

Article

Design and Implementation of Nursing-Secure-Care System with mmWave Radar by YOLO-v4 Computing Methods

Jih-Ching Chiu, Guan-Yi Lee, Chih-Yang Hsieh and Qing-You Lin *

Department of Electrical Engineering, National Sun Yat-Sen University, Kaohsiung 80424, Taiwan; chiujihc@ee.nsysu.edu.tw (J.-C.C.); s224485782@gmail.com (G.-Y.L.); ms0664524@gmail.com (C.-Y.H.)

* Correspondence: tranquil076@gmail.com

Abstract: In computer vision and image processing, the shift from traditional cameras to emerging sensing tools, such as gesture recognition and object detection, addresses privacy concerns. This study navigates the Integrated Sensing and Communication (ISAC) era, using millimeter-wave signals as radar via a Convolutional Neural Network (CNN) model for event sensing. Our focus is on leveraging deep learning to detect security-critical gestures, converting millimeter-wave parameters into point cloud images, and enhancing recognition accuracy. CNNs present complexity challenges in deep learning. To address this, we developed flexible quantization methods, simplifying You Only Look Once (YOLO)-v4 operations with an 8-bit fixed-point number representation. Cross-simulation validation showed that CPU-based quantization improves speed by 300% with minimal accuracy loss, even doubling the YOLO-tiny model's speed in a GPU environment. We established a Raspberry Pi 4-based system, combining simplified deep learning with Message Queuing Telemetry Transport (MQTT) Internet of Things (IoT) technology for nursing care. Our quantification method significantly boosted identification speed by nearly 2.9 times, enabling millimeter-wave sensing in embedded systems. Additionally, we implemented hardware-based quantization, directly quantifying data from images or weight files, leading to circuit synthesis and chip design. This work integrates AI with mmWave sensors in the domain of nursing security and hardware implementation to enhance recognition accuracy and computational efficiency. Employing millimeter-wave radar in medical institutions or homes offers a strong solution to privacy concerns compared to conventional cameras that capture and analyze the appearance of patients or residents.

Keywords: mmWave radar; integrated sensing and communication; convolutional neural network; artificial intelligence of things; gesture recognition



Citation: Chiu, J.-C.; Lee, G.-Y.; Hsieh, C.-Y.; Lin, Q.-Y. Design and Implementation of Nursing-Secure-Care System with mmWave Radar by YOLO-v4 Computing Methods. *Appl. Syst. Innov.* **2024**, *7*, 10. <https://doi.org/10.3390/asi7010010>

Academic Editor: Friedhelm Schwenker

Received: 25 November 2023

Revised: 5 January 2024

Accepted: 16 January 2024

Published: 19 January 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/>).

1. Introduction

Due to the issue of increasing numbers of aged people in recent years, there is a trend toward home-care systems. While there are some safety protection devices, such as cameras and phones, on the market today [1,2], they cannot protect a person's privacy, so they cannot be used in bathrooms, toilets, and changing rooms, nor can they be used in dark places without lighting. The elderly are more likely to need assistance in these private spaces and times. For the above reasons, Wi-Fi waves, up to 5 GHz, have been used for gesture recognition with AI computing models to detect human behaviors [3–5]. Due to factors of lower frequency, the effects of electromagnetic interference seriously hinder the recognition rate and then affect the promotion of its application [6]. Increasing the application frequency will help to improve this phenomenon.

Millimeter wave (mmWave), above 20 GHz, is a special radar technology that uses short-wave electromagnetic waves [7]. A study proposes an integration of the Car-to-Car Network-Hierarchical deep neural network (CtCNET-HDRNN) model with Fifth generation (5G) mmWave [8]. However, this linear machine learning approach does not perform well in recognizing two-dimensional images. The integration of sensing functions

is becoming a key feature of 6G Radio Access Networks (RANs) [9], allowing the use of dense small-area infrastructure to build sensing networks. Millimeter-wave radar transmits signals with wavelengths in the millimeter range, which is one of the advantages of the technology. By capturing the reflected signals, the radar system can determine the distance, velocity, and angle of an object to create a micro-Doppler effect, which can be processed to provide a unique data set, such as distance, velocity, and angle, as well as the ability to distinguish between different targets, so that the sensor can detect the characteristics of different objects within the detection range. For example, this data allows sensors to sense echo signals, and modulation effects from tiny motions, including characteristics typical of objects such as the rotational speed of a bicycle wheel, a person's swaying arm, or an animal's running limbs. In this paper, we will build a sensing system that can detect the object position and gesture recognition for nursing-secure care using an AI model, which can be used in embedded systems. We take the MQTT [10] IoT protocols to transmit our identification results to create an intelligent system on a smart embedded platform and run a test system to validate our research.

The general architecture of this paper is as shown in Figure 1. Sensors can detect user gestures or postures, and after computation through our system, commands can be emitted to control designated IoT devices.

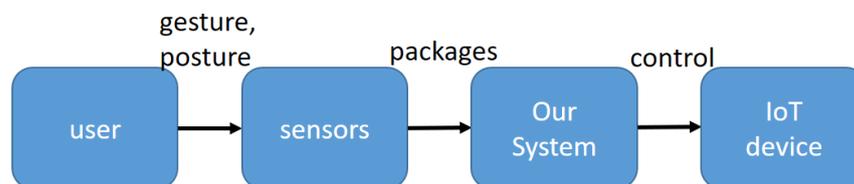


Figure 1. General architecture of the proposed system.

2. Background

2.1. ISAC

The ISAC system mainly integrates sensing and communication [11], and is considered to be one of the most promising technologies to realize the two key requirements in 6G. With the development of networks and the evolution of wireless systems, ISAC has gradually become a hot research topic.

ISAC has recently been proposed for numerous emerging applications, including but not limited to in-vehicle networking, environmental monitoring, remote sensing, IoT, smart cities, and indoor services such as human activity and gesture recognition. More importantly, ISAC was recently identified as an enabling technology for 5G/6G and next-generation Wi-Fi systems.

An important focus in the future of ISAC development is to improve the accuracy, so as to facilitate the communication between UAVs for more complex tasks, and enable simultaneous imaging, mapping, and localization to achieve mutual performance improvements for these functions.

In addition, it is hoped that human senses can be enhanced, such as adding some features to detect things that humans cannot see with their eyes, such as information on blood vessels, organ status, or information on vital signs such as breathing, heartbeat, etc.

