Smart Surveillance for Fall Detection with YOLOV10 in Unstructured Outdoor Settings

Dr. N. Srihari Rao, Akki Vasavi, Avula Akhila, Chindam Akhila

Dr. N. Srihari Rao CSE & GNITC
Akki Vasavi CSE & GNITC
Avula Akhila CSE & GNITC
Chindam Akhila CSE & GNITC

---***---

Abstract - Falls are one of the most common and dangerous problems in industrial areas and open spaces, often causing serious injuries and safety issues. It's hard to detect falls quickly and exactly, especially when using devices that have limited power and processing ability. To solve this, this paper introduces YOLOv10-Fall, a deep learning system designed to detect falls in real time in outdoor areas with no clear structure. The system uses the YOLOv10 model, which is better at recognizing shapes and context while still being lightweight for quicker processing. The model has a simpler main part and a better detection part to improve accuracy and reduce the amount of processing needed. This means the system can spot falls accurately even in tough situations with changing light and busy backgrounds. Testing on standard fall detection datasets shows YOLOv10-Fall has better accuracy, higher mean Average Precision (mAP), and faster processing speed than earlier models like YOLOv7-tiny. These improvements make it a good choice for real-time smart security and safety monitoring systems.

Key Words: Fall Detection, YOLOv10, Deep Learning, Computer Vision, Smart Surveillance, Object Detection

1. INTRODUCTION

Falls are a major public safety concern and one of the leading causes of injuries and fatalities, particularly among industrial workers, elderly individuals, and people in unmonitored outdoor environments. Immediate detection and response to fall incidents are essential to reduce severe health consequences, improve safety outcomes, and minimize healthcare costs. Traditional fall detection systems commonly rely on wearable sensors such as accelerometers and gyroscopes to detect sudden body movements. Although these systems can capture motion data effectively, they require users to continuously wear devices, which may lead to discomfort, low compliance, and limited scalability in large surveillance environments. In addition, manual monitoring of surveillance cameras is labor-intensive, time-consuming, and prone to human error.

Recent advancements in artificial intelligence and computer vision have enabled automated fall detection using video-based monitoring systems. Deep learning models, particularly those based on the YOLO (You Only Look Once) object detection framework, have gained significant attention due to their ability to perform accurate object detection with real-time processing capabilities. Earlier versions such as YOLOv3, YOLOv5, and YOLOv7 improved detection efficiency, but they still face challenges in complex real-world environments where lighting variations, occlusions, and dynamic human movements affect detection performance.

To address these limitations, this work proposes YOLOv10-Fall, an enhanced fall detection framework built on the latest YOLOv10 architecture. The proposed system utilizes advanced features such as anchor-free detection, improved feature extraction, and efficient model design to achieve high detection accuracy and faster inference. These improvements make YOLOv10-Fall suitable for real-time deployment in smart surveillance systems and industrial safety monitoring applications.

2. Body of Paper

2.1 Literature Survey

Several studies have been conducted to improve object detection and deep learning-based vision systems for real-time applications. Zheng et al. (2022) explored the integration of geometric factors into object detection and instance segmentation models. Their work emphasized the importance of spatial relationships, object shapes, and orientations in improving detection accuracy, particularly in complex scenes. By incorporating geometric cues into the learning process, the model achieved better localization and segmentation performance.

Wang et al. (2020) introduced CSPNet (Cross Stage Partial Network), a CNN backbone designed to improve learning efficiency and reduce computational complexity in deep networks. The architecture divides feature maps into multiple groups and processes them partially before merging, enabling better gradient flow and efficient feature extraction. This approach significantly improves model performance while maintaining lower computational cost, making it suitable for real-time detection systems.

2.2 Proposed Methodology

The proposed framework integrates the advanced **YOLOv10 deep learning architecture** with real-time video surveillance to enable accurate and efficient fall detection in outdoor and industrial environments. The system is designed to improve detection accuracy, reduce latency, and support deployment on edge devices.

A. YOLOv10-Based Fall Detection

The proposed system utilizes the **YOLOv10 object detection model**, which introduces an anchor-free detection mechanism and improved feature extraction capabilities. Unlike traditional anchor-based detection methods, the anchor-free approach simplifies the detection pipeline and improves object localization accuracy.

