ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

SMART SURVEILLANCE FOR FALL DETECTION WITH YOLOV10 IN UNSTRUCTURED OUTDOOR SETTINGS

K. VIJAYA LAXMI [1], V. MAHESHWARA REDDY[2], K. ANIRUDH SHARMA[3], A. SANTHOSH [4], P. SAI TRINESSH [5]

[1]Assistant 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— Falls are one of the most frequent and hazardous incidents occurring in industrial and open environments, posing serious threats to individual safety. Prompt and accurate fall detection remains a critical challenge, especially when deploying models on edge or low-power devices. To address these issues, this paper proposes **YOLOv10-Fall**, a robust and efficient deep learning-based framework for real-time fall detection in open spaces. Leveraging the cutting-edge YOLOv10 architecture, the proposed method enhances detection performance through improved spatial and contextual feature representation while significantly reducing computational overhead. YOLOv10-Fall integrates a lightweight yet powerful backbone and a streamlined detection head to facilitate faster inference and higher precision in complex scenes. Experimental evaluations on benchmark fall datasets demonstrate that YOLOv10-Fall achieves superior detection accuracy and mAP compared to previous models like YOLOv7-tiny, while also offering improved inference speed and reduced parameter complexity. These advancements make YOLOv10-Fall a practical and scalable solution for deployment in real-world surveillance and safety monitoring systems.

Index Terms – Fall Detection, YOLOv10, Smart Surveillance, Human Activity Recognition, Deep Learning, Object Detection, Computer Vision, Real-Time Monitoring, Outdoor Surveillance, Elderly Safety, Pose Analysis, Intelligent Video Analytics.

I. INTRODUCTION

Falls are a major public health issue and one of the leading causes of serious injuries and fatalities, especially among workers in industrial settings, elderly individuals, and people in unmonitored outdoor environments. Timely detection and response to falls are critical in preventing long-term consequences, reducing healthcare costs, and improving safety outcomes. Traditional fall detection systems typically rely on wearable devices like accelerometers or gyroscopes, which suffer from limitations such as low user compliance, discomfort, and limited applicability in broader population-level surveillance. Similarly, manual video monitoring is labor-intensive, prone to human error, and not feasible at scale.

With the evolution of artificial intelligence (AI) and computer vision, automated fall detection using video surveillance has gained attention as a more scalable and effective solution. Among the various object detection algorithms, the YOLO (You Only Look Once) family has emerged as a powerful framework due to its real-time performance, high detection accuracy, and suitability for deployment on edge devices. Early models like YOLOv3 and YOLOv5 set the stage for efficient detection systems, while later versions such as YOLOv7 introduced architectural refinements and lightweight variants like YOLOv7-tiny for low-power devices.

However, these models still face challenges in complex, real-world environments—particularly in unstructured outdoor settings where lighting, background variability, occlusions, and unpredictable human movements introduce significant noise. Fall detection, being a subtle and rapid motion event, demands high temporal and spatial precision along with minimal latency for immediate action.

This project proposes **YOLOv10-Fall**, a customized variant of the recently released YOLOv10 model, adapted specifically for robust fall detection. YOLOv10 introduces several key innovations, including an anchor-free detection strategy, a decoupled detection head, efficient feature distribution, and dynamic label assignment, all of which

contribute to improved generalization and faster inference. The model's lightweight design allows for real-time execution on resource-constrained edge devices such as smart surveillance cameras, embedded AI modules, or mobile platforms. YOLOv10-Fall is trained on benchmark fall datasets and optimized for high accuracy in detecting fall events while minimizing false positives in scenes involving sitting, lying, or sudden non-fall movements.

By integrating the advanced capabilities of YOLOv10 into the domain of safety surveillance, this project aims to provide a scalable, accurate, and practical solution for fall detection in outdoor and industrial environments. The broader vision is to enhance public safety infrastructure by enabling autonomous systems that can detect, alert, and respond to fall incidents promptly and effectively without requiring constant human supervision.