2.2. Widar3.0

Wireless devices often use ubiquitous commercial Wi-Fi for sensing systems, which is named DFWS (device-free wireless sensing), called Wi-Fi sensing [5]. The research focuses on how to extract highly identifiable features from channel state information (CSI). In order to obtain more CSI identification features in cross-domain gesture recognition, a system named Widar3.0 was proposed in mid-2019. It combines the advantages of convolutional neural networks and long short-term memory networks into a joint CNN and LSTM in a model. The spatial features learned by the CNN are used as the inputs for the LSTM to simulate the temporal features. Widar3.0 can be used directly through existing

equipment without retraining. However, after the actual test, it is found that after the noise information in the environment is eliminated, redundant echoes will still be generated, which will directly affect the recognition rate after passing through CNN, and this method is not suitable for subtle gesture recognition.

2.3. YOLO-v4 Machine Learning Model

In this paper, we use the YOLO-v4 machine learning model as the recognition tool for AI approaches and apply the detection methods of millimeter wave to create a secure-care system with mmWave radar. The YOLO (You Only Look Once) series, an excellent object detection model, is based on the Convolutional Neural Network (CNN) architecture, known for its high accuracy and speed. However, the YOLO [12] series still faces significant challenges when dealing with embedded systems or resource-constrained environments. The YOLO-v4 and YOLO v4-tiny models utilize Darknet [13] capabilities for neural network construction, weight initialization, forward propagation, and backward propagation, facilitating the processes of training and recognition.

The YOLO-v4 network architecture is shown in Figure 2. YOLO-v4 is roughly composed of four parts: Darknet, SPP, PANet, and YOLO-output, with a total of 161 layers.

- CSPDarknet53: The CSPDarknet53 layer is the entrance of the whole network, as part of the Backbone. (the blue frame part)
- SPP: Feature maps given before the last layer Concat of CSPDarknet53, as part of the Neck. (green frame part)
- PANet: The actions of sitting down and sampling and upsampling in PANet are also used here as part of the Neck.
- YOLO-Output: Finally, YOLO-Output outputs the final results, including the target position of the prediction frame and the reliability of the detection target. (Yellow frame part)

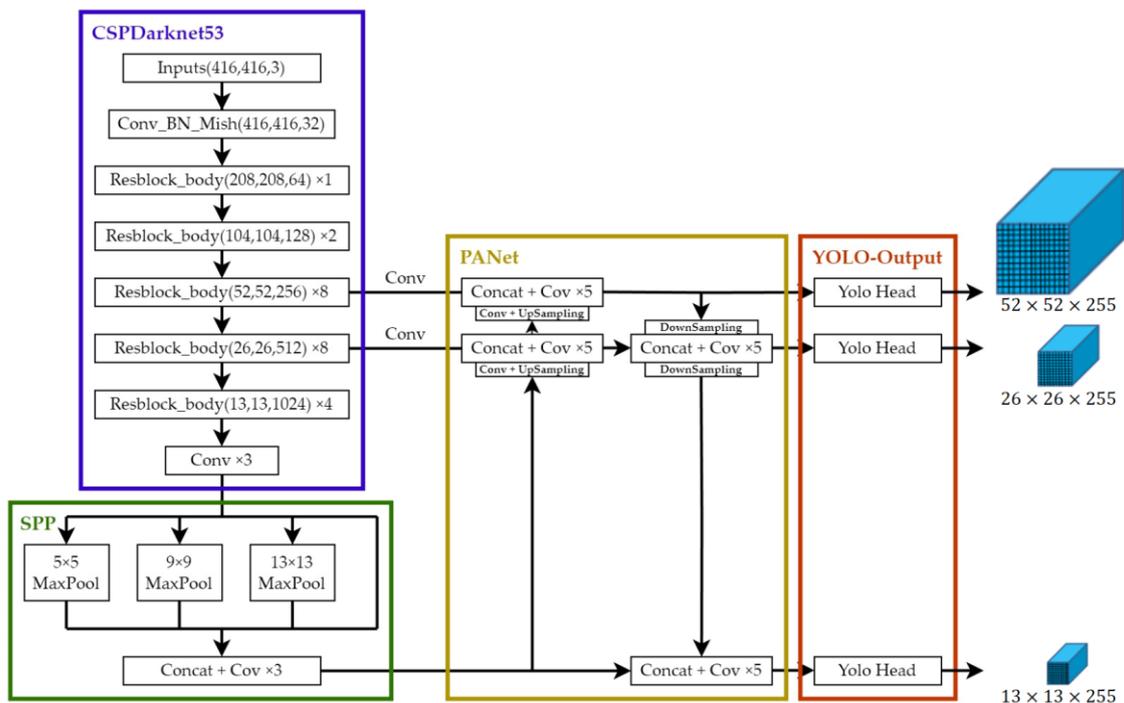


Figure 2. YOLO-v4 Architecture.

Currently, YOLO series networks are often processed using Graphics Processing Units (GPUs) or custom hardware designs, such as Field Programmable Gate Arrays (FPGAs) that enable high degrees of parallelism for computations [14,15]. However, in the context of embedded systems or resource-limited scenarios, CNN-based models still face challenges

due to data computation latency and limited data access bandwidth. To address this, some studies have started applying hardware accelerators to CNN models to enhance computational efficiency. For specific image data, such as sensor data images with lower information content, optimizing the number of layers without sacrificing model accuracy can directly impact the model’s speed.

In deep learning, a convolutional neural network (CNN) is widely used for image recognition. To accurately identify similar images, the number of convolutional layers can be increased to obtain and abstract features of the image. However, as the number of layers increases, a large number of weight tables are generated, which increases the demand for computing resources, computing complexity, and storage space required for the weight tables, thus limiting the performance of embedded systems using CNNs. To solve related problems, many related studies have explored various methods, such as processing computing data, computing directly in memory, or designing a dedicated CNN model to retrain weight tables [16,17], to increase computing speed and reduce storage space. As the CNN calculation results indicate whether a feature is prominent, the classification process finds the category with the highest value among all the categories as the classification result. Based on this principle, this study demonstrates that as long as the relationships between the magnitudes of the calculation results are maintained, the accuracy can be nearly lossless.

2.4. YOLO v4-Tiny Machine Learning Model

YOLO v4-tiny [18], as a lightweight object detection model of YOLO-v4, continues the advantages of the YOLO series and has higher accuracy and fast identification capabilities. It provides effective and accurate object detection in resource-constrained environments and is suitable for a variety of application scenarios, including embedded systems, mobile devices, and real-time vision applications.

Compared with YOLO-v4, YOLO v4-tiny has only one-tenth of the weight parameter, has considerable advantages in speed and storage space, and is easier to adapt to resource-limited situations such as embedded systems. YOLO v4-tiny is an object detection model based on the convolutional neural network (CNN), with a total of 38 layers and can be divided into three parts, Backbone, Neck, and YOLO head, as shown in Figure 3.

