YOLOv11s Based Real Time Fall Detection for Autonomous Surveillance

Final Project Report Submitted to
The Department of Computer Science
Faculty of Computer and Information Technology
Jordan University of Science and Technology
In Partial Fulfillment of the Requirements for the Degree of Bachelors of Science in Computer Science

Prepared by:

Fares Aljbour [164821]
Zaid Alshouha [152565]
Mohammad Khazalh [161843]
Mahmoud Obeidat [162007]

Supervisor:
Ahmad Batainah

January 2026

نموذج حقوق الملكية الفكرية لمشاريع التخرج في قسم علوم الحاسوب

يتم قراءة وتوقيع هذا النموذج من قبل الطلاب المسجلين لمشاريع التخرج في قسم علوم الحاسوب

تعود حقوق الملكية الفكرية لمشاريع التخرج ونتائجها (مثل براءات الاختراع أو أي منتج قابل للتسويق) إلى جامعة العلوم والتكنولوجيا الأردنية، وتخضع هذه الحقوق إلى قوانين وأنظمة و تعليمات الجامعة المتعلقة بالملكية الفكرية وبراءات الاختراع.

بناءا على ما سبق أوافق على ما يلي:

1) أن أحفظ كافة حقوق الملكية الفكرية لجامعة العلوم والتكنولوجيا الأردنية في مشروع التخرج.
2) أن ألتزم بوضع اسم جامعة العلوم والتكنولوجيا الأردنية و أسماء جميع الباحثين المشاركين في المشروع على أي نشرة علمية للمشروع كاملا أو لنتائجه. و يشمل ذلك النشر في المجلات و المؤتمرات العلمية عامة أو النشر على المواقع الإلكترونية أو براءات الاختراع أو المسابقات العلمية.
3) أن ألتزم بأسس حقوق التأليف المعتمدة في جامعة العلوم والتكنولوجيا الأردنية.
4) أن أقوم بإعلام الجهة المختصة في الجامعة عن أي اختراع أو اكتشاف قد ينتج عن هذا المشروع و أن ألتزم السرية التامة في ذلك و أن أعمل من خلال الجامعة على الحصول على براءة الاختراع التي قد تنتج عن هذا المشروع.
5) أن تكون جامعة العلوم والتكنولوجيا الأردنية هي المالك لأي براءة اختراع قد تنتج عن هذا المشروع و تشمل هذه الملكية حق الجامعة في إعطاء التراخيص و التسويق و البيع كمؤسسة راعية و داعمة لكافة الأنشطة البحثية. و يكون حق للطالب شمول اسمه على براءة الاختراع كأحد المخترعين، و في حال تم إعطاء تراخيص أو تسويق و بيع لأي من منتجات المشروع يمنح المخترعون بما فيهم الطالب نسبة من الإيرادات حسب تعليمات البحث العلمي في جامعة العلوم والتكنولوجيا الأردنية.

إسم الطالب	التوقيع
إسم الطالب	التوقيع
إسم الطالب	التوقيع
إسم المشرف	**التوقيع**

تاريخ ....

Real-Time Fall Detection System Using Transfer Learning with YOLOv11s

Fares Aljbour
dept. of Computer Science
Jordan University of Science and Technology
Zarqa, Jordan
faaljboor22@cit.just.edu.jo

2nd Zaid Alshouha
dept. of Computer Science
Jordan University of Science and Technology
Irbid, Jordan
zmalshouha21@cit.just.edu.jo

3rd Mohammad Khazalh
dept. of Computer Science
Jordan University of Science and Technology
Mafraq, Jordan
mmalkhazaleh22@cit.just.edu.jo

4th Mahmoud Obeidat
dept. of Computer Science
Jordan University of Science and Technology
Irbid, Jordan
msobeidat222@cit.just.edu.jo

***Abstract*—Falls are a leading cause of injury and hospitalization among the people especially the elderly creating a critical need for effective and immediate detection systems. .**

