A Fallen Person Detector with a Privacy-Preserving Edge-AI Camera

Kooshan Hashemifard[1] [a], Francisco Florez-Revuelta[1] [b] and Gerard Lacey[2] [c]
[1]*Department of Computing Technology, University of Alicante, San Vicente del Raspeig, Spain*
[2]*Department of Electronic Engineering, Maynooth University, Maynooth, Ireland*

Keywords: Ambient-Assisted Living (AAL), Privacy-Preserving Camera, Fallen Person Detection, Edge-AI.

Abstract: As the population ages, Ambient-Assisted Living (AAL) environments are increasingly used to support older individuals' safety and autonomy. In this study, we propose a low-cost, privacy-preserving sensor system integrated with mobile robots to enhance fall detection in AAL environments. We utilized the Luxonis OAK-D Edge-AI camera mounted on a mobile robot to detect fallen individuals. The system was trained using YOLOv6 network on the E-FPDS dataset and optimized with a knowledge distillation approach onto the more compact YOLOv5 network, which was deployed on the camera. We evaluated the system's performance using a custom dataset captured with a robot-mounted camera. We achieved a precision of 96.52%, a recall of 95.10%, and a recognition rate of 15 frames per second. The proposed system enhances the safety and autonomy of older individuals by enabling the rapid detection and response to falls.

1 INTRODUCTION

The ageing population has led to a growing demand for technologies that can support independent living and enhance the safety of older individuals in their homes. Ambient-Assisted Living (AAL) environments, where smart devices and sensors are used to assist with daily tasks and monitor safety, have gained significant attention as a potential solution. In AAL environments, sensor systems and mobile robots can play a crucial role in ensuring the well-being and extending the independence of older persons.

The widespread adoption of these technologies has been hindered by cost and privacy concerns. To address these challenges, it is essential to develop low-cost and privacy-preserving solutions that can be integrated into existing AAL environments.

Responding quickly to falls in AAL environments is of the utmost importance. Falls are a leading cause of injury and death among older adults. Additionally, older adults who fall and cannot get up on their own may face a long wait for help to arrive. This can result in physical and emotional distress and also risk further health deterioration.

In this study, we describe the development of a low-cost, privacy-preserving fallen person detection system on an edge AI camera, Luxonis OAK-D. Our focus is not to detect the fall while it occurs but to detect when a person has fallen and cannot get up on their own. The system's performance was evaluated using a robot-mounted camera, and the results demonstrate the feasibility of the proposed solution.

The paper is organized as follows: In Section 2, we review existing research related to fallen person detection in AAL environments. Section 3 outlines the design and implementation of the system including details on data collection, network architecture, and training. In Section 4, the results of the evaluation are presented, along with a discussion. Finally, Section 5 provides a conclusion of the paper, highlighting its key contributions and offering directions for future work.

2 BACKGROUND

In this section, we discuss research on fallen person detection in AAL environments, emphasising AAL environments and highlighting some privacy-preserving approaches.

2.1 Fall Detection Technologies

Several technologies have been developed for detecting falls, including wearable devices, pressure mats,

[a] https://orcid.org/0000-0001-5086-3064
[b] https://orcid.org/0000-0002-3391-711X
[c] https://orcid.org/0000-0002-1923-6852

Hashemifard, K., Florez-Revuelta, F. and Lacey, G.
A Fallen Person Detector with a Privacy-Preserving Edge-AI Camera.
DOI: 10.5220/0012037200003476
In *Proceedings of the 9th International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2023)*, pages 262-269
ISBN: 978-989-758-645-3; ISSN: 2184-4984

and other assistive technologies.

Pendants, also known as personal emergency response systems (PERS), are wearable devices that can be worn around the neck or wrist and equipped with a fall detection button (Mann et al., 2005). When a fall is detected, an alert is automatically sent to a designated caregiver or emergency services. Pendants are easy to use, portable, and reliable, making them a popular choice for older adults who are at risk of falling. However, they can be easily misplaced or damaged, and the fall detection button may not be within reach if the person falls in a different position.

Wearable devices, such as smartwatches, activity trackers, and mobile phones, can also detect falls by analyzing movement patterns (Ramachandran and Karuppiah, 2020). The benefits of wearable devices include the ability to monitor health and activity levels, as well as provide fall detection. However, their effectiveness may be limited in complex and cluttered environments, leading to errors and frustration for users.