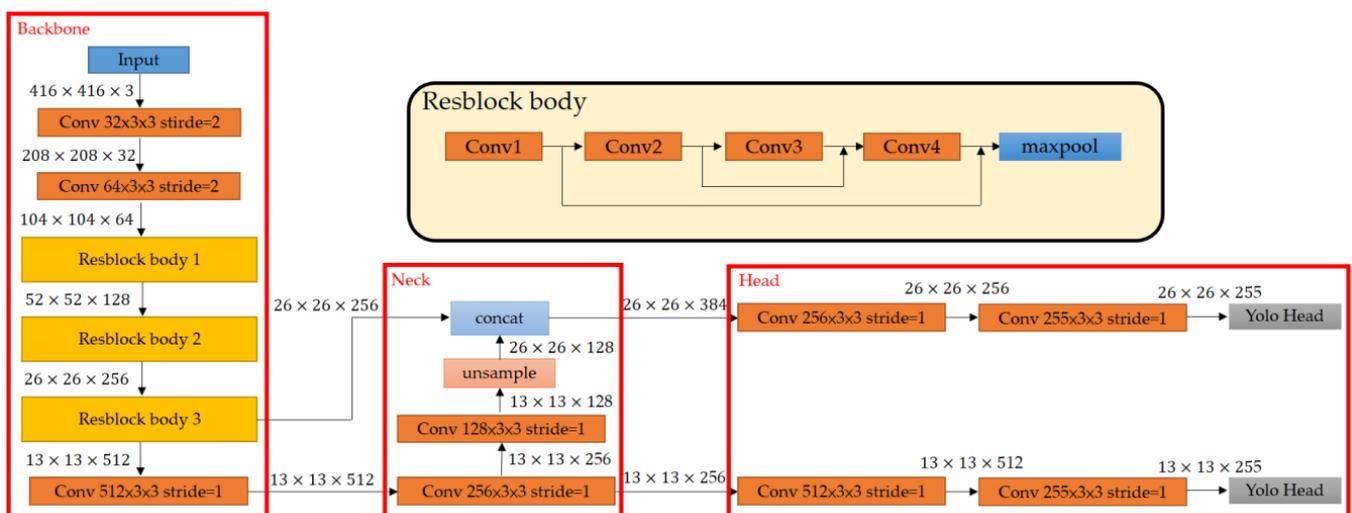


Figure 3. YOLO v4-tiny model architecture.

- Backbone

The backbone consists of a series of convolutional layers and a Resblock body [19]. A Resblock body is one of the key parts of the backbone. This structure retains low-level features while extracting deeper features, effectively increasing the depth of the model

and helping the convolutional layer to capture features of different scales, thereby better-capturing targets of different sizes to improve the accuracy of the model.

- Neck

The features of feature maps of different scales are fused through upsampling and convolution layers to improve the detection capabilities of targets of different scales without increasing excessive calculations.

- YOLO head

YOLO v4-tiny uses two detection layers of different scales, which are responsible for bounding boxes of three specific scales. YOLO v4-tiny converts the received feature maps into target detection results through the detection layer and predicts the corresponding bounding boxes with classified labels and confidence scores.

3. Materials and Methods

The comparison of the advantages and disadvantages of the previously introduced methods with those of our method are shown in Table 1. To safeguard user information and offer comprehensive security protection across all time frames and areas, establishing a stable signal to enhance recognition rates is essential. Due to the increasingly complex electromagnetic environment and signals, traditional identification methods struggle to achieve desired recognition rates. This paper proposes a deep learning-based approach that analyzes the received signals to generate point cloud diagrams, enabling classification algorithms to more effectively differentiate between data points.

The architecture of the nursing-secure-care system is shown in Figure 4, which is constructed by a Millimeter sensor, gesture recognition system, and IoT communication system. In this paper, we focus on how to make the point-cloud image, how to build the recognition system with YOLO, and how to make the MQTT commands to control secure devices for the embedded systems. The confirmed posture results are sent via MQTT to the Topic within the Broker, which in this case is Pi-4. Subsequently, messages are sent separately based on the subscribed content of the subscribers, enabling the rapid transmission of the current posture to users' devices. Millimeter sensors transmit electromagnetic wave signals, which are reflected by objects, similar to radar systems. By capturing reflected signals, radar systems can determine the distance, velocity, and angle of objects for micro-Doppler effects, which can be processed to provide unique data sets such as, distance, velocity, and angle, as well as the ability to distinguish different targets, which enables the sensor to detect the characteristics of different objects within the detection range. These data, for example, allow sensors to sense the echo signals—modulation effects from tiny movements—that include the typical features of objects such as, the spinning speed of a bicycle wheel, the swaying arms of a person, or the running limbs of an animal. In this project, we will build a sensing system with an AI model that can be used in the embedded systems. To approach these goals, we will go through this project as follows. First, we will generate the pixel coordinates by calculating the echo signals of the Doppler effects in period. Second, the pixel coordinates will be used to build the pixel cloud images with layer-coloring methods. Noise can be filtered with mathematical morphologies, such as erosion and dilation methods. Third, we classify the pixel cloud images and train them with machine learning models, such as YOLO-v4, to obtain weight tables for sensing event recognition. Fourth, we simplify the machine learning models and build up the performance-oriented programs to ensure that we can run the mmWave sensing system in the embedded systems.

Table 1. A comparison of the previously introduced method with those of this paper.

Name	Type of Sensor	Application	Recognition Technology	Disadvantages	Advantages	Recognition Rate
Traditional hand/face recognition [19]	Optical camera	hand/face recognition	Various types of CNN models	only suitable for static objects, incapable of posture movement or changes, dependent on light sources	actual images are obtained, the highest recognition rate	about 90~100%
Multi-sensor [2]	Optical/depth camera Radar	hand recognition	DNN combining of Con3D	higher interdependence among sensors affected by environmental conditions	enhance a certain level of recognition accuracy without environmental influence	about 75–93%
Widar3.0 [5]	Wi-Fi	hand recognition, Person localization	CNN-LSTM combining of Con3D	Environmental noise reduces recognition rates, especially for subtle gestures.	Capable of using existing devices without the need for retraining gestures.	about 92.7%
This paper	mmWave	pose/gesture recognition, Person localization, heartbeat detection	YOLO-v4 YOLO-tiny	need to involve moving objects, overlapping objects are harder to distinguish	unaffected by environmental conditions, can expedite computations through quantization methods	about 92–95%

