

Article

AI-Driven Privacy in Elderly Care: Developing a Comprehensive Solution for Camera-Based Monitoring of Older Adults

Chang-Yueh Wang * and Fang-Suey Lin

Graduate School of Design, National Yunlin University of Science and Technology, Yunlin 64002, Taiwan; linfs@yuntech.edu.tw

* Correspondence: d10930015@gmail.yuntech.edu.tw; Tel.: +886-5-534-2601 (ext. 6040)

Featured Application: An artificial intelligence-driven elderly monitoring system that preserves the visual privacy of the subjects under surveillance while providing essential visual context and essential wellness status to caregivers.

Abstract: The need for privacy in elderly care is crucial, especially where constant monitoring can intrude on personal dignity. This research introduces the development of a unique camera-based monitoring system designed to address the dual objectives of elderly care: privacy and safety. At its core, the system employs an AI-driven technique for real-time subject anonymization. Unlike traditional methods such as pixelization or blurring, our proposed approach effectively removes the subject under monitoring from the scene, replacing them with a two-dimensional avatar. This is achieved through the use of YOLOv8, which facilitates accurate real-time person detection and pose estimation. Furthermore, the proposed system incorporates a fall detection algorithm that utilizes a residual causal convolutional network together with motion features of persons to identify emergency situations and promptly notify caregivers in the event of a fall. The effectiveness of the system is evaluated to emphasize its advanced privacy protection technique and fall detection capabilities using several metrics. This evaluation demonstrates the system's proficiency in real-world applications and its potential to enhance both safety and privacy in elderly care environments.



Citation: Wang, C.-Y.; Lin, F.-S. AI-Driven Privacy in Elderly Care: Developing a Comprehensive Solution for Camera-Based Monitoring of Older Adults. *Appl. Sci.* **2024**, *14*, 4150. <https://doi.org/10.3390/app14104150>

Academic Editor: Yoshiyasu Takefuji

Received: 13 April 2024

Revised: 11 May 2024

Accepted: 12 May 2024

Published: 14 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: artificial intelligence; computer vision; elderly care; elderly monitoring; fall detection; machine learning; privacy protection; video surveillance

1. Introduction

The global demographic landscape is experiencing a notable shift, with a rapidly growing elderly population [1]. This demographic change also brings an increase in health-related risks among elderly people, most notably a heightened incidence of fall accidents. The World Health Organization reports that around 28–35% of elderly persons suffer a fall annually [2], posing significant risks, especially for those living alone. In such cases, a fall can lead to severe, potentially life-threatening consequences if assistance is not promptly provided. Therefore, in recent years, many systems have been introduced for effective monitoring and rapid response systems to ensure the safety and well-being of the elderly population.

The development of various monitoring technologies, predominantly focused on fall detection and prevention [3–5], has been a key response to this challenge. These systems are categorized into wearable device-based, radar-based, and camera-based solutions [6]. While camera-based solutions are popular and easy to install, their implementation raises substantial privacy concerns. With the elderly population valuing their independence and dignity, the intrusive nature of continuous monitoring can be a source of discomfort and stress. Therefore, smart home technologies like camera monitoring systems, increasingly adopted to support aging in place, must balance effective monitoring with privacy protection [7].

In this context, the preferences of older adults for privacy-protective surveillance methods, particularly those utilizing avatar-based systems, become highly relevant. According to one study [7], older adults naturally have concerns regarding both visual and behavioral privacy, with a stronger inclination towards maintaining visual privacy. A significant majority of healthy older adults show a preference for technologies that transform their actual images into avatars, rather than more invasive methods. These avatar-based systems effectively maintain the functionality of surveillance—essential for safety and security—while addressing the crucial concerns of visual privacy. By replacing real images with avatars, such systems can reduce the feelings of intrusion and discomfort associated with being constantly monitored, thereby aligning with elderly people's expectations for dignity and independence. Such innovative approaches in smart home technologies not only support the practical aspects of aging in place but also prioritize the emotional and psychological well-being of elderly individuals, thus promoting greater acceptance among older adults.

Building on these challenges, our research presents a novel contribution to the field of elderly care, particularly in addressing the delicate balance between effective monitoring and preserving privacy. First, we introduce an advanced artificial intelligence (AI)-driven camera-based monitoring system based on the YOLOv8 model, employing innovative techniques for real-time subject anonymization. This approach significantly mitigates privacy concerns by replacing traditional invasive video feeds with non-intrusive two-dimensional (2D) avatars. Second, our system incorporates a fall detection algorithm developed using temporal variation in various motion features and a causal convolutional network classifier, ensuring timely responses to emergencies. Our work contributes to the technological and ethical advancements in elderly care and thoughtfully addresses the critical ethical considerations of privacy and dignity. This research represents a significant step towards blending AI technology with a respect for the privacy and dignity of elderly people, emphasizing the importance of safeguarding personal space and autonomy in the development of elderly care systems.

2. Related Works

Elderly monitoring systems have significantly benefited from advancements in sensors, surveillance systems, and the Internet of Things (IoT), which facilitate real-time monitoring and enhanced connectivity. These technologies provide a detailed overview of an elderly individual's environment and health status [5]. Following this, developments in AI, specifically in person detection [6], pose estimation [7], and action recognition [8], have greatly improved the accuracy of these systems. These AI-driven techniques allow for precise identification and automated analysis of elderly movements, aiding in early detection of potential health risks.

2.1. Camera Surveillance Systems

Camera surveillance systems have emerged as a vital component in elderly monitoring, leveraging advanced image processing and AI algorithms. The integration of high-resolution cameras with real-time video analytics has enabled more detailed observation of elderly activities, ensuring safety by providing prompt alerts to caregivers [8–10]. Innovations in this domain include the use of infrared or thermal imaging to enhance monitoring capabilities during low-light or no-light conditions [11–15]. Additionally, the implementation of machine learning models, such as convolutional neural networks (CNNs), has significantly improved the ability of these systems to recognize and interpret complex human behaviors [16,17]. This progress has been vital in not just detecting emergencies like falls or unusual inactivity, but also in understanding daily patterns and deviations that might indicate health or well-being concerns.

Furthermore, the adoption of edge computing in camera surveillance systems marks a crucial development in the area of elderly surveillance systems. By processing data locally, smart camera systems are more cost-effective, reduce latency, and decrease response time,

which is essential for prompt interventions in emergency situations [18–21]. Moreover, edge computing aids in addressing privacy concerns, as sensitive data can be processed and stored locally, minimizing the need for transmission to remote servers. This feature is particularly important in elderly care, where privacy and data security are paramount. However, recent advances in edge computing rely on minimal data transmission, meaning that only essential information, such as alerts or behavioral data, is transmitted to servers and caregivers. While this method effectively addresses visual privacy concerns [7], it does not provide caregivers with a complete picture of a scene. This limitation is of the utmost concern, since fall detection and behavioral analysis models, although useful, are not infallible and are prone to false alarms and missed incidents.

2.2. Artificial Intelligence in Elderly Care

In elderly care, the application of AI has seen a transformative breakthrough, particularly through the deployment of sophisticated algorithms and neural network architectures for pose detection, fall detection, and activity monitoring. Human pose estimation [22,23], serving as the core component of such automated analysis systems, lays the foundation for subsequent analyses like fall detection and activity monitoring. State-of-the-art algorithms such as PoseNet [24] and BlazePose [25] have revolutionized pose estimation with 3D and 2D image data using CNN architectures, significantly impacting research and practical applications in this field. These algorithms excel in accurately identifying and tracking human body positions in real time, which is critical for monitoring elderly behavior.

Fall detection models rely heavily on object detection or pose estimation models. These models typically utilize a combination of spatial and temporal data processed through advanced machine learning algorithms like CNNs and Long Short-Term Memory networks (LSTMs) [26,27]. The spatial aspect involves analyzing the posture and position of an individual, while the temporal component examines movement patterns over time. However, some models also use direct classification approaches using pose/skeleton data, as presented in [28–30], where the pose of a person is used to classify whether or not a fall has occurred. Moreover, other research incorporates object detection models and bounding box techniques for fall classification, as shown in [31,32]. Recent research has employed diverse and innovative strategies [33,34] to enhance the reliability and accuracy of fall detection, a critical component in ensuring the safety of elderly people.

Pose estimation models have also led research in the direction of activity recognition. Recent advancements in machine learning (ML) allow these models to utilize skeletal information derived from pose estimation [35], coupled with spatial-temporal data and/or other robust features, to effectively classify human activities [17,36]. As shown in the analysis presented in [6], vision-based activity recognition models have shown greater robustness compared to motion sensor-based monitoring systems. This is primarily due to the distinct visual features that can be separated into a variety of actions. In contrast, motion sensor features may display similar speeds and patterns across different actions, leading to potential ambiguities in classification. Therefore, several elderly monitoring systems utilize these models to automate the behavioral monitoring of elderly subjects [12]. Some research also combines pose data with handcrafted features to achieve fall detection classification, as presented in the method proposed in [37], where distances and angle features between pose keypoints between consecutive frames were used to detect falls. Similarly, in [38], other biometric features from pose keypoints were extracted and used to train a random forest model to classify falls.

Despite the advancements in skeleton-based fall detection methods [39], there remain significant gaps in the technology, particularly in terms of privacy and user-centric approaches for elderly surveillance. Enhancing the privacy aspects of monitoring systems, for instance, does not just require adhering to ethical standards but also increasing the acceptance and comfort level of elderly individuals who are being monitored. Privacy-preserving methods, such as anonymization of video data, can foster greater trust and willingness among elderly users [40–44].

2.3. Subject Anonymization

In elderly care, privacy-preserving monitoring systems often utilize ambient sensors or wearables to provide passive monitoring, thereby addressing the privacy concerns posed by camera surveillance [42]. These technologies capture essential data regarding the physical well-being and activity levels of elderly people without the invasiveness [45–49] associated with video recordings. However, while sensors and wearables offer a less intrusive means of observation, they may not always provide the comprehensive visual context necessary for certain caregiving tasks. Furthermore, several vision-based monitoring systems [10,26,33,50,51] have proven to be more capable of detecting emergency situations, such as falls, making them more popular among caregivers and less inconvenient for elderly people than wearables.

Subject anonymization is a technique widely used in computer vision for enhancing visual privacy and making it harder to identify subjects under surveillance. Early works in this field focused on methods such as face identification and warping [52] to render the faces of subjects unidentifiable. Some common anonymization methods [53] in video surveillance, including blanking, obfuscation, abstraction, and encryption, safeguard personal identities by first detecting individuals in video feeds and then applying privacy-enhancing techniques to these areas. Blanking replaces sensitive regions with background imagery, obfuscation (via pixelation, blurring, or warping) and abstraction (employing edge detection, avatars, or silhouettes) modify visual details to prevent recognition, and encryption ensures data privacy by restricting access to selected areas with decryption keys, offering strong privacy safeguards. Recent innovations in visual privacy techniques [54,55] focus on whole-body or face detection/segmentation, altering visual representations through methods like blurring, pixelation [56], and embossing; replacing them with silhouettes, skeleton representations, and three-dimensional avatars; or complete invisibility, further advancing privacy protection. Yet the literature indicates that many of these methods might necessitate manual image manipulation or entail substantial processing time, limiting their applicability in real-time systems.

