



Article

FD-YOLO: A YOLO Network Optimized for Fall Detection

Hoseong Hwang ¹, Donghyun Kim ¹ and Hochul Kim ^{1,2,*}

¹ Department of Medical Artificial Intelligent, Eulji University, Seongnam-si 13135, Gyeonggi-do, Republic of Korea; vole91@gmail.com (H.H.); kdh11191224@gmail.com (D.K.)

² Department of Radiological Science, Eulji University, Seongnam-si 13135, Gyeonggi-do, Republic of Korea

* Correspondence: tiger1005@gmail.com

Abstract: Falls are defined by the World Health Organization (WHO) as incidents in which an individual unintentionally falls to the ground or a lower level. Falls represent a serious public health issue, ranking as the second leading cause of death from unintentional injuries, following traffic accidents. While fall prevention is crucial, prompt intervention after a fall is equally necessary. Delayed responses can result in severe complications, reduced recovery potential, and a negative impact on quality of life. This study focuses on detecting fall situations using image-based methods. The fall images utilized in this research were created by combining three open-source datasets to enhance generalization and adaptability across diverse scenarios. Because falls must be detected promptly, the YOLO (You Only Look Once) network, known for its effectiveness in real-time detection, was applied. To better capture the complex body structures and interactions with the floor during a fall, two key techniques were integrated. First, a global attention module (GAM) based on the Convolutional Block Attention Module (CBAM) was employed to improve detection performance. Second, a Transformer-based Swin Transformer module was added to effectively learn global spatial information and enable a more detailed analysis of body movements. This study prioritized minimizing missed fall detections (false negatives, FN) as the key performance metric, since undetected falls pose greater risks than false detections. The proposed Fall Detection YOLO (FD-YOLO) network, developed by integrating the Swin Transformer and GAM into YOLOv9, achieved a high mAP@0.5 score of 0.982 and recorded only 134 missed fall incidents, demonstrating optimal performance. When implemented in environments equipped with standard camera systems, the proposed FD-YOLO network is expected to enable real-time fall detection and prompt post-fall responses. This technology has the potential to significantly improve public health and safety by preventing fall-related injuries and facilitating rapid interventions.

Keywords: falls; artificial intelligence; computer vision; attention block; deep learning



Academic Editor: Pedro Couto

Received: 8 November 2024

Revised: 1 January 2025

Accepted: 3 January 2025

Published: 6 January 2025

Citation: Hwang, H.; Kim, D.; Kim, H. FD-YOLO: A YOLO Network Optimized for Fall Detection. *Appl. Sci.* **2025**, *15*, 453. <https://doi.org/10.3390/app15010453>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Fall accidents, as defined by the World Health Organization (WHO), occur when an individual unintentionally moves to the ground or a lower level. These incidents are recognized as significant public health issues rather than isolated accidental events. Globally, falls result in over 684,000 fatalities annually, ranking as the second leading cause of unintentional injury-related deaths after road traffic accidents [1]. They account for more than 90% of all hip fractures, with nearly 20% of affected individuals succumbing to related complications within a year [2]. These figures highlight the fatal risks posed by falls, particularly to the elderly, and the substantial medical burden they impose on

healthcare systems. Annually, approximately one-third of individuals aged 65 years and older experience a fall, accounting for more than 50% of hospital admissions for unintentional injuries in this demographic [1,3]. These data underscore the extensive societal burden. It is anticipated that the incidence of falls will increase alongside the global aging trend. By 2050, the population aged 65 and older is projected to reach approximately 2 billion, constituting approximately 22% of the global population; consequently, injury and mortality rates associated with falls are expected to rise significantly in aging societies, underscoring the pressing need for preventative technological interventions.

Moreover, the rise in single-person households has reduced the likelihood of receiving immediate assistance after a fall, particularly among older adults. This delay can lead to the “long-lie” phenomenon, where individuals remain immobilized for over an hour, significantly increasing secondary risks such as pressure ulcers, dehydration, hypothermia, and mortality. Thus, there is an escalating demand for accurate fall detection systems capable of facilitating timely interventions [2].

Early fall detection systems predominantly utilized wearable sensors, such as accelerometers and gyroscopes, integrated into personal emergency response systems (PERSs) [4–6]. These sensors effectively monitored body movements but had critical limitations, including the need for precise placement on the body, which could shift during movement, compromising accuracy. To address these challenges, recent advancements focus on integrating sensor data with artificial intelligence (AI) to improve calibration and analysis [7–9]. AI offers the potential to significantly improve fall detection accuracy by compensating for sensor data imperfections and learning movement patterns to obtain more reliable fall detection. Current studies aim to enhance the accuracy of sensor-based fall detection through advanced data preprocessing and algorithm optimization while striving to develop cost-effective equipment.

Nonetheless, a fundamental limitation of these systems was the need to attach external sensors to the body. Although this method might be effective for high-risk individuals such as those predisposed to frequent falls, it had notable limitations when attempting to detect sudden fall incidents in the general population. Moreover, although sensor-based systems can detect falls, they relied on wireless communication to alert others, which introduced additional challenges [10]. For instance, if the communication network malfunctions or if the user was alone during a fall, timely intervention could not occur. Furthermore, in countries such as South Korea, the transmission of bodily data over the Internet could violate privacy protection laws, complicating the possibility of a rapid response. To solve these problems, researchers have increasingly used smartphones as an alternative. This approach involved transmitting sensor data to a smartphone, where AI processed the information and sent alerts externally if necessary [11]. Although this method addressed some of the communication challenges associated with traditional systems, it did not resolve the issue of physically attaching sensors to the body. Consequently, there was an increasing need to develop novel technological approaches to address the limitations inherent in sensor-based systems. To overcome the limitations of sensor-based fall detection systems, research on contactless vision-based fall detection had been actively pursued. The rapid advancements in computer vision had significantly enhanced the accuracy and applicability of such systems. Past computer vision techniques for fall detection had been grounded in mathematical models that captured and analyzed body movements and the spatial relationship between the body and its environment. These approaches included methods for analyzing changes in body posture [11,12], tracking head position [13], and using three-dimensional (3D) technology to assess vertical body distribution [14]. However, despite their potential, these methods failed to capture specific fall movements or exhibit reduced accuracy under certain environmental conditions. To address these limitations, researchers

had explored the use of Kinect cameras, which could provide depth information to detect falls more accurately [15]. The depth-sensing capability of the Kinect camera offered the advantage of improved detection of interactions between the body and the environment, enhancing the precision of fall detection. Nonetheless, this approach presented practical challenges, such as the need for additional hardware installation, high costs, and limitations in the range of accurate depth measurement, which made real-world applications difficult.

Similar to the trajectory of sensor-based fall detection research, vision-based studies that increasingly incorporated AI algorithms in the vision base were developed to mitigate these limitations and improve accuracy. These algorithms analyzed complex body movements with enhanced precision, and deep learning techniques demonstrated strong capabilities for processing unstructured data effectively. However, most AI algorithms for vision-based fall detection were developed for cloud-based platforms [16,17] or integrated into embedded systems [18,19]. Although cloud-based systems provided high computational power and accuracy, they also incurred additional costs and introduced challenges such as transmission delays and significant concerns regarding data privacy. Conversely, leveraging existing closed-circuit television (CCTV) systems that were already widely deployed for crime prevention and monitoring presented a practical alternative. By utilizing these systems, the need for new hardware installation was mitigated, thereby expanding the potential applicability of AI vision-based fall detection systems in real-world environments. This approach substantially reduced costs and simplified implementation while still allowing for the integration of advanced AI techniques to improve detection accuracy.

This study aims to develop an AI network that can be implemented in CCTV systems, which are more cost-effective alternatives to cloud-based platforms, while enhancing fall detection performance. By leveraging advanced neural network techniques, such as attention blocks, this study seeks to address the limitations of existing sensor and vision methods, enabling more accurate and real-time fall detection and transmission. The proposed system not only detects falls in real time but also immediately transmits alerts of fall incidents via CCTV networks, facilitating rapid responses in emergency situations. Consequently, this approach is expected to contribute significantly to reducing fall injuries and protecting life and safety through timely intervention. Furthermore, the development of AI technological advancements will serve as a critical tool for mitigating the risks associated with falls, particularly in the context of an aging society. By improving both the accuracy and efficiency of fall detection, this system has the potential to play a pivotal role in ensuring individual health and safety, promoting swift responses, and ultimately enhancing the quality of life of vulnerable populations.