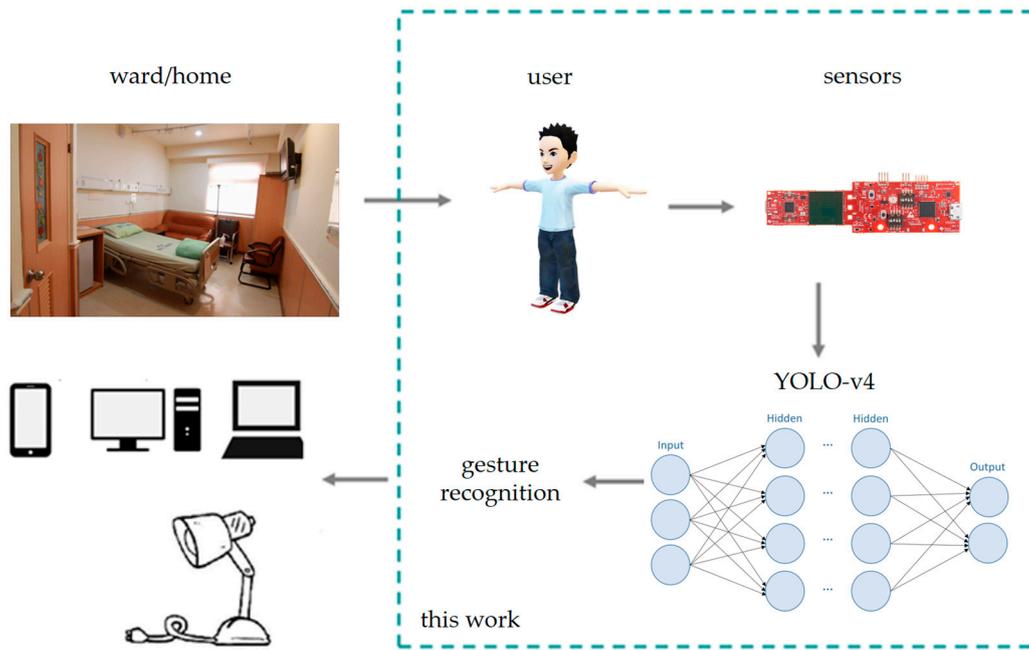


Figure 4. The architecture of the proposed system.

This paper employs the TI IWR6843AOP single-chip mmWave sensor as shown in Figure 5. This chip operates within the frequency range of 60 GHz to 64 GHz and functions adequately in general environments ranging from -40 degrees to 105 degrees Celsius. It consists of 4 receiver (RX) and 3 transmitter (TX) antenna modules. The receiver operates at a Baud rate of 115,200 while the transmitter can reach up to 921,600, facilitating high-speed and precise data transmission. It offers a transmission speed of 50 ms/frame. In this study, we utilize the UART interface to connect with a computer. The received data is analyzed by the computer and plotted on the canvas (Mat window) within the program, enabling our observation.

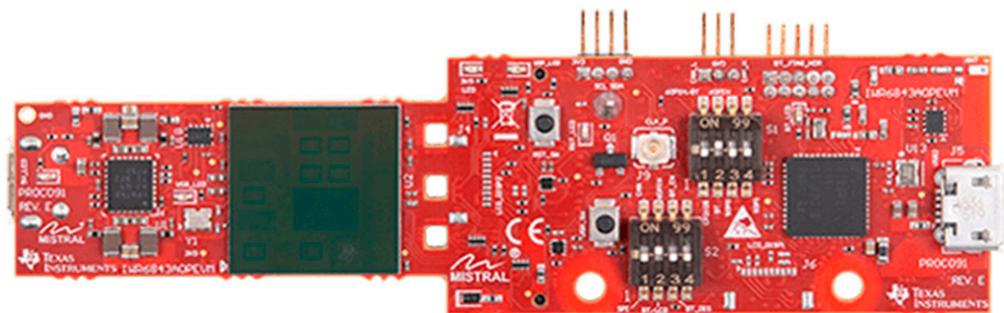


Figure 5. mmWave IWR6843AOP module.

To make it easier for users to use our entire system, we assign colors to the point cloud according to the distance. When the position is far away from what we need to identify, we will lighten the color. In this way, we will tell the user where the setting needs to be adjusted. to find the best distance for attitude recognition. Through the results, it is found that this method has the following advantages. From walking (Stand), sitting (Sit), lying down (for sleep time: Lie), falling (abnormal motion detection: Fall), switching lights (Light), and help (distress: Help) gestures, such as those shown in Figure 6, all can be completed within the complete set of safety protection behavior system in our designs.

The gesture recognition results						
	Stand	Sit	Help		Gesture /Posture	Semi-coloring Point Clouds
Gesture / Posture					Fall 	
Semi-coloring Point Clouds					Lie 	
					Light 	

Figure 6. The gestures for nursing-secure-care system.

When the mmWave detects objects by the micro-Doppler effect, the parameters of elevation, azimuth, and Doppler velocity, will be obtained. We will utilize Equations (1)–(3) to compute coordinates, thereby converting the mmWave spherical coordinates into Cartesian coordinates, as shown in Figure 7, as follows:

$$X = R \times \cos(\text{elevation}) \times \sin(\text{azimuth}) \tag{1}$$

$$Y = R \times \cos(\text{elevation}) \times \cos(\text{azimuth}) \tag{2}$$

$$Z = R \times \sin(\text{elevation}) \tag{3}$$

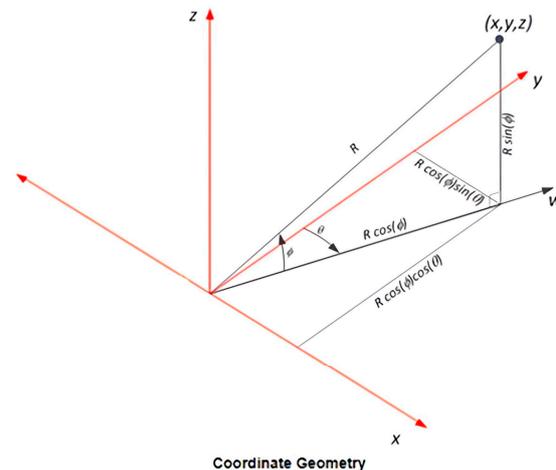


Figure 7. mmWave point cloud coordinates.

To increase the gesture recognition rate, we propose the layer-coloring method. These points will be made as semi-coloring pixel cloud images according to the height, distance, and direction. The coloring mapped table for height is shown in Table 2. We use pixel cloud images, as illustrated in Figure 8, to create classifying and labelling as training target objects for the machine learning model. After training is complete, we can use the weight tables to make an event recognition system. To obtain the time sequence of the scenario, we grab the pixel cloud images of each frame and overlap them by an adaptive time paragraph, such as 0.2 s. The semi-coloring pixel cloud images become the input source data of the mmWave sensing system.