This project is motivated by the need for an automated, non-intrusive surveillance solution that ensures immediate detection without the discomfort of wearable devices.
The project methodology used in this study is the state-of-the-art "You Only Look Once" which is short for (YOLOv11s). We implemented transfer learning by fine-tuning a pre-trained model on a dataset of 17,973 images, divided into training, test and validate sets and their annotations. The system was trained to distinguish between "fall" and "non-fall" classes by drawing a bounding box around the person in the image.
Experimental results demonstrates the robustness of this approach, achieving a Mean Average Precision (mAP@0.5) of 87.6%, a precision of 83.7% and a recall of 83.4%. These finding indicate adapting pre-trained models is highly effective strategy for developing autonomous safety surveillance system.
In conclusion, this research confirms that deep learning-based object detection is a viable and effective tool for enhancing elderly safety. The proposed system offers a robust, automated alternative to traditional sensors, with potential for future integration into autonomous robotics smart home environment..

***Index Terms*—Fall Detection, Object Detection, YOLOv11, Computer Vision.**

I. PROJECT GOALS AND OBJECTIVES

In this project, our primary goal is to develop a system that can detect human falls with high accuracy.

1) To curate and preprocess a balanced dataset of fall and non-fall images.
2) To implement the YOLOv11s object detection model using transfer learning and minimize training time and computational costs.
3) To evaluate the system's performance using standard metrics (mAP, precision, Recall) and ensure it meets safety-critical standards
4) To provide an alternative to wearable sensors for elderly monitoring.

II. INTRODUCTION

Falling are a leading cause of accidental death and injury among human. The rapid detections of a fall can significantly reduce the risk of long-term health complications. While traditional solutions rely on wearable accelerometers, these devices are often rejected by users due to discomfort or forgetfulness. In this paper we present a computer vision solution using YOLOv11s.

1) **Motivation:** the global aging population is increasing, and falls represent a major health concern. Eliminate the need for hardware solutions that require active users patriation.
2) **Previous Work and Research Gap:** While earlier vision-based approaches existed, they often require training from scratch, which is computationally expensive and requires massive dataset to avoid overfitting. This research addresses these limitations by applying transfer learning to the state-of-the-art YOLOv11s architecture, significantly reducing training time while maintaining superior detection accuracy.
3) **Methodology:** In this study, YOLOv11s, an object detection model, was initialized using pre-trained weights, which were likely obtained from COCO. The technique used involves reconfiguring the output head of this model for binary classification, namely Fall or No Fall.

After that, it fine-tunes pre-trained weights on another dataset.

4) **Results:**the proposed model achieved mAP@0.5 of 87.6% on the test set, effectively distinguishing between “fall” and “non-fall”.
5) **Conclusion:** This project present automated fall detection project based on images. The proposed approach focused on visional information, aiming to capture falls in various environment. The results showed that the system is capable of distinguish between real falls and daily life activity’s. The developed model shows potentials for real-time applications such as nursing homes, health care monitoring, smart surveillance system. Future work will include videos, which will add significant improvement because it adds time to the game (temporal dimension) which helps a lot in detecting the movement of falling.
6) **Paper Structure:** The remaining of this paper is organized as follows: Section III describes Approach and Methodology, detailing the dataset preparation, the YOLOv11s architecture, and the transfer learning process used to train the model. Section III.B presents the Results, providing a comprehensive analysis of performance metrics, including confusion metrices and loss curves, along with qualitative visual outputs.

A. REVIEW AND ANALYSIS OF RELATED WORK

Here comes the previous work that relates to your work. It goes like this:
In [1] the author used YOLOv8, The model development process commences with the utilization of a dataset comprising item bounding boxes and corresponding annotations. The YOLOv8 methodology is subsequently employed to train the dataset. The study dataset consists of 2,788 raw images that have been annotated and processed using Roboflow technology. The images are categorized into three groups: the training set comprises 77% of the data, totaling approximately 2,146 images; the validation set constitutes 12%, or about 338 images; and the test set accounts for 11%, roughly 304 images. Data augmentation methods were used in the fourth stage of the Roboflow platform to increase data diversity, resulting in 19,000 images. The ideal value for improving model performance is one hundred epochs, which is how long model training was run. The model testing outcomes, carried out in the same setting as the training, show a mean average accuracy (mAP) of 90.97% and an overall accuracy of 95.36%.