The primary contributions of the proposed model in this study are delineated as follows:

- **Development of an Optimized Object Detection Network [20]:** Through a comprehensive evaluation of various object detection networks, the study identified and optimized a network tailored specifically for the fall detection dataset. This optimization process led to the development of an efficient and high-performing object detection framework.
- **Incorporation of the Swin Transformer for Enhanced Feature Representation [21]:** The proposed model integrates the Swin Transformer module, which leverages global contextual information, to effectively capture and reinforce critical human positional features. This enhancement addresses the limitations of existing approaches, significantly reducing detection failures associated with fall events.
- **Integration of an Attention Mechanism for Performance Enhancement:** To further augment the model's effectiveness, an attention module was incorporated into the

architecture. This addition resulted in superior performance compared to conventional object detection frameworks, underscoring the efficacy of the proposed approach.

The structure of this manuscript is organized as follows. Section 2 provides a detailed review of the relevant literature and background on deep learning-based object detection methodologies. Section 3 describes the composition and characteristics of the fall detection dataset utilized in this study. Section 4 elaborates on the architectural details of the object detection networks employed, including the proposed modules. Section 5 outlines the experimental methodology, performance metrics, and a comparative analysis of the proposed model against baseline approaches. Section 6 discusses the experimental findings and their implications, while Section 6 concludes the study by summarizing its contributions and proposing directions for future research.

2. Related Works

Recent advancements in deep learning have led to significant progress in algorithms applied to various computer vision tasks, including image classification, image segmentation, and object detection. Object detection involves both the recognition of object categories and the precise localization of these objects within an image. As foundational components of AI-driven computer vision systems, deep learning-based object detection algorithms have attracted substantial research attention. Broadly, these algorithms can be categorized into transformer-based and convolutional neural network (CNN)-based approaches, based on their underlying architectural principles. Transformer-based object detection models, such as the Detection Transformer (DETR) [22], exemplify the capabilities of transformers in visual tasks. Despite their promising performance, these models are often constrained by high computational complexity, substantial memory requirements, and extended training durations, rendering them less feasible for deployment in industrial applications with stringent efficiency requirements. Conversely, CNN-based object detection methods are well-established and are typically divided into two-stage and one-stage frameworks. Two-stage approaches, including Region-based Convolutional Neural Networks (R-CNN) [23], Mask R-CNN [24], and Faster R-CNN [25], are renowned for their high detection accuracy. These methods employ a sequential process, where candidate regions are first proposed, and then boundary box regression and object classification are performed using CNNs. While these algorithms have demonstrated effectiveness in complex detection scenarios, their intricate architectures and extended inference times pose limitations for real-time applications, such as fall detection systems, where a rapid response is critical. Following this, Fast R-CNN employed a multi-task loss function, enabling simultaneous optimization of classification and localization tasks, but still suffered from latency issues. Faster R-CNN further refined the process by introducing a Region Proposal Network (RPN) that shared convolutional features between the detection and proposal stages, streamlining the pipeline. However, the architectural complexity of two-stage models remains a challenge. One-stage object detection algorithms, such as the Single Shot Detector (SSD) [26] and the You Only Look Once (YOLO) series [27], adopt a more efficient paradigm. These models perform direct classification and localization on feature maps derived from input images, circumventing the need for separate region proposal steps. SSD achieves multi-scale object detection by leveraging feature layers of varying depths, balancing speed and accuracy comparable to Faster R-CNN, although it struggles with the detection of small objects. In contrast, the YOLO framework eliminates region proposal entirely, predicting bounding boxes and object categories concurrently across the entire image [27]. While YOLO outperforms many state-of-the-art algorithms in computational efficiency, its detection accuracy is slightly lower than that of two-stage models like Fast R-CNN. Given their reduced computational demands, one-stage algorithms are better suited for real-time applications,

particularly in scenarios requiring rapid processing and lightweight model architectures. Over the years, the YOLO series has emerged as a benchmark for state-of-the-art object detection, finding widespread application in industrial and practical contexts.

3. Materials and Methods

In this study, the data were constructed using three fall datasets [28–30], and the YOLO artificial intelligence network was applied for detection. Notably, recent advancements in the YOLO series have achieved significant improvements in detection accuracy, making these algorithms fast, efficient, highly accurate, deployable, and user-friendly. The continuous evolution of the YOLO series, which has now reached YOLOv11, has incorporated various enhancements, leading to substantial upgrades in network performance, further solidifying the suitability of YOLO for real-time applications.

3.1. Datasets

In this study, three open-source fall detection datasets were used to accurately model fall scenarios. This approach was adopted to prevent the model from becoming biased toward specific environments, addressing the potential issue of overreliance on a particular dataset. By integrating these distinct datasets, collected under various conditions and environments, the model was trained to detect fall scenarios in a more generalized manner.

All datasets were labeled dichotomously, distinguishing between the normal and fall states. However, given that the primary objective of this study was to improve fall detection accuracy, we determined that training the model on normal-state images would not contribute significantly to this goal. Consequently, to minimize potential training errors and optimize model performance, the labeling structure was modified to focus exclusively on a single class, that is, falls. This restructuring aims to enhance the capacity of the model to identify fall incidents with greater precision and reduce misclassification rates, thereby improving the overall detection accuracy.

This study utilized a total of 17,661 images, of which 10,596 were allocated for training, 3532 for validation, and the remaining 3533 for final accuracy evaluation. The details are presented in Table 1. The Roboflow dataset [28] included 10,793 images, Kaggle dataset [29] contained 485 images and the CAUCAFall dataset [30] comprised 6383 images. Figure 1 illustrates examples from these datasets. This division of datasets ensures a balanced approach to model development, enabling robust training, thorough validation, and accurate evaluation of model performance across diverse scenarios.

Table 1. Datasets.

Dataset	Total Numbers	Train	Validation	Test
Roboflow	10,793	6493	2150	2150
Kaggle	485	291	97	97
CAUCAFall	6383	3812	1285	1286
total	17,661	10,596	3532	3533

3.2. YOLOv9

In this study, we used the YOLO network, which is a widely recognized AI object detection framework, for fall detection. Detecting falls in real time is crucial; however, it is equally important to accurately identify the location and number of falls to ensure timely responses. Therefore, we selected the YOLO network, which is known for its advantages in both object localization and real-time processing. After optimizing various versions of the YOLO network, each of which demonstrated continuous performance improvements

on other open-source datasets, we trained the network on our fall detection dataset to determine the best-performing version.

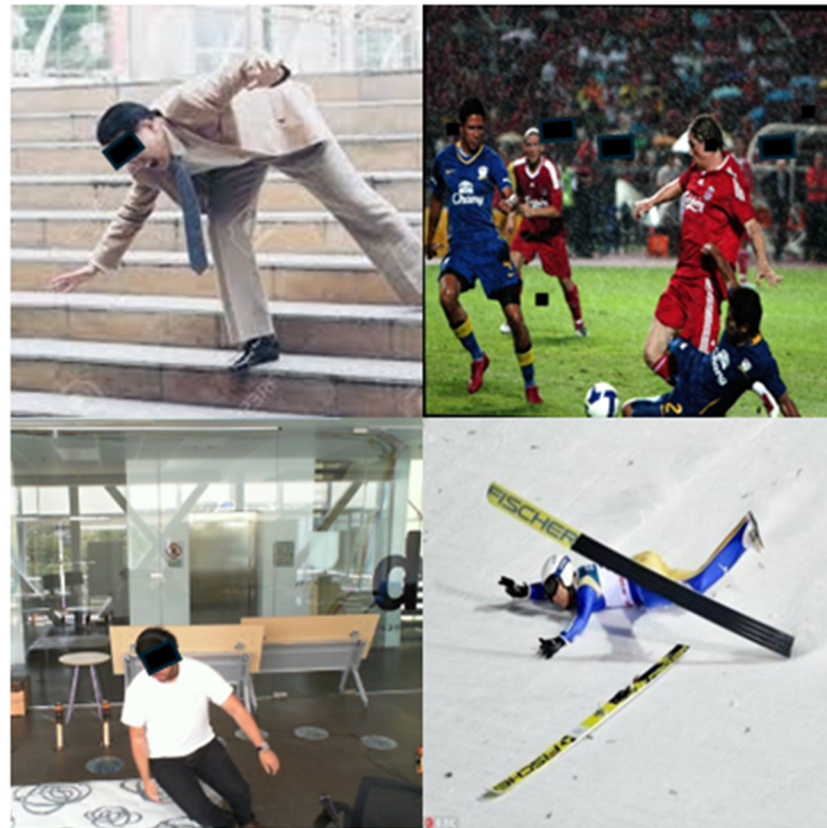


Figure 1. Fall dataset.