Table 2. The coloring mapped table for height.

Coloring Mapped Table				
Height	0~60 cm	60~110 cm	110~220 cm	Over 220 cm
color	Red	Green	Blue	Yellow

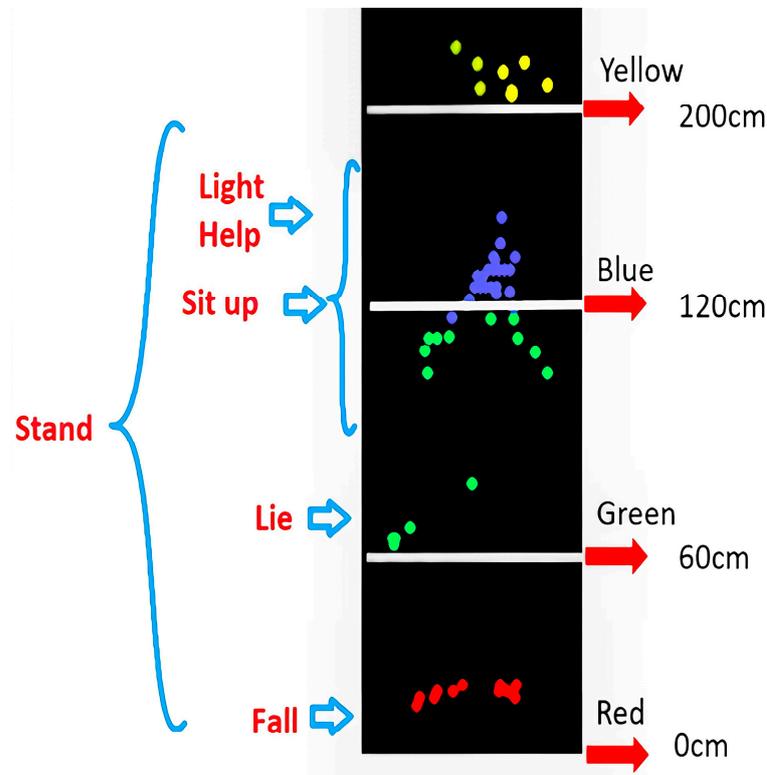


Figure 8. mmWave semi-coloring pixel cloud images.

From Table 3, the recognition rates without semi-coloring illustrate that misjudgments are serious, such that the gesture Fall can be recognized as the gesture Lie, the gesture Light, and the gesture Fall, making it impossible to accurately identify the correct one.

Table 3. The recognition rates without semi-coloring.

	Stand	Sit	Lie	Help	Light	Fall
Stand	60%	30%	0%	10%	0%	0%
Sit	30%	60%	0%	8%	2%	0%
Lie	0%	0%	40%	0%	20%	40%
Help	10%	8%	0%	82%	0%	0%
Light	0%	2%	20%	0%	50%	28%
Fall	0%	0%	40%	0%	28%	32%

From Table 4, the recognition rates with semi-coloring illustrate that the error rate decreases and approaches 0%, which means that the correct posture can be accurately identified. Therefore, the semi-coloring method is an excellent approach to recognizing the gestures.

Table 4. The recognition rates with semi-coloring.

	Stand	Sit	Lie	Help	Light	Fall
Stand	99%	1%	0%	0%	0%	0%
Sit	1%	98%	0%	0%	1%	0%
Lie	0%	0%	97%	0%	1%	2%
Help	0%	0%	0%	97%	3%	0%
Light	0%	1%	1%	3%	94%	1%
Fall	0%	0%	2%	0%	1%	97%

In deep learning, the convolutional neural network (CNN) is a type of deep neural network, which is the most common mode in current applications and is best at image processing. It is inspired by the human visual nervous system and is designed using a variant of multilayer perceptron that requires minimal preprocessing, based on their shared weight architecture and translation-invariant features.

The CNN method has two major characteristics:

1. It can effectively reduce the dimensionality of pictures with large amounts of data into small amounts of data.
2. It can effectively retain image features and conform to the principles of image processing.

The first problem solved by CNN is to simplify complex problems. It reduces the dimension of a large number of parameters into a small number of parameters and then processes them. It retains the characteristics of the image in a visual-like way, and when the image is flipped, rotated, or positionally changed, it can also effectively identify similar images. The YOLO-v4 model is a machine learning operation model with an optimization strategy in the CNN field. The YOLO v4-tiny model is often used in embedded environments due to its large number of layers. We will deeply survey and implement the mmWave sensing system in the YOLO-v4 model in the embedded platform. To speed the edge computing, three target problems will be studied in depth:

1. Domain Quantization for saving storage and improving computing performance.
2. CNN layers to be reduced, based on the YOLO v4-tiny model as the specific light CNN model to speed up object recognition computing.
3. Data Parallelism programming method to be used for coding the CNN model to approach power-efficient computing in embedded systems.

In this paper, we will focus on four objective tasks: (1) Preparation of precise pixel cloud images. (2) Build a mmWave sensing system on an embedded system to detect object location, recognize gesture and posture, distinguish life signs, and track movement of objects. (3) Simplify the machine learning models and create performance-oriented programs to enable the running of a mmWave sensing system in the embedded systems. (4) Test and verify our design. When this project is completed, the effective mmWave sensing system will be created on a smart embedded platform and a demon system will be run to verify our studies.

4. Results

4.1. The Proposed Quantization Mechanism

To improve the complex problem of YOLO-v4 computing, we propose a quantization method to increase computation speed and reduce storage space. To reduce time wasting in the process of quantization, we established a network model without retraining, and a YOLO-v4 identification method that can maintain a certain accuracy. We convert the input image data and weight data into fixed-point representation through our quantization method to improve the huge computing process with floating-point representation. In this way, the computational complexity and the required computational resources can be reduced.

The proposed quantization mechanism is shown in Figure 9. In the initial stage of floating-point quantization, the initial value of the integer part is set to 1 because of the hidden bit of the floating-point number. To retain the maximum value of the decimal part, we initiate the decimal part of the quantization format with the 9th to 14th digits of the decimal part at the floating-point number. In the second step, we subtract 127 from the exponent part of the value to determine the displacement value (N), which represents the displacement direction and displacement amount of the integer decimal boundary. When N is greater than 0, it means that the integer part of this value needs more than 2 bits, so the integer-decimal boundary is right-shifted; when N is less than 0, the value has no integer part, so the integer-decimal boundary is left-shifted to preserve the maximum decimal bit precision. In the third step, before the dynamic quantization of floating-point numbers, the image data and the distribution range of the weight data will be integrated separately, and the displacement data of the respective floating-point numbers will be found and stored. In the fourth step, the quantized integer bits will be aligned according to the size of the displacement data N, and the quantization will be completed. The quantized value can be used for direct calculation.