In [2] they were aiming to find a fall detection tasks, a fall detection algorithm based on Yolov7 was proposed. ODConv-ELAN module was constructed in Yolov7 backbone network to replace the original ELAN structure and enhance the ability of extracting target features. Secondly, the more advanced EIoU function is used as the new boundary frame loss function, which improves the convergence speed and efficiency of the prediction frame in the process of model training. Finally, CA attention mechanism is introduced into the output terminal of the network to improve the detection performance of human fall behavior. In addition, a fall detection dataset in the campus environment was created. The accuracy P of the improved algorithm in this data set reached 94.34%, the recall rate R reached 92.34%, and the average accuracy mAP reached 94.65%, which realized the demand for more accurate human fall detection.

In [3] the asymmetric convolution blocks (ACB) convolution module is used in the Backbone network to replace the existing basic convolution to improve the feature extraction capability. Then, the spatial attention mechanism module is added to the residual structure of the Backbone network to extract more feature location information. Finally, the feature layer structure is improved to remove the feature layer for small targets so that the network can pay more attention to the semantic level information, and at the same time, the classifier is set. The proposed algorithm is trained on the URFD public dataset, and the test set is used for verification. The experimental results show that the average accuracy of all categories of the algorithm reaches 97.2%, which is increased by 3.5% compared to YOLOv5s. Thus, the proposed algorithm can accurately detect the fall behavior of the elderly.

In [4] the authors proposed that the images taken by the camera must be transmitted through a network to the back-end server for calculation. As the demand for Internet of Things increases, this architecture faces such problems as high bandwidth costs and server computing overload. Emerging methods reduce the workload of servers by transferring certain computing tasks from cloud servers to edge computing platforms. To this end, this study developed a fall detection system based on neuromorphic computing hardware, which streamlines and transplants the neural network model of the back-end computer to the edge computing platform. Through the neural network model with integer 8 bit precision deployed on the edge computing platform, the object photos obtained by the camera are converted into human motion features, and a support vector machine is then used for classification. After experimental evaluation, an accuracy of 96% was reached, the detection speed of the overall system was 11.5 frames per second, and the power consumption was 0.3 W. This system can monitor the fall events of older adults in real time and over a long period. All data were calculated on the edge computing platform. The system only reports fall events via Wi-Fi, thereby protecting the privacy of the user. they used YOLO-LW to get the bounding boxes, then they pass to get to features from the bounding boxes the first one is the Shape Aspect Ratio (SAR) and it works by dividing the Width by the Hight (W/H) if it small (tall box) the person is standing if it large (wide box) the person is lying (falling). Then these features passed to a SVM classifier to spiffy the posture (Standing, Bending, Falling), after that to a sliding window to decide if it’s a real fall or not.

[5] they see that Falls are one of the main causes of elderly injuries. If the faller can be found in time, further injury can be effectively avoided. In order to protect personal privacy and improve the accuracy of fall detection, this paper proposes a fall detection algorithm using the CNN-Casual LSTM network based on three-axis acceleration and three-

axis rotation angular velocity sensors. The neural network in this system includes an encoding layer, a decoding layer, and a ResNet18 classifier. Furthermore, the encoding layer includes three layers of CNN and three layers of Casual LSTM. The decoding layer includes three layers of deconvolution and three layers of Casual LSTM. The decoding layer maps spatio-temporal information to a hidden variable output that is more conducive relative to the work of the classification network, which is classified by ResNet18. Moreover, we used the public data set SisFall to evaluate the performance of the algorithm. The results of the experiments show that the algorithm has high accuracy up to 99.79Comparaison: It outperformed other sensor-based models like standard LSTM (99.58%) and FD-CNN (97.47%)

In [6] they work on this model YOLOv7-W6-Pose It's just not to detect a bounding box, but also to use geometric algorithm that can find: 1.Length Factor to normalize different body sizes. 2.Vertical Speed to distinguish a sudden fall from laying down. 3.Angles. then it sends Alerts integrated with Telegram bot for Real-Time notifications. The datasets consist of 1.Online Stock Video Dataset: It has a Falls and ADLs videos used to fine tune the algorithm to determine position 2.Le2i Fall Dataset: (130 video, 99 Fall, 31 ADL) Used for testing 3.Real-Time webcam Recording they preformed live tests using web cam (INSTA360 GO3)