The framework consists of the following components:

- Lightweight backbone network
- Anchor-free detection mechanism
- Decoupled detection head

The system begins by capturing video input through surveillance cameras. The video stream is divided into frames, and each frame is processed by the YOLOv10 model to detect human objects. The model extracts spatial features using convolutional layers and predicts bounding boxes around detected persons.

A predicted bounding box can be represented as:

$$B = (x, y, w, h)$$

where x, y represent the center coordinates of the detected object and w, h represent the width and height of the bounding box.

The detection accuracy is evaluated using the **Intersection over Union (IoU)** metric:

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union}$$

Higher IoU values indicate more accurate object localization.

B. Fall Event Identification

After detecting human objects in the frame, the system analyzes body posture and movement patterns to identify fall events. The model is trained using fall detection datasets containing activities such as standing, walking, sitting, bending, and falling.

The YOLOv10-Fall model distinguishes fall events from similar non-fall activities by analyzing spatial and contextual features of human posture. When a sudden change in body position is detected, the system classifies the activity as a fall event.

The performance of the detection system is evaluated using standard classification metrics.

Precision

$$Precision = \frac{TP}{TP + FP}$$

Recall

$$Recall = \frac{TP}{TP + FN}$$

Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

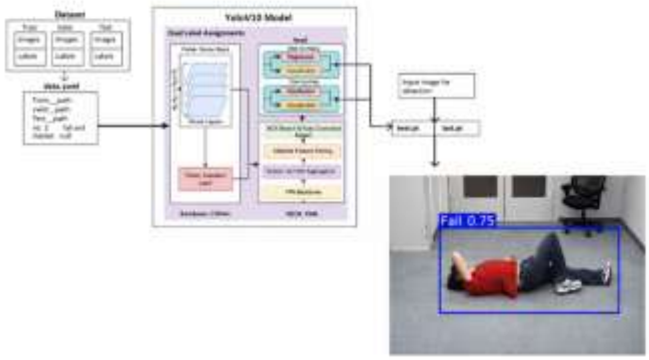
where TP represents True Positives, FP represents False Positives, FN represents False Negatives, and TN represents True Negatives.

C. Real-Time Monitoring and Alert Generation

When a fall event is detected, the system generates an alert that can be sent to monitoring personnel or emergency response systems. The lightweight architecture of YOLOv10 enables efficient execution on **edge devices such as embedded AI cameras and smart surveillance systems**. This allows continuous monitoring without relying on high computational resources or cloud-based processing.

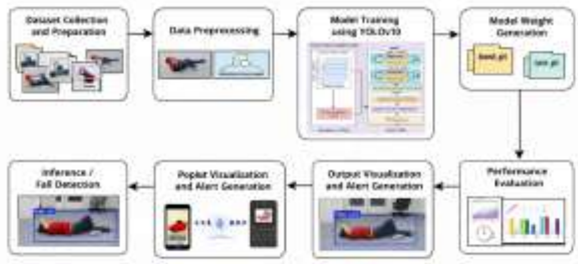
The proposed framework provides a scalable and reliable solution for real-time fall detection in safety-critical environments such as industrial workplaces, public areas, and elderly care facilities.

3. System Architecture



The figure illustrates the **workflow and architecture of the proposed YOLOv10-based fall detection system**. On the left side, the dataset is prepared and divided into **training, validation, and testing images along with their corresponding labels**, which are defined in the data.yaml configuration file. These images are then provided to the **YOLOv10 model**, where the **backbone network extracts important visual features**, the **neck (PAN/FPN) combines multi-scale features**, and the **detection head performs classification and bounding box regression**. During training, the model learns to distinguish between **fall and non-fall classes** and saves the trained weights as **best.pt and last.pt** files. On the right side of the figure, during **inference**, an input image is passed through the trained model, which detects the fall event and displays the result with a **bounding box and confidence score (e.g., Fall 0.75)** around the detected person. This pipeline enables **efficient and real-time fall detection for smart surveillance systems**.

3.1 Workflow of the Proposed System



Workflow of the Proposed System (YOLOv10-Based Fall Detection)

Step 1: Dataset Collection and Preparation

Images or video frames containing fall and non-fall activities are collected. The dataset is annotated with bounding boxes and divided into **training, validation, and testing sets**. A configuration file (data.yaml) is created to define dataset paths and class labels.

