft thresholding, depthwise separable convolution, and attention modules to mitigate noise and occlusion effects. The review [6] focuses on occlusion handling and multi-scale pedestrian detection using part-based and attention-guided mechanisms. The study [7] introduces a Multi-Focus Detection Network using light-field imaging for small pedestrian detection, distinguishing them from anthropomorphic negative samples. The paper [8] reviews deep learning methods for occluded and multi-scale pedestrian detection, highlighting two-stage and single-stage frameworks, attention mechanisms, and dataset usage for performance evaluation. The survey [9] addresses occlusion in pedestrian detection, describing traditional component-based approaches and modern CNN-based solutions including attention mechanisms and repulsion loss functions. The paper [10] applies YOLOv5 for real-time pedestrian detection and age-based classification, demonstrating improved accuracy with data augmentation. The study [11] explores UAV-based YOLOv5 pedestrian detection for social distancing monitoring, emphasizing real-time performance and multi-class detection. The paper [12] presents a scaleinvariant Mask R-CNN for pedestrian detection and suspicious activity recognition in academic settings. The study [13] proposes the Selective Alignment Network (SAN) for cross-domain pedestrian detection, addressing domain shifts between source and target datasets using image- and instance-level adaptation. The paper [14] develops a modified YOLOv3 framework for real-time surveillance systems, optimizing anchor boxes and using a lightweight backbone for efficient detection. Finally, the study [15] introduces PiFeNet, a 3D pedestrian detector using LiDAR point clouds, with Pillar Aware Attention and Mini-BiFPN modules to improve detection accuracy and real-time performance.

Methodology:

The pedestrian detection system is designed around YOLOv8x as the core detection backbone, integrated with a custom-built tracking framework. This combination enables robust pedestrian monitoring across multiple zones with high accuracy. The system follows a modular architecture, where detection, tracking, and region management operate independently but communicate seamlessly.

1. Model Architecture



YOLOv8x features a modular architecture composed of three primary components: the Backbone, Neck, and Head. The Backbone is responsible for extracting rich features from input images using a CNN-based design optimized for faster computation and enhanced feature representation, effectively capturing details at multiple levels to detect pedestrians of varying sizes. The Neck aggregates and refines these features using advanced feature pyramid techniques like the Path Aggregation Network (PANet), enabling efficient fusion of spatial and semantic information to improve detection performance across different object scales and positions. Finally, the Head generates the model's output by producing bounding boxes, class probabilities, and confidence scores through an anchor-free detection approach, which predicts object locations and categories directly without relying on predefined anchor boxes, resulting in greater flexibility and accuracy in object detection.

2. Pre-processing

The pre-processing pipeline for YOLOv8x involves several crucial steps to ensure consistent and efficient model training. Input image resizing standardizes all images to a resolution of 640 × 640 pixels, ensuring uniform input dimensions across the dataset. Normalization follows, where pixel values are scaled to the range [0, 1] to stabilize gradients and accelerate the learning process. Finally, data augmentation techniques such as random cropping, flipping, scaling, and color adjustments are applied to enhance the model's robustness and generalization, enabling it to perform effectively under diverse lighting conditions, orientations, and object variations.

3. Training Process

The training process of YOLOv8x integrates an efficient combination of loss computation, optimization, and batch processing to achieve high performance and generalization. The loss function comprises three main components: the Localization Loss, which measures errors in predicted bounding boxes using Complete IoU (CIoU) or Distance IoU (DIoU) to account for positional and shape discrepancies; the Classification Loss, which assesses the accuracy of predicted object categories using cross-entropy or focal loss to handle class imbalance; and the Confidence Loss, which minimizes false detections by ensuring precise confidence score estimation. The model parameters are updated using adaptive optimizers such as AdamW or SGD, supported by a dynamic learning rate scheduler that adjusts learning rates throughout training for optimal convergence. To maximize computational efficiency, images are processed in batches, allowing better GPU utilization, while mixed-precision training accelerates computation and conserves memory. Finally, the model is trained over multiple epochs, continuously monitoring validation datasets to fine-tune performance and prevent overfitting.