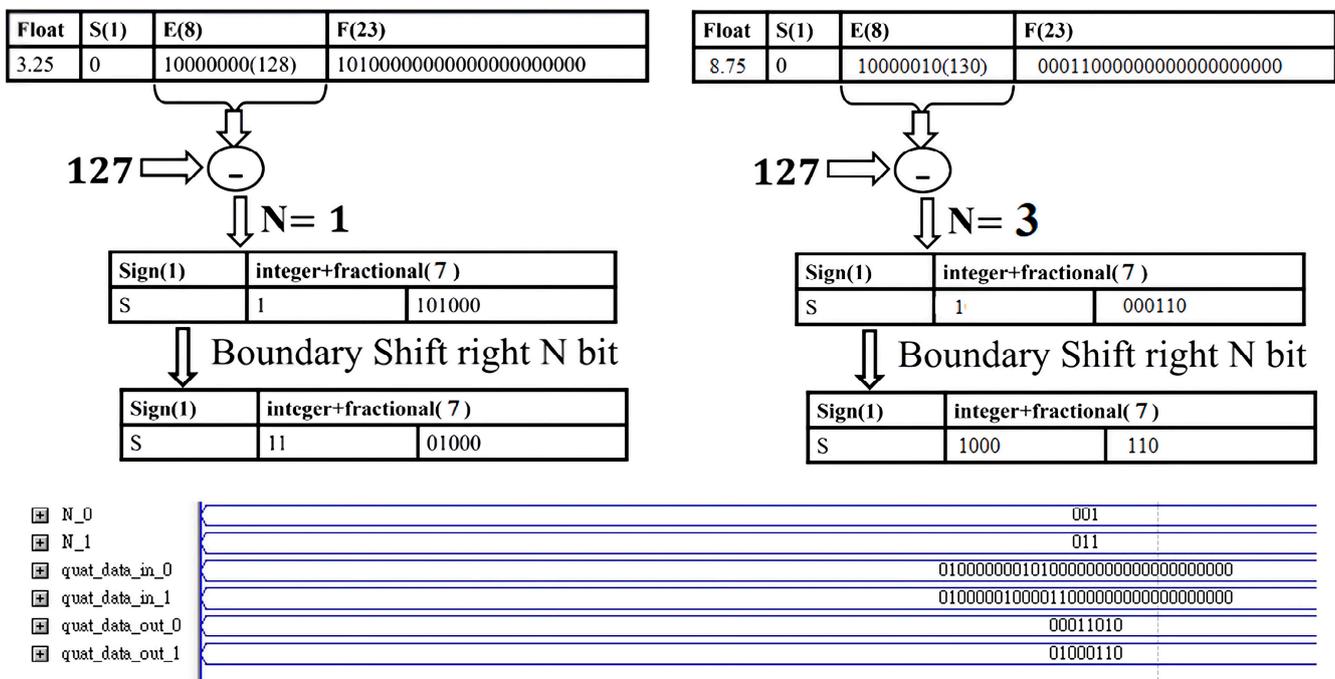


Figure 9. The proposed quantization mechanism.

4.2. Results of Quantization

The analysis results, as shown in Table 5, illustrate that the computing speed has been greatly improved after using our proposed quantization method on the YOLO-v4 model. Compared with the performance of personal computers, the original 32-bit floating-point number is converted into an 8-bit fixed-point number, and the time to identify a photo is approximately decreased from 3000 ms to about 1500 ms, an improvement of about 2 fold. In the embedded system (Raspberry Pi4), the execution time is reduced to 11,700 ms, which is about 3.2 times faster. We also start to test by the In the YOLO v4-tiny model results, the computing speed can be improved by about 3 times in both the personal computer and the embedded system. Importantly, the recognition rate is lower by only 0.04 times, which is within an acceptable range. The line chart of time comparisons is shown in Figure 10.

Table 5. Comparison of results before and after YOLO-v4 model quantization.

Model	Computer Type	Data Type	Time (per Picture)	Accuracy	Improvement
YOLO-v4 (161 layers)	Computer Intel i7-6700	float32	3051 ms	98.7%	1556 ms
		int8	1495 ms	98%	Up to 2.04 times
	Notebook Intel i5-5200	float32	4587 ms	99.2%	2487 ms
		int8	2100 ms	99.1%	Up to 2.18 times
	pi-4 ARM Cortex-A72	float32	17,144 ms	99.2%	11,793 ms
		int8	5351 ms	99.1%	Up to 3.2 times
YOLO-v4 tiny (38 layers)	Computer Intel i7-6700	float32	406 ms	95.2%	269 ms
		int8	137 ms	94.1%	Up to 2.96 times
	Notebook Intel i5-5200	float32	649 ms	95.3%	400 ms
		int8	249 ms	92.8%	Up to 2.6 times
	pi-4 ARM Cortex-A72	float32	1944 ms	95.4%	1338 ms
		int8	606 ms	92.8%	Up to 3.2 times

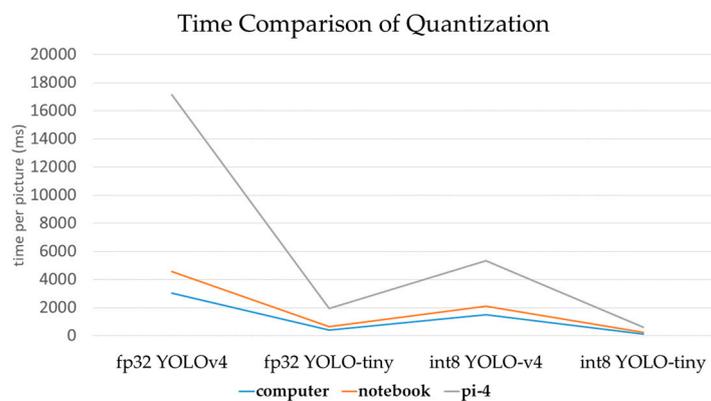


Figure 10. Comparison of results before and after the quantization processing of YOLO-v4 model.

Based on Figures 11 and 12, we can conclude that the computation time of the proposed quantification method and model simplification in this paper can indeed enhance the competitiveness of the YOLO model’s computation, especially in embedded systems (YOLO v4-tiny model + int8).