Other privacy techniques specifically designed for elderly monitoring systems include the use of depth images, rather than solely RGB images, to develop fall detection systems [40,57]. Depth images, by only revealing spatial information, enable accurate fall detection without compromising visual privacy. Infrared/thermal camera-based monitoring systems [11,12,58,59], on the other hand, capture heat signatures, further reducing visible details and enhancing privacy for elderly people. While these methods offer effective monitoring while prioritizing individual privacy, they also restrict the visual information often needed for contextual understanding in caregiving.

Given the outlined challenges and advancements in privacy-preserving technologies for elderly monitoring, there is a pressing need for automated systems capable of providing efficient subject anonymization while ensuring effective surveillance. The system proposed in this paper uses the latest advancements in computer vision and AI to automatically protect individuals' identities while still keeping a close eye on their well-being. It is designed to dynamically adapt to the complexities of real-world scenarios, ensuring that privacy does not compromise the capacity to identify and react to emergencies. With a core focus on the fundamental objectives of surveillance—safety and well-being—the proposed system introduces a privacy-aware approach that retains the integrity of elderly monitoring. This innovative solution effectively meets the dual demands of preserving privacy and delivering comprehensive care, creating an ethical surveillance system through the application of state-of-the-art technology.

3. Materials and Methods

This section details the systematic approach undertaken to design, implement, and evaluate the proposed surveillance system. It describes the technological frameworks and procedural steps that collectively facilitate the achievement of a privacy-preserving elderly monitoring system.

3.1. System Design

The system design, as shown in Figure 1, integrates a Wi-Fi-enabled Internet protocol (IP) camera (TP-Link, Hong Kong), a high-speed Wi-Fi router (TP-Link, Hong Kong), and a desktop server computer (ASUS, Taipei City, Taiwan) equipped with an NVIDIA RTX 3050 graphics processing unit (GPU) (Gigabyte Technology, New Taipei City, Taiwan), an Intel i5 processor (INTEL, Hong Kong) with 12 cores, and 32 GB of random-access memory (RAM) (Micron, Singapore). This setup captures and wirelessly transmits a video stream for real-time processing, including a machine learning model and a privacy-preserving anonymization technique. The processed video stream and elderly status are then securely uploaded to the cloud and made accessible to caregivers through a smartphone application.

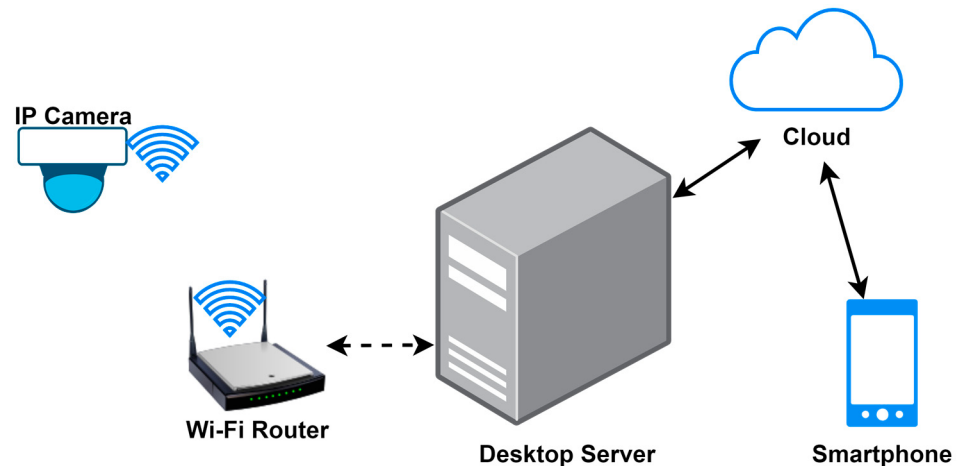


Figure 1. The architecture of the proposed surveillance system.

The system begins with an IP camera capturing live video in the living space of an elderly individual, as shown in Figure 2. This video stream is processed using YOLOv8 [60], a machine learning model that identifies the presence of a person in the scene. Once detection occurs, the process splits into two main functions: ensuring the subject's privacy and analyzing their posture for safety concerns.

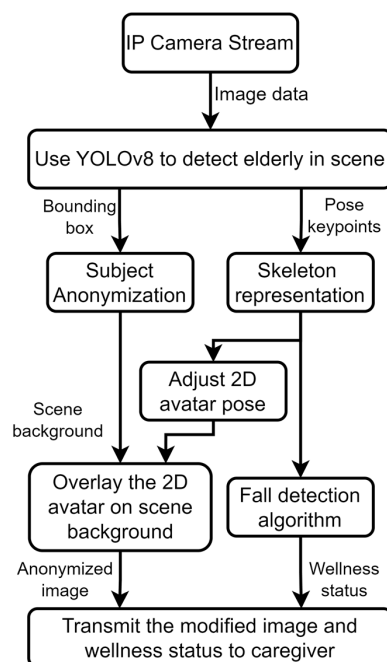


Figure 2. The flowchart of the proposed system's working principle.

To protect privacy, the system employs subject anonymization by using the bounding box information of the detected individual and replacing their visible image with background. This method maintains the individual's privacy by hiding their appearance in the video feed. Concurrently, pose keypoints are extracted using YOLOv8 to form a skeleton representation of the person. This skeleton aids in adjusting the 2D avatar's posture to accurately reflect the person's movements, ensuring a realistic representation.

The skeleton representation is crucial for the fall detection algorithm, which analyzes the person's movements and posture. If the algorithm detects a pattern indicative of a fall, it immediately triggers an alert for assistance. Finally, the system securely transmits the anonymized video stream, now featuring the 2D avatar, along with any alerts from the fall detection algorithm to the caregiver through the cloud. In the event of a fall, the system sends an alert to the caregiver that includes both the fall alarm and a real-time image of the scene without modifications. This ensures that the caregiver receives comprehensive information about the fall and the actual condition of the monitored person, enabling a prompt response.

3.2. Person Detection and Subject Anonymization

For a system focused on crucial safety tasks like monitoring elderly people, picking the right machine learning model is important. It needs to be not only real-time and reliable but also computationally efficient. After looking into several options, this study chose YOLOv8 for its strong ability to detect people and estimate their poses in real time. Developed by Ultralytics and released in January 2023, YOLOv8 [60] has quickly become a favorite in both practical applications and academic research, due to its reliable and accurate detection and classification capabilities. Several researchers have used YOLOv8 for multiple tasks, including multi-object pedestrian tracking [61], student behavior detection in classroom [62], road defect detection [63], safety helmet detection [64], and human pose estimation [65].

The YOLOv8 architecture, shown in Figure 3, is designed to be efficient and not demanding on system resources, meaning that it can run smoothly on various types of hardware, from personal computers to more powerful GPU-based systems. This flexibility is vital for ensuring that the proposed elderly monitoring system can operate effectively across different settings without needing specialized equipment. The YOLOv8's architecture consists of IP camera-to-fine (c2f) modules, which allow the model to first identify larger, more general areas where a person might be (the "coarse" part) and then refine its focus to detect finer details within those areas (the "fine" part). This approach is particularly effective for identifying not just the bounding boxes that indicate where a person is in a video but also the specific pose keypoints or body joints.

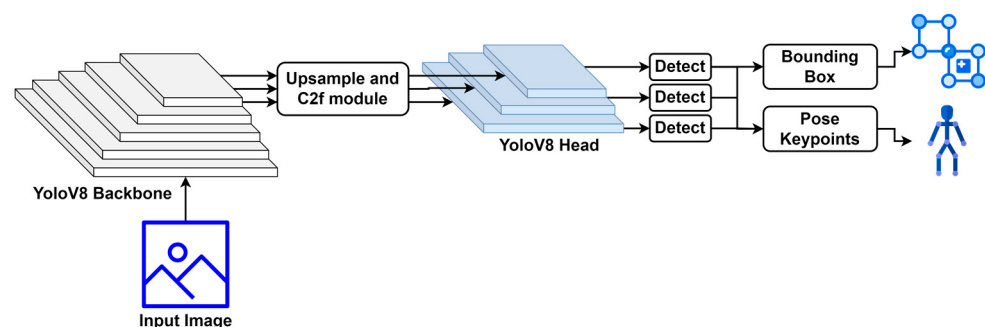


Figure 3. YOLOv8 model used for person bounding box detection and pose keypoint estimation.

In our proposed system for elderly monitoring, the subject anonymization process was developed using the YOLOv8 model's capabilities. The algorithm, as shown in Figure 4, initiates by setting a background image filled entirely with black pixels, creating an empty buffer for the background of subsequent frames. As the system processes each frame of the video stream, it employs YOLOv8 (yolov8s-pose.pt) to detect bounding boxes for persons

in the frame. The original model is converted to support the open neural network exchange (ONNX) runtime to increase the inference speed across CPU and GPU architectures. This step identifies the regions within the frame that require anonymization or privacy protection. Following detection, the algorithm assesses whether any individuals have been identified. If they have, it proceeds to selectively update the regions within these bounding boxes using content from a predefined or updated background buffer image, effectively anonymizing the subject by replacing their visual representation with backgrounds of previously acquired frames. Conversely, if no individuals are detected within a frame, the algorithm updates the entire background buffer with the current frame, refreshing the scene in the absence of significant foreground activity. This dynamic updating mechanism allows the system to maintain an up-to-date background representation of the monitored environment.