4. Inference

During inference, YOLOv8 executes a streamlined detection pipeline to produce accurate and reliable results. It begins with bounding box prediction, where the model generates bounding boxes around detected objects along with corresponding confidence scores that indicate the likelihood of a correct detection. Next, class prediction assigns a specific class label to each detected object, identifying what the model has recognized within the image. Finally, Non-Maximum Suppression (NMS) is applied to eliminate redundant or overlapping bounding boxes, retaining only the most confident and precise predictions. This process ensures that YOLOv8 delivers clean, non-overlapping, and high-confidence detections in real time.

5. Performance Optimization

To enhance deployment efficiency and real-time performance, YOLOv8 employs several model optimization techniques. Quantization reduces the model's size and accelerates inference by converting high-precision weights to lower formats such as FP16 or INT8, making the model more suitable for edge and embedded devices. Pruning further optimizes performance by eliminating redundant weights, filters, or layers, thereby decreasing computational overhead and memory usage while maintaining near-original accuracy. Additionally, YOLOv8 supports hardware acceleration, being optimized for modern GPUs, TPUs, and edge devices like NVIDIA Jetson and Google Coral TPU, enabling efficient execution across a wide range of deployment environments—from cloud servers to real-time embedded systems.

6. Evaluation Metrics

The performance evaluation of YOLOv8 is conducted using key metrics that comprehensively measure its accuracy and efficiency. The Mean Average Precision (mAP) serves as the primary metric, assessing detection accuracy across all object classes and various Intersection over Union (IoU) thresholds, providing a balanced evaluation of both localization and classification performance. Inference Time measures the model's real-time processing capability, indicating how quickly it can analyze and predict objects per image—an essential factor for applications like surveillance and autonomous systems. Lastly, the F1 Score offers a combined assessment of precision and recall, ensuring that the model maintains a strong balance between correctly identifying true positives and minimizing false detections, thus reflecting its overall reliability and robustness in real-world scenarios.

Objective:

- High-accuracy pedestrian detection using YOLOv8x, capable of handling crowded scenes and varying lighting conditions.
- Real-time processing optimized for traffic monitoring, surveillance, and public safety applications.
- Selective detection of pedestrians (class_IDs = [0]), reducing false positives from other objects.
- Robust frame-by-frame analysis, ensuring consistent detection in dynamic environments with moving crowds or vehicles.
- Built in Python with a modular design, making it scalable and easy to integrate features like path tracking or pedestrian counting for large-scale deployments.

Results

- Detection Accuracy: The YOLOv8 pedestrian detection system achieved strong performance, with precision around 81–82% and recall around 87–88%, indicating reliable detection with a balanced trade-off between false positives and false negatives.
- Temporal Consistency: The system maintained good frame-to-frame coherence with a temporal consistency of 78.2%, though the negative detection stability (-0.872) highlights occasional fluctuations in detection confidence across frames.
- Frame-Level Performance: Variability in precision, recall, and F1-score across frames suggests challenges in crowded scenes, occlusions, and motion blur, pointing to areas where temporal context or improved feature extraction could enhance performance.
- Limitations and Improvement Areas: Intermittent detection gaps, under-detection in certain frames, and limited exploitation of video sequence continuity indicate potential for better temporal modeling and scenario coverage to improve robustness in real-time applications.

Conclusion

YOLOv8 demonstrates strong performance in pedestrian detection tasks, achieving impressive metrics with precision around 81.4%, recall at 78.7%, and an F1 score of 79.4%. These results highlight the model's robust architecture and efficient detection capabilities, particularly its ability to maintain high precision while effectively identifying most pedestrians. The temporal consistency score of 78.2% further validates YOLOv8's capacity to provide reliable detections across consecutive video frames, which is crucial for real-time applications. The balanced trade-off between precision and recall, along with its real-time processing efficiency, underscores YOLOv8's suitability for practical deployment in autonomous vehicles, traffic monitoring, and smart city applications. Its single-stage detection approach and advanced architectural improvements in version 8 allow it to handle complex and dynamic environments with significant robustness. Looking forward, integrating temporal smoothing, enhanced tracking mechanisms, and multi-sensor fusion could further elevate YOLOv8's detection performance, particularly in challenging conditions such as rain, fog, and snow. Overall, the results validate YOLOv8 as a highly capable solution for real-time object detection, providing a solid foundation for continued optimization and deployment in real-world autonomous systems.

