energies

Article

Solar-Powered Deep Learning-Based Recognition System of Daily Used Objects and Human Faces for Assistance of the Visually Impaired

Bernardo Calabrese [1], Ramiro Velázquez [1], Carolina Del-Valle-Soto [2], Roberto de Fazio [3], Nicola Ivan Giannoccaro [3,*] and Paolo Visconti [1,3]

[1] Facultad de Ingeniería, Universidad Panamericana, Aguascalientes 20290, Mexico; bernardo.calabrese@up.edu.mx (B.C.); rvelazquez@up.edu.mx (R.V.); paolo.visconti@unisalento.it (P.V.)
[2] Facultad de Ingeniería, Universidad Panamericana, Zapopan 45010, Mexico; cvalle@up.edu.mx
[3] Department of Innovation Engineering, University of Salento, 73100 Lecce, Italy; roberto.defazio@unisalento.it
* Correspondence: ivan.giannoccaro@unisalento.it; Tel.: +39-0832-297813

Received: 4 October 2020; Accepted: 19 November 2020; Published: 21 November 2020

Abstract: This paper introduces a novel low-cost solar-powered wearable assistive technology (AT) device, whose aim is to provide continuous, real-time object recognition to ease the finding of the objects for visually impaired (VI) people in daily life. The system consists of three major components: a miniature low-cost camera, a system on module (SoM) computing unit, and an ultrasonic sensor. The first is worn on the user's eyeglasses and acquires real-time video of the nearby space. The second is worn as a belt and runs deep learning-based methods and spatial algorithms which process the video coming from the camera performing objects' detection and recognition. The third assists on positioning the objects found in the surrounding space. The developed device provides audible descriptive sentences as feedback to the user involving the objects recognized and their position referenced to the user gaze. After a proper power consumption analysis, a wearable solar harvesting system, integrated with the developed AT device, has been designed and tested to extend the energy autonomy in the different operating modes and scenarios. Experimental results obtained with the developed low-cost AT device have demonstrated an accurate and reliable real-time object identification with an 86% correct recognition rate and 215 ms average time interval (in case of high-speed SoM operating mode) for the image processing. The proposed system is capable of recognizing the 91 objects offered by the Microsoft Common Objects in Context (COCO) dataset plus several custom objects and human faces. In addition, a simple and scalable methodology for using image datasets and training of Convolutional Neural Networks (CNNs) is introduced to add objects to the system and increase its repertory. It is also demonstrated that comprehensive trainings involving 100 images per targeted object achieve 89% recognition rates, while fast trainings with only 12 images achieve acceptable recognition rates of 55%.

Keywords: assistive technology; convolutional neural networks (CNN); deep learning; faster R-CNN; mobile computing; object recognition; person recognition; wearable system

1. Introduction

Globally, the World Health Organization (WHO) estimates that there are about 235 million people with severe visual impairment or complete blindness, for which the vision pathology cannot be corrected with the use of standard glasses or surgery [1]; the same considerations and almost similar numerical values have also been reported in [2].

Visual impairment interferes with the person's ability to perform everyday activities, such as environment understanding, urban mobility, reading, computer access, and object finding, among others [3,4]. Trying to adapt, to a greater or lesser extent, to the world is a constant challenge. Many works have addressed the environment understanding [5,6] and the urban navigation problem [7,8], by developing assistive devices based on smart sensors or artificial vision [9]; some others have proposed solutions for reading [10,11] and computer access [12,13] by exploiting the same devices and technologies. However, few works have focused on assisting visually impaired (VI) people in finding daily used objects. Object detection and recognition in a scene could ease VI people's lives: finding and reaching items in the surrounding space increases the quality of life and safety. Not knowing what is around could lead to frustration, anxiety, and involve hazardous situations that might lead to trips, falls, burns, and injuries. Automatic object detection and identification for VI people requires a flexible, adaptable, and computationally efficient approach that continuously learns and increases its knowledge upon use.

Recent progresses in Neural Networks and Deep Learning have contributed to advances in the field of computer vision. Deep Neural Networks (DNN), especially Convolutional Neural Networks (CNNs) have proven to be very effective in areas such as image recognition and classification [14,15]. In particular, the VGG16 (16-layer CNN) and VGG19 (19-layer CNN) architectures have been widely used for these tasks, since they require a moderate amount of training time; nevertheless, they are not efficient for real-time applications with high image processing speeds.

To address the computational burden that limits the CNNs classification speed, the R-CNN (Region-based CNN) was the first approach to be proposed; it selects several regions from the image and then uses CNN to extract features from each region [16]. However, selecting several regions requires the CNN to perform a significant number of computations; consequently, the computing load makes R-CNNs inefficient for real-time applications.