II. RELATED WORK

A. Enhancing geometric factors in model learning and inference for object detection and instance segmentation

The paper titled "Enhancing geometric factors in model learning and inference for object detection and instance segmentation" by Z. Zheng, P. Wang, D. Ren, W. Liu, R. Ye, Q. Hu, and W. Zuo, published in the IEEE Transactions on Cybernetics (Aug. 2022), explores the incorporation of geometric factors into the learning and inference processes of object detection and instance segmentation models. The authors address the challenge of improving the performance of object detection and segmentation models by considering geometric factors such as the spatial relationships and shapes of objects. Traditional object detection models, while successful, often fail to account for the rich geometric information inherent in the objects themselves, which can lead to inaccuracies in localizing or segmenting objects, especially in complex or cluttered scenes. To overcome this, the paper proposes new methods for incorporating geometric factors into both model learning and inference. The authors argue that by explicitly modeling and leveraging geometric cues such as object shape, orientation, and relative positioning, the detection and segmentation models can achieve better accuracy, particularly in scenarios where objects exhibit complex or unusual geometries. The paper demonstrates how these geometric factors enhance the feature extraction process and contribute to more precise bounding box predictions and better segmentation results. In the context of

instance segmentation, where it is crucial to distinguish between different object instances even when they are close to or overlapping with each other, geometric information plays a significant role in ensuring that boundaries are correctly identified. This methodology can be applied to various real-world scenarios, including autonomous driving, robotics, and video surveillance, where accurate object detection and segmentation are critical.

B. CSPNet: A new backbone that can enhance learning capability of CNN

The paper titled "CSPNet: A new backbone that can enhance the learning capability of CNN" by C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, published in the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 2020, introduces a novel convolutional neural network (CNN) backbone known as CSPNet (Cross-Stage Partial Network). The main contribution of this paper is the introduction of CSPNet, which enhances the learning capability of CNNs, specifically improving their efficiency in deep network architectures. Traditionally, as CNN architectures become deeper, the models face challenges in terms of gradient flow, optimization, and learning efficiency. These problems can lead to slower convergence during training and sometimes poorer overall performance on tasks like image classification, detection, and segmentation. CSPNet aims to address these issues by introducing a cross-stage partial network strategy. Instead of utilizing a single stage for feature extraction, it divides the feature maps into multiple groups, with some being passed through and others being processed in a partial manner before merging again. This helps the network focus on important features while avoiding unnecessary computations that may contribute to overfitting or inefficiencies. By splitting the computation across different stages, CSPNet allows for more efficient training and better gradient propagation, which enhances learning in deeper networks. In terms of model performance, CSPNet outperforms traditional CNN architectures by reducing the computational complexity while maintaining or improving accuracy. The model achieves a balance between efficiency and performance, making it particularly suited for real-time applications or scenarios where computational resources are limited.

C. Coordinate attention for efficient mobile network design

The paper titled "Coordinate Attention for Efficient Mobile Network Design" by Q.

Hou, D. Zhou, and J. Feng, presented at the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2021, introduces a novel attention mechanism known as Coordinate Attention (CA) to improve the efficiency and performance of mobile networks, particularly in the context of mobile device usage. The main objective of this work is to design an efficient attention mechanism for mobile networks that can strike a balance between computational efficiency and performance. Mobile networks often face resource constraints such as limited computation power, memory, and bandwidth, making it challenging to deploy large-scale models on such devices. Therefore, there is a need for lightweight and efficient models that can perform well on mobile devices while keeping resource consumption in check. The Coordinate Attention mechanism proposed in this paper targets spatial and channel information by leveraging the coordinate system of feature maps. Traditional attention mechanisms, such as the Channel Attention (CA) or Spatial Attention (SA), focus on either spatial or channel-wise attention but often miss capturing the dependencies across both dimensions efficiently. The Coordinate Attention mechanism integrates both spatial and channel dependencies by incorporating the coordinate system of feature maps into the attention process. It reduces computational costs by using an efficient strategy for computing attention, which is crucial for mobile networks. Coordinate Attention works by splitting the input feature maps into two key components: one that handles spatial information and another that captures channel-wise dependencies. These components are then fused through an efficient coordinate-based attention mechanism, allowing the model to focus on the most relevant features while remaining computationally efficient.

