

Article

# Autonomous Thermal Vision Robotic System for Victims Recognition in Search and Rescue Missions

Christyan Cruz Ulloa <sup>\*</sup>, Guillermo Prieto Sánchez , Antonio Barrientos  and Jaime Del Cerro 

Centre for Automation and Robotics (CAR) Universidad Politécnica de Madrid—Consejo Superior de Investigaciones Científicas, 28006 Madrid, Spain; guillermo.prieto.sanchez@alumnos.upm.es (G.P.S.); antonio.barrientos@upm.es (A.B.); j.cerro@upm.es (J.D.C.)

\* Correspondence: ccruz@etsii.upm.es

**Abstract:** Technological breakthroughs in recent years have led to a revolution in fields such as Machine Vision and Search and Rescue Robotics (SAR), thanks to the application and development of new and improved neural networks to vision models together with modern optical sensors that incorporate thermal cameras, capable of capturing data in post-disaster environments (PDE) with rustic conditions (low luminosity, suspended particles, obstructive materials). Due to the high risk posed by PDE because of the potential collapse of structures, electrical hazards, gas leakage, etc., primary intervention tasks such as victim identification are carried out by robotic teams, provided with specific sensors such as thermal, RGB cameras, and laser. The application of Convolutional Neural Networks (CNN) to computer vision is a breakthrough for detection algorithms. Conventional methods for victim identification in these environments use RGB image processing or trained dogs, but detection with RGB images is inefficient in the absence of light or presence of debris; on the other hand, developments with thermal images are limited to the field of surveillance. This paper's main contribution focuses on implementing a novel automatic method based on thermal image processing and CNN for victim identification in PDE, using a Robotic System that uses a quadruped robot for data capture and transmission to the central station. The robot's automatic data processing and control have been carried out through Robot Operating System (ROS). Several tests have been carried out in different environments to validate the proposed method, recreating PDE with varying conditions of light, from which the datasets have been generated for the training of three neural network models (Fast R-CNN, SSD, and YOLO). The method's efficiency has been tested against another method based on CNN and RGB images for the same task showing greater effectiveness in PDE main results show that the proposed method has an efficiency greater than 90%.

**Keywords:** robotic systems; thermal images; convolutional neural networks; computer vision; search and rescue robots; ROS; Unitree A1



**Citation:** Cruz Ulloa, C.; Prieto Sánchez, G.; Barrientos, A.; Del Cerro, J. Autonomous Thermal Vision Robotic System for Victims Recognition in Search and Rescue Missions. *Sensors* **2021**, *21*, 7346. <https://doi.org/10.3390/s21217346>

Academic Editors: Luis Payá and Oscar Reinoso García

Received: 1 October 2021

Accepted: 31 October 2021

Published: 4 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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

According to a report by the UNDRR (United Nations Office for Disaster Risk Reduction) [1], 7348 natural disasters have occurred in the last two decades. Most of them are caused by floods, storms, and earthquakes, accounting for 40%, 28%, and 8%, respectively [2]. In disasters such as major storms, fatalities occur in ~10% of those affected, while in earthquakes, the number of fatalities is ~49%.

In these scenarios, robotics helps the speed and efficiency of their management [3–5]. Search and rescue robots specialized in victim recognition and search are designed to assist in locating victims who are out of sight or in inaccessible locations [6,7]. In addition, the use of these robots also prevents operators from exposing themselves to hazards such as landslides [8–10].

In recent years, with the rise of Machine Vision due to new techniques which employ neural networks for tasks such as object classification and detection in images and videos,

their application in multiple areas has been explored and investigated. One of these areas is thermography [11,12], some applications such as surveillance and face recognition [13–18] have been developed.

There is also a branch of research on people detection with very low-resolution thermal cameras or even from infrared sensors [19–22]. Note that some works are currently emerging, such as the one by Perdana et al. [20], in which the first steps are taken in the victim detection using neural networks and thermal imagery.

The main developments for people detection in Search and Rescue scenarios using robots, focus on using RGB cameras (additive color mode Red, Green, and Blue) [23–25]. The methods based on thermal images are very limited to classical computer vision techniques applications and primitive neural networks. In addition, data acquisition is performed in open areas or using drones for data capturing in daytime conditions [26–30].

The TASAR (Team of Advanced Search And Rescue Robots) project focuses on using terrestrial Search and Rescue Robots for Humanitarian Assistance and Disaster Relief (HADR missions) [31]. A robotic system was implemented to validate this proof of concept. It uses the Robot Unitree A1 robot equipped with a sensory system (Thermal Camera Optris PI640, Real-Sense, RPLidar) for data capture and transmission. This quadruped robot was used due to its remarkable adaptability to move in unstructured environments [32].

The main contribution of this work focuses on developing an integrated system, which uses the Unitree Robot (teleoperated) inside a PDE, capable of transmitting data (thermal images, RGB images) for its processing in real-time; in an early detection phase it checks for the presence of victims within the environment. The location of victims is executed by processing the thermal images through a Convolutional Neural Network—YOLO (You Only Look Once)—to later issue an alert and generate a trace with the location of the victim. After a comparative experimental phase between the Faster R-CNN, SSD, and YOLO networks, the YOLO network model has been defined as the most efficient model.

This method was contrasted with conventional RGB image methods, using another trained neural network (with an RGB dataset captured simultaneously with the thermal image dataset) capable identify victims in PDE. However, in scenarios with bad lighting conditions or obstructive materials, its efficiency is low concerning the proposed thermal method.

The tests were performed to detect victims in environments with poor light conditions (both day–night and indoors–outdoors) and use different materials to cover the victims. The main results show an efficiency of the system upper than 90% in the early detection of victims in PDE.

The algorithms execution and information flow between the field robot, sensors, and the central control computer has been developed using ROS.

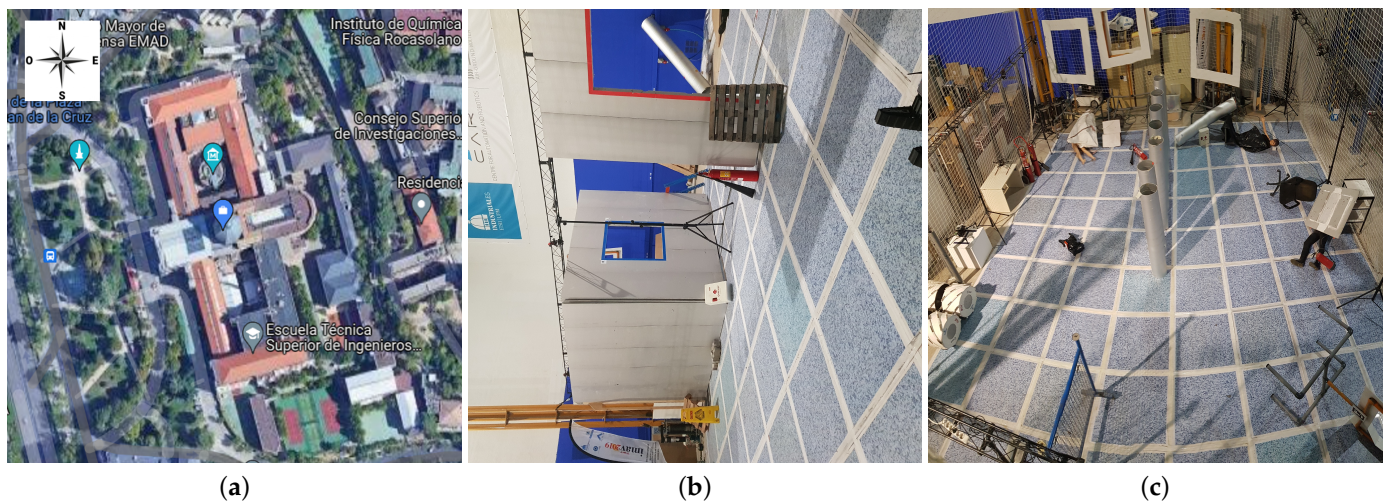
This paper is structured as follows. In Section 2, the materials and methods used in this work are introduced in detail, followed by the Results and Discussion in Section 3. To conclude, Section 4 summarizes the main findings.