In [7] they present an improved YOLOv8 model for fast and highly accurate human fall detection. The improvements include reducing the number of layers in the backbone of YOLOv8 and incorporating the attention mechanism in the head of the network. For training and evaluation, the CAU-CAFall dataset is used consist of 100 videos. The original YOLOv8 and improved YOLOv8 are evaluated on the same dataset, achieving mAP of 0.995. Also, the training time of the improved YOLOv8 is 0.457 hours, faster than other YOLO models. Therefore, the presented model is a practical approach for human fall detection and can serve in public places for safeguarding the elderly of our society against potential injuries. It has 2.15 million parameters, which is 64% fewer than comparable state-of-the-art models.

In [8] they proposed This work aims at proposing an affordable, non-wearable system to detect falls of people in need of care. The proposal uses artificial vision based on deep learning techniques implemented on a Raspberry Pi4 4GB RAM with a High-Definition IR-CUT camera. The CNN architecture classifies detected people into five classes: fallen, crouching, sitting, standing, and lying down. When a fall is detected, the system sends an alert notification to mobile devices through the Telegram instant messaging platform. The system was evaluated considering real daily indoor activities under different conditions: outfit, lightning, and distance from camera. Results show a good trade-off between performance and cost of the system. Obtained performance metrics are: precision of 96.4%, specificity of 96.6%, accuracy of 94.8%, and sensitivity of 93.1%. Regarding privacy concerns, even though this system uses a camera, the video is not recorded or monitored by anyone, and pictures are only sent in case of fall detection. This work can contribute to reducing the fatal consequences of falls in people in need of care by providing them with prompt attention. Such a low-cost solution would be desirable, particularly in developing countries with limited or no medical alert systems and few resources.
The comparison table of these papers "(I)"

We also Implemented other search and review papers: In [9] they done this: Reviewed a vision-based Fall detection systems. They analyzed 81 papers published between 2015 and 2020 retrieved from datasets like IEEE xplore, ScinceDirect, and PubMed. they also classified system based on Characterization (Global vs. Local vs. Depth descriptors) and Classification (Discriminative vs. Generative models). They found that there's a Technology shift from global description to local description using Deep Learning (CNNS and RNNs/LSTM). Found that Deep Learning and Depth sensors make system much more resistant to environmental noise and occlusion. The biggest flaw is unrealistic datasets simply because the datasets include young actors to simulate the falls. Also, the disconnect between the developers and elderly people actual needs. Real-world data is hard to get due the privacy concerns.

In [10] they analyzed vision-based fall detection papers published between 2020 and 2024 they categorized methods into three main types based on sensors: (Single RGB Camera, Multiple Camera, and Depth Camera). They found that the best algorithms found is CNNs are currently the most widely used and effective in classification algorithm due to their superior feature extraction capabilities They found that multi-camera system improved accuracy by reducing the blind spots (occlusion), while depth cameras effectively handle low- light conditions and privacy concerns. they find that data is rare and relays on simulated data which limits real-world applicability. Also, Occlusion and viewpoint, privacy issue RGB cameras are big privacy concerns and cost and complexity.

In these two papers they found that the main problem in fall detection is the lack of real-world-data because of privacy concerns.

B. *Significance of work*

Significance of this work can multi-faceted, addressing critical challenges in healthcare, technology, and social welfare.
1. Enhancing Elderly safety and Survival Rates Falls are major global health concern, recognized by the World Health Organization as the second leading cause of unintentional injury death worldwide, particularly among adults over 65. The "lie time" –the duration of person remains on the floor after falling—is crucial factor in recovery. Long lie times are associated with sever complications such as muscle breakdown, dehydration, and pressure scores. By providing real time detection and immediate alerts, this system significantly reduces response times, directly contributing to higher survival rates and faster recovery for fall victims.
2. Overcoming Limitations of Traditional Devices Current commercial solutions often rely wearable sensors. However, these devices suffer from a major issue: user just don't use them after while. Elderly individuals often forget to wear them,