Pressure mats are another type of fall detection technology that can detect changes in pressure and alert a caregiver or emergency services when a fall is detected (Ariani et al., 2010). Pressure mats are easy to use and can detect falls even when someone is not wearing any other assistive technology. However, they can generate false positives when objects are placed on the mat and create a trip hazard in some AAL settings.

Another type of fall detection technology is the use of infrared sensors, which are commonly used in motion detectors and can detect changes in body heat and movement. These sensors can be placed throughout a home or care facility to monitor activity and detect falls. However, they are not as accurate as other types of technology and may generate false positives when detecting movement unrelated to a fall.

2.2 Computer Vision for Fallen Person Detection

Computer vision has been increasingly used to detect falls and fallen individuals. Survey papers such as (Alam et al., 2022) and (Gutiérrez et al., 2021) provide an overview of the different approaches used in fallen person detection. These approaches can be broadly categorized into two types: those that use body posture analysis through techniques like OpenPose (Cao et al., 2018) and those that use object detection methods like YOLO (Redmon et al., 2015) and its later versions. To evaluate the performance of these systems, researchers can use benchmark datasets such as VFP290K (An et al., 2021), IASLAB-RGBD (Antonello et al., 2017), Multicam (Auvinet et al., 2010), Le2i (Charfi et al., 2013), FPDS (Maldonado-Bascón et al., 2019), and its extension E-FPDS (Lafuente-Arroyo et al., 2022). These datasets vary in terms of their focus, with some focusing on spatio-temporal aspects like Le2i, while others incorporate depth information like Multicam or focus on large-scale outdoor settings like VFP290.

Skeleton segmentation is a common approach, and in (Asif et al., 2020) it achieves an accuracy of 84-97 % on the Multicam and Le2i datasets, albeit with a high computational load. Another study by (Antonello et al., 2017) uses a Kinect mounted on a robot to detect fallen individuals and develops the IASLAB-RGBD Fallen Person Dataset. This study claims 90% accuracy when training in one lab environment and testing on another using two SVMs to classify the skeletons of fallen persons.

(Maldonado-Bascón et al., 2019) developed a YOLO-based people detector with an SVM-based fallen person classifier with high accuracy based on their FPDS dataset. This study used a low-cost robot but required sending images to a cloud server for image processing and fall detection, serving as the predecessor to our work. (Solbach and Tsotsos, 2017) provides a more comprehensive context for fall detection using wearable and CCTV video. They used fixed versus mobile cameras and Lidar. They used ground plane analysis and pose analysis to detect fallen individuals. This study achieved 93% true positive accuracy in an office setting and 91% true positive accuracy in home settings. However, the false positive rate was mentioned but not enumerated, and the system had a high computational load.

In (Feng et al., 2020), YOLOv3 is used to detect individuals, and an LSTM is used to determine if they are falling. Similarly, (Iuga et al., 2018) uses YOLO for person detection with UAV. Another study by (Lafuente-Arroyo et al., 2022) uses the Nvidia Jetson TX2 on a robot to perform image processing and classification. They use +/- ROT90 for face detection and two fall detection approaches and test their results on the "IASLAB-RGBD" and "UR-Fall Detection Dataset (URFD)." This paper also provided the E-FPDS dataset used in our research.

2.3 Privacy Preserving Sensors

The privacy-preserving approach in image processing can be achieved by moving the processing tasks to the edge of the network. This means that the image processing is done locally on the device, such as a wearable device or a camera, before only high-level information is transmitted via the network. This approach, known as edge-AI, has gained attention from researchers, who have demonstrated its potential in various applications.

Sarabia et al. (Sarabia-Jácome et al., 2020) explored the use of wearable 3-axis accelerometers to capture motion information, which was then processed on the device to extract high-level features such as posture and activity recognition. By doing the processing on the edge of the network, they were able to protect the privacy of the user's personal information, while still achieving high accuracy in the recognition of activities.

Chen et al. (Chen et al., 2020) proposed a camera network based on Raspberry Pi (RPi) devices, but they used a server to carry out the image processing. They found that the approach was effective but not privacy-preserving since the images had to be sent across the network for processing.