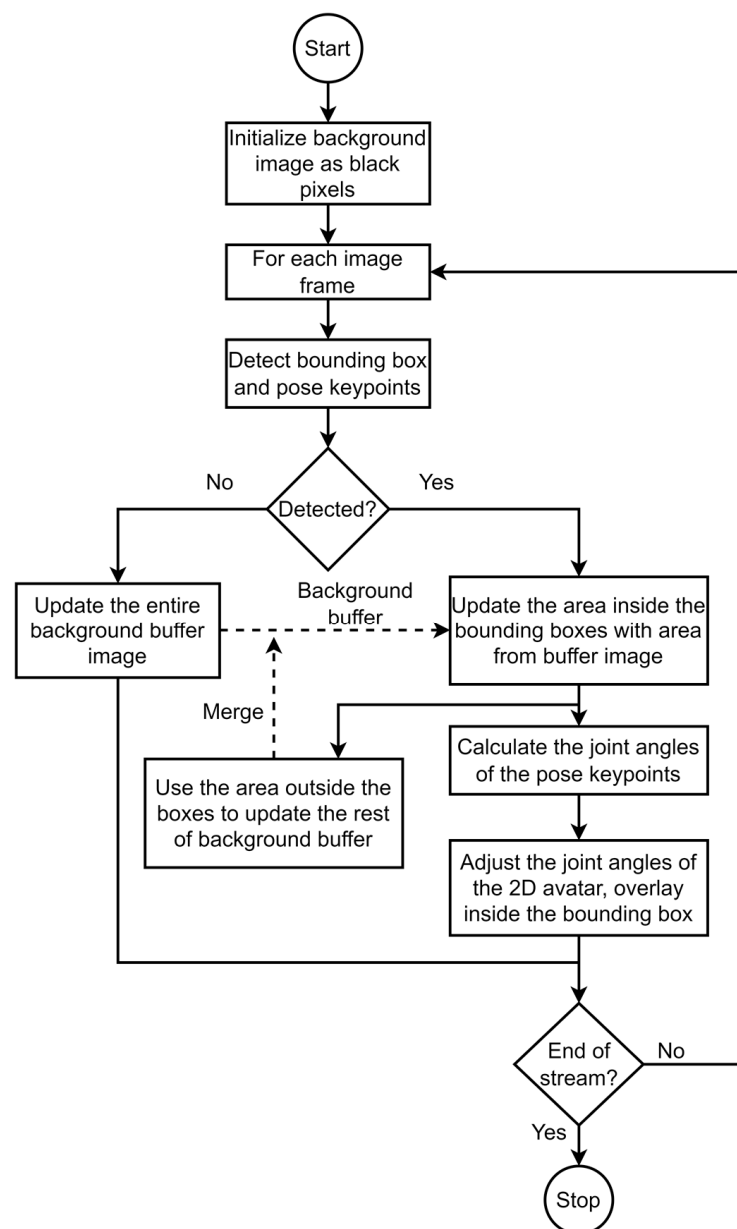


Figure 4. Flowchart of subject anonymization technique used for elderly privacy protection.

For frames where individuals are detected, the system further uses human pose keypoint data to extract the joint angles of the skeleton representation. Leveraging this information, the algorithm adjusts a 2D avatar to match the detected pose, overlaying this avatar within the bounding box in place of the person's image. Moreover, the area

outside the detected bounding boxes is utilized to update the existing background buffer image, ensuring that the scene information remains current in each frame. This step not only preserves the anonymity of the monitored individual but also retains a visual cue of their activities and movements, which is crucial for monitoring well-being and detecting potential emergencies. This process works in an iterative way, with the system continuously checking whether the video stream has concluded. If the stream is ongoing, the algorithm advances to the next frame, repeating this procedure until the end of the stream, ensuring comprehensive and continuous monitoring throughout.

For animating the 2D avatar, this study utilizes Panda3D, an open-source game engine that is supported in the Python programming language. Pose data, extracted in the form of pixel coordinates from the YOLOv8 model, identify 17 keypoints that represent a human figure's posture within a frame, as shown in Figure 5a. To align these keypoints with the rig joints of the 2D avatar, the raw keypoint data are transformed into joint angle data, ensuring that the avatar's movements accurately mirror those of the detected human posture, as shown in Figure 5b. Additionally, two extra keypoints are generated at the midpoint between keypoints {5, 6} and {11, 12} to determine the upper torso center and lower torso center coordinates, respectively. These coordinates are used for calculating the neck and spine angles, aligning with the 2D avatar's rig. To ensure smooth and natural transitions in the avatar's movements across consecutive video frames, a 3rd-order polynomial fitting is applied to the extracted joint angle data to prevent any sudden jumps or jerk motions.

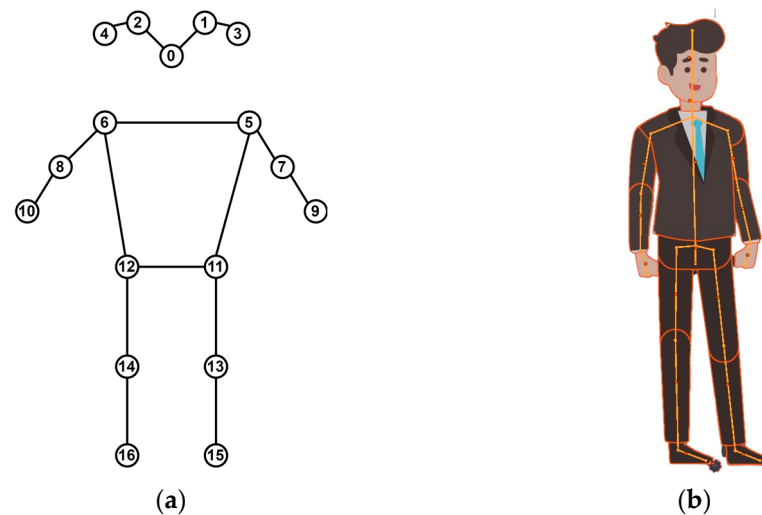


Figure 5. (a) Pose keypoints extracted from the YOLOv8 model. (b) A 2D avatar and skeletal rig.

Given three keypoints, A , B , and C , where B is the joint for which the angle is calculated and A and C are adjacent joints, the angle at B can be calculated using the dot product formula. Using vectors $\overrightarrow{BA}$ and $\overrightarrow{BC}$ from the pixel coordinates of these points, the angle θ for joint B can be calculated as shown in Equation (1):

$$\theta = \arccos \left(\frac{\overrightarrow{BA} \cdot \overrightarrow{BC}}{\|\overrightarrow{BA}\| \|\overrightarrow{BC}\|} \right) \quad (1)$$

However, considering the relative orientation of joints and given the planar nature of 2D avatar animation, this study uses a more direct approach to calculate the angles in 2D space, which can be represented as Equation (2), below:

$$\theta = \arctan2(u_y, u_x) - \arctan2(v_y, v_x) \quad (2)$$

where u_y and u_x are the x and y axis components of $\overrightarrow{BA}$ and v_y and v_x are the x and y axis components of $\overrightarrow{BC}$. This method significantly enhances 2D avatar animation by accurately accounting for vector orientation, which is crucial for replicating human movements. Utilizing the $\arctan2$ operator, which assesses both vector magnitude and direction, ensures that angles are measured across the full 360° range, from 180° to -180° . This capability allows for capturing motion in all four quadrants, an important feature for animating complex movements with lifelike accuracy.

The study applies a 3rd-order polynomial fitting to smooth the transitions of these angles over time, ensuring natural movement of the 2D avatar. Given a set of joint angles θ over time t , the polynomial fitting can be represented using Equation (3):

$$\hat{\theta}(t) = a_0 + a_1t + a_2t^2 + a_3t^3 \quad (3)$$

where a_0 , a_1 , a_2 , and a_3 are the coefficients determined by fitting the polynomial to the historical data of joint angles using the least squares method.

3.3. Fall Detection Method

This study implements a hierarchical architecture for fall detection, as shown in Figure 6, using bounding boxes and pose keypoints from the YOLOv8 model, originally used for subject anonymization in Section 3.2. These outputs are used for the extraction of motion features, including the average velocity of keypoints, the body's center, the bounding box's center, and the aspect ratio of the bounding box. To accommodate frame rate variations across different video acquisition systems, motion features are normalized by calculating respective pixel/second velocities.

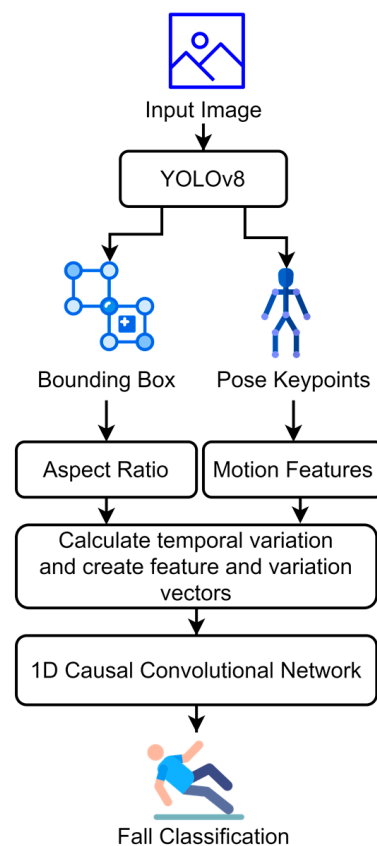


Figure 6. Proposed fall detection method.

Thereafter, a 1D causal CNN is employed for binary classification, determining “fall” or “not fall” scenarios in each frame. This network uses the extracted features and their

temporal variations. The causal nature of the network guarantees that predictions are made based on present and past data only. The size of the training and inference window is set to correspond to 1 s or the number of frames per second, allowing for a consistent temporal scope in analyzing the video data.

The average pose keypoint velocity is calculated by measuring the change in the position of each keypoint across consecutive frames and dividing it by the time elapsed between these frames. If $P_{i,t}$ and $P_{i,t+1}$ represent the positions of keypoint i at times t and $t + 1$, respectively, the velocity V_i for keypoint i is calculated as shown in Equation (4):

$$V_i = \frac{\|P_{i,t+1} - P_{i,t}\|}{\Delta t} \quad (4)$$

The average velocity (V_{avg}) across all N keypoints can be calculated as shown in Equation (5):

$$V_{avg} = \frac{1}{N} \sum_{i=1}^N V_i \quad (5)$$

The body center (C_{body}) is determined using the positions of the shoulders and hips. The midpoints $M_{shoulder}$ and M_{hip} are found by averaging the positions of the left and right shoulders (P_{LS} and P_{RS}) and the left and right hips (P_{LH} and P_{RH}), respectively, as shown in Equations (6)–(8):

$$M_{shoulder} = \frac{P_{LS} + P_{RS}}{2} \quad (6)$$

$$M_{hip} = \frac{P_{LH} + P_{RH}}{2} \quad (7)$$

$$C_{body} = \frac{M_{shoulder} + M_{hip}}{2} \quad (8)$$

The center of the bounding box (C_{bbox}) is calculated by taking the average of the x and y coordinates of the bounding box's corners. If B_{tl} and B_{br} represent the top-left and bottom-right corners of the bbox, respectively, then C_{bbox} can be calculated using Equation (9):

$$C_{bbox} = \frac{B_{tl} + B_{br}}{2} \quad (9)$$

The aspect ratio (AR) of the bounding box is determined by the ratio of its width to its height. The width can be calculated as the difference between the x components of the left (B_{tl}) and right (B_{tr}) corners, while height can be calculated from the y components of the top (B_{tl}) and bottom (B_{bl}) corners. Therefore, the equation for calculating AR can be formulated as in Equation (10):

$$AR = \frac{B_{tr}(x) - B_{tl}(x)}{B_{bl}(y) - B_{tl}(y)} \quad (10)$$

The temporal variation (σ) of the features is calculated over a sliding window containing the sequence of feature values $[F_t, F_{t+1}, F_{t+2}, \dots, F_{t+n}]$ over time t . Within this window, the temporal variation is calculated using Equation (11):

$$\sigma = \sqrt{\frac{1}{k} \sum_{i=t-k+1}^t (F_i - \mu)^2} \quad (11)$$

where F_i represents the feature value at each frame within the window, μ is the mean of these values, and k is the total number of frames in the window. This equation quantifies the variability of feature values around the mean, providing a measure of temporal variation across the video frames.