TABLE I
THE SUMMARY OF RELATED WORK

Ref	Methodology	Results	Drawbacks
[1]	Model: YOLOv8	mAP: 92.10%	Limited Interpretability
[2]	Model: ODConv-ELAN	mAP: 94.65%	Dataset Size and Source
[3]	Model: YOLOv5s	mAP@0.5: 96.4%	Static Detection
[4]	YOLO-LW (slimed YOLOv3-tiny) / SVM	Acc: 96%	Background blending
[5]	Hybrid CNN + Casual LSTM with ResNet18	Acc: 99.79%	High computational complexity
[6]	HModel: YOLOv7-W6-Pose	Acc: 95.7%	Hardware Requirements
[7]	Model: YOLOv8	Acc: 99.5%	High computational complexity
[8]	SSD-MobileNet-v2	Acc: 94.8%	Distance Sensitivity (drops after 3m)

find them uncomfortable, or are unable to activate them during a medical emergency. This project offers a passive monitoring solution that requires no active interactive from the user. By utilizing computer vision, the system ensures continuous projection without intruding on the user's daily routine or requiring them to manage battery-powered wearable devices.
3. Supporting Independent Living and Reducing Caregiver Burden This system facilitates "aging in place" by automatically calling for help during falls, reducing the reliance on overwhelmed assisted living facilities. It significantly alleviates the constant monitoring anxiety and workload for families and caregivers, ensuring the elderly can live independently with greater security and peace of mind.
4. Advancing Cost-Effective Healthcare Technology: Rapid fall detection minimize the sever medical costs associated with delayed discovery and long-term rehabilitation. By leveraging standard cameras and efficient YOLO models, this solution offers scalable, affordable alternative to expensive proprietary hardware, effectively reducing the economic burden on public health systems while broadening access to safety monitoring.

III. APPROACH AND METHODOLOGY

A. Methodology

The methodology of this research utilize the YOLOv11s (You Only Look once) deep learning architecture, chosen for it's superior balance between inference speed and detection accuracy in real time applications. The model was trained on COCO dataset which has a 330K images splinted into three sets: Train2017 ,Val2017, Test2017.
YOLOv11s is a single stage object detection model composed of three main parts: a backbone, a neck, and a detection head. The backbone uses improved C3K2 (Cross-Stage Partial) blocks to efficiently extract hierarchical features from the input images while reducing computational cost.
The neck employs SPPF (Spatial Pyramid Pooling - Fast) to find multi scale features, allowing the model to detect objects of different sizes. "(1)"
The detection head is anchor free decoupled, predicting bounding box coordinates and class probabilities through separate branches, which improves localization accuracy and classification performance.
In the end, YOLOv11s is optimized for real time detection with a good balance between speed and accuracy.
Also, he achieved the following: "(2)" Then our work comes,

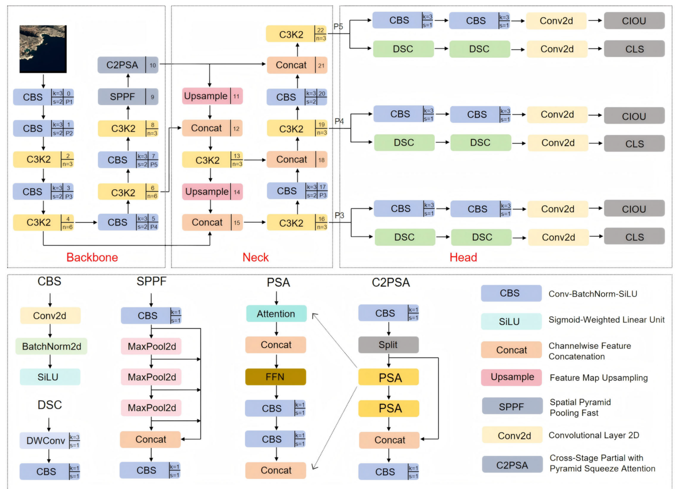

Fig. 1. YOLOv11s architecture.

