RESEARCH ARTICLE

Dual Attention-Driven Optimized YOLOV5 Framework for Accurate Fall Detection in Visual Monitoring Systems

Muhammad Jamal Ahmad[1], **Arshad Khan**[2], **Taimur Ali Khan**[3] and **Bilal Ahmad**[4,*]

[1] Departamento de Sistemas Informaticos, Universidad Politécnica de Madrid, Madrid 28031, Spain
[2] Yoobee Colleges of Creative Innovation, Auckland 1010, New Zealand
[3] Department of IT, Saudi Media Systems, Riyadh 11482, Saudi Arabia
[4] Department of Computer Science, Govt Degree College Lalqilla Maidan Dir Lower, Pakistan

Abstract

Fall detection (FD) systems are an important part of healthcare monitoring, especially for elderly populations, where quick intervention can prevent serious injuries. This paper introduces an optimized YOLOV5-based framework that combines dual attention mechanisms for improved FD in real-time edge deployment situations. The proposed design integrates the Convolutional Block Attention Module (CBAM) and Squeeze-and-Excitation (SE) blocks within the YOLOv5 backbone, along with an improved Focus module that uses slice-based feature extraction. These enhancements allow the model to effectively capture both spatial and channel-wise dependencies, which are essential for distinguishing fall events from normal human activities. Ablation studies confirm the individual contribution of each component, with more notable improvements observed on the challenging DiverseFALL10500 dataset, which features diverse environmental conditions. The framework maintains computational efficiency suitable for edge deployment while offering robust detection performance across different camera angles, lighting conditions, and complex backgrounds. A thorough evaluation on the CAUCAFall and DiverseFALL10500 benchmark datasets shows superior performance compared to existing YOLO variants.

Keywords: fall detection, optimized YOLOv5, attention mechanism, healthcare monitoring, image analysis.

1 Introduction

Rapid progress in artificial intelligence (AI) is transforming society, and healthcare monitoring is among the most pressing areas, as it necessitates real-time analysis and prompt responses. Falls can occur at home, in hospitals, and outdoors, often caused by medical issues, environmental hazards, or mobility challenges. Too often, they lead to serious injuries or even death. The impact is especially severe for older adults, where a single fall can cause lasting physical damage and reduce independence. According to the CDC in the USA [1], falls affect more than one in four adults 65 years and older each year,

Academic Editor:
Xue-Bo Jin

Submitted: 01 September 2025
Accepted: 29 September 2025
Published: 26 November 2025

Vol. 3, **No.** 1, 2026.
10.62762/TIS.2025.559776

***Corresponding author:**
✉ Bilal Ahmad
bilalahmadcs03@gmail.com

Citation
Ahmad, M. J., Khan, A., Khan, T. A., & Ahmad, B. (2025). Dual Attention-Driven Optimized YOLOV5 Framework for Accurate Fall Detection in Visual Monitoring Systems. *ICCK Transactions on Intelligent Systematics*, 3(1), 1–10.

resulting in approximately 3 million emergency visits and 1 million hospitalizations due to serious injuries. Therefore, the urgent need for reliable fall detection (FD) systems has become increasingly important, as these technologies can greatly impact individual well-being, especially among elderly populations. Accurate and prompt feedback significantly improves safety in healthcare settings, enhancing care quality, optimizing resource utilization, and improving patient outcomes. Advanced FD technologies have the potential to transform healthcare delivery, promoting greater independence for vulnerable groups and ultimately saving lives.

Contemporary FD networks generally follow three paradigms, each with unique benefits and challenges [2]. Ambient sensing uses environmental sensors, such as acoustic, vibration, and pressure detectors, for discreet monitoring; however, its coverage is limited and expansion is expensive in multi-room environments [3]. Wearable sensors, often using accelerometers, gyroscopes, and magnetometers, provide direct motion and physiological tracking, yet their effectiveness is reduced by issues of comfort, charging, and long-term user adherence [4, 5]. In contrast, vision-based systems utilize cameras and machine learning to analyze movement, thereby avoiding user compliance issues and providing rich contextual information. However, they must balance practicality with privacy concerns [6].

Within deep learning (DL)–based FD, visual intelligence-based methods dominate research, spanning conventional RGB surveillance and advanced depth-sensing systems [7, 8]. In parallel, sensor fusion frameworks integrate visual inputs with accelerometer and gyroscope data to enhance detection, albeit at the cost of added computational and algorithmic complexity, which limits deployment in resource-constrained settings [9–12]. By contrast, single-modality vision systems are attractive for scalable ambient monitoring. Traditional computer vision approaches struggled with environmental variability and occlusion [13, 14]. However, the emergence of unified detection models, especially the YOLO family, has transformed the field with faster inference and higher accuracy across diverse conditions [15, 16]. Early work utilized YOLOv2 with MS-COCO–trained features for human detection [17], followed by YOLOv3 frameworks that incorporate multi-person tracking and posture analysis. Subsequent advances include robotic-assist systems utilizing monocular vision for proactive fall management [19], YOLOv4 applied to datasets such as UR FD [8], and the recent YOLOv7-fall model, which achieves state-of-the-art performance [20]. Our study builds on this trajectory by proposing architectural improvements to address current challenges in scalability and robustness.