Building on previous research on fall detection systems, this research aims to develop a classification model that can be seamlessly implemented across various scenes. While numerous studies have demonstrated that analyzing variations in features allows for the creation of rule-based methods to detect falls and non-falls within a monitored environment, it is crucial to acknowledge that these rules necessitate a detailed examination of body and motion features, along with expert knowledge of biodynamics. Consequently, this study opted for a 1D causal convolutional neural network (CNN) with residual connections as the model of choice for learning and classification.

One-dimensional CNNs were first used for sequence-to-sequence generation of audio data, as shown in [66]. Over time, various adaptations, including causal and dilated causal CNNs, were developed to cater to a broad spectrum of time-series and sequence datasets necessitating the maintenance of strict causal relationships during training and inference, as shown in [67–70]. Notably, 1D convolutional networks have demonstrated significant efficacy in applications related to human motion modeling [71–73] and classification of complex patterns within 1D time-series datasets [74–77]. This versatility and robust performance highlight the potential of 1D CNNs to capture intricate temporal dynamics, making them a compelling choice for studies focusing on sequence data regression and classification, eliminating the need for complex feature engineering.

This study employs 1D dilated causal CNN layers with residual connections, as illustrated in Figure 7, to construct the classification model. Given that body motion features and bounding box attributes have been previously extracted and their temporal variations are used for training, the model architecture is intentionally simplified. It consists of three residual blocks followed by a fully connected layer with 32 neurons with sigmoid activation at the network's tail. The output node consists of a single neuron with sigmoid activation for the binary classification of fall and non-fall scenarios. Each residual block in the model starts with a dilated causal CNN layer, which uses a specific number of filters and a dilation parameter, followed by a Scaled Exponential Linear Unit (SELU) [78] activation layer. This setup allows for the initial expansion of the network's receptive field, enabling it to capture wider nonlinear temporal patterns and without a significant increase in computational complexity. After this, another causal CNN layer is applied, using the same number of filters but without dilation, and this is also followed by an SELU activation layer. To integrate learned features while preserving information from earlier layers, each block includes a skip connection that undergoes a 1×1 convolution. This convolution is designed to match the number of filters to the block's final layer before adding it to the block's output. The model is structured with three residual blocks, each set to progressively increase the number of filters (16, 32, and 64) and dilation parameters (1, 2, and 3, respectively). This progressive design allows the network to hierarchically extract and refine features from simpler to more complex patterns. Incrementally expanding the number of filters and dilation rates ensures that the model can learn from different motion characteristics to distinguish between normal activities and fall incidents based on temporal patterns within the data.

The proposed network is implemented using PyTorch version 2.1.0, a popular deep learning framework among machine learning researchers. During the training phase, the Adam [79] optimization algorithm is employed, chosen for its effective handling of sparse gradients and its adaptability in different scenarios. The initial learning rate is set to 0.001, a standard starting point that balances the speed and stability of the learning process. As training progresses, the Adam algorithm automatically adjusts this learning rate based on the training data's behavior, ensuring optimal convergence rates.

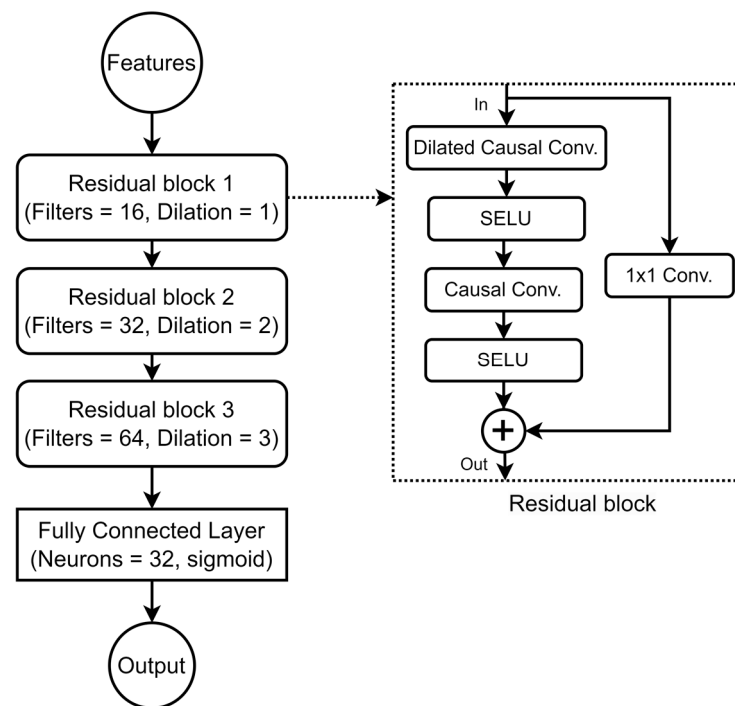


Figure 7. Proposed classification model using residual causal convolutional network.

4. Results

This section presents the outcomes achieved through the implementation of the subject anonymization technique and the fall detection algorithm.

4.1. Dataset

This study used two publicly available datasets for an evaluation of the proposed system. The first dataset used in this study was the University of Rzeszow Fall Detection (URFD) dataset [80], created at the University of Rzeszow, and comprises 70 sequences at a resolution of 640×480 , including 30 falls and 40 activities of daily living, such as sitting, crouching, picking up objects, lying on the floor, and lying on a bed or couch. The dataset features a total of 13,000 images—3000 dedicated to fall sequences and 10,000 to daily activities. They are categorized into three classes: non-falling activities, post-fall scenarios (not used for classification in this study), and active falls (used for classification).

The other dataset was curated by the “Laboratoire Electronique, Informatique et Image” (LE2I) [81] and contains 191 annotated videos from various elderly living environments, including home (60 videos), coffee room (70 videos), office (33 videos), and lecture room environments (28 videos), 130 out of which show falls and the rest of which show daily activities. This dataset includes a range of actions, such as walking, sitting on a chair, sitting on a sofa, and bending over, providing a comprehensive view of typical elderly movements. These activities are particularly valuable as they enable the evaluation of fall detection models to determine whether they can accurately distinguish actions like sitting and bending from actual falls. This dataset was captured using a single camera with a resolution of 320×240 and a frame rate of 25 frames per second. However, in this study, these videos were resized to 640×480 resolution to match the input resolution of YOLOv8.

4.2. Subject Anonymization

The proposed anonymization system was tested using the videos from the LE2I dataset to evaluate the algorithm’s effectiveness. With regard to performance, in terms of detection accuracy, utilizing the YOLOv8 model (yolov8s-pose.pt), the system was able to reliably detect the person’s location in over 99% of the video frames. However, challenges arose in scenarios with significant occlusion, where more than half of the subject’s body

was obscured by environmental objects such as desks or tables. In these instances, the model occasionally encountered difficulties in detecting keypoints with high confidence. Nonetheless, it consistently achieved success in identifying the upper body's bounding box, even under such challenging conditions. Limitations were notably observed in rare cases when parts of the subject exited the camera's field of view, resulting in detection failures. To mitigate these limitations, future implementations could consider the strategic placement of cameras within the environment or the integration of multiple cameras to avoid blind spots, thereby enhancing the system's overall robustness and reliability in diverse settings.

To visually demonstrate the anonymization process utilized in this study, four images are presented to illustrate the stages: an original video frame is shown in Figure 8a, showing a scene prior to any processing. Figure 8b displays the privacy frame, where the person has been removed from the scene using their bounding box. The skeleton representation, as predicted by the YOLOv8 model's pose detection, is shown in Figure 8c. Finally, Figure 8d presents the 2D avatar superimposed onto the scene, representing the person's movements in an anonymized manner. Additionally, another action, a person sitting, is shown in Figure 9. Both Figures 8 and 9 originate from the "lecture room" video of the LE2I dataset. This scene is well-lit, offering a baseline for fair comparison. Furthermore, two frames from the "home" scene of the LE2I dataset, which is moderately lit, are presented in Figures 10 and 11. Despite varying lighting conditions, scene setups, and the actions of the person, the proposed method effectively detected the person in the scenes and superimposed the 2D avatar, demonstrating its ability to perform consistently across different environments and lighting conditions.

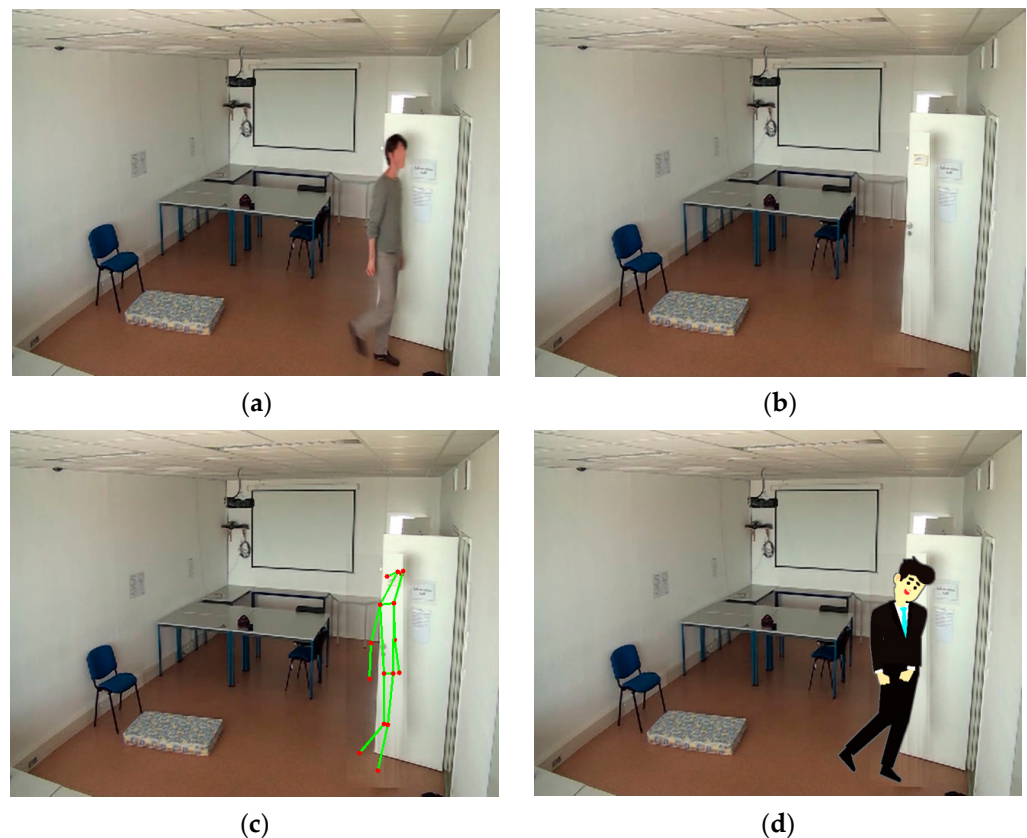


Figure 8. Subject anonymization process in “lecture room” scene with person walking: (a) original frame, (b) foreground removal, (c) skeleton representation, and (d) 2D avatar.