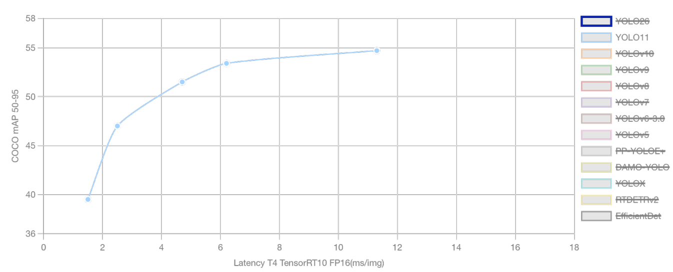

Fig. 2. YOLOv11 Performance

we simply fined tune the YOLOv11s on a dataset called Fall Detection from roboflow, it has 17,973 image that capture variety of daily life activities and real life falls.
The dataset consists of three splits:

1) **test**:
   Fall images: 580.
   Non-fall images: 604.
   Total images: 1184.

2) **train**:
   Fall images: 7890.
   Non-fall images: 7761.
   Total images: 15651.

3) **validate**:
Fall images: 610.
Non-fall images: 504.
Total images: 1114 - 24 missing annotations.

We initially intend to train the model on 50 epochs, but nested we stopped at epoch number 21. We used 0.001 as the learning rate, batch size of 32, image size 640, optimizer "Adamw".
We also, fine tune the model using Google Colab with GPU.

In each iteration, we calculated the following:

1) **Box loss**:Measures how accurate predicted bounding boxes are. The lower the better.

$$\mathcal{L}_{box} = 1 - IoU + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{gt})}{c^2} + \alpha v \quad (1)$$

2) **Classification Loss**: Measures how well object classes are predicted.

$$\mathcal{L}_{cls} = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (2)$$

3) **Distribution Focal Loss (DFL)**: Used for accurate bounding boxes regression.

$$\mathcal{L}_{DLF} = -\left((y_{i+1} - y) \log(S_i) + (y - y_i) \log(S_{i+1})\right) \quad (3)$$

And for evaluation we used the following metrics:

1) **Precision (P)**: Of all detected objects, P% are correct.

$$P = \frac{TP}{TP + FP} \quad (4)$$

2) **Recall (R)**: Of all ground truth objects, R% are correctly detected.

$$R = \frac{TP}{TP + FN} \quad (5)$$

3) **Mean Average Precision at 0.5 (mAP@50)**: Mean Average Precision calculated at IoU = 0.5, measuring detection accuracy with a moderate overlap requirement. We use IoU to find the best bounding box in case there's overlapping.

$$IoU = \frac{Area of Intres}{Area of Union} \quad (6)$$

$$AP = \int_0^1 P(R)\, dR \quad (7)$$

$$mAP@0.5 = \frac{1}{N} \sum_{i=1}^{N} AP_i^{IoU=0.5} \quad (8)$$

4) **Mean Average Precision at 0.5-0.95 (mAP@50-95)**: Mean Average Precision averaged over IoU thresholds from 0.5 to 0.95.