We hypothesized that, although a particular network may exhibit strong general performance, its effectiveness could vary depending on the characteristics of the dataset, leading to the expectation that a specific YOLO version would be optimally suited for fall detection. Consequently, we tested all available versions from the earliest to the most recent iterations and selected the network that yielded the highest performance. In this comprehensive evaluation, YOLOv9 delivered the best results. This superior performance can be largely attributed to the integration of the auxiliary stage into its intermediate layers, which helps prevent information loss as it propagates through the network, thus enhancing the test data performance [31]. Additionally, the inclusion of the Repeated Normalized Cross-Stage Partial (RepNCSP) structure, a refinement of the cross-stage partial bottleneck with two convolutions (C2f) structure introduced in YOLOv8, further contributed to improved fall detection accuracy, suggesting that these architectural innovations played a role in optimizing fall detection in our study.

3.3. Attention Module

To enhance the performance of AI algorithms, detection capabilities are often improved by increasing the network depth, integrating multi-scale feature fusion, and adding detection branches. However, increasing the network depth inherently expands the model size, which poses a limitation for methods in which real-time detection is essential. Consequently, techniques such as attention blocks, which exploit the characteristics of feature maps, have been employed to improve accuracy without increasing the number of layers. This approach significantly enhances the performance while requiring fewer computational resources compared with traditional deep networks.

Detecting falls from images requires distinguishing between falling, sitting, and walking. However, the human body's structure makes it difficult to clearly differentiate these actions. For example, bending over can appear similar to falling. In fall scenarios, key body parts like the nose, shoulders, elbows, and ankles are crucial for distinguishing actions [32]. Falls also involve interaction with the floor, making the body's center of gravity and its angle with the floor important factors [33,34]. This information reflects global spatial interactions in videos. In this study, it is crucial to preserve and extract these global features.

A notable technique for improving performance is the convolutional block attention module (CBAM) [35]. CBAM enhances the ability of the model to focus on features by combining channel attention and spatial attention mechanisms. This allows the network to emphasize the most important regions of the input image, effectively improving the detection accuracy without the need for a more complex network architecture. The first step in the CBAM is the channel attention mechanism, which enables the object detection network to learn the relative importance of each channel. By assigning adaptive weights to individual channels, the model emphasizes feature maps that are for object recognition, while diminishing less relevant or redundant information. This selective emphasis improves the overall performance of the model by focusing its resources on the most informative data. The second step involves the spatial attention mechanism, which accentuates the positional information of specific regions within the image. This mechanism allows the network to analyze the spatial layout of an image, highlighting regions that are crucial for object detection. By focusing on the spatial arrangement, the model can more accurately detect objects of varying sizes or positions, especially in complex scenarios where distinguishing between objects and the background is challenging. The spatial attention mechanism enhances the ability of the model to learn detailed positional information, thereby improving object localization and recognition, particularly in cases of significant variability in object placement or appearance.

By prioritizing the processing of more important and relevant information, this approach reduces the likelihood of missed or falsely detected objects and enhances the detection and recognition accuracy by effectively focusing on specific objects. This selective focus is particularly beneficial in challenging scenarios involving diverse backgrounds, small objects, or intricate patterns where traditional detection methods may struggle to distinguish between objects and the surrounding environment [36]. A schematic of this approach is shown in Figure 2.

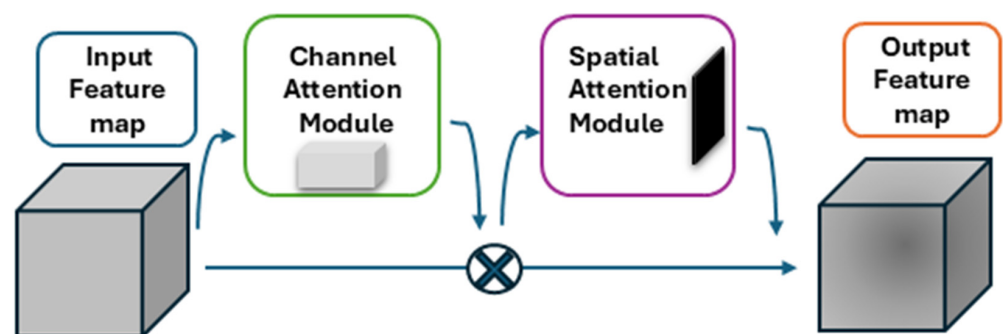


Figure 2. Attention block.

Falls occur in various forms, and effectively extracting global information from images is crucial for emphasizing the relevant features for accurate detection. In some fall scenarios, parts of the body may be occluded by objects in the image, leading to a potential loss of key fall-related features as the layers process the image. To address this limitation, we seek to improve detection accuracy by incorporating a global attention block (GAB) that enhances the existing CBAM structure by adding a global computation mechanism [37]. Similar

to the CBAM, the GAB accounts for both channel and spatial information. However, the GAB goes a step further by considering interactions across the entire feature map when calculating the relationships between different positions, thereby applying both channel and spatial attention on a global scale. This global perspective allows the model to capture more comprehensive contextual information, improving its ability to detect falls, even in cases where critical parts of the body are partially occluded or when fall-related features are dispersed across the image.

CBAM methods often use max or average pooling to extract features, which can result in capturing only specific features of the human body while losing other important details. To solve this, the GAM (Global attention mechanism) uses two convolutional layers to preserve and combine spatial information, reducing the loss of structural details and improving fall detection accuracy. This needs to be validated with experimental results. Pooling methods summarize data by averaging or taking the maximum value, which limits the ability to capture global context but is efficient in terms of parameters. The GAM removes pooling and uses convolutional layers to maintain information and enhance global interactions. While removing pooling may lower parameter efficiency, this was addressed with additional computations. The calculation of the channel attention mechanism is outlined in Equations (1) and (2):

$$M_c(F_1) = \sigma(W_2 * \text{ReLU}(W_1 * \text{Pooling}(F_1))) \quad (1)$$

$$F_2 = M_c(F_1) \otimes F_1 \quad (2)$$

σ is a sigmoid activation function. W_1 and W_2 are the weights matrix. For the input feature map (F_1), the intermediate channel map (F_2) is obtained through element-wise multiplication between the channel attention map (M_c) and the input feature map (F_1). Similarly, the calculation of the spatial attention mechanism is depicted in Equations (3) and (4):

$$M_s(F_2) = \sigma(\text{Conv}_1(\text{Conv}_2(F_2))) \quad (3)$$

$$F_3 = M_s(F_2) \otimes F_2 \quad (4)$$

For the intermediate channel map (F_2), Conv_1 , Conv_2 is convolution operation. The final feature map (F_3) is produced via element-wise multiplication between the spatial attention map (M_s), and the input feature map (F_2).

The distinguishing characteristic of this approach is its ability to compute interactions across all positions on the feature map globally, allowing the model to capture more comprehensive feature relationships. However, this increased consideration of positional interactions also leads to heightened computational complexity. To address this, we conducted experiments using both the CBAM and GAB to compare their effectiveness in improving detection accuracy under these conditions [38]. The overall architecture of the system is illustrated in Figure 3.

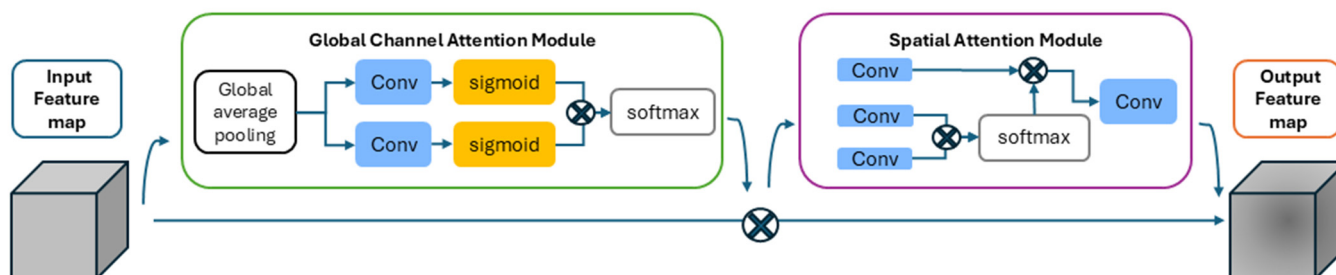


Figure 3. Global attention mechanism.

3.4. Swin Transformer Module

To further optimize the performance and enhance the accuracy of fall detection, we integrated transformer-based modules into an existing CNN-based model. As previously noted, the ability to capture global information is particularly crucial in scenarios in which parts of the body may be occluded by objects, potentially hindering detection accuracy. To capitalize on the strength of transformers in learning and extracting global information, we incorporated a modified Swin Transformer module into the YOLO network [39]. By integrating the Swin Transformer, which is designed to efficiently process global dependencies in visual data, the capability of the model to capture contextual relationships across the entire image is improved. This modification aims to address the limitations posed by partial occlusions and improve the ability of the model to detect falls in diverse and complex environments. The hierarchical design of the Swin Transformer, combined with its ability to capture both local and global information, allows for enhanced feature representation, further contributing to overall performance improvements in fall detection.