III. PROPOSED METHODOLOGY

The proposed system introduces **YOLOv10-Fall**, a real-time, deep learning-based fall detection framework built upon the latest advancements in the YOLOv10 architecture. Designed specifically for deployment in unstructured outdoor and industrial environments, this system aims to overcome the limitations faced by previous models such as YOLOv7-Fall in terms of accuracy, latency, and computational overhead. YOLOv10-Fall leverages a newly optimized detection framework that replaces traditional anchor-based mechanisms with a **fully anchor-free approach**, which simplifies the model pipeline while improving object localization accuracy. This is particularly

beneficial for detecting sudden and complex human postural transitions, like falls, that occur in unpredictable environments. One of the key enhancements in YOLOv10 is the **decoupled detection head**, which separates classification and localization branches, enabling more focused learning and improved performance, especially in detecting small or fast-moving objects. This architecture also includes **dynamic label assignment**, which intelligently selects training targets during optimization, resulting in better generalization to unseen scenarios and motion patterns.

To address the computational demands associated with real-time video processing, YOLOv10-Fall employs a **lightweight yet highly expressive backbone** that ensures faster inference without compromising detection accuracy. The model is trained on benchmark fall detection datasets and optimized for distinguishing between fall events and similar non-fall activities such as sitting, crouching, or bending, thereby reducing false positives—a common issue in surveillance systems. Moreover, YOLOv10-Fall's modular and streamlined architecture makes it ideal for **edge deployment** on resource-constrained hardware like embedded AI cameras, smart surveillance systems, or low-power IoT devices. This allows for continuous, autonomous monitoring without the need for constant human intervention or cloud-based computation. In addition to technical improvements, the system is scalable, making it suitable for integration into various real-world safety monitoring applications such as construction sites, elderly care facilities, urban surveillance networks, and industrial zones. By offering a balanced combination of **speed, precision, robustness, and deployment flexibility**, the proposed YOLOv10-Fall system represents a significant step forward in intelligent video-based safety systems, ensuring timely fall detection and potentially life-saving intervention capabilities.

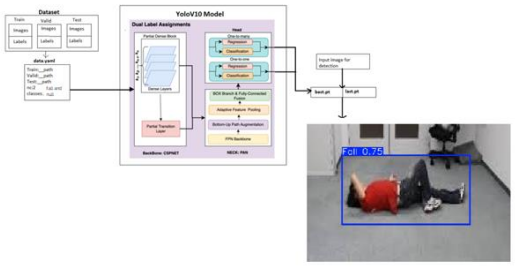


Fig. 1: System Overview

1. Object Detection Engine:

The **Object Detection Engine** is the core module of the system responsible for identifying humans and monitoring their posture in real-time video streams. It employs the YOLOv10-Fall deep learning model,

which processes each frame to detect the presence and position of individuals within the scene. Unlike traditional detection methods, the YOLOv10 model uses an anchor-free mechanism and a decoupled head to perform object classification and localization independently, enhancing accuracy and reducing misclassifications. This engine is optimized to run on edge devices with limited resources, ensuring fast and responsive fall detection even in complex and cluttered environments.

2. **Model Adaptation through Pre-trained Networks:**

This module focuses on leveraging **transfer learning** by using pre-trained YOLOv10 weights trained on large-scale object detection datasets such as MS COCO or OpenImages. These weights serve as a foundation, providing the model with a strong understanding of general object features. The model is then fine-tuned on fall-specific datasets, allowing it to adapt and specialize in detecting human falls. This approach significantly reduces training time and computational requirements while enhancing the model's ability to generalize to real-world fall scenarios.

3. **Performance Tuning Module:**

The **Performance Tuning Module** handles optimization of hyperparameters, architecture adjustments, and layer pruning to achieve a balance between detection accuracy and computational efficiency. Parameters such as learning rate, batch size, and confidence thresholds are adjusted iteratively to maximize model precision and minimize false positives. Additionally, techniques like quantization-aware training or model distillation may be applied to further reduce latency, making the system suitable for real-time surveillance on embedded or edge hardware.