Fast R-CNN improves on the R-CNN by computing the image as a whole. Instead of feeding the set of regions to the CNN, Fast R-CNN takes the input image directly into the CNN to generate a single convolutional feature map. From this map, the proposed regions are identified using selective search algorithms. Fast R-CNN drastically reduces the training and detection times compared to R-CNN [17]. Still, its main inconvenience is that it requires the generation of many proposed regions to obtain precise object recognition. Hence, the bottleneck of this architecture is the selective search algorithm. Their successor, Faster R-CNN, replaces the selective search with a region proposal network [18]. This method allows to reduce the number of proposed regions while ensuring computational efficiency and precise object detection. Figure 1 compares the computing time of the above R-CNN based architectures for performing the objects' recognition in an image. Note that Faster R-CNN surpasses its predecessors; in fact, while R-CNN and Fast R-CNN take 50 and 2.3 s (the latter value corresponding to 4.6% of the time spent by R-CNNs), respectively, Faster R-CNN executes the same task in just 0.2 s (namely, only 200 ms), representing only the 0.4% of the time consumed by R-CNNs [17].

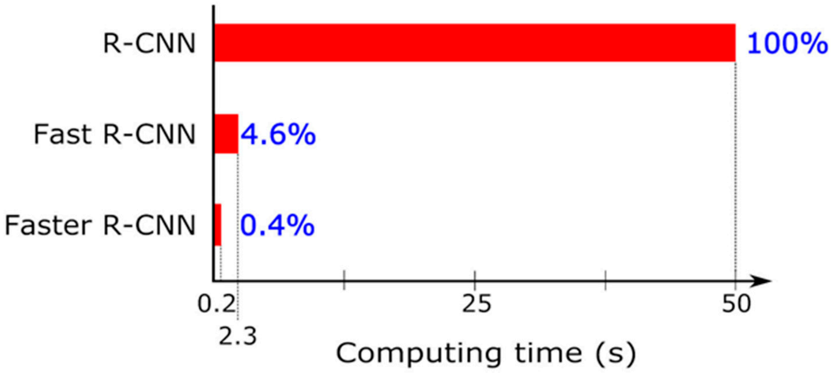


Figure 1. Comparison of performance of Region-based CNN approaches for object recognition.

This work presents a functional approach for the design and implementation of simple low-cost and efficient deep learning-based system, powered by a photovoltaic harvesting solution, for object finding for VI people by employing Faster R-CNNs. In addition to object recognition, the realized solar-powered system assists the user in positioning the objects in the surrounding space by integrating an ultrasonic sensor and a script in the main code. In CNNs, the more images used for training, the better result; however, in our proposal, quick trainings providing good results with an acceptable recognition rate (i.e., a value as high as 55%) are possible with only 12 images.

The novelty of this proposal lies in presenting a complete solution integrating deep learning structures in wearable hardware that is low-cost, reliable, highly performant, and truly affordable for the end-users. Moreover, the platform exhibits great flexibility by allowing the addition of new objects to its internal database in a simple manner. The device's recognition capabilities can be progressively increased and respond to the user's changing needs. Furthermore, the power consumption of the system has been discussed and optimized according to the working environment; then, a user-wearable solar harvesting system, based on flexible solar panels, has been designed and tested obtaining a NiMH battery life-time up to 6.1 h (outdoor operation). To our knowledge, such a system devoted to target the needs of the VI community has not been reported before in literature.

The rest of the paper is organized as follows: Section 2 provides a detailed literature review on systems providing object detection and recognition, in particular, those devoted to assist VI people. Section 3 overviews the system components and details the major concepts involved. Section 4 presents the experimental results obtained with the proposed system, whereas in Section 5 an in-depth discussion of the obtained experimental results is provided together with a comparison with those related to other scientific works. Finally, Section 6 concludes the paper summarizing the main contributions and giving the future work perspectives.

2. Related Works

Afterward, an overview of innovative assistive technology (AT) based devices for helping VI or blind users is reported, providing, at first, a classification of the different technologies employed in these applications. Furthermore, several innovative visual recognition methods are analyzed, classifying them based on their application.

2.1. State of the Art on Innovative Assistive Technology Devices for Aiding VI People

Recently, different wearable devices were reported in the scientific literature for aiding VI people, thus improving their life quality and reducing the likelihood of accidents.

Different technologies are available for object recognition, obstacle detection, or navigation support, etc.; in particular, the classification of main technologies employed for this typology of the device, based on their applications, is reported in the following Table 1.