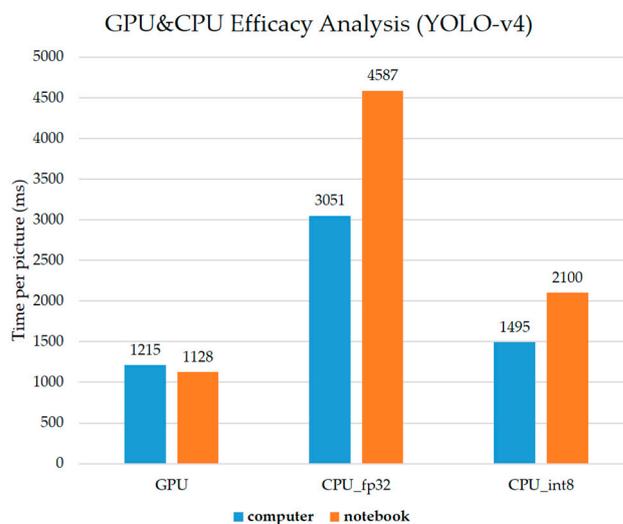


Figure 11. GPU vs. CPU only (YOLO-v4).

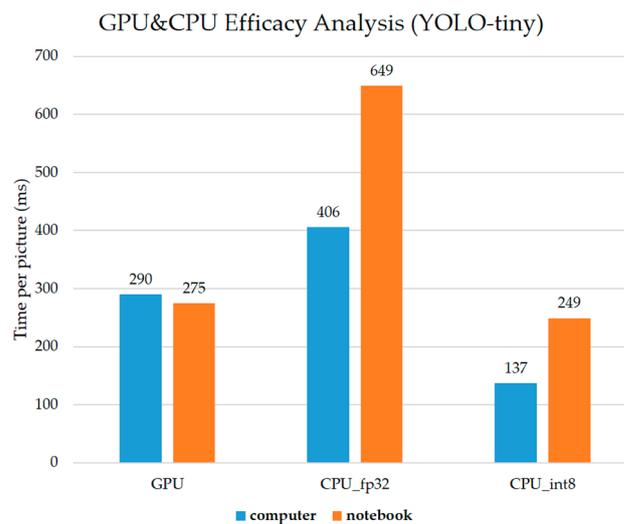


Figure 12. GPU vs. CPU only (YOLO v4-tiny).

4.3. Results of the Optimization of YOLO v4-Tiny Architecture

mmWave point cloud images are different from ordinary photos. The amount of information contained in the image is created using a coloring algorithm. Therefore, we believe that during training and recognition, we can optimize the YOLO v4-tiny architecture layer by layer to obtain the same recognition effect. YOLO v4-tiny contains three Resblock bodies. After replacing the Resblock bodies with the convolution layer in the sequence case, we analyzed the impact on the accuracy of the point cloud image recognition after training and used mAP (mean average precision) [20] to judge the recognition ability. According to the evaluation results, shown in Table 6, for the identification of mmWave point cloud images, replacing Resblock body 1 with a convolutional layer will cause a sharp drop in mAP. Replacing Resblock body 2 with 1 convolution operation and Resblock body 3 with 2 convolution operations will not have a drastic impact on mAP. Therefore, we can conclude that when identifying mmWave point cloud images, Resblock body 2 and Resblock body 3 should be replaced by 3 convolutional layers to reduce the number of layers, which do not significantly affect the identification ability.

Table 6. mAP changes after the convolutional layer replaces the Resblock body.

YOLO-v4 Tiny Architecture	mAP(@0.50)
Original	0.562309
Reduce the Resblock body 1 to 1 convolution operation	0.170274
Reduce the Resblock body 2 to 1 convolution operation	0.477842
Reduce the Resblock body 3 to 2 convolution operation	0.548834
Reduce the Resblock body 2 and 3 to 3 convolution operation	0.421765

5. Discussion

Although the integration of sensing tools and communication systems has been recognized as a pivotal area, there is a deficiency in comprehensive exploration aimed at simplifying these models to achieve heightened efficiency and reduced resource consumption. There are limited comprehensive studies showcasing the effective amalgamation of these components.

This paper aims to integrate sensing tools with communication systems to facilitate efficient data transmission and processing. It addresses the challenge of simplifying intricate machine learning models, specifically the YOLO v4-tiny model, in order to improve performance, while minimizing resource utilization. Additionally, the research focuses

on the development of hardware-based quantization techniques designed to convert data from floating-point to fixed-point number formats. This endeavor is intended to expedite computation processes and reduce storage requirements.

Simplified machine learning models and hardware-based quantization techniques can benefit scientists and researchers by providing efficient methods for processing data and reducing computational resources, thereby accelerating the pace of research in machine learning and related fields. Moreover, for society, these advancements can lead to the creation of more efficient and accurate systems for healthcare, security, and surveillance, contributing to improved safety measures, healthcare monitoring, and technological advancements that benefit society at large.

We analyze the effects of the proposed quantization method to be used in YOLO models, which are computed on PC CPU-only (tagged as CPU) stations and CPU + GPU (tagged as GPU) stations. Figure 9 illustrates that computing performance with the proposed quantization method on CPU is lower by a time difference of 280 ms compared with GPU in YOLO-v4 models. Figure 10 shows that computing performance with the proposed quantization method on CPU is higher by a time difference of 153 ms compared with GPU in YOLO v4-tiny models. Following the above results, the proposed quantization method is an important mechanism for the computing of the YOLO-v4 model. In particular, the YOLO v4-tiny model with the proposed quantization method on CPU has excellent performance compared with GPU. This result proves that the proposed quantization method is suitable for use on the YOLO v4-tiny model in the embedded systems to create a smart IOT system. According to the evaluation results, we can replace Resblock body 2 and Resblock body 3 with convolutional layers to simplify the number of layers. In this way, the total reduced convolution operations are 5 layers and the computing performance of the YOLO v4-tiny model will be improved by up to 20%.

Currently, performance testing relies primarily on C language for estimating the prediction time of images. In the future, the aim is to implement the entire framework onto a chip and establish a complete computing architecture. Regarding the utilization of YOLO-v4 in this paper, there is a hope to adjust the architecture to design a learning model that better suits the goals of this research with enhanced performance.

6. Conclusions

This work integrates AI with mmWave sensors in the domain of nursing security. Millimeter-wave radar solely detects object movements and aids in preventing patients or occupants from falling without notice or assistance. Employing millimeter-wave radar in medical institutions or homes can offer a strong solution to privacy concerns compared to conventional cameras that capture and analyze the appearance of patients or residents.