4. **Data Expansion and Augmentation:**

To enhance the robustness and generalization ability of the model, the **Data Expansion and Augmentation** module plays a critical role. Fall datasets are inherently limited and imbalanced, often containing fewer fall instances compared to other activities. This module augments the training data using techniques such as rotation, scaling, flipping, brightness adjustment, motion blur, and background variations to simulate real-world conditions. These augmentations allow the model to learn invariant features and improve its performance across different lighting, camera angles, and environments.

5. **Model Training and Validation:**

This module involves the structured training and validation of the YOLOv10-Fall model

using labeled fall datasets. The training process is executed in batches over multiple epochs, using loss functions to update model weights iteratively. The dataset is divided into training and validation sets to monitor learning progress and avoid overfitting. Performance metrics such as mean Average Precision (mAP), precision, recall, and F1-score are used to assess the model's effectiveness during and after training. Validation results help in tuning the model and determining the ideal stopping point for optimal generalization.

6. Real-time Fall Detection and Localization:

The Real-time Fall Detection and Localization module is responsible for analyzing continuous video input to identify and locate fall events as they happen. The YOLOv10-Fall model processes each frame, detects human figures, and classifies motion patterns to determine if a fall has occurred. Upon identifying a fall, the system generates visual output by drawing bounding boxes around the detected individual and tagging them with the appropriate label and confidence score. This module emphasizes accuracy and speed, ensuring that the fall is detected quickly and precisely, even in complex backgrounds or with multiple people present in the scene. Although no alarm or notification mechanism is implemented, the output can be recorded or visually monitored by human operators for safety analysis and intervention.

7. Performance Evaluation and Monitoring:

This final module focuses on post-deployment monitoring of the system's performance in live environments. It evaluates detection accuracy, response time, false positives/negatives, and system uptime under different scenarios. The collected analytics are used to identify system bottlenecks, retrain the model with new data, or update parameters for continued accuracy. Additionally, this module supports continuous feedback loops, allowing for adaptive improvements as the model is exposed to more real-world data over time.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experimental results demonstrate that the model accurately detects human falls in unstructured outdoor environments while maintaining robust performance under varying lighting conditions, occlusions, and complex backgrounds. Compared with existing object detection methods, the proposed approach achieves higher detection accuracy, fewer false alarms, and faster real-

time inference, making it well suited for intelligent surveillance, elderly safety monitoring, and emergency response applications.