1.1 Key Contributions

This work presents four primary contributions that advance the state of the art in vision-based FD systems for healthcare applications.

- **Enhanced YOLOv5 Backbone with Attention Mechanisms**: We propose an improved YOLOv5 backbone network that includes the Convolutional Block Attention Module (CBAM) and Squeeze-and-Excitation (SE) blocks to boost feature discrimination and enhance FD accuracy in complex visual situations with varying lighting and backgrounds.
- **Optimized Focus Module with Slice-based Feature Extraction**: We present a new Focus module enhancement that employs strategic slice operations and channel-wise concatenation via CBS (Conv-BN-SiLU) blocks, enabling more efficient feature representation and improved computational performance for edge deployment scenarios.
- **Comprehensive Cross-dataset Validation**: We perform extensive experiments on two different FD datasets (CAUCAFall and DiverseFall10500) to demonstrate the generalization and robustness of our proposed system across various demographic groups, environmental conditions, and fall scenarios, proving its practicality for real-world healthcare monitoring.

2 Related Literature

The proliferation of DL methodologies has significantly transformed vision-based FD research [21]. YOLO architectures have emerged as particularly influential due to their exceptional real-time processing capabilities [22, 23]. Early approaches demonstrated YOLOv2's effectiveness for human detection tasks, utilizing transfer learning from pre-trained models and subsequently fine-tuning on custom datasets comprising 500 manually labeled images [17]. Building on this foundation, researchers developed YOLOv3-based solutions addressing multi-person detection scenarios [18]. Innovative applications have extended beyond traditional detection paradigms.