Table 1. Summarizing table with the main reported technologies used in AT devices and relative applications.

Technology	Application
Visual recognition system	Object recognition, Face Recognition, Navigation, Positioning, Access control, Text recognition
Radio-frequency identification (RFID) technology	Object recognition, Positioning, Access Control
Global Positioning System (GPS) Navigation system	Navigation, Positioning
Ultrasound detection system	Navigation, Access control, Obstacle detection
Laser Imaging Detection and Ranging (LIDAR)/Optical Time-of-Flight distance sensors	Navigation, Obstacles detection

Table 1. *Cont.*

Technology	Application
Vibrotactile Interfaces	User interface
Audio Interfaces	User interface

Below, an overview of scientific works involving the technologies reported in the previous table is proposed; in several cases, the combination of multiple technologies has been integrated into these systems, to increase their functionalities or improve the reliability.

Meshram et al., in [19], proposed an innovative AT device, called the Nav-Cane, to aid VI people in orientation and navigation in indoor and outdoor environments, identifying objects or obstacles present along the user's path. The developed device is equipped with a radio-frequency identification (RFID) reader, ultrasonic sensors, placed at different heights, a GPS receiver, an inertial sensor, and a water sensor, all integrated inside a cane-like structure. The RFID reader is used to recognizing objects previously tagged, as well as the ultrasonic sensors provide direct feedback to the user, through tactile warnings produced by a vibration motor, related to the presence of obstacles at different levels (i.e., foot, knee, waist, chest); besides, a wet floor sensor warns the user about the presence of a wet floor, and an alarm button is added for alerting, through an email or Short Message Service (SMS), rescue teams indicating the user's GPS coordinates. The experimental results, carried out on 80 VI subjects in a controlled environment, demonstrated the effectiveness of the Nav-Cane in obstacle detection, descending or ascending stairs, as well as to aid the user to identify objects.

Similarly, in [20], the authors proposed the Assistor, a smart walking cane for helping blind or VI people to recognize obstacles, and thus navigating till the destination. This device employs three ultrasonic sensors, two at the top and two at the bottom to cover a wider coverage angle, and a miniaturized camera (model Pixy CMUCam5, manufactured by Charmed Labs, Austin, TX, USA), integrating a powerful processor supporting several communication protocols to transmit acquired images to a host microcontroller. The data from sensors are sent via Bluetooth communication to a smartphone, where they are processed to drive the servo-motor used to move the walking cane. The smartphone GPS receiver and Google map application guide the user toward the destination, whereas the sensing section identifies obstacles employing the Speeded Up Robust Features (SURF) algorithm, supported by an image-object database.

In [21], the authors presented an obstacle detection system based on multiple sonar devices, constituted by several ultrasound sensors, and distributed to sense a large area in the field of view (FOV) of the user; an actuator is associated to each sensor, represented by a vibrating motor (model SAM-A300, manufactured by Samsung, Seoul, South Korea), and installed to provide vibro-tactile feedback to the user concerning the obstacle position. The device is managed by a PIC18F6720 microcontroller, which collects the data from the ultrasound sensors and processes them in order to drive the actuators, according to a previous calibration carried out on the user conditions. Five different users tested the developed device; the experimental results indicate a reduction of 50% of the time required by the user to pass through a series of obstacles.

Chen et al. proposed a smart wearable system for image recognition, adopting cloud and local cooperative processing [22]. Specifically, the cloud server carries out image processing tasks, whereas the local processing unit only uploads the images on the cloud server and receives the processing results; therefore, low-cost and resources-limited processors can be employed for developing the wearable device, reducing its overall cost. The proposed wearable device includes a micro-camera, an ultrasonic sensor, and an infrared sensor installed on the frame of the glasses; a Raspberry Pi board is used as a local processor, providing to the device wireless connectivity (WiFi or 4G), for sharing the images with the cloud server, exploiting its parallel computing power and storage capacity. A proper algorithm combines the captured images with the data provided by the ultrasonic and infrared sensors to extract the "point of interest" useful for recognition. The experimental results demonstrated the effectiveness of the proposed device in detecting faces, text, and objects in real scenarios.

In [23], the authors proposed a wearable Obstacle Stereo Feedback (OSF) system to assist the navigation of VI users; the system implements a downsampling Random Sample Consensus (RANSAC) algorithm to elaborate the acquired point cloud for detecting obstacles placed along the user's path. Furthermore, the Head-Related Transfer Function (HRTF) is employed to provide an acoustic representation of obstacles as a function of their 3D coordinates.