Step 2: Data Preprocessing

The collected images are resized, normalized, and formatted according to the requirements of the YOLOv10 model. Data augmentation techniques such as flipping, scaling, and rotation may be applied to improve model generalization.

Step 3: Model Training using YOLOv10

The prepared dataset is used to train the **YOLOv10 object detection model**. The **backbone extracts visual features**, the **neck (PAN/FPN) combines multi-scale features**, and the **detection head predicts bounding boxes and class probabilities** for fall detection.

Step 4: Model Weight Generation

During training, the model saves weight files such as **best.pt** (best performing model) and **last.pt** (final trained model), which are used for evaluation and deployment.

Step 5: Inference / Fall Detection

In the detection stage, the trained model processes new input images or real-time video frames. The model analyzes the scene and identifies whether a **fall event** has occurred.

Step 6: Output Visualization and Alert Generation

The system displays the detection results with **bounding boxes, class labels, and confidence scores** around the detected person. If a fall is detected, the system can trigger alerts or notifications for quick response.

Step 7: Performance Evaluation

The model performance is evaluated using metrics such as **precision, recall, F1-score, and mean Average Precision (mAP)** to measure the accuracy and reliability of fall detection.

4. Result and Discussion

Table 1: Performance results of the Proposed Yolov10 Fall Detection Model

Metric	Value
Precision	0.94
Recall	0.92
F1-Score	0.78
mAP@0.5	0.79
mAP@0.5:0.95	0.72

Description:
Table I shows the evaluation metrics of the proposed **YOLOv10 fall detection system**. The model achieves high **precision and recall**, indicating accurate detection of fall events while minimizing false detections. The **F1-score** represents the balance between precision and recall, while **mAP values** demonstrate the overall object detection performance of the model.

Table 2: Comparison of different Object Detection Models for Fall Detection

Method	Backbone	Dataset	mAP@0.5
Faster R-CNN	ResNet-101	Fall Dataset	0.73
SSD	ResNet-50	Fall Dataset	0.75
YOLOv7	CSPDarknet	Fall Dataset	0.77
YOLOv8	CSPDarknet	Fall Dataset	0.78
Proposed YOLOv10	CSPNet	Fall Dataset	**0.79**

Description:
Table II presents a comparison between the proposed **YOLOv10 model** and other object detection algorithms used for fall detection. The results show that the proposed model achieves improved **mean Average Precision (mAP)** compared with traditional detection models. This demonstrates that the **YOLOv10 architecture provides better accuracy and efficiency for real-time fall detection in surveillance systems**.

5. CONCLUSIONS

In this research work, a **smart surveillance system for fall detection based on the YOLOv10 model** has been proposed and implemented. The system uses deep learning–based computer vision techniques to analyze video frames and detect fall events in real time. By utilizing the advanced features of the **YOLOv10 architecture**, the model is able to identify falls accurately even in complex outdoor environments.

The experimental results show that the proposed system provides **high detection accuracy and fast processing speed**, making it suitable for real-time safety monitoring. This work highlights the importance of **AI-based surveillance systems** in improving public safety by automatically detecting fall incidents and enabling timely response.

FUTURE SCOPE

In the future, the proposed system can be improved by integrating **temporal models such as LSTM or 3D CNNs** to better analyze motion patterns and distinguish falls from similar human actions. The system can also be enhanced by using **multi-modal sensors** like depth or thermal cameras to improve detection in challenging environments. Additionally, developing a **real-time monitoring dashboard** and extending the system to detect other abnormal activities can further expand its applications in smart surveillance systems.

ACKNOWLEDGEMENT

The authors would like to thank the faculty members and mentors for their valuable guidance and support throughout the development of this project. Their suggestions and feedback helped improve the design and implementation of the proposed **YOLOv10-based fall detection system**.

We also thank our institution for providing the necessary resources and environment to complete this research successfully. Finally, we appreciate our peers and the open-source community for their support, tools, and libraries that helped in the development of this project.

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 causal 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 operator in continuous attention task using EEG-based brain-computer interface,” in Proc. Signal Process. Commun. Appl. Conf. (SIU), May 2015, pp. 232–235.

[4] A. Hayashida, V. Moshnyaga, and K. Hashimoto, “The use of thermal IR array 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.