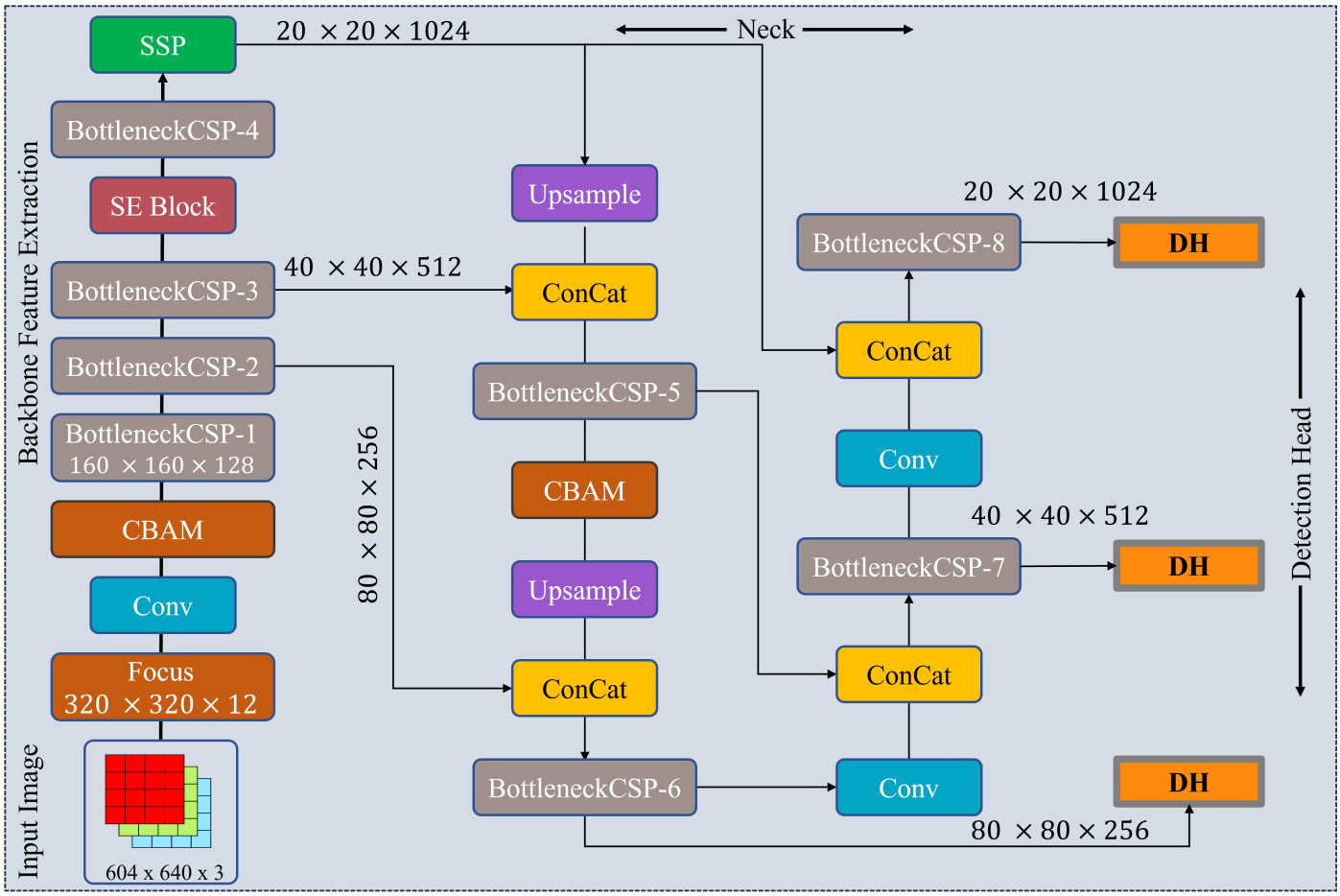

Figure 1. Proposed optimized YOLOV5 architecture with dual attention mechanisms, CBAM, SE blocks, and enhanced Focus module for real-time FD.

One notable system integrated monocular vision with robotic assistance for proactive fall risk assessment [19]. This approach evaluated environmental disorder levels through single-camera monitoring, employing socially acceptable robots to facilitate human-system interaction and demonstrating high adaptability for seamless integration into smart living environments. Subsequently, YOLOv4-based frameworks were engineered specifically for practical deployment using the UR FD dataset, which contains 1691 fall instances and 1731 normal activity samples [8]. This sensor-free approach achieved automatic fall recognition through standard RGB cameras, extracting features directly from self-annotated visual data. Recent architectural innovations have focused on YOLOv5 enhancements for elderly care applications [24]. These modifications replaced conventional convolutions with asymmetric blocks in the backbone network and integrated spatial attention mechanisms within residual structures to improve localization accuracy. Contemporary developments have introduced YOLOv7-fall architectures claiming enhanced feature extraction with reduced computational complexity [20]. However, training limitations persist, with datasets containing only 4,016 images, which constrains model generalization. These collective observations underscore the critical need for comprehensive datasets and systematic architectural enhancements to advance state-of-the-art vision-based FD systems.

3 Proposed Method

3.1 Overview

The proposed FD framework presents a comprehensive, vision-based system designed for real-time edge deployment in healthcare monitoring environments. Our approach improves the YOLOv5 architecture by strategically integrating attention mechanisms and enhancing feature extraction. The system comprises three main components: an enhanced backbone network with dual attention mechanisms, an optimized feature fusion neck, and multi-scale detection heads. The key innovation involves systematically incorporating the Convolutional Block Attention Module (CBAM)

and Squeeze-and-Excitation (SE) blocks within the YOLOv5 backbone, along with an improved Focus module that uses slice-based feature extraction. This architectural enhancement captures both spatial and channel-wise dependencies essential for distinguishing fall events from normal human activities.

3.2 Modified YOLOv5 Architecture

Our modified YOLOv5 architecture introduces improvements across multiple network components to optimize FD performance, as shown in Figure 1. The network employs dual attention mechanisms at key feature extraction points, with the original CSP bottleneck blocks enhanced by attention modules for improved feature representation. The neck architecture utilizes improved feature pyramid networks with strategic concatenation operations, enabling multi-scale feature fusion that is essential for detecting falls across different person sizes and distances. The detection head maintains the original three-scale output structure while incorporating refined anchor strategies specifically designed for human pose variations during fall events. These changes improve the model's ability to capture spatial characteristics specific to fall scenarios while maintaining computational efficiency for edge deployment.

3.3 Dual Attention with CBAM Integration

The CBAM provides dual attention mechanisms that capture both channel and spatial dependencies [11, 25, 26]. CBAM operates through a sequential attention process, first computing channel attention weights and then refining spatial attention. The channel attention component employs average and max pooling operations to generate channel statistics, processed through a shared multi-layer perceptron (MLP) to produce channel attention weights:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \tag{1}$$

The spatial attention module processes channel-refined features to create spatial attention maps that highlight important spatial locations. This component uses pooling operations across the channel dimension, concatenates the results, and applies a convolutional layer with sigmoid activation. The spatial attention mechanism allows the model to focus on body regions and poses that indicate fall events. CBAM integration occurs at multiple bottleneck stages after CSP modules, ensuring thorough attention-driven feature refinement throughout the extraction pipeline while keeping computational efficiency.

3.4 Squeeze-and-Excitation Blocks

Squeeze-and-Excitation (SE) blocks provide complementary channel attention capabilities, focusing on channel-wise feature recalibration. The SE mechanism operates through global information embedding via squeeze operations and adaptive recalibration through excitation transformations.