Neto et al. in [24] presented a face recognition system for VI users' assistance; the developed system employs a Kinect-Based sensor bar installed on a helmet and a KNN algorithm, based on histogram-oriented descriptors compressed by the principal component method. The experimental results demonstrate that the proposed detection algorithm requires lower computational resources compared to other techniques reported in the literature, still keeping an excellent accuracy under different operative conditions, such as background, illumination, and point of view.

Katzschmann et al., in [25], presented the Array of Lidarsand Vibrotactile Unit (ALVU), a wearable device, thought for blind and VI users, for detecting obstacles and their physical boundaries. The device includes a sensor belt and a haptic strap; the first one is equipped with time-of-flight distance sensors, for measuring the distance between the user and the obstacles, whereas the haptic strap provides feedback to the user, by a series of vibratory motors placed on the user's abdomen. Other relevant applications involving real-time object recognition can be found in the literature: Chen et al. introduced in [26] the Glimpse system: a real-time object recognition system for smartphones running on external server machines. Experiments on road signs show a recognition accuracy from 75% up to 80%. Viola and Jones proposed in [27] a face detection framework capable of processing images extremely rapidly while achieving high detection rates (95%). Jauregi and coworkers addressed door identification for robot navigation in indoor spaces [28]; by employing a three-stage algorithm, they achieved 98% recognition rates.

Few systems have targeted VI people needs; Niu et al. described, in [29], a wearable system that detects doorknobs and human hands to help the blind people locate and use doors. Panchal et al. presented, in [30], a new approach for recognizing text from scene images and to convert it into speech so that it can assist VI people. Jabnoun and colleagues reported in [31] an object recognition system based on SIFT (Scale Invariant Features Transform) and SURF algorithms to assist VI people during navigation. Ciobanu et al. introduced, in [32], a method for detecting indoor staircases by employing IMU (inertial measurement unit) sensors and processing depth images in order to aid VI people in unfamiliar environments.

In this context, this paper takes a step forward to the field of object recognition for assisting the VI targeting daily used objects such as doorknobs, power outlets, white canes, light switches, among many others as well as person recognition. The designed and tested solar-powered system captures real-time video, locates the objects of interest, recognizes them, tracks them frame by frame, and finally provides audible descriptive feedback to the user.

2.2. Overview of Applications and Innovative Methods for Visual Recognition

Object recognition has become an important topic in computer vision. It has been already explored in video surveillance [33], robot navigation [34], medical imaging [35], smart homes [36], and even tourism [37]. In this paragraph, innovative methods for object recognition have been reported and detailed, also providing a comparative analysis between the reported algorithms.

In [33], the authors proposed an innovative method to determine the performances of object recognition algorithms in the videos, highlighting specific features of the particular method, such as region splitting or merging; the method relies on the comparison between the output of the recognition algorithm and correct split segment extracted with 1 frame/s sampling rate. Similarly, Lu et al., in [38], introduced a real-time object detection framework for video, employing the You Only Look Once (YOLO) network, with an improved convolution method for speeding up the elaboration, and thus object detection. Through a preprocessing, the effects of the background are removed, as well as the processing noise. The experimental results indicate that the proposed method obtains better performances compared to the initial YOLO algorithm, reaching higher detection speed and accuracy. Furthermore, object recognition algorithms can find application for navigation systems in Autonomous Vehicle (AV) and robotic fields; Hernández et al. proposed an object recognition system, based Support Vector Machine (SVM) classification method applied on RGB images [34]; two segmentation approaches have been tested based on geometric shape descriptors and bag of word method, respectively.

These algorithms can be used also for medical imaging applications, aiding VI users to carry out tasks otherwise would not be able to perform; for instance, in [35], the authors proposed a mobile application to allow blind users to read a text. The Camera Reading for Blind People project employs Optical Character Recognition (OCR) and Text to Speech Synthesis (TTS) techniques, integrated on a smartphone, for acquiring pictures from a text, and vocally synthesizing the recognized text.

Furthermore, the automated systems for smart homes can widely exploit object recognition systems; Baeg et al. developed a smart home environment for service robots, equipped with a camera, RFID reader, and a wireless communication module [36]. Upon the use of the RFID reader, the robot obtains course information about its position and encountered objects, then, by a recognition system based MPEG-7 visual descriptors, the robot can determine the exact position of the object for grasping it. The visual recognition system can be also used to recognize places, monuments, statues, paintings, etc., in order to provide information and data to the tourists; in [37], the authors proposed a mobile vision system for automatic object recognition applied to the images acquired using a camera phone. The system allows determining places of tourist interest, to provide detailed information related to the architecture, history, or cultural context of historical or artistic relevance.