Similarly, Maldonado et al. (Maldonado-Bascón et al., 2019) used an RPi camera mounted on a robot to capture images, which were then transmitted across the network for processing. However, in their more recent work, they implemented edge-AI image processing using an Nvidia Jetson TX2, which allowed them to do the processing on the device itself, thus protecting the privacy of the images.

Overall, edge-AI is a promising approach to privacy-preserving image processing since it allows for local processing of data, reducing the risk of data breaches and ensuring that only high-level information is transmitted across the network.

3 DESIGN & IMPLEMENTATION

There is a growing interest in assistive robots for monitoring and aiding older people in Ambient-assisted living (AAL) environments. However, privacy concerns related to the cloud processing of CCTV images from the home are still a barrier to their widespread adoption. Recent advances in smart cameras can address this issue by performing image analysis in the camera. The Luxonis OAK-D camera is one example of this approach. It uses deep learning models and real-time computer vision to provide a low-latency, privacy-preserving solution. We used the OAK-D cameras with custom object detection models fine-tuned on fall datasets to detect fallen persons in real-time with high accuracy.

In this study, transfer learning techniques were used to train the algorithms on the E-FPDS fallen person dataset. The trained models were then converted into the DepthAI format and deployed on the Luxonis camera. The performance of the models was evaluated in real-life conditions and compared in terms of accuracy, resource requirements, and processing speed. The findings and insights from this evaluation will be discussed in the next section.

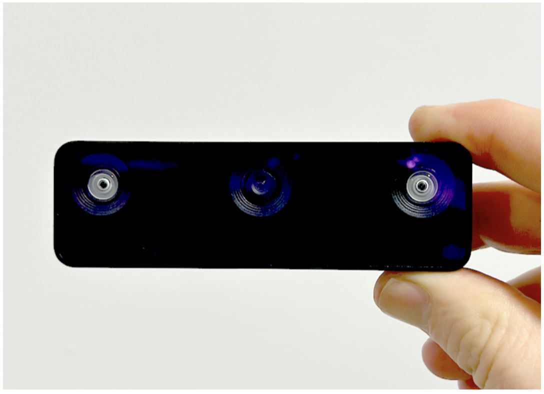

Figure 1: Luxonis OAK-D camera.

It is important to note that not only is accuracy highly critical, but we also need the camera to quickly detect a fallen person. This is particularly relevant for edge devices, which, although more affordable compared to full processors and common GPUs utilized in deep learning, still have limitations in terms of computational resources. Additionally, as a single device may be responsible for managing multiple concurrent tasks, efficiency becomes even more valuable.

The remainder of this section will cover the selection of the network architecture and the transfer learning to detect fallen persons. To achieve a high framerate on the camera using a compact network, we used a knowledge distillation step to simplify the model. We will then discuss the steps required to transfer the model onto the camera.

3.1 Selection of the Backbone Network and Pre-Trained Model

The first step in the methodology involves selecting a pre-trained network for fine-tuning. Various models and architectures for general object detection, trained on large datasets such as ImageNet (Deng et al., 2009) and COCO (Lin et al., 2014), are available in TensorFlow model zoo or PyTorch repositories. These models are usually evaluated and compared based on metrics such as speed, mean average precision, and output type.

Although all these pre-trained models can recognize the "person" class, their accuracy evaluations may not be entirely reliable because detecting fallen individuals is the primary focus of this work. Moreover, fallen individuals can assume various positions, such as twists, fetal positions, or abnormal poses, which may not be commonly represented in normal human detection datasets, leading to lower detection rates for these poses.

To select the best initial network and address this issue, experiments were conducted to evaluate the detection rates of several candidate networks on a fallen person dataset. The candidate networks included well-known and popular architectures such as MobileNet SSD (Howard et al., 2017), EfficientDet (Tan et al., 2020), CenterNet (Tan et al., 2020), and YOLO (Duan et al., 2019).