The Swin Transformer operates by applying self-attention across sliding windows, which allows it to process the input data by dividing them into patches, which are then arranged linearly for transformer processing. Initially, self-attention is performed through the window multihead self-attention (W-MSA) mechanism, which computes the relationships within patches confined to a specific window. Subsequently, the shift window multihead self-attention (SW-MSA) mechanism shifts the window boundaries to analyze the relationships between patches located in different windows. These two complementary processes enable the model to capture the correlations between patches across different regions of the input image. The calculation formula for self-attention is the same as Equation (5):

$$\text{Attention}(Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{d}} + B\right)V \quad (5)$$

Here, Q , K , and V represent the matrix for Query, Key, and Value, respectively, while B denotes the matrix for relative positional bias. The matrix V corresponds to the value matrix with attention weights applied, and the similarity between Query and Key is computed using the formula $(\frac{QK^T}{\sqrt{d}})$. The Shifted Window mechanism facilitates interaction between adjacent windows by alternating window positions across consecutive blocks. This process is represented mathematically as in Equation (6) where \hat{z}^l is the output of block l , W-MSA denotes Window-based Multi-head Self-Attention, and LN refers to Layer Normalization.

$$\hat{z}^l = \text{W-MSA}\left(\text{LN}\left(z^{l-1}\right)\right) + z^{l-1} \quad (6)$$

The Self-Attention in the Shifted Window is calculated as shown in Equation (7), where SW-MSA indicates Shifted Window-based Multi-head Self-Attention.

$$z^l = \text{SW-MSA}\left(\text{LN}\left(\hat{z}^l\right)\right) + \hat{z}^l \quad (7)$$

By leveraging this approach for fall detection, the model emphasizes spatial location information during hierarchical feature extraction. This helps enhance the detection of falls by considering both local and global dependencies between image patches, even when parts of the body are occluded or positioned in challenging ways. We anticipate that this method will yield more effective and accurate fall detection results. The basic structure of this process is illustrated in Figure 4.

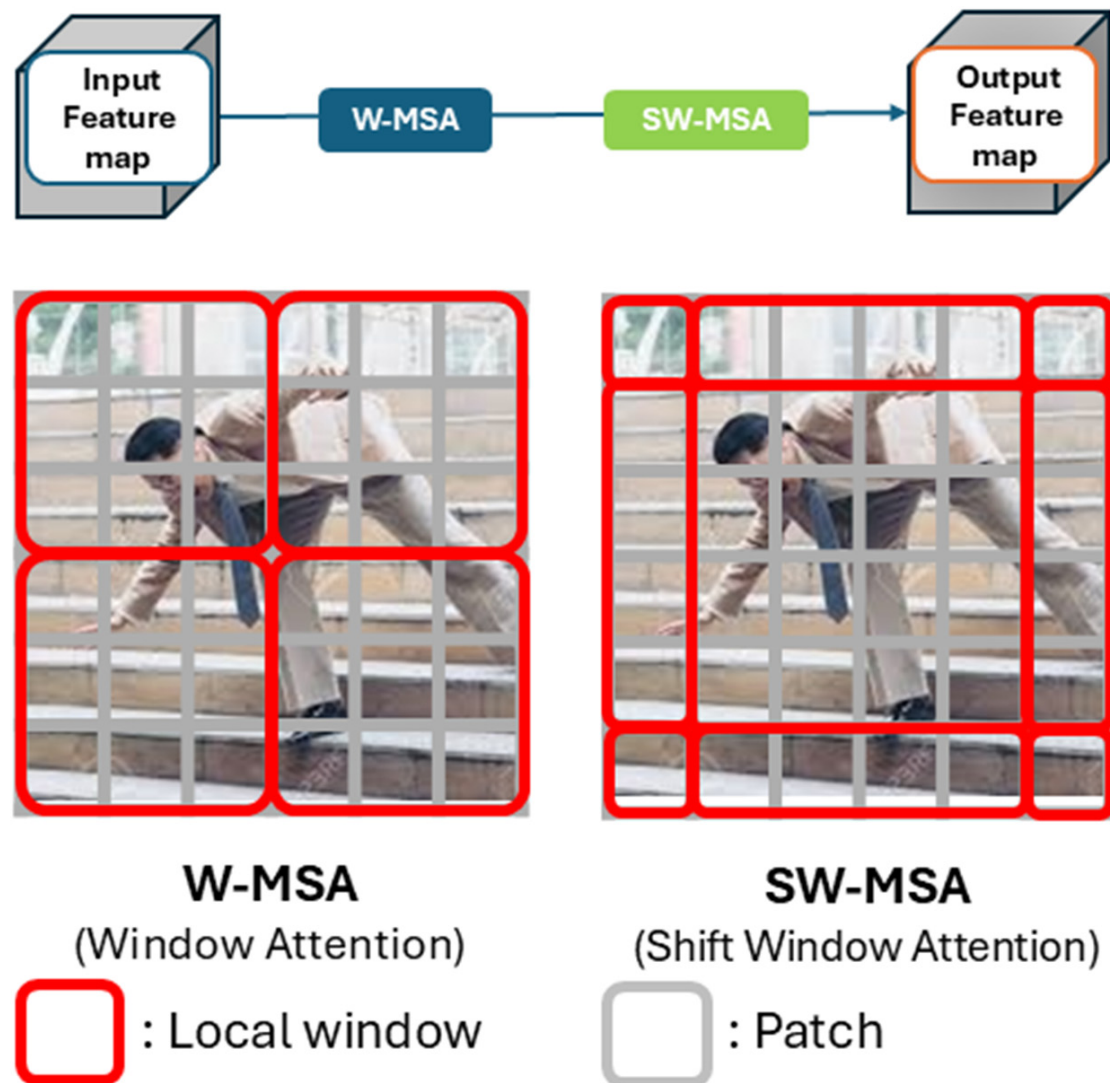


Figure 4. Swin Transformer.

3.5. FD-YOLO

The dual-head feature of YOLOv9 was employed to mitigate the loss of fall-related features in scenarios in which falls may be occluded by objects during the feature map extraction stage. This mechanism ensures that even partially obscured falls are detected. To preserve this critical functionality, the auxiliary stage, which incorporates a dual-head structure, is maintained without modification. Additionally, we integrate the Fall Detection (FD) module along with the global attention module (GAM) and CBAM into the opposite feature extraction path to further refine the detection performance.

To enhance accuracy while minimizing structural modifications within the network, the FD module, which modifies the structure of the RepNCSP module, is utilized to replace the original RepNCSP in YOLOv9, and its structure is illustrated in Figure 5.

This modification allows for improved feature retention and accuracy in fall detection. The final network architecture, illustrated in Figure 6, reflects a comprehensive design approach aimed at minimizing unpredictable effects while ensuring that all performance enhancement considerations are fully addressed.

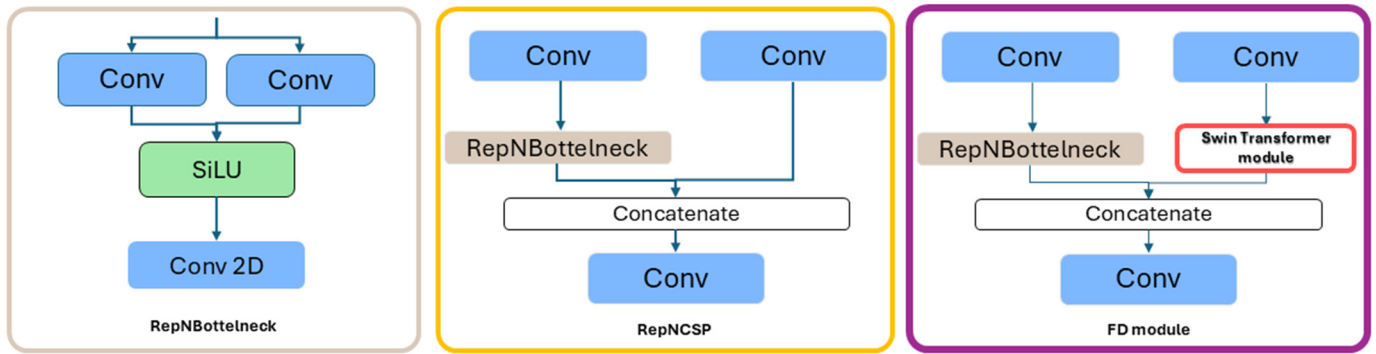


Figure 5. FD module.