The squeeze operation aggregates spatial information using global average pooling, generating channel-wise statistics:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i, j) \tag{2}$$

The excitation component employs a gating mechanism with two fully connected layers, utilizing ReLU and sigmoid activations, to generate channel-wise scaling factors that modulate the importance of feature channels. This enables the model to emphasize channels that contain discriminatory fall-related features while suppressing less relevant information. SE blocks integrate at strategic locations following convolutional operations in the backbone network, ensuring effective channel-wise refinement without disrupting hierarchical feature extraction. The lightweight design makes them suitable for edge deployment where computational efficiency remains critical.

3.5 Enhanced Focus Module

The enhanced Focus module improves the efficiency of feature extraction while maintaining computational performance for edge deployment [27]. Our module employs strategic slice operations that partition input feature maps into multiple sub-regions, enabling granular feature analysis and the enhanced preservation of spatial information. The slice-based approach divides input tensors along spatial dimensions, creating feature patches that capture fine-grained spatial details often lost in conventional downsampling:

$$S_i = Focus_{slice}(X, stride = 2)[i], \text{ where } i \in \{0, 1, 2, 3\} \tag{3}$$

Following slice operations, channel-wise concatenation through CBS (Conv-BN-SiLU) blocks creates enriched feature representations combining

Table 1. Performance comparison of YOLOv3-v5 variants on FD datasets (ordered by mAP performance).

S No	Model	CAUCAFall				DiverseFALL10500			
		mAP	Precision	F1-score	Recall	mAP	Precision	F1-score	Recall
1	yolov4	0.9896	0.9901	0.9898	0.9913	0.825	0.818	0.801	0.794
2	yolov5x	0.9898	0.9892	0.9903	0.9906	0.834	0.769	0.797	0.829
3	yolov5m	0.9911	0.9914	0.9923	0.9927	0.848	0.850	0.825	0.807
4	yolov3	0.9912	0.9904	0.9916	0.9931	0.842	0.848	0.827	0.809
5	yolov5n	0.9924	0.9921	0.9935	0.9942	0.870	0.810	0.839	0.852
6	yolov5l	0.9926	0.9925	0.9934	0.9942	0.858	0.805	0.813	0.837
7	yolov5s	0.9932	0.9943	0.9954	0.9961	0.906	0.895	0.878	0.845
8	**Optimized YOLOv5**	**0.9942**	**0.9938**	**0.9945**	**0.9952**	**0.912**	**0.908**	**0.895**	**0.887**

Table 2. Ablation study of our network with the integration of different modules on the CAUCAFall dataset.

Model	mAP	Precision	F1-score	Recall
YOLOv5s	0.9932	0.9943	0.9954	0.9961
YOLOv5s + Focus	0.9936	0.9946	0.9958	0.9964
YOLOv5s + CBAM	0.9938	0.9948	0.9959	0.9966
YOLOv5s + SE	0.9940	0.9950	0.9961	0.9968
Optimized YOLOv5 (Ours)	**0.9942**	**0.9938**	**0.9945**	**0.9952**

information from multiple spatial regions. The CBS blocks provide efficient feature transformation with batch normalization and SiLU activation functions, ensuring stable training dynamics. The enhanced Focus module incorporates residual connections and attention-guided feature fusion, enabling selective emphasis on relevant spatial regions. Integration occurs at the network input stage, replacing the original Focus component to establish a strong foundation for subsequent attention mechanisms throughout the architecture.

4 Experimental Results

4.1 Implementation Details and Experimental Setup

The proposed Optimized YOLOv5 framework was implemented using a PyTorch environment on a system equipped with an NVIDIA RTX 4090 GPU with 24GB of memory. Training was conducted with a batch size of 10 over 200 epochs using the SGD optimizer with an initial learning rate of 0.01, momentum of 0.937, and weight decay of 0.0005. Data augmentation techniques, such as random horizontal flipping, mosaic augmentation, and mixup, were employed during training to enhance the model's generalization. The input image resolution was set to 640×640 pixels to balance detection accuracy with computational efficiency for edge deployment. All experiments were conducted using identical hyperparameters to ensure fair comparison across different model variants as empirically validated in [28]. The assessment of the proposed network is based on essential object detection metrics, including mAP, precision, recall, and F1-score. These metrics are well acknowledged in the field of object detection and are precisely specified mathematically, as explained in [29].

4.2 Datasets

Two benchmark datasets were utilized for a comprehensive evaluation of the proposed FD system [28, 30]. The CAUCAFall dataset contains high-quality video sequences captured in controlled indoor environments with consistent lighting conditions, providing a standardized evaluation benchmark. The DiverseFALL10500 dataset presents a more challenging scenario with diverse environmental conditions, varying camera angles, different lighting situations, and complex backgrounds that better represent real-world deployment scenarios. Both datasets include annotated fall events with precise temporal boundaries, allowing for a robust quantitative assessment of detection performance. The datasets are split into training, validation, and testing sets according to the ratio highlighted in the study [28].