## 2. Materials and Methods

### 2.1. Materials

The experiments for the development of this research were carried out at the facilities of the Centro de Automática y Robótica located in (40°26′23.4″ N–3°41′21.7″ W) as shown in Figure 1, both in outdoor (Figure 1a) and indoor (Figure 1b,c) scenarios. The latter has been recreated based on the NIST (National Institute of Standards and Technology) standardized environments for disaster environments; specifically in the yellow zone (debris and moderate obstacles) [33].





**Figure 1.** Indoor and outdoor scenarios are used for test development. (a) ETSII-UPM Outdoor Testing Environment. (b) Scenarios recreated for indoor testing. (c) Scenarios recreated for indoor testing—top view. Source: Authors.

Table 1 shows the equipment used for this research development.

**Table 1.** Evaluation of the neural network parameters.

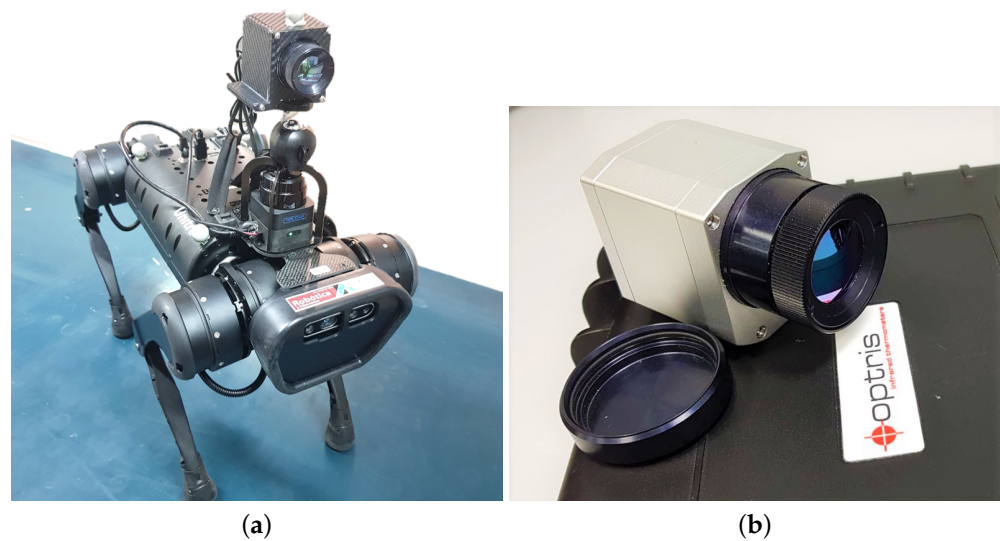
Component	Amount	Description
Unitree A1	1	Quadruped Robot
Real-Sense	1	RGB-Depth Sensor
Optris Pi640	1	Thermal Camera
Nvidia Jetson Xavier-NX	1	Embedded On-board System
MSI Laptop	1	External Core System

The robotic system used for the autonomous detection of victims by means of thermal images is mainly integrated by the Unitree A1 robot (Figure 2a) which has integrated into its front part a Real-Sense and the thermal camera (Figure 2b) coupled by means of mechanical support made of a 3D-printed coupling capable of absorbing vibrations emitted due to the robot's movement due to a ball joints and springs system.

For the processing of the neural network, robot control, and data flow, an MSI computer with an Intel i10 processor and NVIDIA GEFORCE GTX 1660Ti GPU has been used, which supplies the process for its execution in real-time. The computer has been connected by a 5G Wireless Network with the Robot in the field through the ROS Master-Slave (PC MSI-Unitree Robot) Communication System. Real-time requirements were managed by the *controller\_manager* packets, obtaining a latency of 10 ms. The interface used for information management was RVIZ (ROS Visualization).

The infrared camera model Optris PI 640 of the company OPTRIS was used to obtain thermal images. This camera operates in the spectral range of 8 to 14  $\mu\text{m}$ , being able to detect temperatures within the range of  $-20$  to  $900$   $^{\circ}\text{C}$ ; it has a weight of 320 g. It has a thermal sensitivity of 75 mK and an accuracy of  $\pm 2^{\circ}\text{C}$  or  $\pm 2\%$ , whichever is greater.

The robot used (Unitree A1 of the company UNITREE) has been selected due to its great versatility, agility, and ability to move in environments with unstructured floors (slopes, slopes, debris, etc.), thanks to its system of locomotion by legs, sensors, and real-time processing. The main features of this robot are a maximum speed of 3.3 [m/s], autonomy of one to two hours, great stability, 12 degrees of freedom (DOF), integrated sensors (RPLidar, Real-Sense, IMU, and pressure sensors on the legs), and 3 on-board computers.



**Figure 2.** Robot and instrumentation used for the proposed method validation. (a) Unitree A1 Robot equipped with thermal camera and Real-Sense. (b) Optris Pi640 Thermal Camera.

## 2.2. Interaction between Subsystems

The communications architecture of the implemented method (Figure 3) with the robot in the field, the remote station, and the data processing has been developed as part of the TASAR project entirely using ROS.

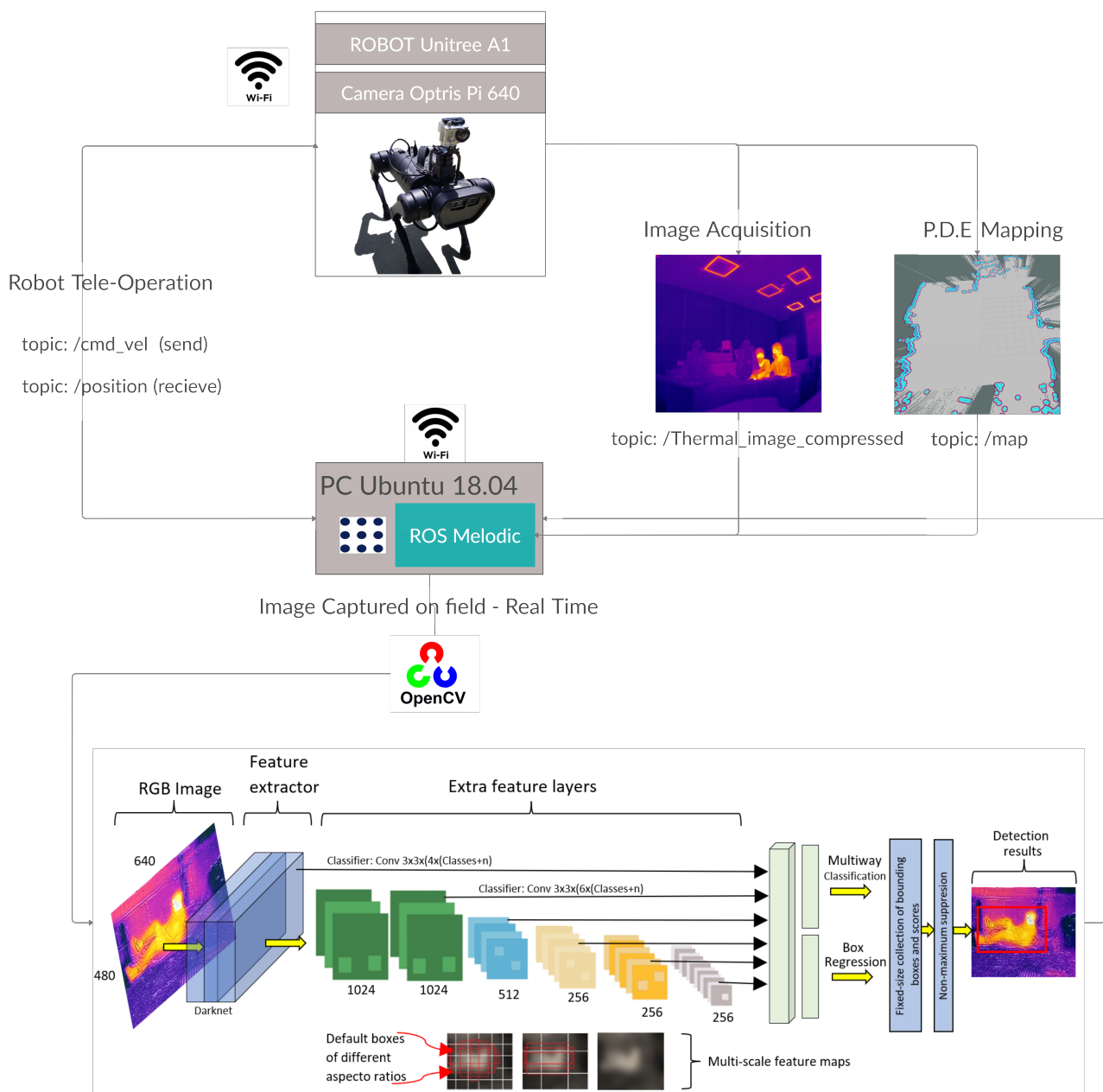
The overall system consists of two subsystems, shown in Figure 3: the first one is a high-power central computer that sends the velocity commands to the robot in the field to execute the displacements along the PDE. On the other hand, it receives the position data, thermal, and rgb images from the robot to process them through a convolutional neural network in real time.