References:

1. Iftikhar, S., Zhang, Z., Asim, M., Muthanna, A., Koucheryavy, A., & Abd El-Latif, A. A. (2022). Deep learning-based pedestrian detection in autonomous vehicles: Substantial issues and challenges. Electronics, 11(21), 3551.

2. Islam, M. M., & Karimoddini, A. (2022). Pedestrian detection for autonomous cars: inference fusion of deep neural networks. IEEE Transactions on Intelligent Transportation Systems, 23(12), 23358-23368.

3.] Cao, J., Pang, Y., Xie, J., Khan, F. S., & Shao, L. (2021). From handcrafted to deep features for pedestrian detection: A survey. IEEE transactions on pattern analysis and machine intelligence, 44(9), 4913-4934.

4. Li, K., Zhuang, Y., Lai, J., & Zeng, Y. (2023). PFYOLOv4: An improved small object pedestrian detection algorithm. IEEE Access, 11, 17197-17206.

5. Li, F., Li, X., Liu, Q., & Li, Z. (2022). Occlusion handling and multi-scale pedestrian detection based on deep learning: A review. IEEE Access, 10, 19937-19957.

6. Zhao, Y., Shi, F., Zhao, M., Zhang, W., & Chen, S. (2020). Detecting small scale pedestrians and anthropomorphic negative samples based on light-field imaging. IEEE Access, 8, 105082-105093.

7. Xiao, Y., Zhou, K., Cui, G., Jia, L., Fang, Z., Yang, X., & Xia, Q. (2021). Deep learning for occluded and multi-scale pedestrian detection: A review. IET Image Processing, 15(2), 286-301.

8. Ning, C., Menglu, L., Hao, Y., Xueping, S., & Yunhong, L. (2021). Survey of pedestrian detectionwith occlusion. Complex & Intelligent Systems, 7, 577-587.

9. Sukkar, M., Kumar, D., & Sindha, J. (2021, July). Real-time pedestrians detection by YOLOv5. In 2021 12th international conference on computing communication and networking technologies (ICCCNT) (pp. 01-06). IEEE.

10. Shao, Z., Cheng, G., Ma, J., Wang, Z., Wang, J., & Li, D. (2021). Real-time and accurate UAV pedestrian detection for social distancing monitoring in COVID-19 pandemic. IEEE transactions on multimedia, 24, 2069-2083.

11. Gawande, U., Hajari, K., & Golhar, Y. (2023). Real-time deep learning approach for pedestrian detection and suspicious activity recognition. Procedia Computer Science, 218, 2438-2447.

12. Jiao, Y., Yao, H., & Xu, C. (2021). SAN: Selective alignment network for cross-domain pedestrian detection. IEEE transactions on image processing, 30, 2155-2167.

13. Xu, M., Wang, Z., Liu, X., Ma, L., & Shehzad, A. (2022). An efficient pedestrian detection for realtime surveillance systems based on modified YOLOv3. IEEE Journal of Radio Frequency Identification, 6, 972-976.

14. Le, D. T., Shi, H., Rezatofighi, H., & Cai, J. (2022). Accurate and real-time 3D pedestrian detection using an efficient attentive pillar network. IEEE Robotics and Automation Letters, 8(2), 1159-1166.

15. Zhang, H., Pop, D. O., Rogozan, A., & Bensrhair, A. (2021). Accelerate high resolution image pedestrian detection with non-pedestrian area estimation. IEEE Access, 9, 8625-8636.