4.3 Comparative Analysis

Table 1 provides a detailed comparison between our proposed Optimized YOLOv5 and existing YOLO variants on benchmark datasets. The results show that our method outperforms others across all evaluation metrics. On the CAUCAFall dataset,

Figure 2. Qualitative results of the proposed Optimized YOLOv5 on DiverseFALL10500 and CAUCAFall datasets.

Table 3. Ablation study of our network with the integration of different modules on the DiverseFALL10500 dataset.

Model	mAP	Precision	F1-score	Recall
YOLOv5s	0.906	0.895	0.878	0.845
YOLOv5s + Focus	0.908	0.897	0.881	0.849
YOLOv5s + CBAM	0.909	0.901	0.886	0.862
YOLOv5s + SE	0.910	0.903	0.889	0.875
Optimized YOLOv5 (Ours)	**0.912**	**0.908**	**0.895**	**0.887**

our approach achieves an mAP of 0.9942, which is 0.0010 higher than the best baseline YOLOv5s model. Although the improvement appears modest, it indicates that the CAUCAFall dataset, being collected in controlled environments, has inherent performance ceilings that limit further significant gains. More notable improvements are observed on the more challenging DiverseFALL10500 dataset, where our method achieves 0.912 mAP, compared to 0.906 for YOLOv5s, representing a meaningful 0.6% gain. The improved results on this diverse dataset confirm the effectiveness of our dual attention mechanisms in managing complex real-world scenarios with changing environmental conditions. Importantly, our approach maintains steady precision (0.908) and recall (0.887), demonstrating reliable detection across various fall scenarios. The comparison indicates that traditional YOLO variants struggle with the variety in DiverseFALL10500, with performance drops from 0.825 (YOLOv4) to 0.906 (YOLOv5s) in mAP. Our optimized design narrows this gap by utilizing strategic attention and extracting better features.

4.4 Ablation Study

To validate the contribution of each component, we performed systematic ablation studies on both datasets. Tables 2 and 3 demonstrate the progressive performance improvements achieved through incremental module integration. Starting from the YOLOv5s baseline, the improved Focus module yields initial gains of 0.0004 mAP on CAUCAFall and 0.002 mAP on DiverseFALL10500. The CBAM integration yields additional gains of 0.0002 and 0.001 mAP, respectively, while incorporating the SE block provides further improvements of 0.0002 and 0.001 mAP. The complete integration of all components in our Optimized YOLOv5 achieves the highest performance on both datasets. The ablation results demonstrate more significant improvements on the challenging DiverseFALL10500 dataset, where the combined effect of all components yields a 0.006 mAP increase over the baseline. This pattern confirms that our attention mechanisms are particularly effective in handling complex environmental variations and diverse fall scenarios. The precision and recall metrics show consistent improvements, with the complete model achieving an optimal balance between detection accuracy and false positive reduction.

4.5 Optimizer Comparison

Table 4 presents the performance analysis of different optimization algorithms on our proposed architecture.