The second subsystem is the robot in the field, which thanks to all the instrumentation is in charge of collecting images to send them and, by receiving speed commands, to move through the PDE. The robot in this experimental phase of the system is tele-operated from a remote station by an operator or a previously trained rescuer, sending speed peaks to the robot (`/cmd_vel`).

Thermal image reception and processing are performed in real-time at the remote station, so that the operator can know beforehand if there is a victim (even if trapped in debris) in need of primary assistance and its location within the environment based on the position of the robot.

For the capture and subsequent transmission of thermal images from the camera integrated in the robot, the drivers for ROS (in the Jetson NX), developed by the manufacturer Optris, have been installed. The method based on fuzzy filters has been used to eliminate noise in thermal images as it is the most recommended for thermal images instead of the conventional median filter methods [34].

For real-time processing, it has been previously configured in the Anaconda virtual environment with the ROS and OpenCV packages, so that, with the network already trained, the inference can be performed. The neural network architecture used is also detailed (this neural network methodology has been applied in other developments such as precision agriculture for detecting fruits in highly noise-contaminated environments [30]). This process can be detailed in the lower part of the Figure 3.



**Figure 3.** Subsystems integration for the detection of victims in PDE. Source: Authors.

### 2.3. Materials and Termography

This section is focused on highlighting the importance of materials and their interaction and influence on thermal images. It seeks to take advantage of the emissivity of some materials such as plastic or others in order to identify victims that are fully covered.

The fraction of the incident energy that must be captured by the thermal camera to measure the temperature of an object is the emitted energy. Emissivity plays a fundamental role in estimating this energy based on the energy received. Different materials have different emissivity values, leading to more or less accurate temperature measurements. This dimensionless value is quantized from [0–1].

Organic materials generally have a high emissivity, so their measurement is usually straightforward. Materials such as paper, ceramics, wood, soil, plants, sand, rubber, stone, paints, dark or matte coatings, etc. have an emissivity of  $\sim 0.95$  in the spectral range of 8 to 14  $\mu\text{m}$ . For urban, building, and industrial environments, there are differences in emissivity between different common materials. For building materials such as concrete

(0.93), normal brick (0.92–0.94), glazed brick (0.94–0.89), glass (0.95–0.98), and asphalt (0.98), the emissivity is very high and thus the temperature is easily measurable. Table 2 shows the emissivity measurements for different materials.

**Table 2.** Elements and components of the mobile platform [35,36].

Material	Temperature (°C)	$\epsilon$
Aluminum, glossy laminated	170	0.04
Asphalt	20	0.93
Concrete	25	0.93
Lead, rusted	20	0.28
Ice	0	0.97
Iron, frosted	20	0.24
Iron, shiny	150	0.13
Iron, rusted	20	0.85
Soil	20	0.66
Glass	90	0.94
Silver	20	0.02
Wood	70	0.94
Plastic (PE, PP, PVC)	20	0.94

For plastics, the thickness of the object is a significant factor. Thick plastics usually have a high emissivity of approximately 0.86–0.95. The problem arises with thin plastic films. These have a very high transmissivity, so that when measuring the temperature with a temperature camera, what is shown is not the temperature of the film itself, but that of the objects placed behind it, as shown in the Figure 4.



**Figure 4.** Example of thin-film transmissivity. Source: Authors.

#### 2.4. Neural Networks and Environments

One of the main objectives of this work is to compare the performance and accuracy in detecting victims of a number of architectures implementing neural networks. Three networks have been chosen: Faster R-CNN [37], SSD [38], and YOLOv3 [39]. This YOLO version has been used to optimize computational efficiency and achieve real-time development (under the conditions the experiments have developed), avoiding latency in processing. The hyperparameters used for this proposal are Initial Learning rate: 0.001, Learning Rate Schedule: (burn\_in = 1000, steps = 400,000, scales = 0.1), Batch size: 16, and Training Epochs: 100.

Each neural network has a number of advantages and disadvantages that distinguish one over the other in different areas, so it is the subject of study to find out which one provides the best characteristics for victim detection.

For the development of this research, multiple virtual environments were set up in Python using Anaconda [40]. Multiple packages such as Tensorflow [41], Pytorch [42], and Open CV [43] were configured and installed in these environments, depending on the network.



For the Faster R-CNN and SSD networks, the Tensorflow library was used using the object detection API developed by Google [44]. For the YOLOv3 network, the pytorch library was mainly used.

### 2.5. Data Collection and Datasets

To train the neural networks, multiple videos were recorded to obtain their corresponding frames in the form of images. Appendix A contains a sample of the generated datasets.

The different Datasets generated for the training have been generated, including images that contain victims under different circumstances, such as partial coverage due to rubble and different materials such as wood or concrete.

It is of utmost importance to obtain robust models by making recordings that show a wide variety of situations and facets.

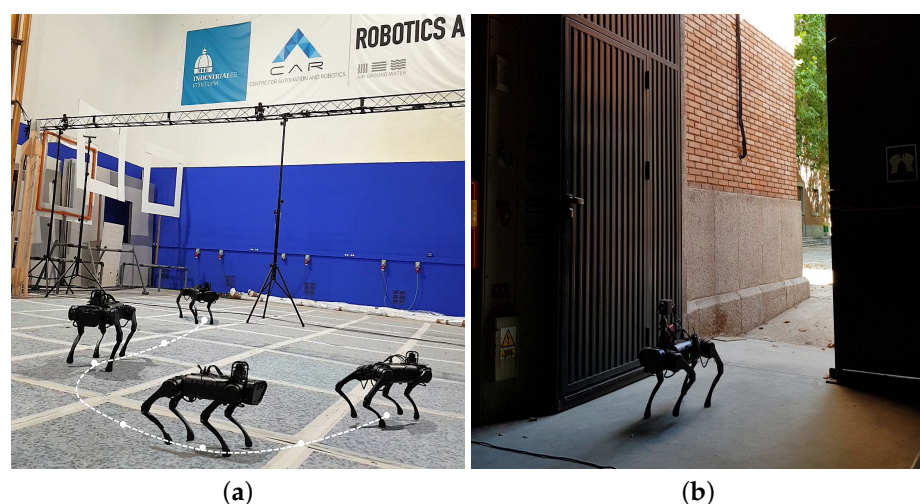
This is why videos have been recorded in which simulated victims appeared in different positions, at different times of the day, and in which the victim's body was not shown in its entirety, but only parts of the body. It is of utmost importance to obtain robust models by making recordings that show a wide variety of situations and facets.

Furthermore, the victim's entire body was not shown, but rather specific body parts such as heads, legs, arms, etc. In addition, materials typical of different environments, such as building materials, were used. Different datasets for the training were generated, including images that contain victims under different circumstances, such as partial coverage due to rubble and different materials such as wood or concrete.

Different environments, such as building site, or urban environment materials, such as metal and wood plates or plastic and fabric films, were used to both simulate the environment and to partially or totally cover the bodies.

### 2.6. Tests and Experiments

As previously mentioned, each neural network has a series of advantages and disadvantages concerning others, such as the precision and speed. A series of tests have been devised in order to corroborate some theoretical aspects and to find the network that works best for victim detection. Figure 5 shows the Unitree Robot in different scenarios during the execution of the field tests. Appendix B shows the videos of the robot's movement in the scenarios.