$$mAP@0.5:0.95 = \frac{1}{10} \sum_{k=0}^{9} mAP_{0.5+0.05k} \quad (9)$$

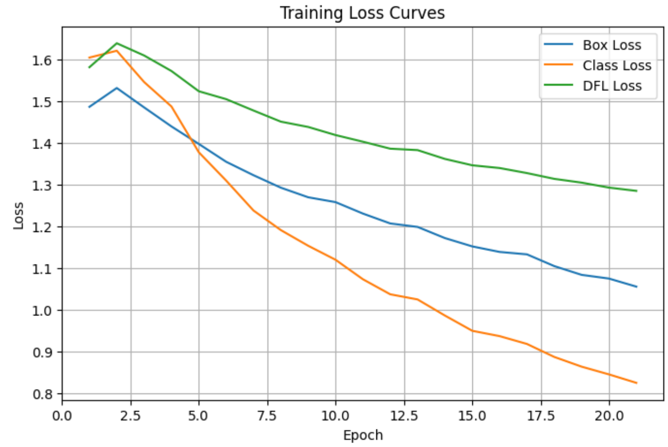

Fig. 3. Training Loss

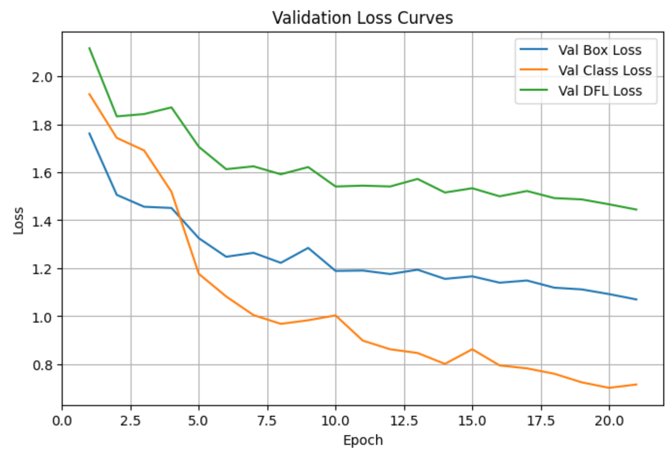

Fig. 4. Validation Loss

B. Results:

1) **Training convergence** : The training stability was noticed using loss curves as in "(3)"
we notice that the loss decreasing smoothly and stabilizing at the epoch number 19, as well as for the validation loss "(4)"
2) **Quantitative Evaluation**: The model's classification performance is detailed in the Confusion Matrix "(5)". The model achieved a high true positive rate for the Fall class, with minimal confusion against the background.
we can see the Precision and Recall curves in this figure "(6)"
3) **Qualitative Analysis**: Here is some the model predictions on some images in different environment, postures, lightning conditions and indoor-outdoor situations "(7)".

C. LOCATION AND SAFETY CONSIDERATIONS

Location of deployment the proposed fall detection system is designed primarily for indoor environment where elderly reside, such as private homes, nursing facilities, and hospital wards.