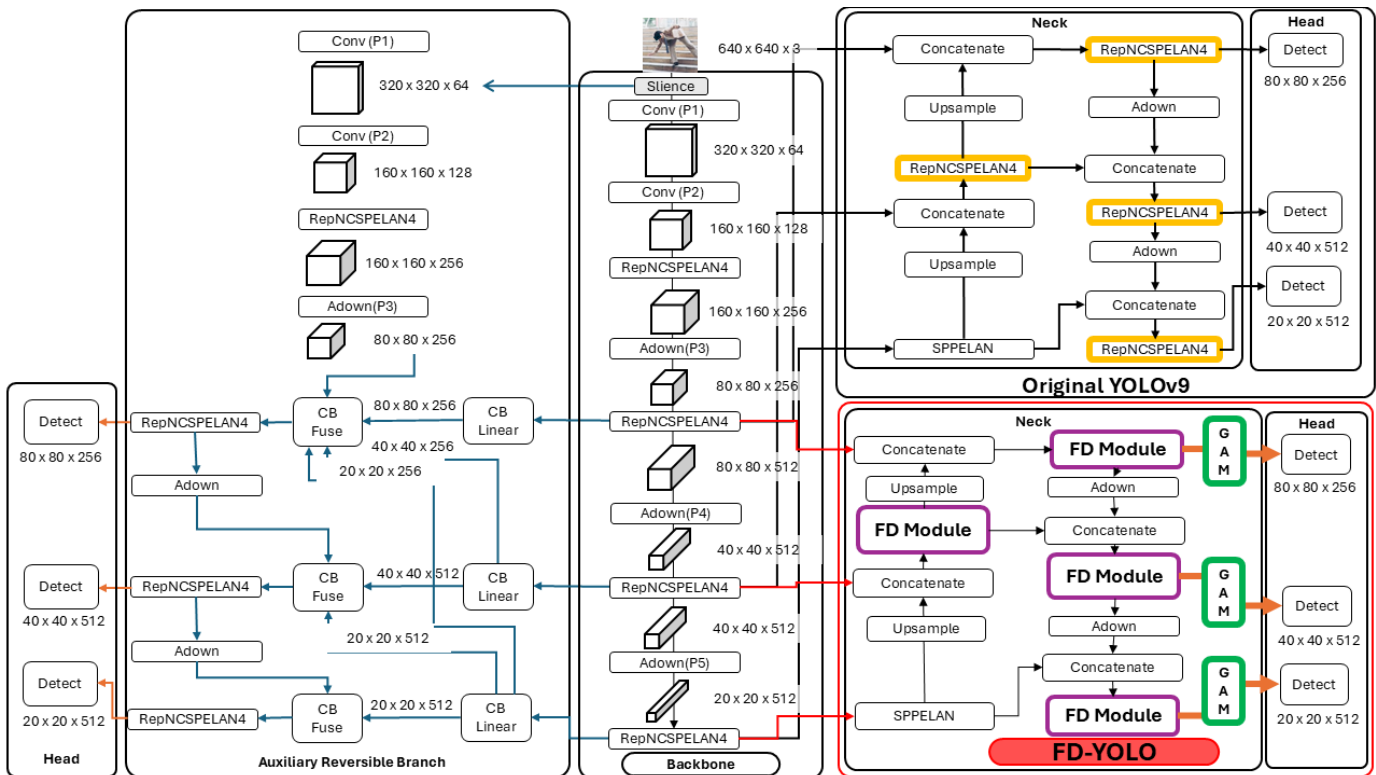


Figure 6. FD-YOLO structure.

3.6. Evaluation Metrics

The primary objective of this study was to detect falls accurately using images. Because undetected falls can prevent timely intervention, we placed particular emphasis on analyzing the false negative (FN) rate because missed detections can have serious consequences. Given that the YOLO network is designed for real-time image processing and can detect falls in video-based scenarios, we assumed that falls would be detected within the standard 60 frames of video footage. Thus, our goal was to develop a network capable of maintaining an FN count at or below 60, ensuring reliable fall detection across various scenarios.

The evaluation of object detection networks commonly involves performance metrics derived from a confusion matrix, including recall, precision, accuracy, and mean average precision (mAP). The confusion matrix is used to assess the predictive performance of a network and consists of four key values: true positive (TP), true negative (TN), false positive (FP), and FN. These values were used in different combinations to calculate the performance metrics used in this study.

The performance metrics evaluated in this study are described below:

Recall, as defined in Equation (8), is the proportion of actual positive cases (falls) correctly identified by the model. It is a crucial metric in fall detection as it directly relates to minimizing missed detections (false negatives):

$$\text{Recall(R)} = \frac{TP}{TP + FN} \quad (8)$$

Precision, as shown in Equation (9), represents the proportion of correctly identified positive cases among all predicted positives. This metric emphasizes the model's ability to reduce false positives:

$$\text{Precision(P)} = \frac{TP}{TP + FP} \quad (9)$$

Accuracy, expressed in Equation (10), measures the overall proportion of correctly classified instances (both positive and negative) among all predictions made by the model:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \quad (10)$$

The mean average precision (mAP), defined in Equation (11), evaluates the model's performance by averaging the precision scores over all object classes. It is derived by integrating the precision–recall curve for each class and computing the mean across all classes:

$$\text{Average Precision (AP)} = \int_0^1 P dR, \text{mAP} = \frac{1}{N} \sum_{i=1}^N AP_i \quad (11)$$

In object detection, Intersection over Union (IoU), as defined in Equation (12), is another critical metric for performance evaluation:

$$\text{Intersection over Union (IoU)} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (12)$$

IoU measures the accuracy of the predicted object's location by quantifying the degree of overlap between the predicted bounding box and the ground truth bounding box. Specifically, IoU is defined as the ratio of the overlapping area between the predicted and actual bounding boxes to the total area of their union. This metric provides a precise and quantitative assessment of how well the predicted bounding box aligns with the actual object. IoU is widely utilized in object detection tasks to refine model predictions and evaluate detection accuracy. The overlapping area refers to the portion where the predicted and actual bounding boxes intersect, whereas the union area represents the total combined area of both the predicted and actual bounding boxes. The IoU value ranges from 0 to 1, where an IoU of 1 indicates perfect alignment between the predicted and actual bounding boxes, and an IoU of 0 indicates no overlap. The IoU is commonly used in conjunction with mAP to assess the performance of object detection models.

Typically, the mAP is evaluated at different IoU thresholds. For instance, mAP@0.5 refers to the mAP when the IoU threshold is set to 0.5, which means that a prediction is considered valid only if the IoU between the predicted and actual bounding boxes is 0.5 or higher. Additionally, mAP@0.5–0.95 calculates the mAP across multiple IoU thresholds, ranging from 0.5 to 0.95 in increments of 0.05 (i.e., 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95). This method provides a more comprehensive evaluation of the localization accuracy of a model at various levels of overlap precision.

3.7. Learning Environment

In this study, the hardware setup included an i9-13900 CPU, 128GB RAM, and an RTX 4090 GPU. The operating system used was Ubuntu 22.04 LTS. The software environment consisted of Python 3.8, PyTorch 2.4.1, and CUDA 12.1. The hyperparameter configuration employed in this study was meticulously designed to optimize the training dynamics and maximize the performance of the proposed network. The model was trained for 300 epochs with a batch size of 16, utilizing an input image resolution of 512 pixels to balance computational efficiency and feature extraction capabilities.

4. Results

In this study, the primary focus was on detecting the occurrence of falls, rather than pinpointing their exact location with high precision. Given the need to prioritize identifying falls over exact localization, mAP@0.5 was selected as the primary evaluation metric, because it allows for some tolerance in bounding box overlap while still ensuring the reliable detection of fall events.

In this study, we evaluated the fall detection performance by integrating various versions of the YOLO network with transformer-based attention blocks. The experiments were conducted by measuring a range of performance metrics during the training and validation phases of the model; the training outcomes are presented in Figure 7. The experimental results confirmed that the YOLOv9e Swin network demonstrated the highest performance, significantly enhancing the fall detection accuracy.