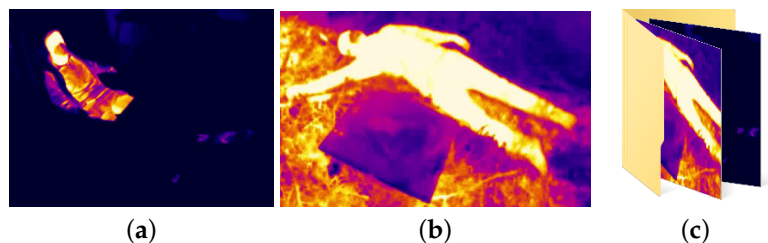


**Figure 5.** Robot Unitree in different scenarios. (a) Robot in indoors (good light conditions). (b) Robot in outdoors (bad light conditions). Source: Authors.

The first test aims to study the influence of temperature contrast on victim detection. This contrast is given between the difference in temperatures captured by the thermal camera of the person or victim to be detected and the environment. At ambient temper-



atures similar to the average temperature of people outdoors, approximately 20–25 °C, the ambient-person contrast is very low. Therefore, it is intended to test the influence of the thermal contrast on the image by performing tests using three different datasets: one which uses images recorded at night, one which uses images recorded at day, and another which use both of them (both for indoors and outdoors), as shown in Figure 6.

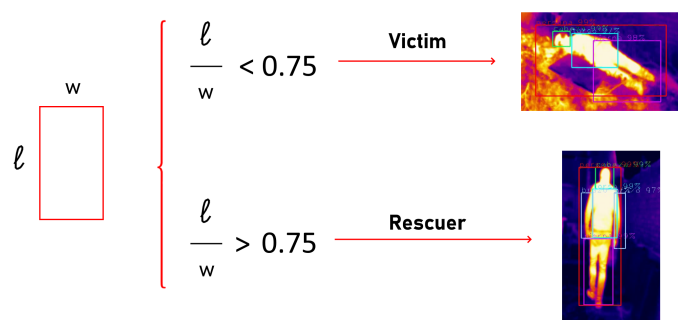


**Figure 6.** Different datasets used in training. (a) Night dataset. (b) Day dataset. (c) Combined dataset. Source: Authors

The second test is related to body parts detection. Due to the fact that in accidents or disasters victims may be trapped or buried in debris, the camera may be able to see a single limb or extremity, but not the silhouette of the entire body. Therefore, the networks will be trained so that, apart from detecting people, they can detect different parts of the body. For this reason, in each image of the different datasets, in addition to the ‘person’ label, additional labels will be used: ‘head’, ‘arm’, ‘leg’, and ‘torso’.

The third test consists of distinguishing between rescuers and victims. The most logical method to distinguish rescuers from victims would be to label in the datasets rescuers as ‘rescuer’ and victims as ‘victims’. However, due to the similarity they would have to each other, there would be a very high classification error. A simple method that can be effective is to analyze the dimensions of the bounding box. Accident victims will normally be lying on the ground or in similar positions. This fact can be taken advantage of by measuring the width and length of the of the bounding boxes of the images.

Several images showing both rescuers and victims have been analyzed and it has been concluded that the victims have a length-to-width ratio of less than 0.75, as shown in Figure 7.



**Figure 7.** Length-to-width ratio to detect victims. Source: Authors.

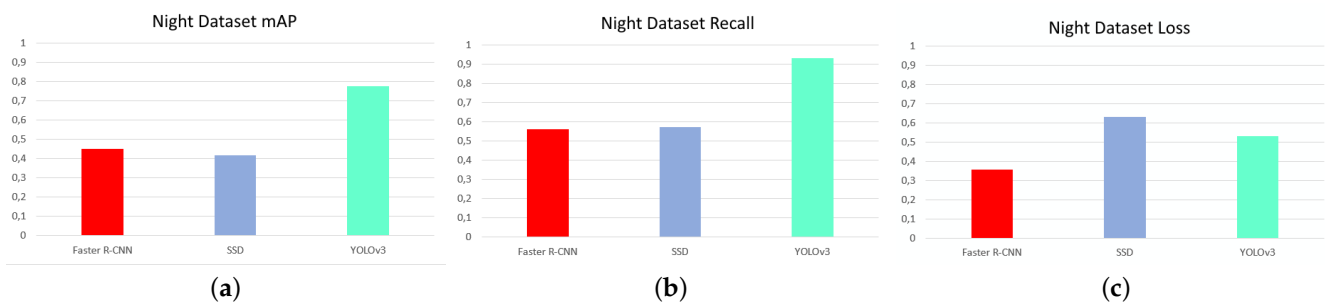
The proposed method robustness has been contrasted with conventional methods of using RGB images; based on this, different indicators have been obtained, such as detection in night environments, daytime environments, debris, indoor and outdoor environments, and radiation emitted by specific areas. Therefore, the results show that the proposed method relevant to conventional methods.

### 3. Results and Discussion

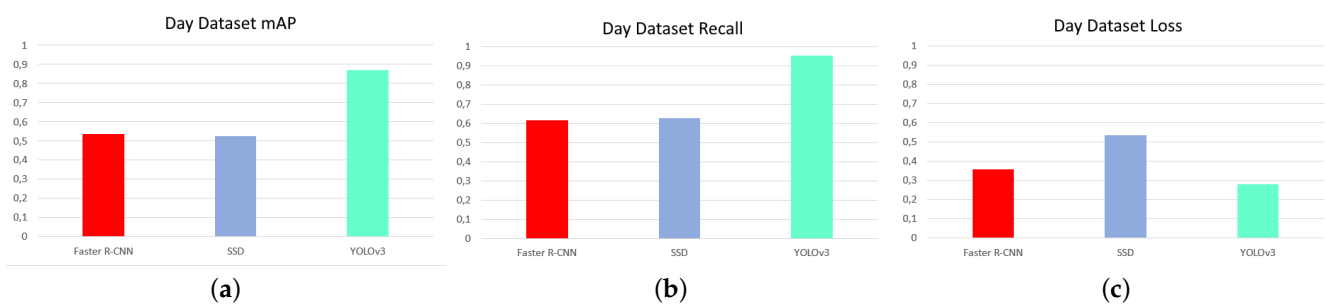
#### 3.1. Analysis of Neural Networks Implemented

##### 3.1.1. Comparison of Faster R-CNN, SSD, and YOLO

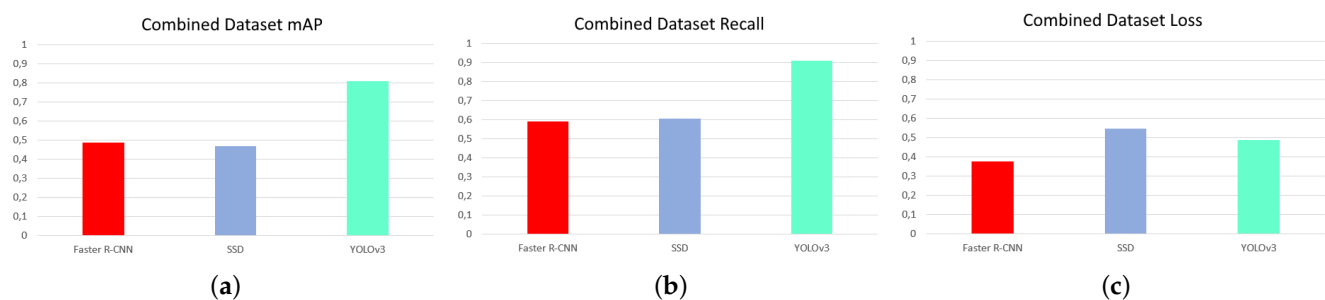
The first test was to measure the influence of temperature contrast on victim detection, with the aim of finding the ideal network and dataset combination. Figures 8–10 are shown below, showing the values of accuracy (mAP), recall, and loss for the networks for each dataset.