3.2 Transfer Learning for Fallen Person Detection

While pre-trained object detector networks have the ability to detect more than 90 different object classes, the specific objective of this study is to identify individuals and differentiate them based on their fallen or upright status. Some existing approaches deal with one-class detection and rely on binary classifiers or aspect ratio measurements of the bounding box for classification (Vaidehi et al., 2011; Charfi et al., 2012). In contrast, this study takes an end-to-end approach where the detector output directly produces results as two classes: fallen and upright individuals. To accomplish this, the top layer of the object detector network was replaced with two output neurons for each status (fallen and upright) while retaining the pre-trained parameters, except for the last layer. The model was trained on fallen person images using the best-performing networks from the previous step and evaluated using COCO metrics.

3.3 Knowledge Distillation for Efficiency Improvement

Deep learning has greatly advanced object detection, but state-of-the-art CNN-based networks can be computationally expensive and difficult to deploy on smaller devices, especially in real-time and multi-tasking scenarios (Zhou et al., 2019). To address this challenge, knowledge distillation has emerged as a promising approach to directly learn compact models by transferring knowledge from a large model (teacher network) to a smaller one (student network) while reducing computational costs without sacrificing validity. In this work, we investigate the use of Fine-grained Feature Imitation (Wang et al., 2019) for object detection, which is based on the idea that the local features in the object region and near its anchor location contain important information and are more crucial for the detector and how the teacher model tends to generalize. These regions are estimated, and the student model imitates the teacher on them to improve its performance.

The objective of this work is to enhance the performance of object detection for fallen person dataset by applying the knowledge distillation method. The approach comprises two stages: first, training a smaller network conventionally, and second, fine-tuning a smaller student detector by incorporating knowledge from multiple large candidate models in the previous section, which were trained on the fall dataset and referred to as teacher networks.

The smaller student detector is trained by using both ground truth supervision and feature response imitation on object anchor locations from the teacher networks. The performance of the student detector is then evaluated by comparing its results before and after knowledge distillation. The aim of this study is to demonstrate the effectiveness of knowledge distillation in improving object detection for fallen person detection, and thereby achieving the same or acceptable performance with a smaller and more efficient network.

3.4 Deploying on Luxonis Camera

When deploying a trained model to a camera device, it is important that the model is compatible with the supported frameworks such as Caffe, MXNet, TensorFlow, TensorFlow 2 Keras, Kaldi, and ONNX. However, these models cannot be used directly by the DepthAI platform. Instead, they need to be converted into a MyriadX format blob file, which optimizes them for the best inference on the MyriadX VPU processor inside the device. The conversion process involves two steps: first, the model is converted to the OpenVINO Intermediate Representation (IR), and then the IR model is compiled into a MyriadX blob file using either an online server or a local conversion tool. It is crucial to ensure that all the layers and loss functions are compatible and supported by OpenVINO to make the conversion successful.

4 EXPERIMENTS

4.1 Dataset

While many fall detection datasets focus on the entire falling process, which may not be practical for a patrol robot to observe, it is crucial to consider datasets that align with the goal of detecting fallen individuals.

For fallen person detection in this work, we used the E-FDPS dataset (Maldonado-Bascon et al., 2019). This dataset contains 6982 images captured in indoor

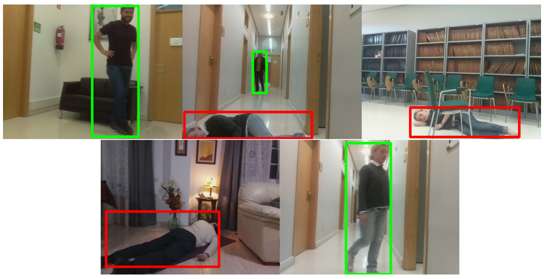

Figure 2: E-FDPS sample images.

environments, with 5023 instances of falls and 2275 instances of non-falls in various scenarios, including variations in pose, size, occlusions, and lighting. The dataset also includes an "Elderly set", which consists of 272 fallen individual images captured in a home environment that is highly relevant to our task. We divided the dataset into different sets for training, validation, and evaluation purposes.

4.2 Results

To identify an effective backbone network for fallen person detection, pre-trained models were evaluated on the Elderly set. The performance of each model was measured by calculating the detection accuracy on the "person" class. Table 1 summarizes the results of this evaluation, comparing the person detection rates of different pre-trained models. The YOLO architecture was found to have the highest average precision and average recall, indicating that it is better equipped to handle the variability in pose and appearance of fallen persons. These findings suggest that pre-trained YOLO models are well-suited for fallen person detection.