Trabelsi et al., in [39], developed a novel multi-modal algorithm for aiding VI users in the recognition of objects into an indoor environment, employing RGB-D (Red Green Blue-Depth) images with new complex-valued representation, in order to overcome the limitations of the traditional techniques. Two methods have been proposed to categorize objects; a Multi-Label Complex-Valued Neural Network (ML-CVNN) is developed, based on an adaptative clustering method applied to multi-label problems solving. The latter, called L-CVNN, uses a CVNN for each considered label to obtain a multi-label vector. The experimental results demonstrate the efficiency of the proposed techniques based on RGB-D images compared to existing methods, such as RGB ML-Real-Valued Neural Network (RVNN), Depth ML-RVNN, RGBD ML-RVNN, in the object classification. Similarly, Malūkas et al. introduced a real-time navigation system for VI and blind people, employing a segmentation framework based on a deep Convolutional Neural Network (CNN) algorithm to recognize objects/features into an image [40]. Three different CNN algorithms have been tested (i.e., AlexNet, GoogLeNet, and VGG-Visual Geometry Group Net), obtaining the best performances in the segmentation with the VGG16 (16-layers) neural network, and reaching 96.1 ± 2.6% accuracy in the recognition of paths, structures, and path boundaries.

Jayakanth proposed a real-time algorithm for object recognition in indoor environments, such as door, stairs, and signs [41]; the algorithm is based on a transfer learning approach for implementing a deep learning model trained using the AlexNet. Furthermore, different texture feature extracting techniques (Local Binary Pattern—LBP, Binarized Statistical Image Features—BSIF, and Local Phase Quantization—LPQ) have been tested in the proposed framework, followed by a machine learning classifier (i.e., K-Nearest Neighbors—KNN, Naive Bayes—NB, and SVM) for the identification and classification of objects. The test results demonstrate that the BSIF and LPQ texture extractors excellently work in object recognition; specifically, the first one in conjunction with the KNN classifier obtain 98.4% accuracy, whereas the second one produces 98.9% and 98.4% when operates with SVM and KNN classifiers, respectively.

Furthermore, Jabnoun et al., in [42], described a visual tool for VI people, based on an object recognition algorithm allowing to determine the dissimilarity between video frames; specifically, the algorithm employs the Real-Valued Local Dissimilarity Map (RVLDM) method, as a measure of frames' dissimilarity, and the Scale-Invariant Features Transform (SIFTS) keypoints extraction, for determining the objects depicted in different frames. By comparing the proposed method with similar visual substitution approaches, optimal performances have been demonstrated, in terms of computational speed in different operative conditions, such as different point-of-view, presence of occlusion, frame rotation, and different illumination. Finally, in [43], the authors described a novel 3D object reconstruction method, based on a modified hybrid artificial neural network, obtaining more precise filling of partial object images and reducing the process noise compared to the YOLOv3 algorithm. Furthermore, the obtained reconstruction is more stable than the results obtained with other reconstruction techniques.

In Table 2, the performances of novel implementations of common object recognition algorithms are reported. As evident, YOLO frameworks reach higher processing speed, but impose strong spatial constraints on bounding box predictions, thus limiting the algorithm ability to discern near objects [44]. Furthermore, region-based CNN algorithms, and specifically Faster R-CNN ones, represent an excellent trade-off between operation speed, accuracy, and resource utilization [45].

Table 2. Summarizing table of the performances of common object recognition algorithms.

Algorithm	Architecture	Mean Average Precision [%]	Mean Computing Time [ms]	Dataset
Hernández [34]	SVM	78.34	n.a.	NYU Depth Dataset V2 [46]
Lu [38]	Fast YOLO	88.45	22	Custom
Trabelsi [39]	ML-CVNN	87.2	n.a.	RGBD [47]
Malūkas [40]	FCN-VGG16	96.1	105	ImageNet [48]
Jayakanth [41]	CNN + SVM	100	451	MCindoor20000 [49]
Redmon [50]	YOLOv3	57.9	51	COCO
Ren [45]	Faster R-CNN	73.2	198	COCO

3. System Overview

With the aim of assisting VI people, we have developed a wearable assistive technology (AT) solar-powered device that recognizes daily used objects and persons and that lets the user know about its presence and location in the surrounding space.

Figure 2 shows the conceptual representation of the AT device and its operation principle. The system encompasses three major elements: a miniature camera, an embedded system on module (SoM), and an ultrasonic sensor; the images acquired by the camera are processed in real-time by the SoM to detect commonly used objects. When an object is detected, the SoM spells out the name of the object through an integrated speaker letting the user know its presence and location in the nearby space. An ultrasonic sensor placed at belt buckle level completes the description providing the distance to the recognized object. The following subsections detail the designed system's elements.