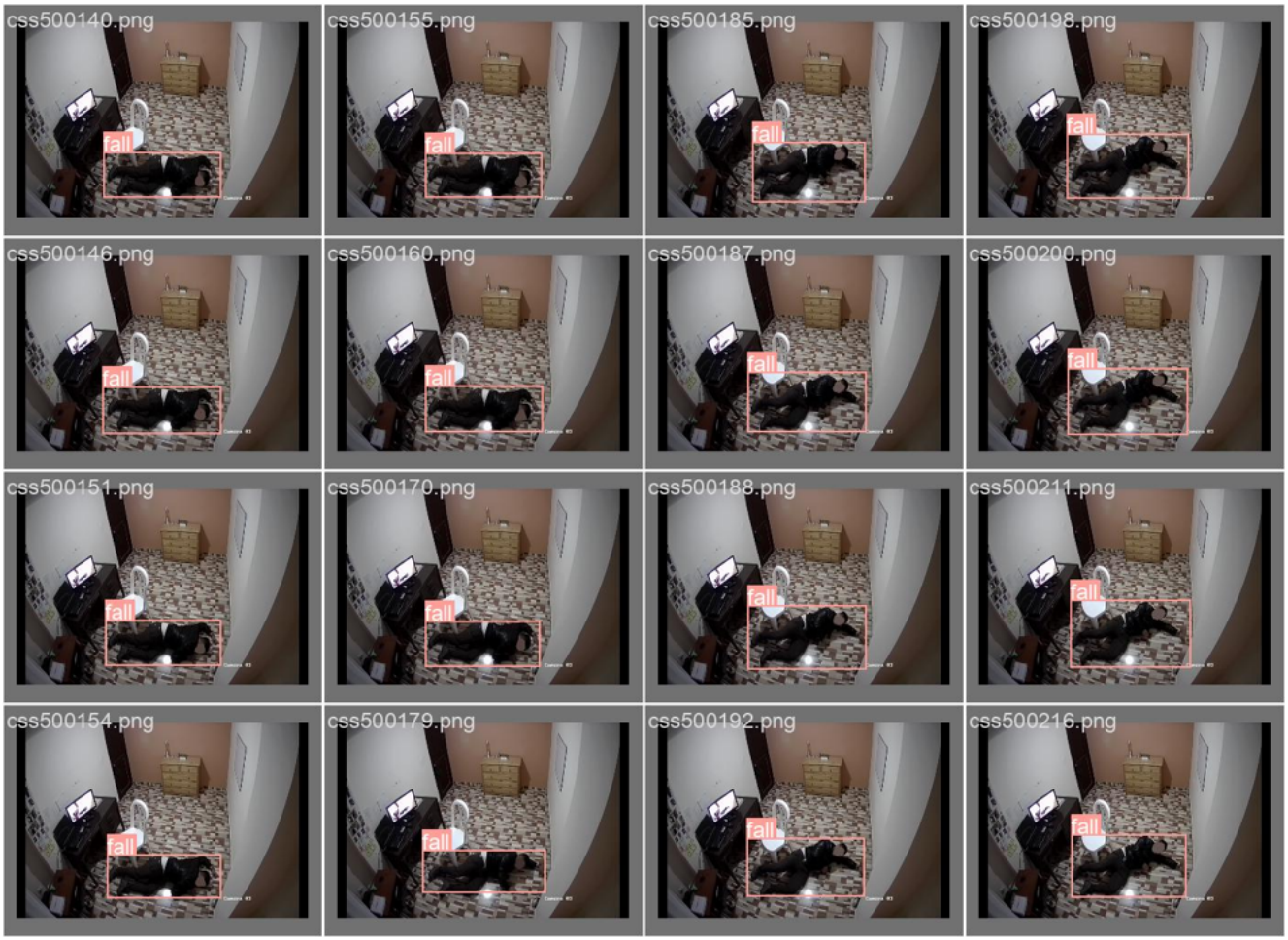

Figure 3. Performance validation of the proposed network on the CAUCAFall benchmark dataset, demonstrating stable learning dynamics and generalization capability.

The SGD optimizer shows better convergence and final performance compared to adaptive methods. While Adam and its variants (AdamW, Nadam) accelerate initial convergence, SGD provides better generalization on both datasets with final mAP scores of 0.9942 and 0.912, respectively. SGD's superior performance is attributed to its ability to find flatter minima during training, resulting in improved generalization on unseen FD scenarios. RMSprop performed the worst with mAP values of 0.9921 and 0.891, indicating suboptimal convergence for our specific architecture and loss landscape.

4.6 Qualitative Results

Figure 2 illustrates the qualitative performance of our proposed system across diverse scenarios present in both datasets. The visualization demonstrates strong detection capabilities across various lighting conditions, camera angles, and environmental settings. Our method effectively detects fall events under challenging situations, including partial obstructions, motion blur, and complex backgrounds that often cause failures in traditional methods. The detection results demonstrate consistent bounding box accuracy and confidence scores for different types of falls, including forward falls, backward falls, and sideways collapses. The model performs reliably regardless of person orientation, clothing differences, or scene complexity, confirming its practical usability for real-world deployment scenarios. Figure 3 presents the validation demonstrating the stable learning dynamics of our proposed network on the CAUCAFall dataset. The challenging detections indicate effective learning without overfitting.

5 Conclusion

In this paper, we present an optimized YOLOV5 framework that integrates dual attention mechanisms (CBAM and SE blocks) with an enhanced Focus module for robust FD. The proposed architecture achieves superior performance on CAUCAFall and DiverseFALL10500 datasets compared to the baseline YOLOv5s. The more significant improvement in DiverseFALL10500 demonstrates the enhanced

Table 4. Comparison of different optimizers on our proposed Optimized YOLOv5 model using FD datasets.

Optimizer	CAUCAFall				DiverseFALL10500			
	mAP	Precision	F1-score	Recall	mAP	Precision	F1-score	Recall
Adam	0.9928	0.9925	0.9931	0.9938	0.898	0.892	0.876	0.863
AdamW	0.9935	0.9932	0.9939	0.9946	0.905	0.899	0.883	0.870
RMSprop	0.9921	0.9918	0.9924	0.9931	0.891	0.885	0.869	0.856
Nadam	0.9932	0.9929	0.9936	0.9943	0.902	0.896	0.880	0.867
SGD	**0.9942**	**0.9938**	**0.9945**	**0.9952**	**0.912**	**0.908**	**0.895**	**0.887**

robustness of the framework in handling complex real-world scenarios with varying lighting conditions, camera perspectives, and environmental diversity. Detailed ablation studies validate the individual contributions of each component, revealing that the enhanced Focus module establishes improved foundations for feature extraction, while the CBAM and SE blocks provide complementary spatial and channel-wise attention refinements. A comprehensive evaluation across various environmental conditions and fall scenarios confirms consistent detection performance. At the same time, qualitative analysis demonstrates reliable bounding box accuracy and confidence scores across various fall types, including forward, backward, and sideways collapses. These findings demonstrate the effectiveness of strategically integrated attention mechanisms in enhancing object detection in healthcare applications, thereby contributing to automated monitoring systems that can prevent severe injuries and save lives through timely FD. In the future, we aim to investigate more YOLO variants with progressive modifications on challenging datasets. Moreover, we plan to introduce controlled uncertainties into the training sets to ensure the model is well-trained on visual complexities.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