In the next experiment, all of the previously evaluated models were fine-tuned on the E-FDPS training set and evaluated on the Elderly set to validate the conclusion that YOLO models are more suitable for detecting fallen persons. The performance of these fine-tuned models was evaluated not only in terms of detecting fallen bodies, but also in terms of correctly classifying them. The results of this evaluation are shown in Table 2, which further supports the use of YOLO models, as they perform more balanced in terms of precision and recall.

On the other hand, although the other networks may perform well on one metric, their results are skewed and unsuitable for real-case scenarios. This highlights the importance of selecting a model that can perform well on precision and recall to achieve accurate and reliable fallen person detection.

This study tested different versions of YOLO models, such as YOLOv5 and 6 in small, medium,

Table 1: Initial performance evaluation of pre-trained models over detecting the "person" class in Elderly set.

Model	Detection Precision	Detection Recall	F1-score	Average IoU
MobileNet SSD	84.02%	71.04%	76.98%	72.92%
EfficientDet	88.64%	76.32%	82.02%	73.41%
CenterNet	83.33%	51.53%	63.68%	**73.57%**
YOLOv6 S	90.61%	74.22%	81.60%	72.72%
YOLOv6 M	**95.50%**	72.35%	82.32%	72.43%
YOLOv6 L	92.50%	**85.71%**	**88.97%**	72.82%

Table 2: Fallen person detection performance on Elderly set by fine-tuned models.

Model	Precision	Recall	F1-score	Average IoU
MobileNet SSD	85.41%	86.52%	85.96%	49.25%
EfficientDet	75.28%	**94.88%**	83.95%	53.28%
CenterNet	92.60%	92.50%	92.54%	67.75%
YOLOv6 S	85.30%	88.5%	86.87%	66.49%
YOLOv6 M	88.10%	86.50%	87.29%	70.41%
YOLOv6 L	**98.42%**	92.25%	**95.23%**	**71.44%**

and large sizes. The evaluation was performed on the test set of the E-FDPS dataset which included 140 non-fall and 719 fall instances. The results, presented in Table 3, show that the YOLOv6 large version generally performs better in terms of performance, but it requires more computational resources. To make the smallest YOLO model more effective, the knowledge distillation approach was used, where the YOLOv6 large model was used as a teacher network. The performance of the YOLOv5 small model was compared before and after the knowledge distillation approach, which showed an improvement in its performance. Given the difference in the number of parameters, this approach makes the small YOLO model more practical for deployment on edge devices with limited resources.

4.3 Implementation Details

In this work, the implementation of YOLO networks was performed using PyTorch framework. On the other hand, the remaining networks were trained using TensorFlow Object Detection API and its corresponding repository. The training was conducted on a single NVIDIA GeForce GTX 3090Ti GPU for 150 epochs, using ADAM optimization algorithm with an initial learning rate of 0.001 and decay rate of 0.96. The duration of the training process varied between 6 to 12 hours depending on the network size.

Table 3: Fallen person detection performance of YOLO versions on E-FDPS test set and paramter numbers.

Model	Precision	Recall	mAP50	mAP50-95	Parameter Number
YOLOv6 L	98.02%	98.25%	98.31%	61.84%	59.6M
YOLOv6 M	96.55%	96.63%	97.41%	62.86%	34.9M
YOLOv6 S	95.71%	96.78%	97.32%	60.04%	18.5M
YOLOv5 L	97.67%	96.29%	98.65%	62.91%	46.5M
YOLOv5 S	93.91%	91.17%	95.11%	56.80%	7.2M
YOLOv5 S-v6 L	96.52%	95.10%	97.22%	61.52%	7.2M

4.4 Evaluation on the in-the-Wild Dataset

A set of in-the-wild videos was employed to assess the model's performance in real-world settings. These videos, captured using the Luxonis OAK-D camera, featured individuals who had fallen and those who had not in an indoor office environment. The dataset comprised 38 videos recorded from various angles by the patrol robot, featuring diverse actors of different ages and genders and displaying different poses and falls from various perspectives. The RoboFlow (Dwyer and Nelson, 2022) platform was used to annotate the videos, resulting in 500 high-quality images, including 368 falls and 132 non-falls. This step was necessary to determine the model's efficacy in real-world conditions while accounting for potential performance reduction caused by changes in the source domain or model conversion. The results are presented in Table 4.