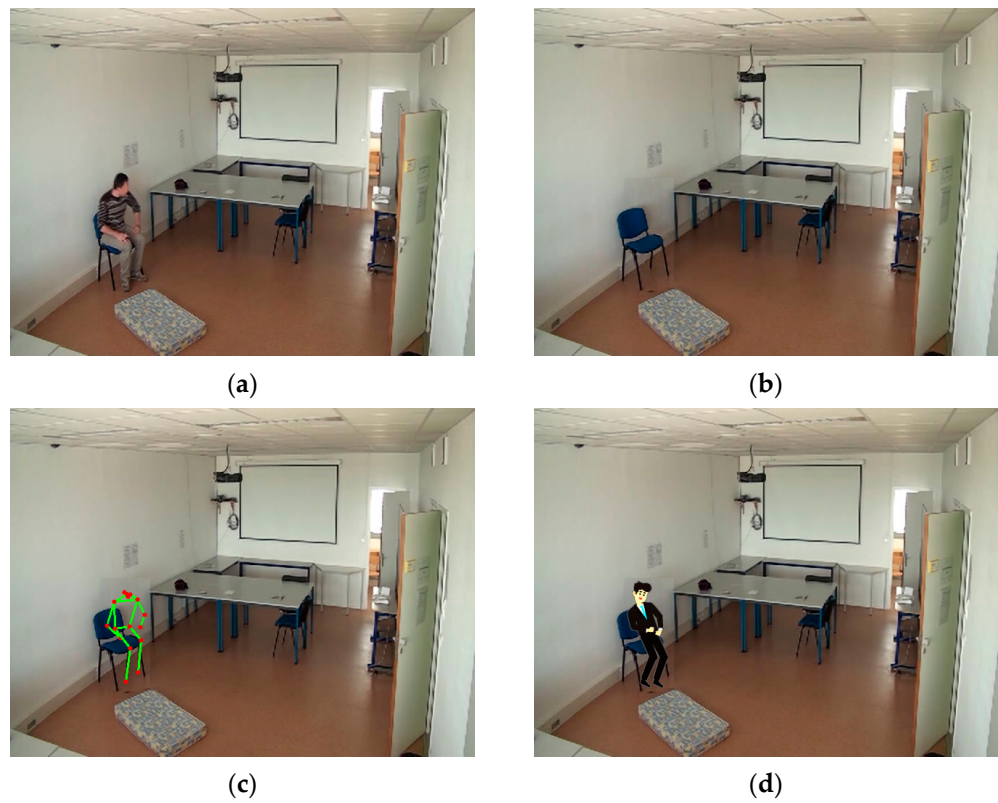


Figure 9. Subject anonymization process in “lecture room” scene with person sitting on a chair: (a) original frame, (b) foreground removal, (c) skeleton representation, and (d) 2D avatar.

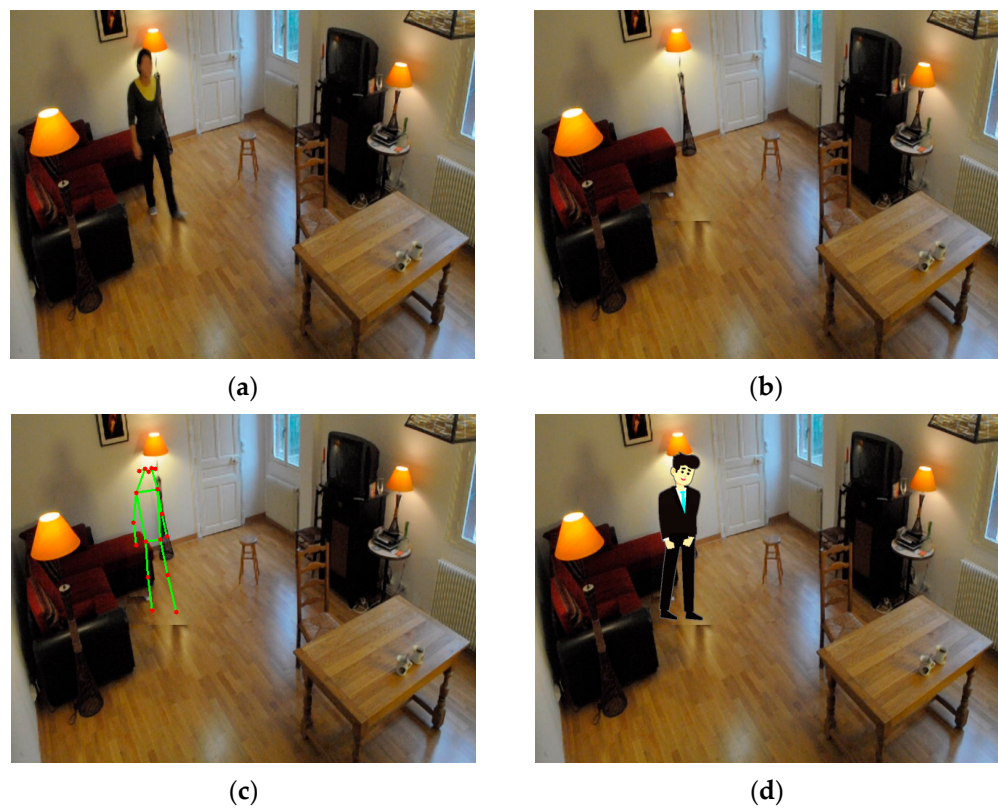


Figure 10. Subject anonymization process in “home” scene with person walking: (a) original frame, (b) foreground removal, (c) skeleton representation, and (d) 2D avatar.

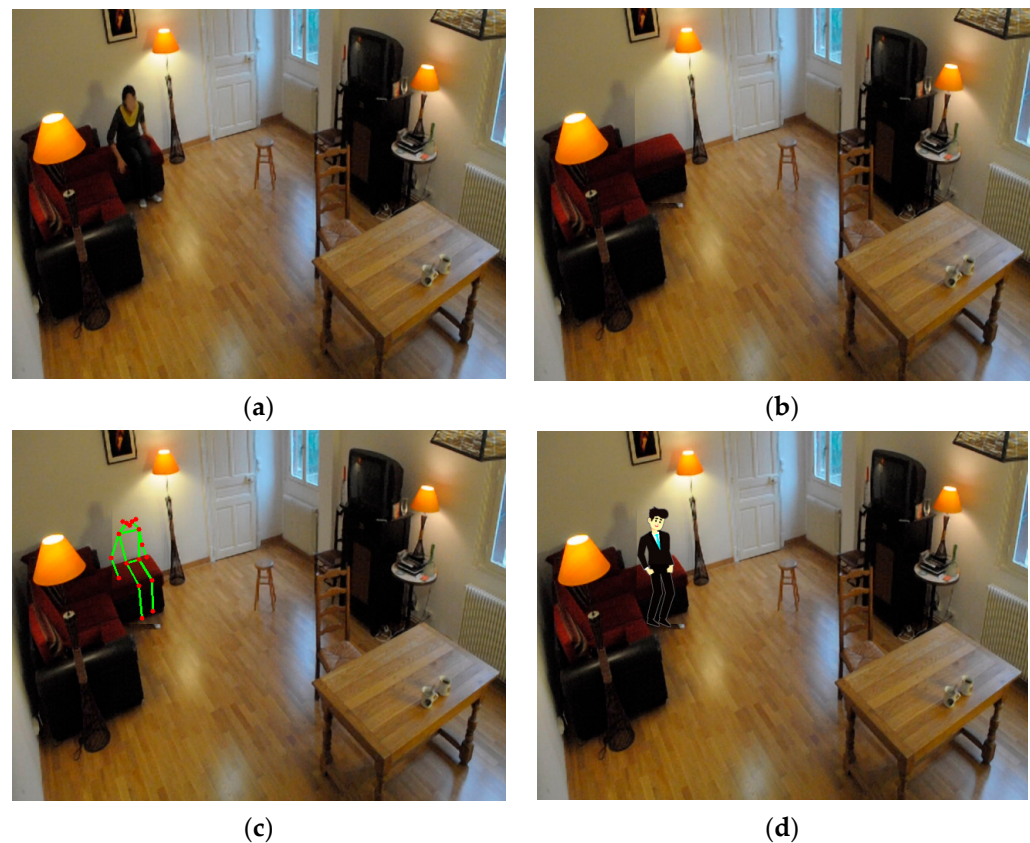


Figure 11. Subject anonymization process in “home” scene with person sitting on a sofa: (a) original frame, (b) foreground removal, (c) skeleton representation, and (d) 2D avatar.

4.3. Fall Detection

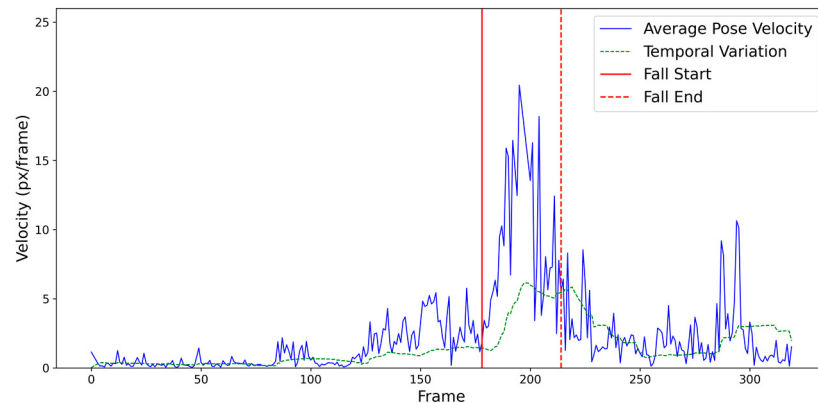
To evaluate the proposed fall detection system, bounding boxes and motion features were extracted from video frames, as detailed in Section 3.3. The training of the proposed 1D residual causal CNN model utilized videos exclusively from the “office” scene of the LE2I dataset. The average pose velocity and its temporal variation for a video containing a fall and another without a fall are shown in Figure 12a and Figure 12b, respectively.

Given that there were only 33 videos for the “office” scene, which is insufficient for training a deep learning model effectively, this study incorporated a data augmentation technique. This approach rolled the feature data along the time-series axis to increase the dataset. Specifically, the time-series data underwent a shift of 25 frames along the time axis for each video frame. Iterating this process and adjusting the frame shift by a factor of 25 frames, the size of the training dataset was substantially increased by 100 times. To illustrate this method, Figure 13 shows (a) the original average pose velocity, (b) the data after a 25-frame shift to the right, and (c) the data following a 25-frame shift to the left, demonstrating the augmentation’s effect on the dataset’s variation in fall location for the time series. Using this augmentation method, the model was trained for 100 epochs, and the model’s weights were saved for inference from the rest of the dataset.