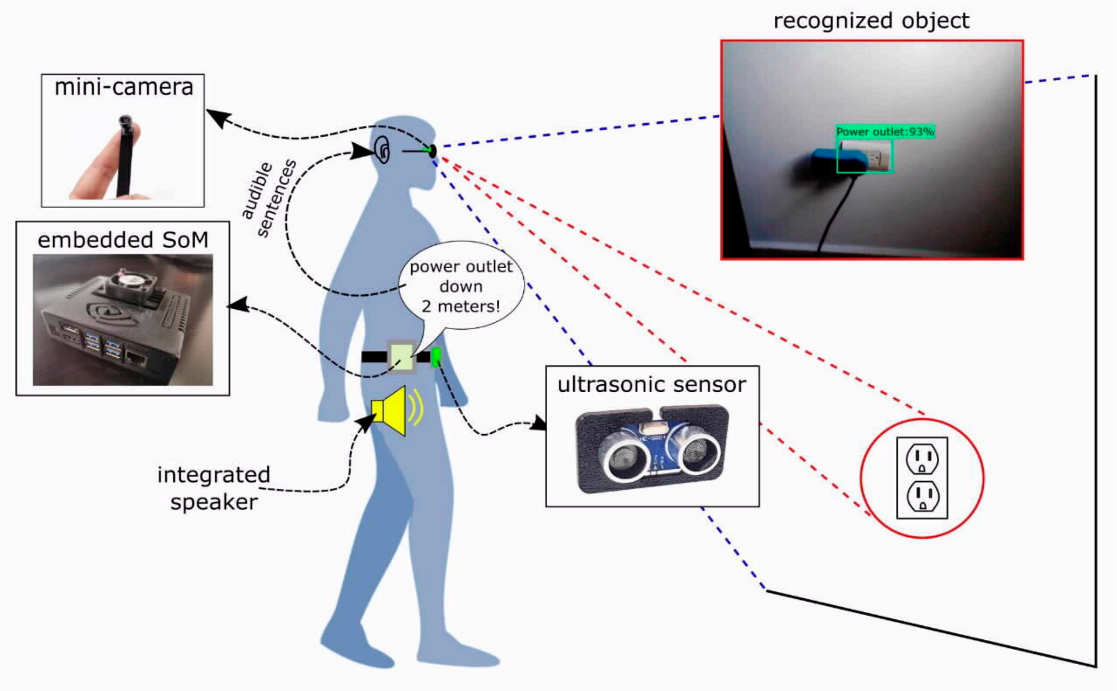


Figure 2. Wearable AT device for VI people; the camera captures an image of a wall with a power outlet, algorithms running on SoM detect power outlet and the ultrasonic sensor provides distance to the object. The system lets user know about its presence, location, and distance via audio form.

3.1. Hardware Section

The employed miniature camera is a CMOS RGB sensor (model 6986154272705, manufactured by Hamswan Company, Shenzhen, China) with an indicative price of 30–35 USD (Figure 3a). It provides images with a resolution of 1280 × 960 pixels and a field of view (FOV) of 120 degrees (as reference, the human FOV is 190 degrees [51]). Its dimensions (12.5 mm × 12.5 mm × 17 mm) and mass (12 g) allow having it placed on the VI glasses frame, so that users can wear the small size and light camera, fully integrated into the glasses, without even noticing it. The miniature camera is directly interfaced with a local processing and communication module, which acquires the frames and wirelessly transmits them to the main processing unit (below described), via peer-to-peer WiFi communication (Figure 3a); in this way, no cables are required, allowing maximum freedom of movement for the user. Its compactness, low cost (30–35 USD), and USB powered feature make it ideal for this application.

The SoM used is the Jetson Nano embedded device (manufactured by NVIDIA Co., Santa Clara, USA) with an indicative price of 120–130 USD. Its 4 Gb LPDDR4 (Low-Power Double Data Rate) RAM, 128-core NVIDIA Maxwell architecture-based graphics processing unit (GPU), and Quad-core ARM A57 Central Processing Unit CPU, running up to 1.4 GHz, allow the execution of algorithms based on artificial intelligence (AI) in real-time [52,53]. The prototyping board includes several interfaces, such as a DisplayPort, HDMI-High Definition Multimedia Interface, four USB-Universal Serial Bus 3.0 ports, two CSI-Camera Serial Interface connectors, Gigabit Ethernet, an M.2 Wifi card slot, and a set of GPIO placed on the side of the board, making it ideal for a variety of AI applications in Autonomous Vehicles (AVs) and robotics (Figure 3b). Its dimensions (70 mm × 45 mm × 25 mm) and mass (240 g) make it wearable; for example, the users can attach it to their belts, as shown in Figure 2. The Jetson Nano board is featured by a 5 V DC power supply voltage, applicable to a common barrel plug or a micro-USB connector. The power consumption of the SoM is in the range between 5–10 W, according to the computational load of the module, which is ensured by a 4600 mAh NiMH battery pack continuously recharged by a solar-based energy harvesting system. As following detailed, by specified experimental tests, an energy autonomy from 1.6 h up to 3.2 h has been obtained depending on the SoM operating mode activated over time, which can be further enhanced, depending on the solar illumination level, by recharging the battery thanks to the solar harvesting system.