Table 4: Fallen person detection performance of YOLO versions on in-the-wild set and the FPS on camera.

Model	Precision	Recall	mAP50	mAP50-95	FPS
YOLOv6 L	99.12%	98.64%	99.45%	65.58%	1
YOLOv6 M	94.95%	98.21%	98.64%	65.33%	5
YOLOv6 S	92.71%	95.22%	98.10%	62.67%	10
YOLOv5 L	93.48%	88.36%	94.89%	58.12%	1
YOLOv5 S	83.67%	86.44%	93.37%	46.77%	15
YOLOv5 S-v6 L	94.46%	91.83%	95.04%	57.29%	15

Figure 3: Samples of in-the-wild recording.

5 DISCUSSION AND CONCLUSIONS

In this study, we have evaluated the effectiveness of different pre-trained and fine-tuned object detection models for detecting fallen persons, with the ultimate goal of enabling assistive robots to detect and respond to fall incidents. Through a series of experiments, we have shown that YOLO-based models generally outperform other models in terms of detection accuracy and balanced precision and recall, especially YOLOv6 large which offers high performance but with increased computational requirements. Moreover, we have demonstrated that through knowledge distillation, we can achieve close to realtime performance levels using smaller models, such as YOLOv5 small, which are more practical for deployment on edge devices with limited resources.

To test the models in real-world scenarios, we used a set of in-the-wild videos recorded by a patrol robot, resulting in 500 high-quality images, including 368 falls and 132 non-falls, and we achieved promising results that highlight the potential for the use of these models in practical applications. Overall, our results show that choosing the appropriate model for detecting fallen persons depends on the specific use case and the trade off between computational speed and error tolerance. In more critical and high-risk scenarios, stronger models may be preferred. In contrast, smaller models with faster computational speed may be more suitable in scenarios where multiple tasks need to be performed by a single edge device.

The present study aimed to evaluate the performance of RGB-based object detection methods for fallen person detection using the Luxonis OAK-D edge AI camera. Although the results obtained were promising, it should be noted that the camera is capable of capturing depth data as well. Therefore, this work could serve as a baseline for future studies that incorporate both RGB and depth-based methods, which could potentially lead to the development of a highly reliable and robust fallen person detection system that could be implemented at an industrial scale.

ACKNOWLEDGEMENTS

This work has been part supported by the visuAAL project on Privacy-Aware and Acceptable Video-Based Technologies and Services for Active and Assisted Living (https://www.visuaal-itn.eu/) funded by the EU H2020 Marie Skłodowska-Curie grant agreement No. 861091. The project has also been part supported by the SFI Future Innovator Award SFI/21/FIP/DO/9955 project Smart Hangar. Thanks also to Luke Casey and Chizubere Lovelyn Ulogwara for their help in deep learning and data capture.

REFERENCES

Alam, E., Sufian, A., Dutta, P., and Leo, M. (2022). Vision-based human fall detection systems using deep learning: A review. *Computers in Biology and Medicine*, 146:105626.

An, J., Kim, J., Lee, H., Kim, J., Kang, J., Kim, M., Shin, S., Kim, M., Hong, D., and Woo, S. S. (2021). VFP290K: A Large-Scale Benchmark Dataset for Vision-based Fallen Person Detection. In *Thirty-Fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.

Antonello, M., Carraro, M., Pierobon, M., and Menegatti, E. (2017). Fast and robust detection of fallen people from a mobile robot. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4159–4166.

Ariani, A., Redmond, S. J., Chang, D., and Lovell, N. H. (2010). Software simulation of unobtrusive falls detection at night-time using passive infrared and pressure mat sensors. In *2010 Annual international conference of the IEEE engineering in medicine and biology*, pages 2115–2118. IEEE.

Asif, U., Mashford, B., Von Cavallar, S., Yohanandan, S., Roy, S., Tang, J., and Harrer, S. (2020). Privacy Preserving Human Fall Detection using Video Data. In Dalca, A. V., McDermott, M. B., Alsentzer, E., Finlayson, S. G., Oberst, M., Falck, F., and Beaulieu-Jones, B., editors, *Proceedings of the Machine Learning for Health NeurIPS Workshop*, volume 116 of *Proceedings of Machine Learning Research*, pages 39–51. PMLR.