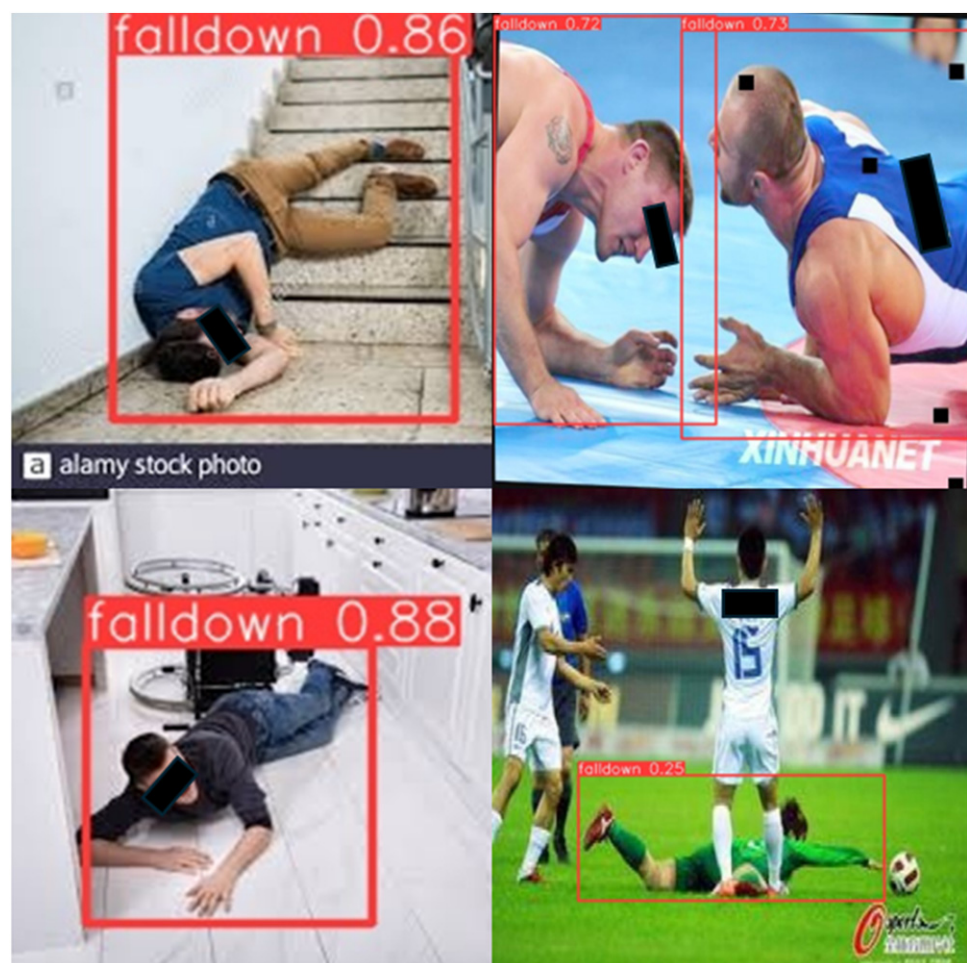


Figure 7. Detection samples.

Notably, the integration of transformer-based attention blocks improved the ability of the network to capture global information, which is a critical factor in detecting fall incidents, particularly in real-time scenarios. This capability allows the model to identify falls more effectively even in complex environments or when parts of the body are occluded. The combination of YOLOv9e and Swin Transformer modules contributes to superior detection accuracy, highlighting the potential of transformer architectures in advancing fall detection technology.

4.1. YOLO Attention Block

First, we compared the basic fall detection performance across various versions of the YOLO network. Each network was trained using the same dataset, with the mAP values measured during the training phase and the TP, TN, FP, and FN values recorded during the testing phase. The experimental results revealed that YOLOv9e achieved the highest TP count and lowest FN count, indicating a superior overall performance in fall detection. However, because each YOLO version demonstrated varying performance levels under specific conditions, we sought to further enhance these results by integrating additional attention blocks into the architecture. In doing so, we aimed to leverage the strengths of transformer-based attention mechanisms in capturing global features and refining the detection process. The results of these experiments detailing the performance improvements with the application of attention blocks are summarized in Table 2.

Table 2. Comparison results by model.

Model	mAP@0.5	FN
YOLOv3	0.981	197
YOLOv5	0.965	211
YOLOv6	0.971	236
YOLOv7	0.920	223
YOLOv8	0.968	221
YOLOv9	0.981	194
YOLOv10	0.974	311
YOLOv11	0.978	211

In this study, mAP values above 0.98 were considered equivalent. While detecting fall objects is important, the main focus was on reducing false negatives (FN), as missing fall situations poses a greater risk. Undetected falls can delay assistance, making it crucial to identify falls quickly and respond promptly. To address this, global information enhancement modules were added to reduce FN and improve the detection of fall situations, which was the key evaluation metric.

4.2. Additional Experiments with Network Attention Blocks

This study aims to reduce FN in fall detection, which occurs when actual falls are missed. Reducing FNs can help prevent critical situations caused by undetected falls. To achieve this, the study uses two modules: SWIN and GAM. These modules analyze global video information to capture limb positions and body structures, even when parts are obscured during a fall. The study shows these modules improve detection accuracy and address challenges in identifying falls. The experimental results indicate that the model utilizing YOLOv9-Swin with the GAB achieved the best performance in terms of mAP@0.5 and FN count. Specifically, this model recorded an mAP@0.5 of 0.982 and an FN count of 134, demonstrating the highest detection accuracy and lowest FN rate for fall detection across all tested configurations. The detailed results of these experiments are summarized in Table 3.

Table 3. The results of experiments.

Swin Transformer	Global Attention	CBAM	mAP@0.5	FN
			0.981	194
✓		✓	0.984	183
✓	✓		0.982	134

The precision–recall (PR) curve and confusion matrix for this model are illustrated in Figures 8–10, respectively, which provide a comprehensive view of the precision, recall, and overall predictive performance of the model. All network speeds were measured as follows: 0.6 ms for preprocessing, 2.1 ms for inference, and 0.4 ms for postprocessing per image.

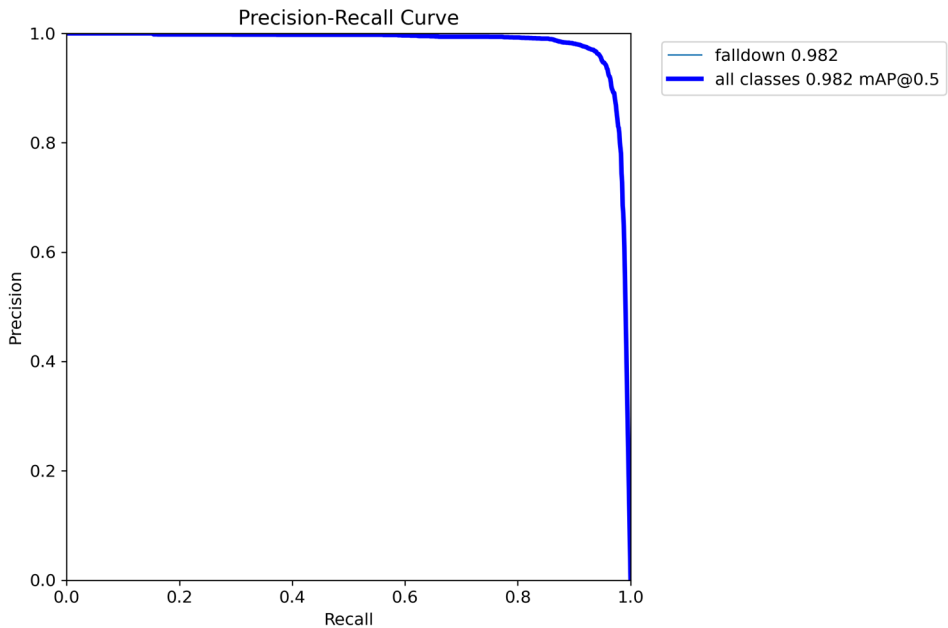


Figure 8. FD-YOLO PR curve.

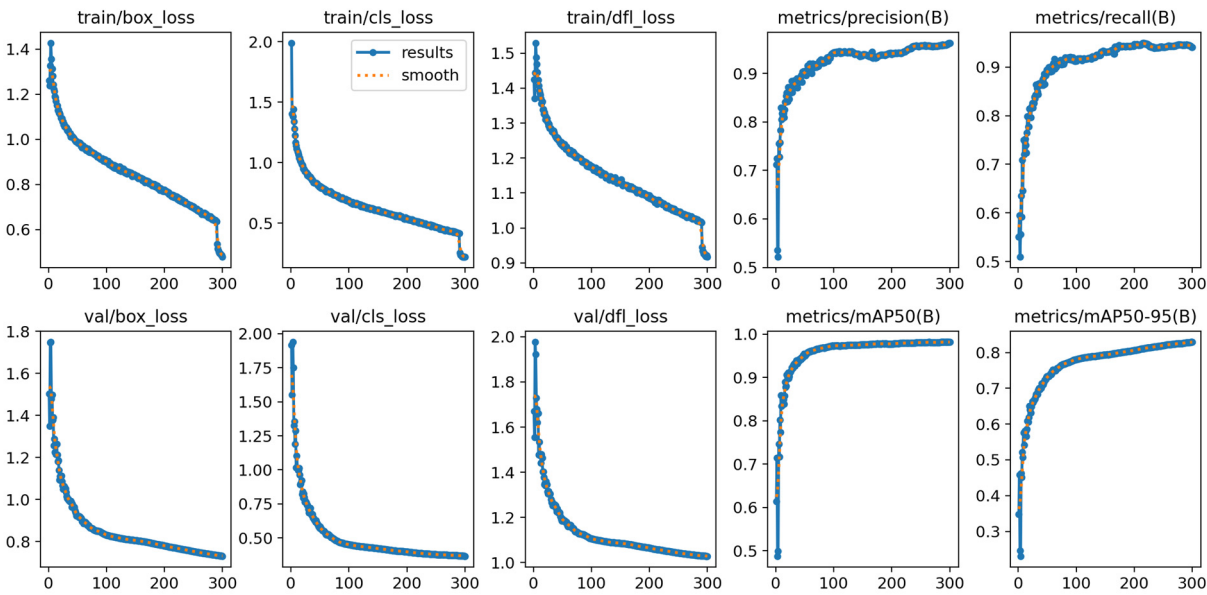


Figure 9. FD-YOLO training curve.