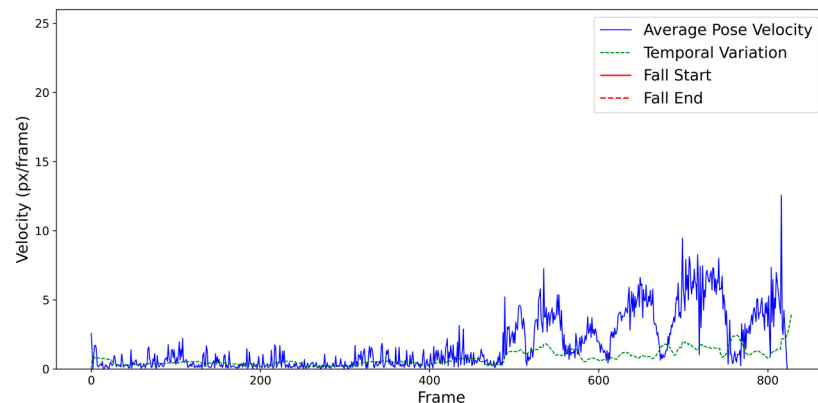
This study utilized the two benchmark datasets mentioned in Section 4.1 to evaluate the proposed model. The model trained on the “coffee” scene of the LE2I dataset was tested on the rest of the LE2I dataset scenes and also on the URFD dataset. Given the binary classification nature of the fall detection, the evaluation metrics employed in this research are detailed below [33]:

- Accuracy: The percentage of items correctly classified, calculated as $(TP + TN)/(TP + TN + FP + FN)$.
- Precision: The proportion of true-positive identifications among all positive predictions, calculated as $TP/(TP + FP)$.

- Recall/sensitivity: The true-positive rate or the model's ability to identify positive instances accurately, calculated as $TP/(TP + FN)$.
- Specificity: Often used together with recall, specificity measures the true-negative rate or the model's ability to correctly identify negative instances, calculated as $TN/(TN + FP)$.
- F1-Score: This denotes the harmonic mean between precision and recall, calculated as $2 \times (\text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall})$



(a)



(b)

Figure 12. Time-series plot of the average pose velocity and its temporal variation (a) for a video containing a fall and (b) for a video containing no fall.

Table 1 presents the frame-wise evaluation results for the LE2I and URFD datasets using the proposed fall detection method. For the LE2I dataset, the method achieved an accuracy of 98.86%, a precision of 90.02%, a sensitivity of 91.11%, a specificity of 99.35%, and an F1-Score of 0.905. In comparison, on the URFD dataset, the method attained an accuracy of 96.23%, a precision of 97.18%, a sensitivity of 92.56%, a specificity of 98.40%, and an F1-Score of 0.948. These results indicate that the proposed method performs accurately across different datasets, excelling particularly in specificity, reducing false alarms while maintaining high accuracy and precision. The results of the proposed method are comparable to those of the methods of the original datasets [80,81] and other state-of-the-art methods [33,34] evaluated with these datasets.

Furthermore, when evaluating the system on a video-wise basis, both accuracy and precision reached 100% for the tested datasets. This suggests that, while frame-wise metrics provide insight into the method's performance at a granular level, the video-wise evaluation showcases the system's ability to accurately classify fall or non-fall events in entire video sequences with perfect precision. This highlights the method's effectiveness in real-world applications, where accurate event classification over live video streams is critical.

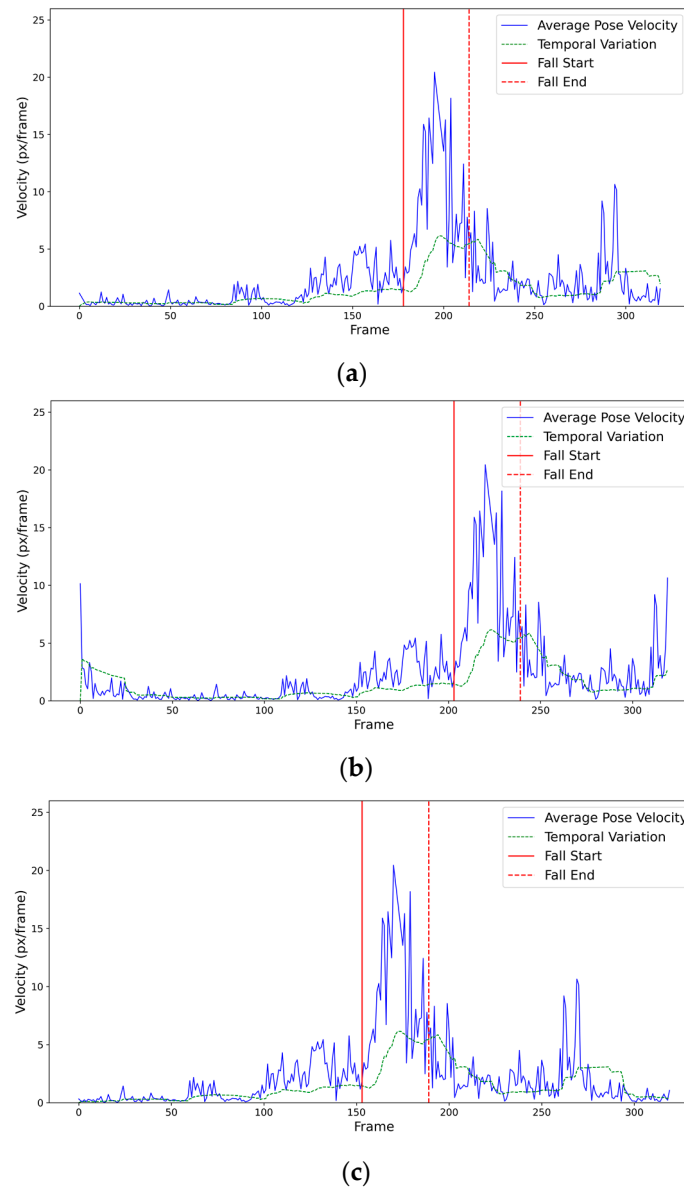


Figure 13. Time-series plot of the average pose velocity and its temporal variation: (a) original time series, (b) with 25-frame shift to the right, and (c) with 25-frame shift to the left.

Table 1. Frame-wise evaluation results for the LE2I and URFD datasets using the proposed fall detection method.

Dataset	Accuracy [%]	Precision [%]	Sensitivity [%]	Specificity [%]	F1-Score
LE2I	98.86	90.02	91.11	99.35	0.905
URFD	96.23	97.18	92.56	98.40	0.948

4.4. System Latency

In the proposed system equipped with an NVIDIA RTX 3050 Ti GPU with an Intel i5 processor (12 cores) and 32GB of RAM, the system's latency demonstrates notable variation depending on the hardware setup. When utilizing the GPU for inference, it achieves up to 72 frames per second (fps), and when solely relying on the CPU, it maintains 43 fps. This performance improvement is due to the ONNX optimization of the YOLOv8 model and the proposed fall detection model. When the system's resources are limited to using only four CPU cores, it still attains a reasonable speed of 14 fps. This performance demonstrates the system's ability to sustain real-time processing capabilities across a variety of hardware

setups, emphasizing its adaptability and potential applicability in end-user scenarios that involve low-cost hardware. Additionally, by employing a smaller YOLOv8 model like YOLOv8-nano, the system may achieve even faster speeds, further enhancing its accessibility and performance in budget-friendly hardware setups. The proposed system architecture may also scale effectively to manage feeds from multiple cameras simultaneously to address different room and blind spot issues without sacrificing processing efficiency. This scalability ensures comprehensive monitoring in various scenarios, ensuring that no blind spots are left unwatched.

5. Conclusions

The results achieved through the proposed fall detection and subject anonymization method, as tested on the LE2I and URFD datasets, demonstrate the system's efficacy in ensuring privacy while maintaining high accuracy in fall detection across various real-world scenarios. The method's performance, with a frame-wise accuracy reaching 98.86% and a precision of up to 97.18%, alongside the video-wise accuracy and precision both reaching 100%, demonstrates its robustness and reliability. Particularly noteworthy is the system's specificity, which significantly reduces false alarms, a crucial factor in elderly monitoring systems to prevent unnecessary panic or stress.

The subject anonymization aspect of this study plays a pivotal role in addressing privacy concerns, a significant barrier to the acceptance of monitoring technologies among elderly people. By effectively anonymizing individuals in the video through the removal of identifiable features and replacing them with 2D avatars, this method respects the privacy of the monitored individuals. This approach not only ensures the utility of the monitoring system for safety and wellness purposes but also promotes its acceptance among elderly users who might be concerned about their privacy and dignity.

In conclusion, while current state-of-the-art technologies in fall detection offer high accuracy, they often require substantial computational resources and typically overlook the ethical and user-centric aspects of monitoring. This lack of attention to user preferences, particularly around visual privacy, can significantly hinder the acceptance and applicability of these technologies among elderly users, who may prioritize dignity and privacy in their care solutions. This study offers a robust solution that effectively balances the dual objectives of accurate fall detection and thorough privacy protection. The high accuracy and low false-positive rate indicate significant progress in elderly care technology. Importantly, this research highlights the innovative application of artificial intelligence not just for utility in fall detection but also in addressing design needs related to privacy concerns. Its emphasis on privacy protection through subject anonymization is likely to encourage greater acceptance among the elderly population, making it a significant step forward in the development of assistive technologies that cater to the sensitive needs of this demographic.

Some of the limitations observed in this study are inherent to using 2D avatar representations to model complex 3D human movements. While these 2D avatars are effective for basic monitoring, they struggle to accurately convey complex actions from certain angles, which can lead to ambiguities in interpreting the subject's actual posture. Future research could focus on integrating optimized 3D modeling techniques, which offer a more detailed and dynamic representation of human poses while preserving real-time performance. Future work will also focus on refining the fall detection and anonymization algorithms for enhanced accuracy and real-time processing, exploring multi-camera setups to reduce occlusions, and integrating advanced privacy-preserving techniques. Additionally, efforts will aim at improving user acceptance through ethical considerations to ensure that the technology not only meets the highest standards of efficiency and privacy but also considers the major aspects of social ethics.

Author Contributions: Conceptualization, C.-Y.W.; methodology, C.-Y.W.; software, C.-Y.W.; validation, C.-Y.W. and F.-S.L.; formal analysis, C.-Y.W.; investigation, C.-Y.W.; resources, C.-Y.W.; data curation, C.-Y.W.; writing—original draft preparation, C.-Y.W.; writing—review and editing, F.-S.L.;

visualization, C.-Y.W.; supervision, F.-S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was by approved by the Research Ethics Committee of the National Chung Cheng University Human Research Ethics Center, Chiayi County, Taiwan (project no.: CCUREC112050502, protocol code: Version 2/26.6, 2023, approval date: 20 July 2023).

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here [<http://fenix.ur.edu.pl/~mkepski/ds/uf.html>, URFD] and [<https://gestion-web2.u-bourgogne.fr:8443/?lang=en>, LE2I].