Auvinet, E., Rougier, C., Meunier, J., St-Arnaud, A., and Rousseau, J. (2010). Multiple cameras fall dataset. *DIRO-Université de Montréal, Tech. Rep*, 1350:24.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., and Sheikh, Y. (2018). Openpose: Realtime multi-person 2d pose estimation using part affinity fields.

Charfi, I., Miteran, J., Dubois, J., Atri, M., and Tourki, R. (2012). Definition and performance evaluation of a robust svm based fall detection solution. In *2012 eighth international conference on signal image technology and internet based systems*, pages 218–224. IEEE.

Charfi, I., Miteran, J., Dubois, J., Atri, M., and Tourki, R. (2013). Optimized spatio-temporal descriptors for real-time fall detection: comparison of support vector machine and Adaboost-based classification. *Journal of Electronic Imaging*, 22(4):041106.

Chen, Y., Kong, X., Meng, L., and Tomiyama, H. (2020). An Edge Computing Based Fall Detection System for Elderly Persons. *Procedia Computer Science*, 174:9–14.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.

Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., and Tian, Q. (2019). Centernet: Keypoint triplets for object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6569–6578.

Dwyer, B. and Nelson, J. (2022). Roboflow (version 1.0) [software].

Feng, Q., Gao, C., Wang, L., Zhao, Y., Song, T., and Li, Q. (2020). Spatio-temporal fall event detection in complex scenes using attention guided LSTM. *Pattern Recognition Letters*, 130:242–249.

Gutiérrez, J., Rodríguez, V., and Martin, S. (2021). Comprehensive Review of Vision-Based Fall Detection Systems. *Sensors*, 21(3):947.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

Iuga, C., Drăgan, P., and Bușoniu, L. (2018). Fall monitoring and detection for at-risk persons using a UAV. *IFAC-PapersOnLine*, 51(10):199–204.

Lafuente-Arroyo, S., Martín-Martín, P., Iglesias-Iglesias, C., Maldonado-Bascón, S., and Acevedo-Rodríguez, F. J. (2022). RGB camera-based fallen person detection system embedded on a mobile platform. *Expert Systems with Applications*, 197:116715.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer.

Maldonado-Bascón, S., Iglesias-Iglesias, C., Martín-Martín, P., and Lafuente-Arroyo, S. (2019). Fallen People Detection Capabilities Using Assistive Robot. *Electronics*, 8(9):915.

Maldonado-Bascon, S., Iglesias-Iglesias, C., Martín-Martín, P., and Lafuente-Arroyo, S. (2019). Fallen people detection capabilities using assistive robot. *Electronics*, 8(9):915.

Mann, W. C., Belchior, P., Tomita, M. R., and Kemp, B. J. (2005). Use of personal emergency response systems by older individuals with disabilities. *Assistive technology*, 17(1):82–88.

Ramachandran, A. and Karuppiah, A. (2020). A survey on recent advances in wearable fall detection systems. *BioMed research international*, 2020.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2015). You only look once: Unified, real-time object detection.

Sarabia-Jácome, D., Usach, R., Palau, C. E., and Esteve, M. (2020). Highly-efficient fog-based deep learning AAL fall detection system. *Internet of Things*, 11:100185.

Solbach, M. D. and Tsotsos, J. K. (2017). Vision-Based Fallen Person Detection for the Elderly. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 1433–1442.

Tan, M., Pang, R., and Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10781–10790.

Vaidehi, V., Ganapathy, K., Mohan, K., Aldrin, A., and Nirmal, K. (2011). Video based automatic fall detection in indoor environment. In *2011 International Conference on Recent Trends in Information Technology (ICRTIT)*, pages 1016–1020. IEEE.

Wang, T., Yuan, L., Zhang, X., and Feng, J. (2019). Distilling object detectors with fine-grained feature imitation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4933–4942.

Zhou, L., Wen, H., Teodorescu, R., and Du, D. H. (2019). Distributing deep neural networks with containerized partitions at the edge. In *Hotedge*.