We use the point cloud coordinates with semi-coloring methods to enhance the recognition rate. Based on the 8-bit fixed-point number representation method to quantize the weights and image data, the YOLO-v4 model can be streamlined to operate with fixed-point number types. This approach reduces the required resources for operations and accelerates the processing speed. Our quantization methods were simulated and validated using various PC CPUs and embedded systems. Prioritizing the retention of the integer part maintained a certain level of precision, successfully accelerating our computational processes. Finally, we present our design system with the data from mmWave millimeter-wave radar sensor to make the form of a point cloud and obtain the identification results through AI computing technology with the Yolo v4-tiny model. After the identification is completed, the results are transmitted to the Broker with the MQTT protocol to control light on/off. To effectively integrate the results into a long-term care application environment, we also built this system in Raspberry Pi4 and simplified the deep learning model to adapt to the embedded system.

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

visualization, C.-Y.H.; supervision, J.-C.C.; project administration, J.-C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

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

References

1. Meng, H.; Freeman, M.; Pears, N.; Bailey, C. Real-time human action recognition on an embedded, reconfigurable video processing architecture. *J. Real-Time Image Process.* **2008**, *3*, 163–176. [[CrossRef](#)]
2. Molchanov, P.; Gupta, S.; Kim, K.; Pulli, K. Multi-sensor system for driver's hand-gesture recognition. In Proceedings of the 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG), Ljubljana, Slovenia, 4–8 May 2015; Volume 1, pp. 1–8.
3. Qian, C.; Wu, Z.; Yang, Z.; Liu, Y.; Jamieson, K. Widar: Decimeterlevel passive tracking via velocity monitoring with commodity Wi-Fi. In Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing, New York, NY, USA, 10–14 July 2017.
4. Li, C.; Liu, M.; Cao, Z. WiHF: Gesture and user recognition with WiFi. *IEEE Trans. Mobile Comput.* **2022**, *21*, 757–768. [[CrossRef](#)]
5. Zhang, R.; Jiang, C.; Wu, S.; Zhou, Q.; Jing, X. Wi-Fi Sensing for Joint Gesture Recognition and Human Identification from Few Samples in Human-Computer Interaction. *IEEE J. Sel. Areas Commun.* **2022**, *40*, 2193–2205. [[CrossRef](#)]
6. Shahzadm, M.; Zhang, S. Augmenting user identification with WiFi based gesture recognition. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *2*, 1–27. [[CrossRef](#)]
7. Zhang, X.; Tao, M.; Tang, T.; Yang, J. Automatic Classification and Recognition Method based on Partially-Connected Differentiable Architecture Search for ISAC Systems. In Proceedings of the 2021 13th International Symposium on Antennas, Propagation and EM Theory (ISAPE), Zhuhai, China, 1–4 December 2021; pp. 1–3. [[CrossRef](#)]
8. Ahmed, T.H.; Tiang, J.J.; Mahmud, A.; Do, D.-T. Proposed CtCNet-HDRNN: A Cornerstone in the Integration of 5G mmWave and DSRC for High-Speed Vehicular Networks. *IEEE Access* **2023**, *11*, 126482–126506. [[CrossRef](#)]
9. De Lima, C.; Belot, D.; Berkvens, R.; Bourdoux, A.; Dardari, D.; Guillaud, M.; Isomursu, M.; Lohan, E.-S.; Miao, Y.; Barreto, A.N.; et al. Convergent Communication, Sensing and Localization in 6G Systems: An Overview of Technologies, Opportunities and Challenges. *IEEE Access* **2021**, *9*, 26902–26925. [[CrossRef](#)]
10. Andy, S.; Rahardjo, B.; Hanindhito, B. Attack scenarios and security analysis of MQTT communication protocol in IoT system. In Proceedings of the 2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Yogyakarta, Indonesia, 19–21 September 2017; pp. 19–21.
11. Zhang, Q.; Wang, X.; Li, Z.; Wei, Z. Design and Performance Evaluation of Joint Sensing and Communication Integrated System for 5G mmWave Enabled CAVs. *IEEE J. Sel. Top. Signal Process.* **2021**, *15*, 1500–1514. [[CrossRef](#)]
12. Pestana, D.; Miranda, P.R.; Lopes, J.D.; Duarte, R.P.; Vestias, M.P.; Neto, H.C.; De Sousa, J.T. A full featured configurable accelerator for object detection with YOLO. *IEEE Access* **2021**, *9*, 75864–75877. [[CrossRef](#)]
13. Darknet Deep Learn Framework. Available online: <https://github.com/AlexeyAB/darknet> (accessed on 25 March 2022).
14. Herrmann, V.; Knapheide, J.; Steinert, F.; Stabernack, B. A YOLO v3-tiny FPGA Architecture using a Reconfigurable Hardware Accelerator for Real-time Region of Interest Detection. In Proceedings of the 2022 25th Euromicro Conference on Digital System Design (DSD), Maspalomas, Spain, 31 August–2 September 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 84–92.
15. Lian, X.; Liu, Z.; Song, Z.; Dai, J.; Zhou, W.; Ji, X. High-Performance FPGA-Based CNN Accelerator with Block-Floating-Point Arithmetic. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **2019**, *27*, 1874–1885. [[CrossRef](#)]
16. Koster, U.; Webb, T.; Wang, X.; Nassar, M.; Bansal, A.K.; Constable, W.; Elibol, O.; Hall, S.; Hornof, L.; Khosrowshahi, A.; et al. Flexpoint: An adaptive numerical format for efficient training of deep neural networks. *arXiv* **2017**, arXiv:1711.02213. Available online: <http://arxiv.org/abs/1711.02213> (accessed on 2 December 2017).
17. Zhou, A.; Yao, A.; Guo, Y.; Xu, L.; Chen, Y. Incremental Network Quantization: Towards Lossless CNNs with Low-precision Weights. *arXiv* **2017**, arXiv:1702.03044.
18. Jiang, Z.; Zhao, L.; Li, S.; Jia, Y. Real-time object detection method based on improved YOLOv4-tiny. *J. Netw. Intell.* **2022**, *7*, 1–11. Available online: <https://arxiv.org/abs/2011.04244> (accessed on 16 January 2023).
19. Wang, H.; Sun, L. Design of static human posture recognition algorithm based on CNN. In Proceedings of the 2022 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), Dalian, China, 20–21 August 2022; pp. 890–894. [[CrossRef](#)]
20. Paul, H.; Ferrari, V. End-to-end training of object class detectors for mean average precision. In Proceedings of the Computer Vision—ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, 20–24 November 2016; pp. 198–213.

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.