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

References

1. Population Division of the Department of Economic and Social Affairs. *World Population Prospects 2022*; United Nations: New York, NY, USA, 2022; ISBN 978-92-1-148373-4.
2. World Health Organization. *WHO Global Report on Falls Prevention in Older Age*; World Health Organization: Geneva, Switzerland, 2007.
3. Osborne, T.F.; Veigulis, Z.P.; Arreola, D.M.; Vrublevskiy, I.; Suarez, P.; Curtin, C.; Schalch, E.; Cabot, R.C.; Gant-Curtis, A. Assessment of a Wearable Fall Prevention System at a Veterans Health Administration Hospital. *Digit. Health* **2023**, *9*, 20552076231187727. [[CrossRef](#)] [[PubMed](#)]
4. Ren, L.; Peng, Y. Research of Fall Detection and Fall Prevention Technologies: A Systematic Review. *IEEE Access* **2019**, *7*, 77702–77722. [[CrossRef](#)]
5. Hamm, J.; Money, A.G.; Atwal, A.; Paraskevopoulos, I. Fall Prevention Intervention Technologies: A Conceptual Framework and Survey of the State of the Art. *J. Biomed. Inform.* **2016**, *59*, 319–345. [[CrossRef](#)] [[PubMed](#)]
6. Rastogi, S.; Singh, J. Human Fall Detection and Activity Monitoring: A Comparative Analysis of Vision-Based Methods for Classification and Detection Techniques. *Soft Comput.* **2022**, *26*, 3679–3701. [[CrossRef](#)]
7. Wang, C.-Y.; Lin, F.-S. Exploring Older Adults' Willingness to Install Home Surveillance Systems in Taiwan: Factors and Privacy Concerns. *Healthcare* **2023**, *11*, 1616. [[CrossRef](#)] [[PubMed](#)]
8. Buzzelli, M.; Albé, A.; Ciocca, G. A Vision-Based System for Monitoring Elderly People at Home. *Appl. Sci.* **2020**, *10*, 374. [[CrossRef](#)]
9. Jansen, B.; Deklerck, R. Home Monitoring of Elderly People with 3D Camera Technology. In Proceedings of the First BENELUX Biomedical Engineering Symposium, Brussels, Belgium, 7–8 December 2006.
10. Feng, W.; Liu, R.; Zhu, M. Fall Detection for Elderly Person Care in a Vision-Based Home Surveillance Environment Using a Monocular Camera. *Signal Image Video Process.* **2014**, *8*, 1129–1138. [[CrossRef](#)]
11. Yang, Y.; Yang, H.; Liu, Z.; Yuan, Y.; Guan, X. Fall Detection System Based on Infrared Array Sensor and Multi-Dimensional Feature Fusion. *Meas. J. Int. Meas. Confed.* **2022**, *192*, 110870. [[CrossRef](#)]
12. Ramanujam, E.; Padmavathi, S. Real Time Fall Detection Using Infrared Cameras and Reflective Tapes under Day/Night Luminance. *J. Ambient Intell. Smart Environ.* **2021**, *13*, 285–300. [[CrossRef](#)]
13. Park, J.; Chen, J.; Cho, Y.K.; Kang, D.Y.; Son, B.J. CNN-Based Person Detection Using Infrared Images for Night-Time Intrusion Warning Systems. *Sensors* **2020**, *20*, 34. [[CrossRef](#)]
14. Cosar, S.; Yan, Z.; Zhao, F.; Lambrou, T.; Yue, S.; Bellotto, N. Thermal Camera Based Physiological Monitoring with an Assistive Robot. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; pp. 5010–5013. [[CrossRef](#)]
15. Riquelme, F.; Espinoza, C.; Rodenas, T.; Minonzio, J.-G.; Taramasco, C. eHomeSeniors Dataset: An Infrared Thermal Sensor Dataset for Automatic Fall Detection Research. *Sensors* **2019**, *19*, 4565. [[CrossRef](#)] [[PubMed](#)]
16. Fernando, Y.P.N.; Gunasekara, K.D.B.; Sirikumara, K.P.; Galappaththi, U.E.; Thilakarathna, T.; Kasthurirathna, D. Computer Vision Based Privacy Protected Fall Detection and Behavior Monitoring System for the Care of the Elderly. In Proceedings of the 2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Vasteras, Sweden, 7–10 September 2021; pp. 1–7. [[CrossRef](#)]
17. Beddiar, D.R.; Nini, B.; Sabokrou, M.; Hadid, A. Vision-Based Human Activity Recognition: A Survey. *Multimed. Tools Appl.* **2020**, *79*, 30509–30555. [[CrossRef](#)]
18. Nikouei, S.Y.; Chen, Y.; Song, S.; Xu, R.; Choi, B.Y.; Faughnan, T.R. Real-Time Human Detection as an Edge Service Enabled by a Lightweight CNN. In Proceedings of the 2018 IEEE International Conference on Edge Computing (EDGE), San Francisco, CA, USA, 2–7 July 2018; pp. 125–129. [[CrossRef](#)]
19. Chen, Y.; Kong, X.; Meng, L.; Tomiyama, H. An Edge Computing Based Fall Detection System for Elderly Persons. *Procedia Comput. Sci.* **2020**, *174*, 9–14. [[CrossRef](#)]

20. Kim, S.; Park, J.; Jeong, Y.; Lee, S.E. Intelligent Monitoring System with Privacy Preservation Based on Edge AI. *Micromachines* **2023**, *14*, 1749. [[CrossRef](#)] [[PubMed](#)]
21. Williams, A.; Xie, D.; Ou, S.; Grupen, R.; Hanson, A.; Riseman, E. Distributed Smart Cameras for Aging in Place. In Proceedings of the ACM SenSys Workshop on Distributed Smart Cameras, Boulder, CO, USA, 31 October 2006.
22. Samkari, E.; Arif, M.; Alghamdi, M.; AlGhamdi, M.A. Human Pose Estimation Using Deep Learning: A Systematic Literature Review. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1612–1659. [[CrossRef](#)]
23. BenGamra, M.; Akhloufi, M.A. A Review of Deep Learning Techniques for 2D and 3D Human Pose Estimation. *Image Vis. Comput.* **2021**, *114*, 104282. [[CrossRef](#)]
24. Kendall, A.; Grimes, M.; Cipolla, R. PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; Volume 2015, pp. 2938–2946.
25. Bazarevsky, V.; Grishchenko, I.; Raveendran, K.; Zhu, T.; Zhang, F.; Grundmann, M. BlazePose: On-Device Real-Time Body Pose Tracking. *arXiv* **2020**, arXiv:2006.10204.
26. Li, S.; Man, C.; Shen, A.; Guan, Z.; Mao, W.; Luo, S.; Zhang, R.; Yu, H. A Fall Detection Network by 2D/3D Spatio-Temporal Joint Models with Tensor Compression on Edge. *ACM Trans. Embed. Comput. Syst.* **2022**, *21*, 1–19. [[CrossRef](#)]
27. Egawa, R.; Miah, A.S.M.; Hirooka, K.; Tomioka, Y.; Shin, J. Dynamic Fall Detection Using Graph-Based Spatial Temporal Convolution and Attention Network. *Electronics* **2023**, *12*, 3234. [[CrossRef](#)]
28. Noor, N.; Park, I.K. A Lightweight Skeleton-Based 3D-CNN for Real-Time Fall Detection and Action Recognition. In Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Paris, France, 2–6 October 2023; pp. 2171–2180. [[CrossRef](#)]
29. Min, W.; Yao, L.; Lin, Z.; Liu, L. Support Vector Machine Approach to Fall Recognition Based on Simplified Expression of Human Skeleton Action and Fast Detection of Start Key Frame Using Torso Angle. *IET Comput. Vis.* **2018**, *12*, 1133–1140. [[CrossRef](#)]
30. Kong, X.; Kumaki, T.; Meng, L.; Tomiyama, H. A Skeleton Analysis Based Fall Detection Method Using ToF Camera. *Procedia Comput. Sci.* **2021**, *187*, 252–257. [[CrossRef](#)]
31. De Miguel, K.; Brunete, A.; Hernando, M.; Gambao, E. Home Camera-Based Fall Detection System for the Elderly. *Sensors* **2017**, *17*, 2864. [[CrossRef](#)] [[PubMed](#)]
32. Lafuente-Arroyo, S.; Martín-Martín, P.; Iglesias-Iglesias, C.; Maldonado-Bascón, S.; Acevedo-Rodríguez, F.J. RGB Camera-Based Fallen Person Detection System Embedded on a Mobile Platform. *Expert Syst. Appl.* **2022**, *197*, 116715. [[CrossRef](#)]
33. Alam, E.; Sufian, A.; Dutta, P.; Leo, M. Vision-Based Human Fall Detection Systems Using Deep Learning: A Review. *Comput. Biol. Med.* **2022**, *146*, 105626. [[CrossRef](#)] [[PubMed](#)]
34. Gutiérrez, J.; Rodríguez, V.; Martín, S. Comprehensive Review of Vision-Based Fall Detection Systems. *Sensors* **2021**, *21*, 947. [[CrossRef](#)] [[PubMed](#)]
35. Hbali, Y.; Hbali, S.; Ballihi, L.; Sadgal, M. Skeleton-Based Human Activity Recognition for Elderly Monitoring Systems. *IET Comput. Vis.* **2018**, *12*, 16–26. [[CrossRef](#)]
36. Nguyen, H.-C.; Nguyen, T.-H.; Scherer, R.; Le, V.-H. Deep Learning for Human Activity Recognition on 3D Human Skeleton: Survey and Comparative Study. *Sensors* **2023**, *23*, 5121. [[CrossRef](#)] [[PubMed](#)]
37. Alaoui, A.Y.; Elfikhi, S.; Thami, R.O.H. Fall Detection for Elderly People Using the Variation of Key Points of Human Skeleton. *IEEE Access* **2019**, *7*, 154786–154795. [[CrossRef](#)]
38. Wang, Y.; Deng, T. Enhancing Elderly Care: Efficient and Reliable Real-Time Fall Detection Algorithm. *Digit. Health* **2024**, *10*, 20552076241233690. [[CrossRef](#)]
39. Hoang, V.H.; Lee, J.W.; Piran, M.J.; Park, C.S. Advances in Skeleton-Based Fall Detection in RGB Videos: From Handcrafted to Deep Learning Approaches. *IEEE Access* **2023**, *11*, 92322–92352. [[CrossRef](#)]
40. Xiao, H.; Peng, K.; Huang, X.; Roitberg, A.; Li, H.; Wang, Z.; Stiefelham, R. Toward Privacy-Supporting Fall Detection via Deep Unsupervised RGB2Depth Adaptation. *IEEE Sens. J.* **2023**, *23*, 29143–29155. [[CrossRef](#)]
41. Cao, Y.; Erdt, M.; Robert, C.; Naharudin, N.B.; Lee, S.Q.; Theng, Y.L. Decision-Making Factors Toward the Adoption of Smart Home Sensors by Older Adults in Singapore: Mixed Methods Study. *JMIR Aging* **2022**, *5*, e34239. [[CrossRef](#)] [[PubMed](#)]
42. Gochoo, M.; Alnajjar, F.; Tan, T.-H.; Khalid, S. Towards Privacy-Preserved Aging in Place: A Systematic Review. *Sensors* **2021**, *21*, 3082. [[CrossRef](#)]
43. Demiris, G.; Hensel, B.K.; Skubic, M.; Rantz, M. Senior Residents’ Perceived Need of and Preferences for “Smart Home” Sensor Technologies. *Int. J. Technol. Assess. Health Care* **2008**, *24*, 120–124. [[CrossRef](#)]
44. Pirzada, P.; Wilde, A.; Doherty, G.H.; Harris-Birtill, D. Ethics and Acceptance of Smart Homes for Older Adults. *Informatics Health Soc. Care* **2022**, *47*, 10–37. [[CrossRef](#)]
45. Gochoo, M.; Tan, T.H.; Velusamy, V.; Liu, S.H.; Bayanduuren, D.; Huang, S.C. Device-Free Non-Privacy Invasive Classification of Elderly Travel Patterns in a Smart House Using PIR Sensors and DCNN. *IEEE Sens. J.* **2018**, *18*, 390–400. [[CrossRef](#)]
46. Uddin, M.Z.; Khaksar, W.; Torresen, J. Ambient Sensors for Elderly Care and Independent Living: A Survey. *Sensors* **2018**, *18*, 2027. [[CrossRef](#)]
47. Camp, N.; Lewis, M.; Hunter, K.; Johnston, J.; Zecca, M.; Di Nuovo, A.; Magistro, D. Technology Used to Recognize Activities of Daily Living in Community-Dwelling Older Adults. *Int. J. Environ. Res. Public Health* **2021**, *18*, 163. [[CrossRef](#)] [[PubMed](#)]