Taimur Ali Khan is an employee of Department of IT, Saudi Media Systems, Riyadh 11482, Saudi Arabia. The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Facts about falls. (2024, June 10). Older Adult Fall Prevention. Retrieved from https://www.cdc.gov/falls/data-research/facts-stats/index.html

[2] Ren, L., & Peng, Y. (2019). Research of Fall Detection and Fall Prevention Technologies: A Systematic Review. *IEEE Access*, *7*, 77702-77722. [CrossRef]

[3] Ma, L., Liu, M., Wang, N., Wang, L., Yang, Y., & Wang, H. (2020). Room-level fall detection based on ultra-wideband (UWB) monostatic radar and convolutional long short-term memory (LSTM). *Sensors*, *20*(4), 1105. [CrossRef]

[4] Wang, K., Zhan, G., & Chen, W. (2019). A new approach for IoT-based fall detection system using commodity mmWave sensors. In *Proceedings of the 2019 7th International Conference on Information Technology: IoT and Smart City* (pp. 197-201). [CrossRef]

[5] Sheng-lan, Z., Yi-fan, Y., Li-fu, G., & Diao, W. (2019, November). Research and design of a fall detection system based on multi-axis sensor. In *Proceedings of the 4th International Conference on Intelligent Information Processing* (pp. 303-309). [CrossRef]

[6] Er, P. V., & Tan, K. K. (2020). Wearable solution for robust fall detection. In *Assistive Technology for the Elderly* (pp. 81-105). Academic Press. [CrossRef]

[7] Alam, E., Sufian, A., Dutta, P., & Leo, M. (2022). Vision-based human fall detection systems using deep learning: A review. *Computers in biology and medicine, 146*, 105626. [CrossRef]

[8] Raza, A., Yousaf, M. H., & Velastin, S. A. (2022). Human fall detection using YOLO: a real-time and AI-on-the-edge perspective. In *2022 12th International Conference on Pattern Recognition Systems (ICPRS)* (pp. 1-6). IEEE. [CrossRef]

[9] Qi, P., Chiaro, D., & Piccialli, F. (2023). FL-FD: Federated learning-based fall detection with multimodal data fusion. *Information Fusion, 99*, 101890. [CrossRef]

[10] Galvão, Y. M., Ferreira, J., Albuquerque, V. A., Barros, P., & Fernandes, B. J. (2021). A multimodal approach using deep learning for fall detection. *Expert Systems with Applications, 168*, 114226. [CrossRef]

[11] Khan, H., Usman, M. T., & Koo, J. (2025). Bilateral feature fusion with hexagonal attention for robust saliency detection under uncertain environments.

Information Fusion, 121, 103165. [CrossRef]

[12] Rassekh, E., & Snidaro, L. (2025). Survey on data fusion approaches for fall-detection. *Information Fusion, 114*, 102696. [CrossRef]

[13] Lin, B. S., Yu, T., Peng, C. W., Lin, C. H., Hsu, H. K., Lee, I. J., & Zhang, Z. (2022). Fall detection system with artificial intelligence-based edge computing. *IEEE Access, 10*, 4328-4339. [CrossRef]

[14] Lv, H., Yan, H., Liu, K., Zhou, Z., & Jing, J. (2022). Yolov5-ac: Attention mechanism-based lightweight yolov5 for track pedestrian detection. *Sensors, 22*(15), 5903. [CrossRef]

[15] Gholami, R., Jahromi, H. D., & Sedaghat, S. (2025). Design and Implementation of a Near Real-Time Human Detection Robot Using YOLO Framework and IoT Technologies. *IEEE Access*. [CrossRef]

[16] Chin, W. H., Tay, N. N. W., Kubota, N., & Loo, C. K. (2020, July). A lightweight neural-net with assistive mobile robot for human fall detection system. In *2020 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-6). IEEE. [CrossRef]

[17] Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7263-7271). [CrossRef]

[18] Zhang, X., Xie, Q., Sun, W., & Wang, T. (2025). Fall detection method based on spatio-temporal coordinate attention for high-resolution networks. *Complex & Intelligent Systems, 11*(1), 1. [CrossRef]

[19] Chen, Y., Li, W., Wang, L., Hu, J., & Ye, M. (2020). Vision-based fall event detection in complex background using attention guided bi-directional LSTM. *IEEE Access, 8*, 161337-161348. [CrossRef]