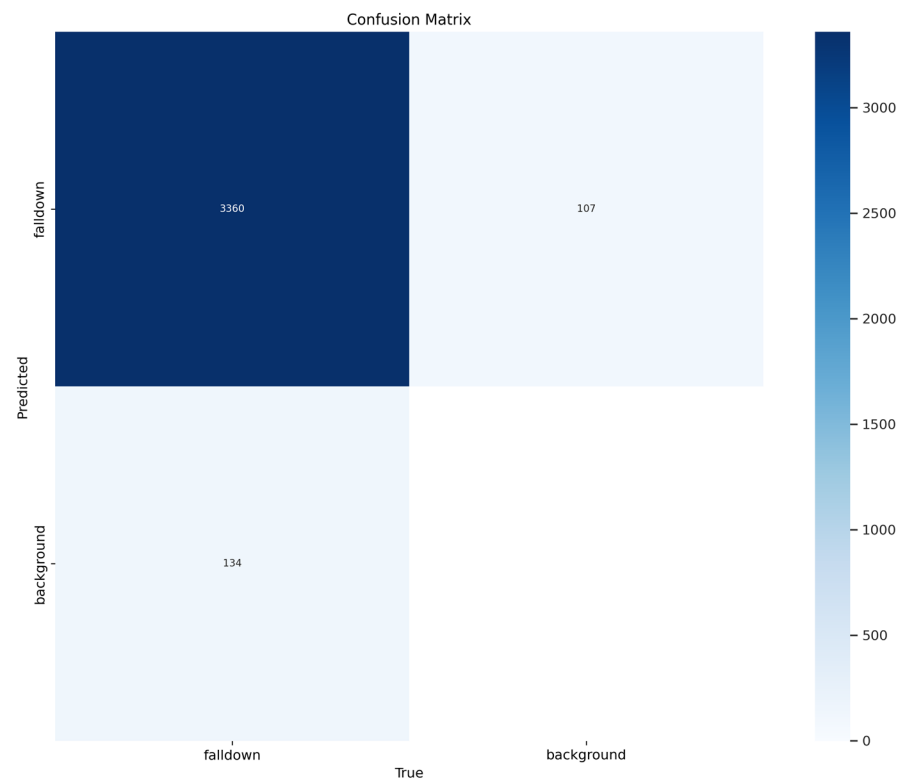


Figure 10. FD-YOLO confusion matrix.

5. Discussion

This study aimed to address several challenges encountered by existing vision-based fall detection methods. Previous research faced limitations in hardware performance, which hindered the consistent achievement of high detection accuracy and real-time processing speeds. Rapid transmission of fall events is essential for effective intervention; however, this critical aspect has not been sufficiently addressed in previous studies. Although certain approaches improve fall detection accuracy by increasing the network depth or incorporating attention mechanisms, these efforts are often constrained by CNN-based attention blocks. Specifically, such blocks have a limited ability to capture global visual information, which is crucial for handling occlusions and other complex scenarios in which parts of the body may be obscured. Consequently, previous methods struggled to maintain detection in situations involving partial visibility or intricate visual environments.

In addition, the widespread availability of CCTV systems confirms that many computers are already in place, connected online, and capable of performing inference. This availability has led to the exploration of a real-time system that, compared with embedded systems, can provide more accurate fall detection by leveraging global information, even in cases of occlusions. Furthermore, previous studies often relied on single fall detection datasets, exposing limitations of the ability of the model to generalize across diverse environments.

Recently, with the development of models such as the generative pretrained transformer (GPT), it has been demonstrated that transformer networks, in addition to CNN-based networks, can achieve higher accuracy in various tasks. Specifically, in vision tasks, transformer networks have demonstrated the capability to learn the structural features of an image's spatial information, leading to improved detection accuracy. However, it was also noted that these networks typically require large datasets to reach optimal performance, because they rely on extensive data to effectively capture and process global spatial information.

Subsequent research has explored the application of transformer networks to existing CNN architectures in the form of attention blocks, culminating in the development of the Swin Transformer. The introduction of Swin Transformer attention significantly enhanced the ability of the network to capture global information, resulting in improved fall detection accuracy. The attention block allows the network to focus on key regions of the image, thereby enabling more effective fall detection, even in scenarios with occlusions or complex background environments. This approach addresses the limitations of traditional CNN architectures, which primarily process local information by allowing the network to learn from the overall image context. Consequently, the combination of a CNN and transformer-based attention provides a more comprehensive method for identifying falls, particularly under challenging visual conditions.

Furthermore, given the importance of not only detecting falls but also identifying their precise location, an object detection network is necessary. The YOLO network, which is recognized for its strength in real-time detection and object localization, was selected as the most suitable network architecture for this task. The ability of YOLO to efficiently perform both detection and localization makes it an ideal candidate for accurately identifying falls and determining their location within a frame, thereby facilitating quicker and more effective responses in real-time scenarios.

In this study, we selected YOLO-based AI networks with varying depths and compared their fall detection performance. Early versions of the YOLO network, while offering faster detection speeds, often suffered from reduced accuracy. To address this issue, we conducted a comparative analysis of the latest versions of YOLO. The results indicated that YOLOv9e demonstrated the best balance between speed and accuracy, offering superior performance compared with earlier versions. This version effectively resolves the trade-off between detection speed and accuracy, making it the most suitable for real-time fall detection.

The focus of this study was to develop a network that integrates transformer-based Swin attention with selected YOLO networks, and train this network using a fused fall detection dataset. Building on prior research, we identified effective methods for incorporating a Swin Transformer into the YOLO network architecture. Subsequent optimization efforts involve adjusting the window size specifically for single-class training on the fall detection dataset and determining the most effective types and positions of attention blocks within the network. These optimizations resulted in high accuracy during the experimental trials, demonstrating the potential of the combined YOLO-Swin model for precise and efficient fall detection.

During the experiments, researchers initially hypothesized that the latest algorithms would deliver the best performance. However, they found that the most recent network architectures did not necessarily provide the highest accuracy, and that deeper networks did not guarantee better fall detection results. Furthermore, the CBAM, which is commonly regarded as effective in enhancing network performance, unexpectedly led to a decline in the accuracy of the fall detection dataset. In contrast, when the modified GAM was applied, this attention block performed better and improved the detection accuracy. These findings highlight the importance of tailoring network structures and attention mechanisms specifically to the dataset and task at hand rather than relying solely on the latest or most complex architectures.

The proposed FD-YOLO network achieved the highest accuracy and lowest FN for fall detection, demonstrating its strong potential for effectively identifying fall incidents in real time. This suggests that it is possible to establish a highly efficient fall detection system without additional hardware installation, even in general CCTV environments. The adaptability and efficiency of this system make it applicable to a wide range of fields, including elder-care safety management systems in aging societies, pedestrian protection

systems in public spaces, and fall prevention systems in hospitals and nursing facilities. The scalability of the FD-YOLO network makes it a valuable tool for enhancing safety and intervention in various critical environments.

FD-YOLO, which proposes a fall detection network, is specifically designed to maintain high performance not only within a controlled environment but also across diverse real-world settings. This adaptability allows the network to be seamlessly integrated into widely deployed CCTV systems, facilitating the continuous monitoring of fall incidents. Furthermore, the ability of a network to transmit detected events rapidly is expected to be of significant value in various applications, ensuring timely responses and interventions. This capability highlights the potential of the FD-YOLO network as an essential tool to enhance safety and prevent falls in real-world environments.

However, this study is limited by its retrospective nature because AI training and network optimization were conducted using specific datasets. In the future, we aim to address this limitation by obtaining data from a variety of environments through multiple experiments conducted in real-world settings and by acquiring prospective data via Institutional Review Board (IRB) approval. This will allow the development of a more robust and generalizable fall detection network. In addition, we plan to conduct follow-up research to assess the practical applicability of the proposed system in real-world scenarios.