48. Pham, S.; Yeap, D.; Escalera, G.; Basu, R.; Wu, X.; Kenyon, N.J.; Hertz-Picciotto, I.; Ko, M.J.; Davis, C.E. Wearable Sensor System to Monitor Physical Activity and the Physiological Effects of Heat Exposure. *Sensors* **2020**, *20*, 855. [\[CrossRef\]](#)
49. Randazzo, V.; Ferretti, J.; Pasero, E. A Wearable Smart Device to Monitor Multiple Vital Parameters—VITAL ECG. *Electronics* **2020**, *9*, 300. [\[CrossRef\]](#)
50. Shu, F.; Shu, J. An Eight-Camera Fall Detection System Using Human Fall Pattern Recognition via Machine Learning by a Low-Cost Android Box. *Sci. Rep.* **2021**, *11*, 2471. [\[CrossRef\]](#)
51. Gaikwad, S.; Bhatlawande, S.; Shilaskar, S.; Solanke, A. A Computer Vision-Approach for Activity Recognition and Residential Monitoring of Elderly People. *Med. Nov. Technol. Devices* **2023**, *20*, 100272. [\[CrossRef\]](#)
52. Korshunov, P.; Ebrahimi, T. Using Warping for Privacy Protection in Video Surveillance. In Proceedings of the 2013 18th International Conference on Digital Signal Processing (DSP), Fira, Greece, 1–3 July 2013; pp. 1–6.
53. Winkler, T.; Rinner, B. Security and Privacy Protection in Visual Sensor Networks: A Survey. *ACM Comput. Surv.* **2014**, *47*, 1–42. [\[CrossRef\]](#)
54. Padilla-López, J.R.; Chaaraoui, A.A.; Flórez-Revuelta, F. Visual Privacy Protection Methods: A Survey. *Expert Syst. Appl.* **2015**, *42*, 4177–4195. [\[CrossRef\]](#)
55. Rakhmawati, L.; Wirawan; Suwadi. Image Privacy Protection Techniques: A Survey. In Proceedings of the TENCON 2018—2018 IEEE Region 10 Conference, Jeju, Republic of Korea, 28–31 October 2018; pp. 76–80. [\[CrossRef\]](#)
56. Fan, L. Image Pixelization with Differential Privacy. In *Data and Applications Security and Privacy XXXII. DBSec 2018. Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2018; Volume 10980, pp. 148–162. [\[CrossRef\]](#)
57. Zin, T.T.; Htet, Y.; Akagi, Y.; Tamura, H.; Kondo, K.; Araki, S.; Chosa, E. Real-Time Action Recognition System for Elderly People Using Stereo Depth Camera. *Sensors* **2021**, *21*, 5895. [\[CrossRef\]](#)
58. Tateno, S.; Meng, F.; Qian, R.; Hachiya, Y. Privacy-Preserved Fall Detection Method with Three-Dimensional Convolutional Neural Network Using Low-Resolution Infrared Array Sensor. *Sensors* **2020**, *20*, 5957. [\[CrossRef\]](#)
59. Rafferty, J.; Synnott, J.; Nugent, C.; Morrison, G.; Tamburini, E. Fall Detection Through Thermal Vision Sensing. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Cham, Switzerland, 2016; Volume 10070, pp. 84–90. ISBN 9783319487984.
60. Terven, J. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716. [\[CrossRef\]](#)
61. Xiao, X.; Feng, X. Multi-Object Pedestrian Tracking Using Improved YOLOv8 and OC-SORT. *Sensors* **2023**, *23*, 8439. [\[CrossRef\]](#)
62. Chen, H.; Zhou, G.; Jiang, H. Student Behavior Detection in the Classroom Based on Improved YOLOv8. *Sensors* **2023**, *23*, 8385. [\[CrossRef\]](#)
63. Wang, X.; Gao, H.; Jia, Z.; Li, Z. BL-YOLOv8: An Improved Road Defect Detection Model Based on YOLOv8. *Sensors* **2023**, *23*, 8361. [\[CrossRef\]](#)
64. Bao, J.; Li, S.; Wang, G.; Xiong, J.; Li, S. Improved YOLOV8 Network and Application in Safety Helmet Detection. *J. Phys. Conf. Ser.* **2023**, *2632*, 012012. [\[CrossRef\]](#)
65. Wang, S.; Zhang, X.; Ma, F.; Li, J.; Huang, Y. Single-Stage Pose Estimation and Joint Angle Extraction Method for Moving Human Body. *Electronics* **2023**, *12*, 4644. [\[CrossRef\]](#)
66. van den Oord, A.; Dieleman, S.; Zen, H.; Simonyan, K.; Vinyals, O.; Graves, A.; Kalchbrenner, N.; Senior, A.; Kavukcuoglu, K. WaveNet: A Generative Model for Raw Audio. *arXiv* **2016**, arXiv:1609.03499.
67. Shen, Z.; Zhang, Y.; Lu, J.; Xu, J.; Xiao, G. SeriesNet: A Generative Time Series Forecasting Model. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–8. [\[CrossRef\]](#)
68. Wang, J.H.; Lin, G.F.; Chang, M.J.; Huang, I.H.; Chen, Y.R. Real-Time Water-Level Forecasting Using Dilated Causal Convolutional Neural Networks. *Water Resour. Manag.* **2019**, *33*, 3759–3780. [\[CrossRef\]](#)
69. Chuya-Sumba, J.; Alonso-Valerdi, L.M.; Ibarra-Zarate, D.I. Deep-Learning Method Based on 1D Convolutional Neural Network for Intelligent Fault Diagnosis of Rotating Machines. *Appl. Sci.* **2022**, *12*, 2158. [\[CrossRef\]](#)
70. Kiranyaz, S.; Avci, O.; Abdeljaber, O.; Ince, T.; Gabbouj, M.; Inman, D.J. 1D Convolutional Neural Networks and Applications: A Survey. *Mech. Syst. Signal Process.* **2021**, *151*, 107398. [\[CrossRef\]](#)
71. Cheng, C.; Zhang, C.; Wei, Y.; Jiang, Y.G. Sparse Temporal Causal Convolution for Efficient Action Modeling. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 592–600. [\[CrossRef\]](#)
72. Hamad, R.A.; Kimura, M.; Yang, L.; Woo, W.L.; Wei, B. Dilated Causal Convolution with Multi-Head Self Attention for Sensor Human Activity Recognition. *Neural Comput. Appl.* **2021**, *33*, 13705–13722. [\[CrossRef\]](#)
73. Hou, S.; Wang, C.; Zhuang, W.; Chen, Y.; Wang, Y.; Bao, H.; Chai, J.; Xu, W. A Causal Convolutional Neural Network for Multi-Subject Motion Modeling and Generation. *Comput. Vis. Media* **2023**, *10*, 45–59. [\[CrossRef\]](#)
74. Jain, P.K.; Choudhary, R.R.; Singh, M.R. A Lightweight 1-D Convolution Neural Network Model for Multi-Class Classification of Heart Sounds. In Proceedings of the 2022 International Conference on Emerging Techniques in Computational Intelligence (ICETCI), Hyderabad, India, 25–27 August 2022; pp. 40–44. [\[CrossRef\]](#)
75. Li, F.; Liu, M.; Zhao, Y.; Kong, L.; Dong, L.; Liu, X.; Hui, M. Feature Extraction and Classification of Heart Sound Using 1D Convolutional Neural Networks. *EURASIP J. Adv. Signal Process.* **2019**, *2019*, 59. [\[CrossRef\]](#)
76. Jiang, Z.; Lai, Y.; Zhang, J.; Zhao, H.; Mao, Z. Multi-Factor Operating Condition Recognition Using 1D Convolutional Long Short-Term Network. *Sensors* **2019**, *19*, 5488. [\[CrossRef\]](#)

77. Chen, C.-C.; Liu, Z.; Yang, G.; Wu, C.-C.; Ye, Q. An Improved Fault Diagnosis Using 1D-Convolutional Neural Network Model. *Electronics* **2021**, *10*, 59. [\[CrossRef\]](#)
78. Klambauer, G.; Unterthiner, T.; Mayr, A.; Hochreiter, S. Self-Normalizing Neural Networks. *Adv. Neural Inf. Process. Syst.* **2017**, *2017*, 972–981.
79. Kingma, D.P.; Ba, J.L. Adam: A Method for Stochastic Optimization. *arXiv* **2017**, arXiv:1412.6980.
80. Kwolek, B.; Kepski, M. Human Fall Detection on Embedded Platform Using Depth Maps and Wireless Accelerometer. *Comput. Methods Programs Biomed.* **2014**, *117*, 489–501. [\[CrossRef\]](#)
81. Charfi, I.; Miteran, J.; Dubois, J.; Atri, M.; Tourki, R. Optimized Spatio-Temporal Descriptors for Real-Time Fall Detection: Comparison of Support Vector Machine and Adaboost-Based Classification. *J. Electron. Imaging* **2013**, *22*, 041106. [\[CrossRef\]](#)

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