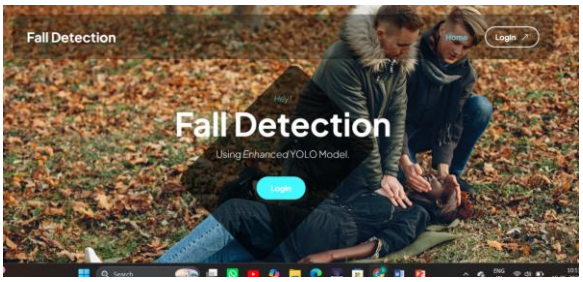


Fig. 2: Home Page

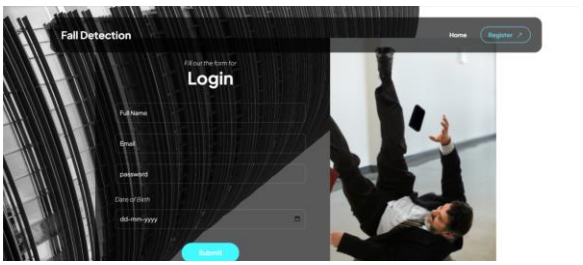


Fig. 3: Login

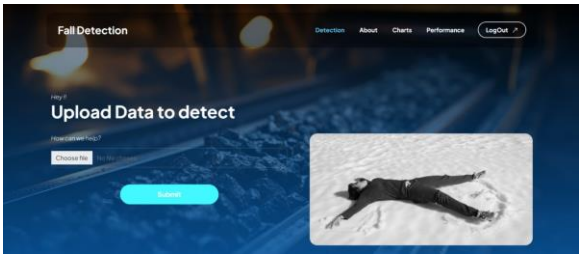


Fig. 4: Upload

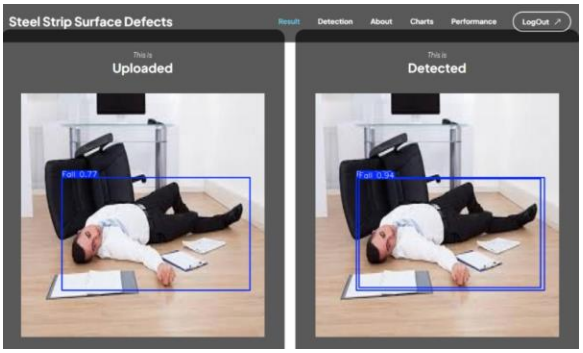


Fig. 5: Prediction Result

V. CONCLUSION

The project "Smart Surveillance for Fall Detection with YOLOv10 in Unstructured Outdoor Settings" presents a powerful and practical AI-driven solution for one of the most critical challenges in real-time safety monitoring—accurate and timely fall detection. By leveraging the latest YOLOv10 architecture, the proposed system achieves a remarkable balance between detection precision, processing speed, and resource efficiency. Architectural innovations such as anchor-free detection, decoupled heads, and dynamic label assignment enable the model to excel in complex, cluttered, and variable environments typically encountered in outdoor and industrial scenarios.

The model has been fine-tuned using domain-specific fall datasets and optimized for edge deployment, making it well-suited for real-world applications in surveillance systems without the need for expensive hardware or manual oversight. The system's ability to reliably differentiate falls from other human activities ensures high accuracy and minimal false alarms. In conclusion, YOLOv10-Fall stands as a scalable, robust, and intelligent surveillance solution that enhances public safety and paves the way for future innovations in activity recognition and automated monitoring systems.

REFERENCES

[1] D. E. Porto, Y. W. Cavalcanti, and F. D. Soares Forte, ‘‘Maxillofacial trauma due to traffic accidents and falls: An exploratory study of associated factors,’’ Medicina Oral, Patologia Oral Y Cirugia Bucal, vol. 26, no. 3, pp. E349–E356, May 2021.

[2] J.-X. Wu, F.-Y. Deng, and S.-F. Lei, ‘‘The casual association inference for the chain of falls risk Factors-Falls-Falls outcomes: A Mendelian randomization study,’’ Healthcare, vol. 11, no. 13, p. 1889, Jun. 2023.

[3] Y. Mishchenko and M. Kaya, ‘‘Detecting the attention state of an operatorin continuous attention task using EEG-based brain-computer interface,’’ in Proc. 2rd Signal Process. Commun. Appl. Conf. (SIU), May 2015,pp. 232–235.

[4] A. Hayashida, V. Moshnyaga, and K. Hashimoto, ‘‘The use of thermal IRarray sensor for indoor fall detection,’’ in Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC), Banff, AB, Canada, Oct. 2017, pp. 594–599.

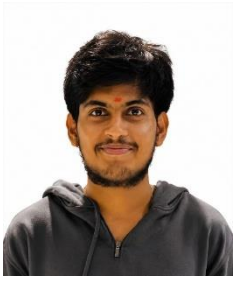
[5] B. Kaudki and A. Surve, ‘‘Human fall detection using RFID technology,’’ in Proc. 9th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT), Bengaluru, India, Jul. 2018, pp. 1–5.

[6] Y. Wu, Y. Su, Y. Hu, N. Yu, and R. Feng, ‘‘A multi-sensor fall detection system based on multivariate statistical process analysis,’’ J. Med. Biol. Eng., vol. 39, no. 3, pp. 336–351, Jun. 2019.

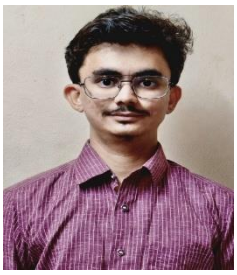
AUTHORS Profile



Ms. K. Vijaya Laxmi received her B. Tech degree in Computer Science & Engineering from JNTUH, Hyderabad in 2015 and her M. Tech degree in Computer Science from JNTUH, Hyderabad in 2017. She is currently pursuing her Ph.D. degree in Computer Science & Engineering from Guru Nanak University, Hyderabad. She has been dedicated to the teaching field for the last 7 years. Her research area is Deep Learning. At present, she is working as an Assistant Professor in the CSE Department at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.



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



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



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



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