[20] Zhao, D., Song, T., Gao, J., Li, D., & Niu, Y. (2024). Yolo-fall: A novel convolutional neural network model for fall detection in open spaces. *IEEE Access, 12*, 26137-26149. [CrossRef]

[21] Islam, M. M., Tayan, O., Islam, M. R., Islam, M. S., Nooruddin, S., Kabir, M. N., & Islam, M. R. (2020). Deep learning based systems developed for fall detection: a review. *IEEE Access, 8*, 166117-166137. [CrossRef]

[22] Wu, P., Li, H., Zeng, N., & Li, F. (2022). FMD-Yolo: An efficient face mask detection method for COVID-19 prevention and control in public. *Image and vision computing, 117*, 104341. [CrossRef]

[23] Tong, K., & Wu, Y. (2022). Deep learning-based detection from the perspective of small or tiny objects: A survey. *Image and Vision Computing, 123*, 104471. [CrossRef]

[24] Chen, T., Ding, Z., & Li, B. (2022). Elderly Fall Detection Based on Improved YOLOv5s Network. *IEEE Access, 10*, 91273-91282. [CrossRef]

[25] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018, September). CBAM: Convolutional Block Attention Module. In *European Conference on Computer Vision* (pp. 3-19). Cham: Springer International Publishing. [CrossRef]

[26] Usman, M. T., Khan, H., Singh, S. K., Lee, M. Y., & Koo, J. (2024). Efficient deepfake detection via layer-frozen assisted dual attention network for consumer imaging devices. *IEEE Transactions on Consumer Electronics*. [CrossRef]

[27] Yar, H., Khan, Z. A., Ullah, F. U. M., Ullah, W., & Baik, S. W. (2023). A modified YOLOv5 architecture for efficient fire detection in smart cities. *Expert Systems with Applications, 231*, 120465. [CrossRef]

[28] Khan, H., Ullah, I., Shabaz, M., Omer, M. F., Usman, M. T., Guellil, M. S., & Koo, J. (2024). Visionary vigilance: Optimized YOLOV8 for fallen person detection with large-scale benchmark dataset. *Image and Vision Computing, 149*, 105195. [CrossRef]

[29] Khan, H., Huy, B. Q., Abidin, Z. U., Yoo, J., Lee, M., Seo, K. W., ... & Suhr, J. K. (2023). A modified yolov4 network with medium-scale challenging benchmark for efficient animal detection. *Proceedings of the Korean Institute of Next Generation Computing*, 183-186.

[30] Guerrero, J. C. E., España, E. M., Añasco, M. M., & Lopera, J. E. P. (2022). Dataset for human fall recognition in an uncontrolled environment. *Data in brief, 45*, 108610. [CrossRef]

Muhammad Jamal Ahmed received his bachelor's degree in Computer Science and IT from the University of Engineering and Technology, Peshawar, Pakistan, in 2016, and then pursued his M.Sc. in Computing Science and Engineering from Kyungpook National University, Daegu, South Korea. He is currently working as an early-stage researcher in the Department of Informatics, Universidad Politécnica de Madrid, Spain. His research interests include Artificial Intelligence, Deep Learning, and Time Series Analysis. (Email: muhammadjamal.a@upm.es)

Dr Arshad Khan received his Bachelor's degree in Computer Science from the Agriculture University Peshawar, Pakistan, in 2013, and the MS degree in computer networks from the Qurtuba University of Information Technology (QUIT), Peshawar, Pakistan, in 2016, respectively. Dr. Arshad Khan received his Ph.D. in Computer Science from the Auckland University of Technology, New Zealand. His research spans a wide range of areas, including blockchain technology, machine learning, information systems, health analytics, cybersecurity, federated learning, cloud computing, IoT security, and medical data analytics. He has contributed to several interdisciplinary projects aimed at improving healthcare infrastructure, secure data sharing, and intelligent automation. Dr. Khan has authored peer-reviewed articles in high-impact journals, including Springer Nature, MDPI Sensors, and Electronics. He is an active member of the IEEE and regularly serves as a reviewer for IEEE conferences and journals.

Taimur Ali Khan holds a Bachelor's degree in Information Technology from the University of Agriculture, Peshawar. He is currently working as a Senior Developer and IT Consultant at Saudi Media Systems. With extensive experience in software development and IT solutions, he integrates academic knowledge with real-world applications. His research interests span machine learning, deep learning, and their applications in intelligent information systems, as well as system architecture, enterprise software development, and emerging technologies. He aims to bridge practical industry expertise with cutting-edge research to develop innovative and scalable AI-driven solutions.

Bilal Ahmad graduated with distinction from the University of Malakand with a BS in computer science. His research focuses on developing and applying advanced algorithms to solve real-world problems. His research interests include Machine Learning, Deep Learning, Computer Vision and Visual Intelligence. He has a strong technical background in programming, data analysis, and AI model development. (Email: bilalahmadcs03@gmail.com)