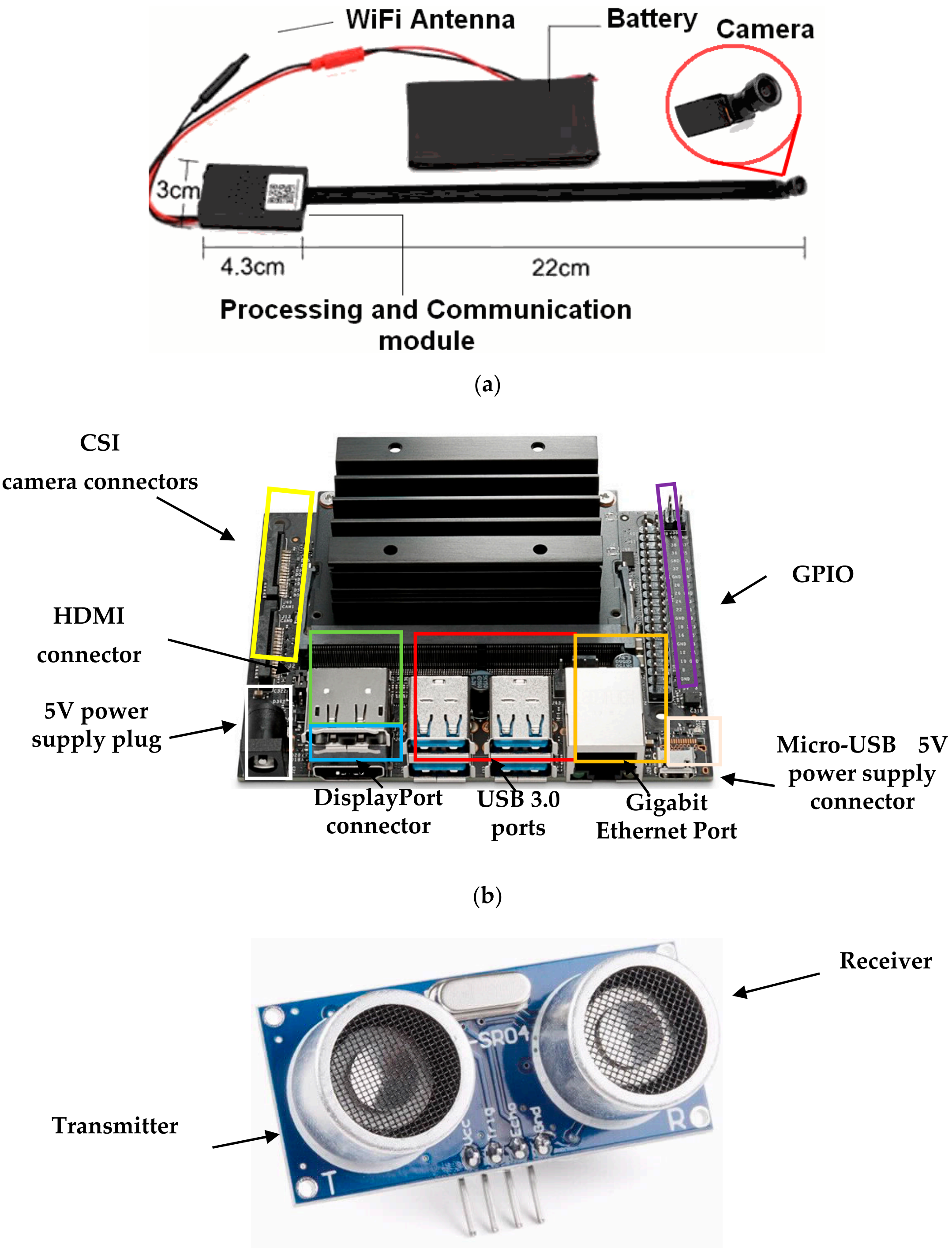


Figure 3. (**a**) The miniature camera used for the proposed wearable devices with highlighted the main section, (**b**) top-view of the Jetson Nano electronic module with highlighted the main interfaces, (**c**) the ultrasonic distance sensor with transmitter and receiver mounted on a single board.