**Figure 8.** mAP, recall, and loss for networks with night dataset. (a) mAP, (b) recall, and (c) loss. Source: Authors.



**Figure 9.** mAP, recall, and loss for networks with day dataset. (a) mAP, (b) recall, and (c) loss. Source: Authors.



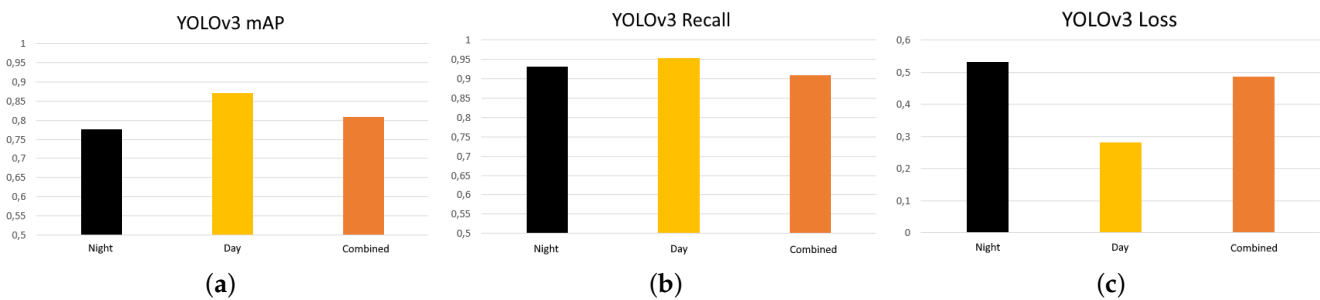
**Figure 10.** mAP, recall, and loss for networks with combined dataset. (a) mAP, (b) recall, and (c) loss. Source: Authors.

Analyzing these diagrams it can be seen that the YOLOv3 network is far superior in MAP and recall for all datasets. It also has the lowest loss values of all, except in the case of the night dataset, where the lowest loss is that of the Faster R-CNN network. It follows that, based on evaluation results, the YOLOv3 network is undoubtedly the best of the three.

Another fact to take into account when choosing one network over another is the speed of inference. There is a big difference in the frames per second in inference for each network. The Faster R-CNN network, despite having good precision and recall values and little loss, is extremely slow in inference, working at very few frames per second. The opposite happens for the SSD network, where it does work at a higher fps rate but has a higher loss. YOLOv3 has good values in both the detection parameters and inference

speed. There are previous studies that compared the inference execution speed of the networks, such as that in [45], which used a GPU with 12 GB of RAM to compare the performance and speed in inference of multiple networks. The results obtained were an average fps rate for Faster R-CNN of 7, for SSD of 19, and for YOLOv3 of 45.

Figure 11 shows the analysis of the YOLOv3 results comparing the datasets.



**Figure 11.** mAP, recall, and loss comparison for the YOLOv3 datasets. (a) mAP, (b) recall, and (c) loss. Source: Authors.

### 3.1.2. YOLO Performance with Different Datasets

Figure 11 shows that the highest precision and recall values and the lowest loss values were obtained for the day, combined and night datasets. This is totally contrary to the expected results, which expected that, as there was higher contrast and less influence of other environmental conditions such as solar radiation, the night dataset would be the ideal one. The networks trained with the combined dataset offer values of precision, recall, and loss not much below those of the daytime dataset, with the advantage that they are more versatile, functioning correctly both day and night.

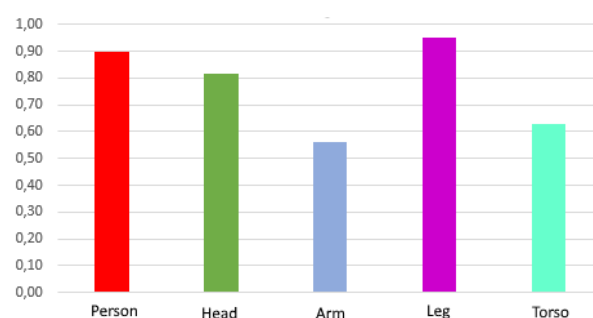
One fact that stands out from the Figures 8–10 diagrams is the high loss value that the SSD network has compared to the rest of the networks for all datasets. This occurs for multiple reasons, but the most likely one is related to the results of the second test, which is the detection of different victim body parts.

### 3.1.3. Class Average Precision for Implemented Neural Network

For this test, the datasets were labeled with five classes: ‘person’, ‘head’, ‘arm’, ‘leg’, and ‘torso’. Due to the similarity of some of these classes to each other, as can be the case for the arms and legs in the thermal images, there is a large increase in classification losses. This same problem is presented to a lesser extent by the Faster R-CNN network but not by the YOLOv3 network.

The inference results are very satisfactory, showing high accuracy and recall and low loss in predictions, making YOLOv3 a viable network for victim detection.

In Figure 12, the average accuracy of each class for the three datasets for YOLOv3 is shown in order to find out which body parts are easier to detect.



**Figure 12.** Average class precision for YOLOv3. Source: Authors.

As can be seen in Figure 12, all classes have good average precision. The classes ‘person’, ‘head’, and ‘leg’ have a very high accuracy, reaching 0.95 for the case of ‘leg’, while the classes ‘arm’ and ‘torso’ have a somewhat lower accuracy, but within acceptable margins.

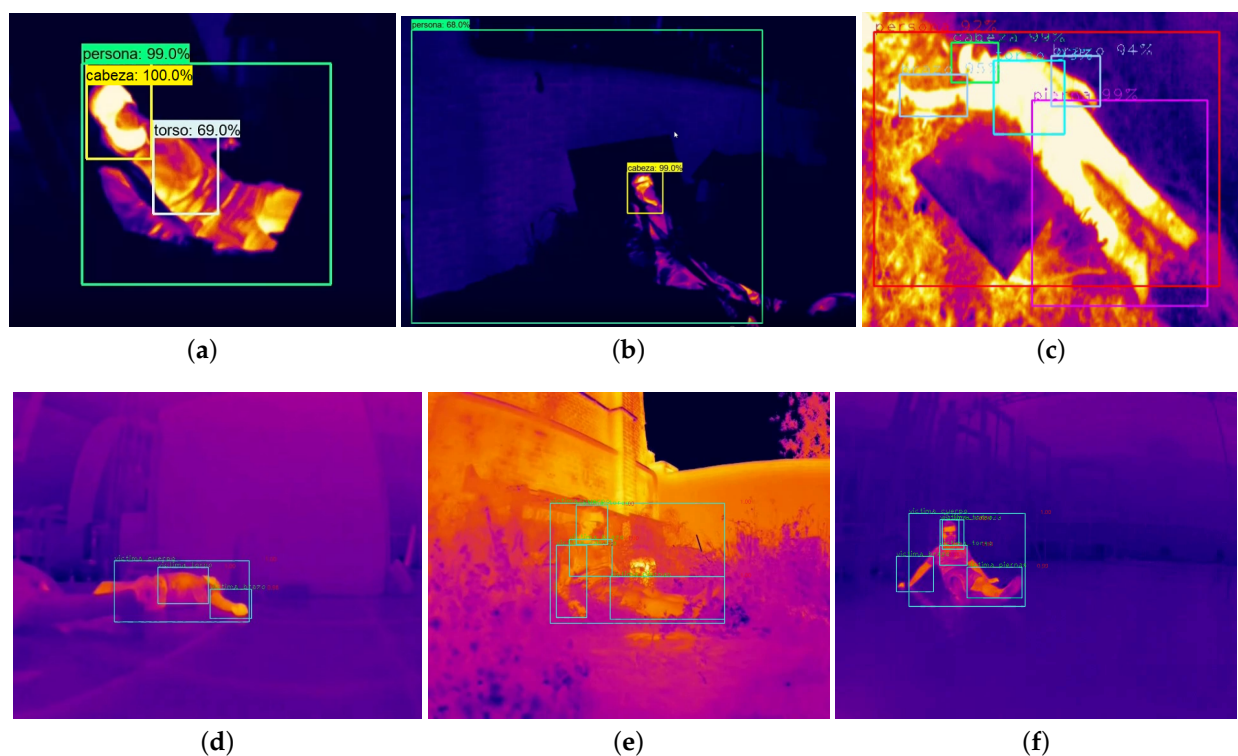
The test aimed at evaluating the distinction between normal people and victims was carried out by evaluating different videos featuring both bystanders and alleged victims on the YOLOv3 network. The effectiveness of the method of analyzing the length-to-width ratio of the detected person to distinguish the two cases was evaluated, and the results obtained were more than satisfactory. Each time a possible victim was detected in the image, a message was displayed on the screen alerting of the situation and the coordinates of the image where the victim was located.