We also observed that other studies achieved high performance by using data preprocessing techniques to analyze skeletal structures and incorporating this information into AI models. As a result, we intend to explore the integration of such methods in network structures, particularly in environments that do not place excessive strain on computational resources, to further enhance the performance of our system in future studies.

6. Conclusions

This study made a significant contribution to the advancement of fall detection technology, with the proposed FD-YOLO network demonstrating substantial potential for widespread application in real-world scenarios. If future research can further enhance the network's performance and validate its effectiveness in diverse real-world environments, a safer and more efficient fall detection system can be realized. Such a system could play a crucial role in addressing pressing social issues, particularly in aging societies, by enhancing safety and ensuring timely interventions for fall-related incidents.

Author Contributions: Conceptualization, H.H. and H.K.; resources, H.K.; validation, D.K., data curation, D.K.; writing—original draft preparation, H.H.; writing—review and editing, H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) No. 2016R1C1B2012888, RS-2016-NR016451.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available in a public, open access repository([<https://universe.roboflow.com/roboflow-universe-projects/fall-detection-ca3o8>], [<https://www.kaggle.com/datasets/uttejkumarkandagatla/fall-detection-dataset>], [<https://data.mendeley.com/datasets/7w7fccy7ky/4>]).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. WHO. Falls. Available online: <http://www.who.int/en/news-room/fact-sheets/detail/falls> (accessed on 6 September 2024).
2. Kerdegari, H.; Samsudin, K.; Rahman Ramli, A.; Mokaram, S. Development of wearable human fall detection system using multilayer perceptron neural network. *Int. J. Comput. Intell. Syst.* **2013**, *6*, 127–136. [\[CrossRef\]](#)
3. Kwolek, B.; Kepski, M. Improving fall detection by the use of depth sensor and accelerometer. *Neurocomputing* **2015**, *168*, 637–645. [\[CrossRef\]](#)
4. Ajerla, D.; Mahfuz, S.; Zulkernine, F. A real-time patient monitoring framework for fall detection. *Wirel. Commun. Mob. Comput.* **2019**, *2019*, 9507938. [\[CrossRef\]](#)
5. Bourke, A.K.; O'Brien, J.; Lyons, G.M. Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm. *Gait Posture* **2007**, *26*, 194–199. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Chen, G.-C.; Huang, C.-N.; Chiang, C.-Y.; Hsieh, C.-J.; Chan, C.-T. A reliable fall detection system based on wearable sensor and signal magnitude area for elderly residents. In Proceedings of the Aging Friendly Technology for Health and Independence: 8th International Conference on Smart Homes and Health Telematics, ICOST 2010, Proceedings 8, Seoul, Republic of Korea, 22–24 June 2010; pp. 267–270.
7. He, Y.; Li, Y.; Bao, S.-D. Fall detection by built-in tri-accelerometer of smartphone. In Proceedings of the 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics, Hong Kong, China, 5–7 January 2012; pp. 184–187.
8. Yuwono, M.; Moulton, B.D.; Su, S.W.; Celler, B.G.; Nguyen, H.T. Unsupervised machine-learning method for improving the performance of ambulatory fall-detection systems. *Biomed. Eng. Online* **2012**, *11*, 9. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Khan, S.S.; Hoey, J. Review of fall detection techniques: A data availability perspective. *Med. Eng. Phys.* **2017**, *39*, 12–22. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Roush, R.E.; Teasdale, T.A.; Murphy, J.N.; Kirk, M.S. Impact of a personal emergency response system on hospital utilization by community-residing elders. *South. Med. J.* **1995**, *88*, 917–922. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Brulin, D.; Benezeth, Y.; Courtial, E. Posture recognition based on fuzzy logic for home monitoring of the elderly. *IEEE Trans. Inf. Technol. Biomed.* **2012**, *16*, 974–982. [\[CrossRef\]](#)
12. Thome, N.; Miguet, S.; Ambellouis, S. A real-time, multiview fall detection system: A LHMM-based approach. *IEEE Trans. Circuits Syst. Video Technol.* **2008**, *18*, 1522–1532. [\[CrossRef\]](#)
13. Rougier, C.; Meunier, J.; St-Arnaud, A.; Rousseau, J. 3D head tracking for fall detection using a single calibrated camera. *Image Vis. Comput.* **2013**, *31*, 246–254. [\[CrossRef\]](#)
14. Auvinet, E.; Multon, F.; Saint-Arnaud, A.; Rousseau, J.; Meunier, J. Fall detection with multiple cameras: An occlusion-resistant method based on 3-d silhouette vertical distribution. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *15*, 290–300. [\[CrossRef\]](#)
15. Bian, Z.-P.; Hou, J.; Chau, L.-P.; Magnenat-Thalmann, N. Fall detection based on body part tracking using a depth camera. *IEEE J. Biomed. Health Inform.* **2014**, *19*, 430–439. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Alam, E.; Sufian, A.; Dutta, P.; Leo, M. Real-Time human fall detection using a lightweight pose estimation technique. In Proceedings of the International Conference on Computational Intelligence in Communications and Business Analytics, Kalyani, India, 27–28 January 2023; pp. 30–40.
17. Zheng, X.; Cao, J.; Wang, C.; Ma, P. A High-Precision Human Fall Detection Model Based on FasterNet and Deformable Convolution. *Electronics* **2024**, *13*, 2798. [\[CrossRef\]](#)
18. Wang, Y.; Chi, Z.; Liu, M.; Li, G.; Ding, S. High-performance lightweight fall detection with an improved YOLOv5s algorithm. *Machines* **2023**, *11*, 818. [\[CrossRef\]](#)
19. Kan, X.; Zhu, S.; Zhang, Y.; Qian, C. A lightweight human fall detection network. *Sensors* **2023**, *23*, 9069. [\[CrossRef\]](#) [\[PubMed\]](#)
20. 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. [\[CrossRef\]](#)
21. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
22. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-end object detection with transformers. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 213–229.
23. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
24. 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.
25. 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.* **2016**, *39*, 1137–1149. [\[CrossRef\]](#)

26. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Proceedings, Part I 14, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
27. Liu, C.; Tao, Y.; Liang, J.; Li, K.; Chen, Y. Object detection based on YOLO network. In Proceedings of the 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 14–16 December 2018; pp. 799–803.
28. Falldown Detection Computer Vision Project. Available online: <https://universe.roboflow.com/roboflow-universe-projects/fall-detection-ca3o8> (accessed on 23 August 2024).
29. Fall Detection Dataset. Available online: <https://www.kaggle.com/datasets/uttejmarkandagatla/fall-detection-dataset> (accessed on 23 August 2024).
30. Guerrero, J.C.E.; España, E.M.; Añasco, M.M.; Lopera, J.E.P. Dataset for human fall recognition in an uncontrolled environment. *Data Brief* **2022**, *45*, 108610. [[CrossRef](#)]
31. Wang, C.-Y.; Yeh, I.-H.; Liao, H.-Y.M. Yolov9: Learning what you want to learn using programmable gradient information. *arXiv* **2024**, arXiv:2402.13616.
32. Wang, B.-H.; Yu, J.; Wang, K.; Bao, X.-Y.; Mao, K.-M. Fall detection based on dual-channel feature integration. *IEEE Access* **2020**, *8*, 103443–103453. [[CrossRef](#)]
33. Ma, X.; Wang, H.; Xue, B.; Zhou, M.; Ji, B.; Li, Y. Depth-based human fall detection via shape features and improved extreme learning machine. *IEEE J. Biomed. Health Inform.* **2014**, *18*, 1915–1922. [[CrossRef](#)]
34. Yang, L.; Ren, Y.; Zhang, W. 3D depth image analysis for indoor fall detection of elderly people. *Digit. Commun. Netw.* **2016**, *2*, 24–34. [[CrossRef](#)]
35. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
36. Yan, Z.; Hao, L.; Yang, J.; Zhou, J. Real-Time Underwater Fish Detection and Recognition Based on CBAM-YOLO Network with Lightweight Design. *J. Mar. Sci. Eng.* **2024**, *12*, 1302. [[CrossRef](#)]
37. Liu, Y.; Shao, Z.; Hoffmann, N. Global attention mechanism: Retain information to enhance channel-spatial interactions. *arXiv* **2021**, arXiv:2112.05561.
38. Song, C.H.; Han, H.J.; Avrithis, Y. All the attention you need: Global-local, spatial-channel attention for image retrieval. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2022; pp. 2754–2763.
39. Cao, X.; Zhang, Y.; Lang, S.; Gong, Y. Swin-transformer-based YOLOv5 for small-object detection in remote sensing images. *Sensors* **2023**, *23*, 3634. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.