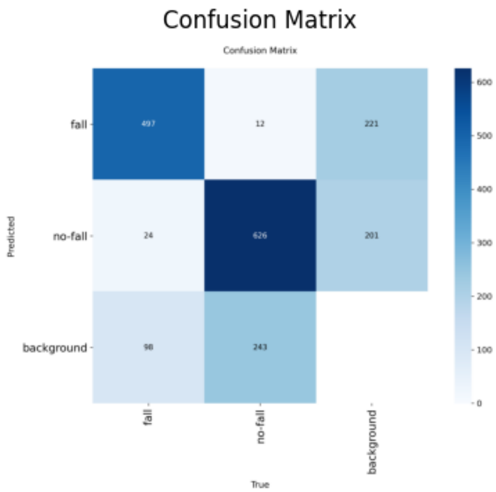

Fig. 5. Confusion Matrix

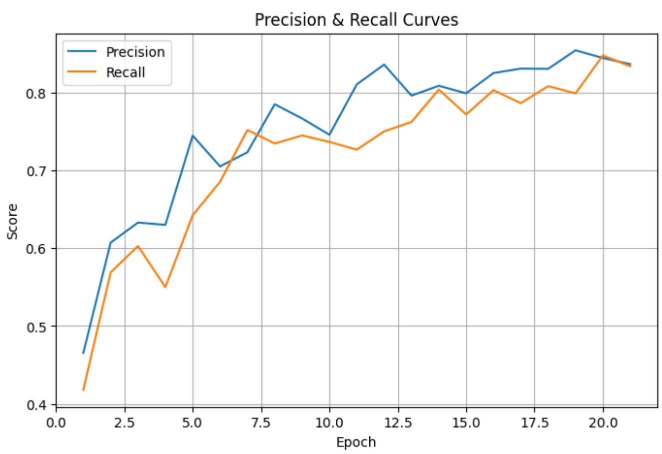

Fig. 6. Precision and Recall curves

Fig. 7. Model Prediction on some images

Camera Placement: To ensure optimal performance and minimize occlusion, the camera should be placed at high point (e.g., ceiling corners) to provide a comprehensive viewpoint of the room. This placement reduce blind spots caused by furniture and ensures the system can monitor the entire area where falls are most likely to happen.
Edge Device: The processing unit (computer running the model) should be located within the room or facility. This allows for "Edge Computing" meaning the video is processed on-site rather than being transmitted to a remote cloud server.
For Safety and Privacy:
Data security: to prevent unauthorized access to camera feed, the system should operate on a secure local network, isolated from the public internet where possible.
Considerations system reliability "Fail-Safe": In a safety critical application like fall detection, a "False Negative" (failing to detect a fall) can have severe health consequences. therefore, the system prioritizes High Recall (currently 83.4%) to ensure that actual falls are rarely missed.
Electrical safety: All hardware components (cameras, GPU/processing units) operate on standard low-voltage power supplies, ensuring there is no risk of electrical injury to the residents.

D. EXPECTED RESULTS/OUTPUTS

Expected Results The primary output of this project a robust, real time fall detection system capable of processing video feeds to identify hazardous. High Performance Detection Model: The model achieved a mean Average Precision (mAP@0.5) of 87.6%. This indicate a high level of reliability in distinguish between fall and non-fall events across the test dataset.
Accurate classification Metrics:The system is expected to maintain a balance between sensitivity and reliability: In precision (83.7%): The system minimizes false alarms, ensuring that when a fall is detected, it's highly likely to be a genuine event, and recall (83.4%): The system successfully identifies the vast majority of actual falls, which is critical for safety dependent applications.
Visual Localization and Alerting: Functionally, the system outputs processed video frames where detected individuals are highlighted with bounding boxes. Each box is labeled with the predicted class (fall, non-fall) and a confidence score, providing immediate visual context to the caregiver.

REFERENCES

[1] S. Ukampan, "Real-time fall detection for elderly care using yolov8 with a custom-built image dataset," *Engineering Access*, vol. 11, no. 2, pp. 271–276, 2025.
[2] H. Cao and J. Xu, "Fall detection algorithm based on improved Yolov7," in *International Conference on Mechatronics and Intelligent Control (ICMIC 2024)* (K. Zhang and P. Lorenz, eds.), vol. 13447, p. 1344735, International Society for Optics and Photonics, SPIE, 2025.
[3] T. Chen, Z. Ding, and B. Li, "Elderly fall detection based on improved yolov5s network," *IEEE Access*, vol. 10, pp. 91273–91282, 2022.
[4] B.-S. Lin, T. Yu, C.-W. Peng, C.-H. Lin, H.-K. Hsu, I.-J. Lee, and Z. Zhang, "Fall detection system with artificial intelligence-based edge computing," *IEEE Access*, vol. 10, pp. 4328–4339, 2022.

[5] J. Wu, J. Wang, A. Zhan, and C. Wu, "Fall detection with cnn-casual lstm network," *Information*, vol. 12, no. 10, 2021.
[6] E. Tîrziu, A.-M. Vasilevschi, A. Alexandru, and E. Tudora, "Enhanced fall detection using yolov7-w6-pose for real-time elderly monitoring," *Future Internet*, vol. 16, no. 12, 2024.
[7] A. R. Khekan, H. S. Aghdasi, and P. Salehpour, "Fast and high-precision human fall detection using improved yolov8 model," *IEEE Access*, vol. 13, pp. 5271–5283, 2025.
[8] V. Vargas, P. Ramos, E. A. Orbe, M. Zapata, and K. Valencia-Aragón, "Low-cost non-wearable fall detection system implemented on a single board computer for people in need of care," *Sensors*, vol. 24, no. 17, 2024.
[9] A. Benkaci, L. Sliman, and H. N. Dellys, "Vision-based human fall detection systems: A review," *Procedia Computer Science*, vol. 241, pp. 203–211, 2024. 19th International Conference on Future Networks and Communications/ 21th International Conference on Mobile Systems and Pervasive Computing/14th International Conference on Sustainable Energy Information Technology.
[10] J. Gutiérrez, V. Rodríguez, and S. Martin, "Comprehensive review of vision-based fall detection systems," *Sensors*, vol. 21, no. 3, 2021.