### 3.2. Efficiency of Victim Detection Using the Proposed Method in PDE

This test sought to evaluate the performance of the YOLOv3 model trained thermal images captured by the Unitree Robot. The result was good, although it presented a slightly higher classification error between leg and arm classes than previous tests. Despite this, the model is perfectly valid, detecting everything correctly on most occasions.

In this way, the effectiveness of YOLOv3 has also been tested in indoor environments. The main results of the identification of victims are shown in Figure 13, in different environments and lighting conditions, with SSD, R-CNN, and YOLOv3.

Appendix B shows the videos of the real-time execution for detecting victims in the scenarios.



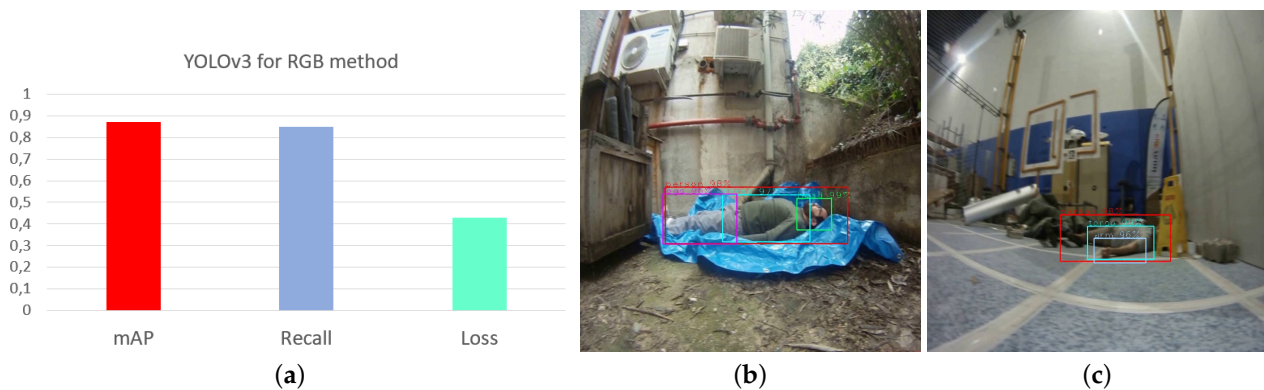
**Figure 13.** Examples of victim detection with Faster R-CNN, SSD, and YOLOv3. The efficiency in the detection of victims respectively is a = 99%, b = 60%, c = 98%, d = 97%, e = 96%, and f = 99%. (a) Faster R-CNN, (b) SSD, (c) YOLOv3 (Day outdoor), (d) YOLOv3 (Day indoor), (e) YOLOv3 (Day outdoor), and (f) YOLOv3 (Night indoor). Source: Authors.

### 3.3. Comparison of Proposed and Traditional Methods for Identifying Victims

For the development of this comparison, a YOLOv3 neural network has been trained for the detection of victims using RGB images, and the results of the training and the validations are shown in the Figure 14. The same labels used in the thermal method have

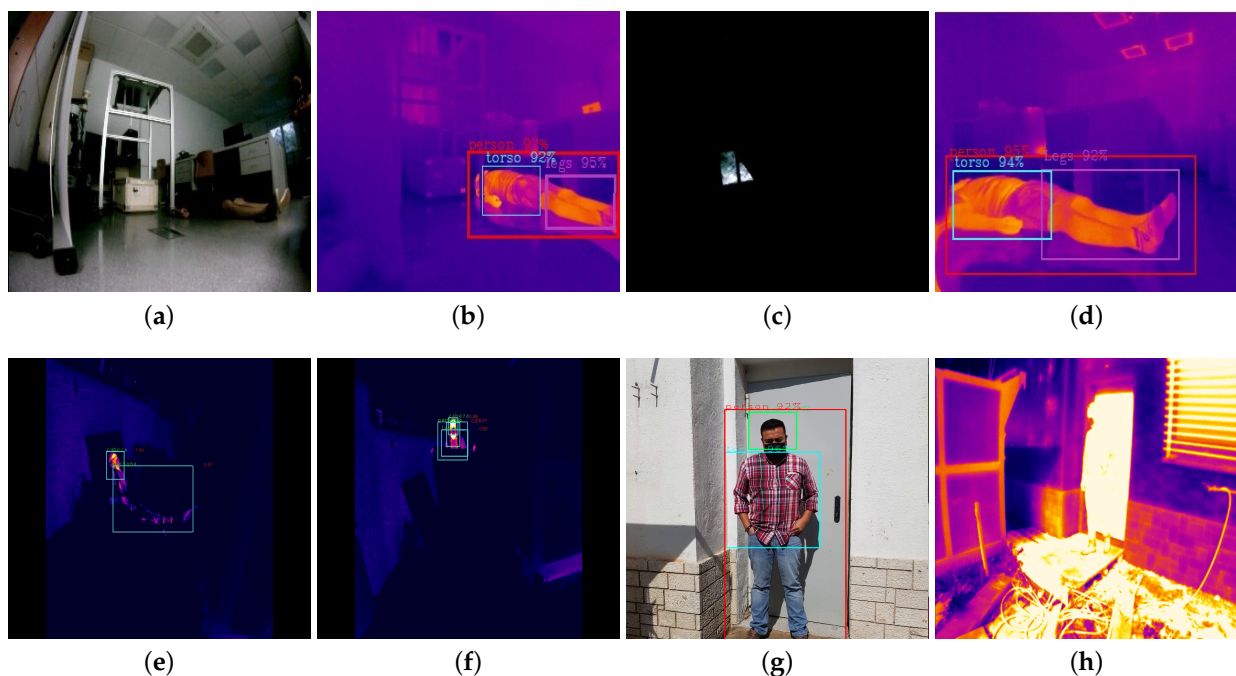
been used. Results in environments with good light conditions have a high percentage of detection efficiency.

The RGB images used for this method have been captured by the real sense that the robot has in its front part. The procedure for real-time processing is similar to that used in the proposed method.



**Figure 14.** Evaluation of the conventional method that uses RGB images for the detection of victims with good lighting conditions, using CNN-YOLOv3. (a) Neural Network Training. (b) Outdoor evaluation. (c) Indoor evaluation. Source: Authors.

Figure 15 shows different situations analyzed in real environments where both methods have been tested. Figure 15a,b corresponds to the RGB and Thermal methods, respectively, as well as Figure 15c,d. In both cases, the thermal images correspond to the RGB images, however the RGB detection is poor due to the very poor lighting conditions.



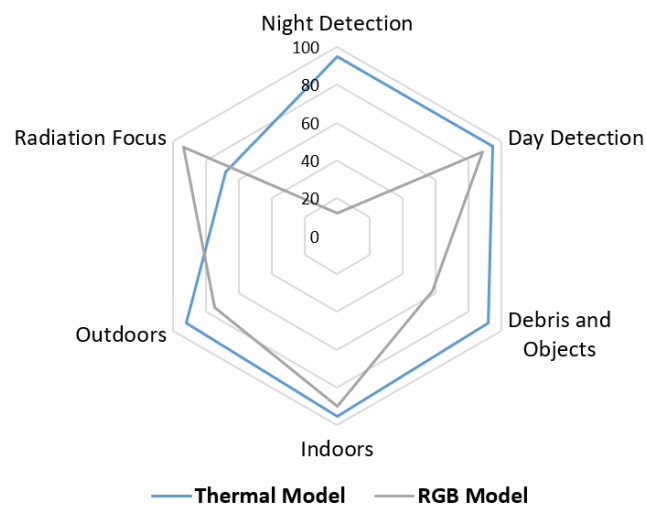
**Figure 15.** Evaluation of the conventional methods in front of proposed method for victims detection with different lighting conditions, using CNN-YOLOv3. (a) Case 1: Bad detection of RGB method in low light. (b) Case 1: Good detection of Thermal method in low light. (c) Case 2: Bad detection of RGB method in absence of light. (d) Case 2: Good detection of Thermal method in absence of light. (e) Case 3: Good detection of victims (fully covered) for the thermal method. (f) Case 4: Good detection of victims (partially covered) for the thermal method. (g) Case 5: Good detection of people in front of heat sources for RGB method. (h) Case 5: Bad detection of people in front of heat sources for Thermal method. Source: Authors.



Figure 15e,f corresponds to thermal images of totally and partially covered victims, the thermal method for this case due to the emissivity of the covering material (plastic), is the best option.

Finally, Figure 15g,h corresponds to a person in front of a door that has accumulated heat during a summer day (heat source). In this case, the thermal method is more effective.

Figure 16 shows the radial graph of the percentage parameters obtained from experimental for both methods, where it can be concluded that for most environments the proposed method is more efficient than the conventional ones. On the other hand, one of the shortcomings of this method falls on the presence of large heat sources. Therefore, a solution would be to combine both methods to obtain a more robust system in the face of these disturbances.

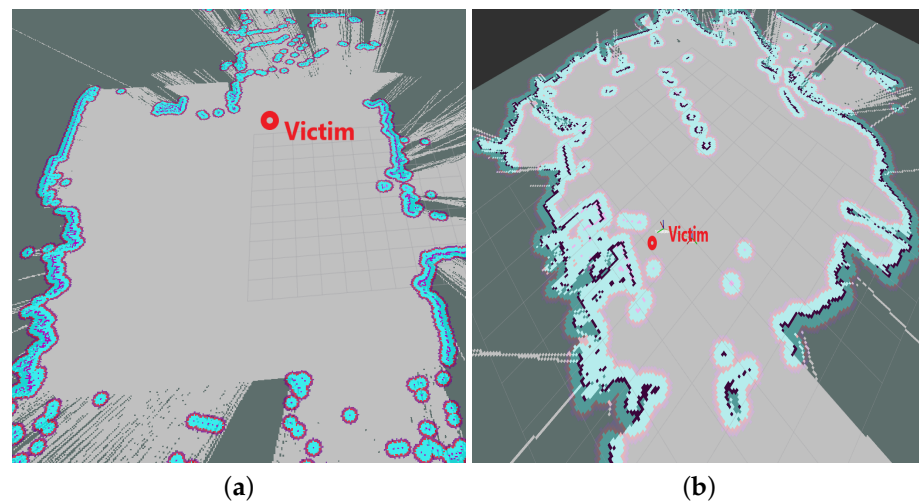


**Figure 16.** Percentage comparison of efficiency of the analyzed methods (Thermal and RGB). Source: Authors.

For applications such as detecting people in specific places, such as public spaces like streets or PDE. Apart from the disadvantages above, visible spectrum imaging has problems related to the resolution of these cameras required for such a task. Conventional RGB cameras require a much higher image resolution than thermal cameras for the same accuracy and performance.

Using thermal images with a much lower resolution than ordinary cameras, processing and storage in memory require much less computational power. In some situations, for cameras with very low resolution, the computational cost is so low that it can be carried out on embedded systems with very limited resources. In these cases, the energy consumption is also lower.

Figure 17 shows the robot exploration result within the explored environment, the victims have been located within the map generated during the exploration of the robot.



**Figure 17.** Victims location in the mapped environment. (a) Victim detection in reconstructed Scenario 1. (b) Victim detection in reconstructed Scenario 2. Source: Authors.

#### 4. Conclusions

This article shows the proof-of-concept of an integrated Robotic System for victim detection in post-disaster environments using a quadruped robot as a medium to capture and transmit thermal images in real-time. Thermal images were processed through neural networks to analyze the existence of victims in an analyzed environment area. To this end, several subsystems were developed and integrated into ROS. The method was executed and validated on a recreated PDE according to NIST regulation.

The use of the Unitree A1 quadruped robot has allowed the exploration area of the PDE to be overcrowded, thanks to its locomotion system by legs that allow adaptability to uneven floors or with rubble, great agility, and speed for movement. The technology that this robot has, combined with ROS, allowed the development of the application in real-time.

The great advances in computer vision that have taken place in the last few years, due to the incorporation of neural networks into vision models, have meant a great leap forward in the field of computer vision. There are many architectures for neural networks, such as Faster R-CNN, SSD, or YOLO, which implement a series of algorithms capable of performing object detection, each of them having a series of advantages and disadvantages. These three models have been tested in this work to analyze their effectiveness in detecting victims in PDE.

The SSD network has numerous drawbacks for the victim detection task, which cause it to have a high loss value, making its correct application impossible. The Faster R-CNN network greatly improves the results obtained concerning SSD, but its slow inference speed makes it practically impossible to run in real-time.

The YOLOv3 network has a much higher MAP and recall than the other two networks, with values of 85% and 95%, respectively, and a slightly lower loss of approximately 35%. In addition, YOLOv3 has the highest inference speed compared to the other networks. Therefore, it has been concluded that this network is the most suitable one for detecting victims.

The method implemented for the distinction between victim and pedestrians, based on measures the length-to-width ratio of the ‘rescuer’ class to distinguish rescuer from victims, is a viable method that provides satisfactory results.

The effectiveness of the proposed victim detection method has been validated through real-life tests in both indoor and outdoor environments, and for day and night environments it has also shown great efficiency even for materials that cover entirely victims, such as plastics obtaining an efficiency superior to 90% in the detection of victims determined experimentally.

A combination of the RGB method and the proposed method is established to obtain an even more robust system within the lines of future work.

**Author Contributions:** Conceptualization, A.B., C.C.U., G.P.S. and J.D.C.; methodology, A.B. and C.C.U.; software, C.C.U. and G.P.S.; validation, C.C.U. and G.P.S.; formal analysis, C.C.U. and A.B.; investigation, C.C.U., A.B., G.P.S. and J.D.C.; resources, A.B. and J.D.C.; data curation, C.C.U. and G.P.S.; writing—original draft preparation, C.C.U., G.P.S. and J.D.C.; writing—review and editing, C.C.U., A.B. and G.P.S.; visualization, A.B. and J.D.C.; supervision, A.B. and J.D.C.; project administration, A.B. and J.D.C.; funding acquisition, A.B. and J.D.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research has been possible thanks to the financing of RoboCity2030-DIH-CM, Madrid Robotics Digital Innovation Hub, S2018/NMT-4331, funded by “Programas de Actividades I+D en la Comunidad Madrid” and cofunded by Structural Funds of the EU and TASAR (Team of Advanced Search And Rescue Robots), funded by “Proyectos de I+D+i del Ministerio de Ciencia, Innovación y Universidades” (PID2019-105808RB-I00). This research was developed in Centro de Automática y Robótica—Universidad Politécnica de Madrid—Consejo Superior de Investigaciones Científicas (CAR UPM-CSIC).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

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

## Abbreviations

The following abbreviations are used in this manuscript:

ROS	Robot Operating System
RVIZ	ROS Visualization
CNN	Convolutional Neural Networks
PDE	Post-Disaster Environments
TASAR	Team of Advanced Search And Rescue Robots
RGB	Red Green and Blue

## Appendix A

Datasets for Neural Network Training <https://mega.nz/folder/T3BBHSqa#QJkfC2Yp5zPx0OCDP2XqA>, accessed on 30 October 2021.

## Appendix B

Videos for Unitree <https://mega.nz/folder/GjB3HIKK#qVg7jyOlc55xJU0hqhKXdw>, accessed on 30 October 2021.

## References

1. UNDRR Home. Available online: <https://www.undrr.org/> (accessed on 23 June 2021).
2. Noticias ONU: Pese al Aumento de las Amenazas de Origen Natural en el Siglo XXI, los Países Siguen “Sembrando las Semillas de su Destrucción”. 2020. Available online: <https://news.un.org/es/story/2020/10/1482242> (accessed on 23 June 2021).
3. Drew, D.S. Multi-Agent Systems for Search and Rescue Applications. *Curr. Robot. Rep.* **2021**, *2*, 189–200. [CrossRef]
4. Delmerico, J.; Mintchev, S.; Giusti, A.; Gromov, B.; Melo, K.; Horvat, T.; Cadena, C.; Hutter, M.; Ijspeert, A.; Floreano, D.; et al. The current state and future outlook of rescue robotics. *J. Field Robot.* **2019**, *36*, 1171–1191. [CrossRef]
5. Queralta, J.P.; Taipalmaa, J.; Pullinen, B.C.; Sarker, V.K.; Gia, T.N.; Tenhunen, H.; Gabbouj, M.; Raitoharju, J.; Westerlund, T. Collaborative multi-robot systems for search and rescue: Coordination and perception. *arXiv* **2020**, arXiv:2008.12610.
6. Pozniak, H. Robots... assemble! Although humans and highly trained dogs will always be critical in search-and-rescue operations, robots are being developed—Taking inspiration from the natural world—That are helping rescue teams save lives. *Eng. Technol.* **2020**, *15*, 67–69. [CrossRef]
7. Shah, B.; Choset, H. Survey on urban search and rescue robots. *J. Robot. Soc. Jpn.* **2004**, *22*, 582–586. [CrossRef]
8. Davids, A. Urban search and rescue robots: From tragedy to technology. *IEEE Intell. Syst.* **2002**, *17*, 81–83.
9. Paoletti, M.; Haut, J.; Plaza, J.; Plaza, A. A Comparative Study of Techniques for Hyperspectral Image Classification. *Rev. Iberoam. Autom. Inform. Ind.* **2019**, *16*, 129–137.

10. Chadwick, R.A. The impacts of multiple robots and display views: An urban search and rescue simulation. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Orlando, FL, USA, 26–30 September 2005; pp. 387–391.
11. John, V.; Mita, S.; Liu, Z.; Qi, B. Pedestrian detection in thermal images using adaptive fuzzy C-means clustering and convolutional neural networks. In Proceedings of the 2015 14th IAPR International Conference on Machine Vision Applications (MVA), Tokyo, Japan, 18–22 May 2015; pp. 246–249. [[CrossRef](#)]
12. Alsafasfeh, M.; Abdel-Qader, I.; Bazuin, B.; Alsafasfeh, Q.; Su, W. Unsupervised fault detection and analysis for large photovoltaic systems using drones and machine vision. *Energies* **2018**, *11*, 2252. [[CrossRef](#)]
13. Tammana, A.; Amogh, M.; Gagan, B.; Anuradha, M.; Vanamala, H. Thermal Image Processing and Analysis for Surveillance UAVs. In *Information and Communication Technology for Competitive Strategies (ICTCS 2020)*; Springer: Singapore, 2021; pp. 577–585.
14. Tamboli, M.S.M.J.; Desai, K.R. Approach of Thermal Imaging as a Facial Recognition. *JournalNX* **2016**, *2*, 1–4.
15. Wang, W.; Zhang, J.; Shen, C. Improved human detection and classification in thermal images. In Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 2313–2316.
16. Ilikci, B.; Chen, L.; Cho, H.; Liu, Q. Heat-Map Based Emotion and Face Recognition from Thermal Images. In Proceedings of the 2019 Computing, Communications and IoT Applications (ComComAp), Shenzhen, China, 26–28 October 2019; pp. 449–453.
17. 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.
18. Portmann, J.; Lynen, S.; Chli, M.; Siegwart, R. People detection and tracking from aerial thermal views. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 1794–1800.
19. Cerutti, G.; Prasad, R.; Farella, E. Convolutional neural network on embedded platform for people presence detection in low resolution thermal images. In Proceedings of the ICASSP 2019-IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019.
20. Gomez, A.; Conti, F.; Benini, L. Thermal image-based CNN's for ultra-low power people recognition. In Proceedings of the 15th ACM International Conference on Computing Frontiers, Ischia, Italy, 8–10 May 2018; pp. 326–331.
21. Cerutti, G.; Milosevic, B.; Farella, E. Outdoor People Detection in Low Resolution Thermal Images. In Proceedings of the 2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech), Split, Croatia, 26–29 June 2018.
22. Jiménez-Bravo, D.M.; Mutombo, P.M.; Braem, B.; Marquez-Barja, J.M. Applying Faster R-CNN in Extremely Low-Resolution Thermal Images for People Detection. In Proceedings of the 2020 IEEE/ACM 24th International Symposium on Distributed Simulation and Real Time Applications (DS-RT), Prague, Czech Republic, 14–16 September 2020; pp. 1–4.
23. Fung, A.; Wang, L.Y.; Zhang, K.; Nejat, G.; Benhabib, B. Using deep learning to find victims in unknown cluttered urban search and rescue environments. *Curr. Robot. Rep.* **2020**, *1*, 105–115. [[CrossRef](#)]
24. Lygouras, E.; Santavas, N.; Taitzoglou, A.; Tarchanidis, K.; Mitropoulos, A.; Gasteratos, A. Unsupervised Human Detection with an Embedded Vision System on a Fully Autonomous UAV for Search and Rescue Operations. *Sensors* **2019**, *19*, 3542. [[CrossRef](#)] [[PubMed](#)]
25. Cebollada, S.; Payá, L.; Flores, M.; Peidró, A.; Reinoso, O. A state-of-the-art review on mobile robotics tasks using artificial intelligence and visual data. *Expert Syst. Appl.* **2021**, *167*, 114195. [[CrossRef](#)]
26. Bañuls, A.; Mandow, A.; Vázquez-Martín, R.; Morales, J.; García-Cerezo, A. Object Detection from Thermal Infrared and Visible Light Cameras in Search and Rescue Scenes. In Proceedings of the 2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), Abu Dhabi, United Arab Emirates, 4–6 November 2020; pp. 380–386. [[CrossRef](#)]
27. Hoshino, W.; Seo, J.; Yamazaki, Y. A study for detecting disaster victims using multi-copter drone with a thermographic camera and image object recognition by SSD. In Proceedings of the 2021 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), Delft, The Netherlands, 12–16 July 2021; pp. 162–167. [[CrossRef](#)]
28. Dawdi, T.M.; Abdalla, N.; Elkalyoubi, Y.M.; Soudan, B. Locating victims in hot environments using combined thermal and optical imaging. *Comput. Electr. Eng.* **2020**, *85*, 106697. [[CrossRef](#)]
29. Madridano, Á.; Campos, S.; Al-Kaff, A.; Garcia, A.; Martín, D.; Escalera, A. Unmanned aerial vehicle for fire surveillance and monitoring. *Rev. Iberoam. Autom. Inform. Ind.* **2020**, *17*, 254–263. [[CrossRef](#)]
30. Vasconez, J.; Delpiano, J.; Vougioukas, S.; Auat Cheein, F. Comparison of convolutional neural networks in fruit detection and counting: A comprehensive evaluation. *Comput. Electron. Agric.* **2020**, *173*, 105348. [[CrossRef](#)]
31. Barrientos, A. TASAR—Team of Advanced Search And Rescue Robots. Available online: <https://www.car.upm-csic.es/?portfolio=tasar> (accessed on 29 June 2021).
32. Kumar, A.; Fu, Z.; Pathak, D.; Malik, J. Rma: Rapid motor adaptation for legged robots. *arXiv* **2021**, arXiv:2107.04034.
33. Yanco, H.A.; Drury, J.L.; Scholtz, J. Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition. *Hum. Interact.* **2004**, *19*, 117–149.
34. Budzan, S.; Wyżgolik, R. Noise Reduction in Thermal Images. In *Computer Vision and Graphics*; Chmielewski, L.J., Kozera, R., Shin, B.S., Wojciechowski, K., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 116–123.
35. Salisbury, J.W.; D'Aria, D.M. Emissivity of terrestrial materials in the 8–14  $\mu\text{m}$  atmospheric window. *Remote Sens. Environ.* **1992**, *42*, 83–106. [[CrossRef](#)]
36. Raman, R.; Thakur, A. Thermal emissivity of materials. *Appl. Energy* **1982**, *12*, 205–220.
37. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [[CrossRef](#)] [[PubMed](#)]

38. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
39. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
40. Quiroga, H. *Anaconda*; Createspace Independent Publishing Platform: North Charleston, SC, USA, 2017.
41. Blokdyk, G. *Tensorflow: A Complete Guide*; 5starcooks: Brendale, Australia, 2018.
42. PyTorch. Available online: <https://pytorch.org/> (accessed on 27 September 2021).
43. Home-OpenCV. 2021. Available online: <https://opencv.org/> (accessed on 27 September 2021).
44. Research/object\_detection at master · tensorflow/models. Available online: [https://github.com/tensorflow/models/blob/master/research/object\\_detection/g3doc/tf2.md](https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf2.md) (accessed on 30 June 2021).
45. Sanchez, S.A.; Romero, H.J.; Morales, A.D. A review: Comparison of performance metrics of pretrained models for object detection using the TensorFlow framework. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *844*, 012024. [[CrossRef](#)]