The ultrasonic sensor used in the AT device is the HC-SR04 from Shenzhen Robotlinking Technology Co. (Shenzhen, China) with an indicative price of 20 USD. To estimate the distance to an object, it uses a transmitter to emit acoustic pulses to the environment and a receiver to capture their echoes (Figure 3c). This sensor offers a stable performance within the range 0.02 to 4 m and a high precision of ±3 mm. The sensing ultrasonic beam is just 15 degrees. The system consumes just 75 mW (5 V, 15 mA) and is powered by the SoM. Its dimensions (45 mm × 20 mm × 15 mm) and mass (20 g) make it easily installable on the SoM's belt at buckle level, thus sensing the space in front of the user. Sensing range, precision, and dimensions make it a good option for this application.

The employed CMOS RGB camera connects via USB port to the Jetson Nano SoM, which automatically recognizes it with no further software configuration (namely, plug and play mode). The interested readers can find the Jetson's nano supported cameras and hardware in [54,55]. Similarly, the ultrasonic sensor connects via cables to the Jetson Nano SoM transmitting the electric signal encoding the sensed distance.

Finally, a standard computer speaker was integrated to the SoM using its input/output ports. A speaker was preferred to headphones to avoid the obstruction of environmental hearing; in fact, VI people strongly rely on environmental cues to navigate and orient themselves in space [3].

The Jetson Nano module integrates several solutions to optimize the power consumption of the board, tailoring them on the specific application. In particular, it has two operating modes with different power consumption, namely the *mode 0* (also called *MaxN* mode, default mode) and *mode 1* (also called 5 W mode); in the first one, the power consumption of 10 W is enabled to obtain the maximum performance, whereas in the latter, the power consumption of the board is limited to only 5 W, by constraining the memory, CPU and Graphical Processing Unit (GPU) clock frequencies, and the number of active cores [56]. Specifically, *mode 1* has maximum CPU and GPU clock frequencies limited to 918 MHz and 640 MHz, respectively, and only two active cores. The Jetson Nano power mode can be changed by the *nvpmode* command, passing as parameter the identifier of the selected modality. Enabling the *mode 1*, the object recognition time increased compared to the elaboration in *MaxN* mode, obtaining 360 ms mean computing time, but leaving unaltered the mean average precision (namely, 86%).

Therefore, the developed recognition system is equipped with a triaxial MEMS (Micro-Electromechanical System) accelerometer (model MMA8451Q, manufactured by NXP, Eindhoven, The Netherland) used to detect the user speed, and thus dynamically adapt the SoM power consumption according to the user condition. Specifically, if the user is stationary or moving slowly, a high recognition speed is not needed, the *mode 1* is thus enabled; on the contrary, if the user walks quickly, a rapid object detection is required, in this condition, the *mode 0* is thus enabled. The power management logic, implemented by the processing board, continuously detects the user speed and if this last is higher than 2 m/s for a time interval greater than 5 s, then the *mode 0* is set; else, the board is configured in *mode 1*. In this way, a reduction in energy consumption of the developed system is obtained, compared to the continuous operation of the device in *MaxN* mode, leaving, from a practical point of view, unaltered the functionality. Therefore, the energy autonomy of the developed wearable AT device is increased, within the range between extreme values in which the system constantly consumes 5 W or 10 W. The MMA8451Q breakout board was placed inside the cover containing the Jetson Nano board, fixed at the belt of the user (as shown in Figure 2), ideal position for detecting the body movements, as reported in [57].

In order to verify possible commercial developments of the proposed novel AT device, extensive research has been carried out for finding similar devices and for comparing the performances. A commercial wearable device for the blind and VI, named Horus [58], has been recently presented on the web market. It is based on the NVIDIA Jetson embedded AI computing platform and it is composed of a wearable headset with cameras and a pocket unit that contains a processor and a long-lasting battery. Horus is presented in [58] as a wearable device that observes, understands, and describes the environment to the person using it, providing useful information; it is able to read texts, to recognize faces, objects and the user can activate each functionality through a set of buttons located both on the headset and the pocket unit.

The architecture, the algorithms, and the software implementation of the proposed device in this research work are optimized and simpler than the Horus prototype. A final cost comparison shows definitely that the proposed solution could be more competitive also considering a commercial development; in fact, the total cost of the proposed prototype has been evaluated in only 200 USD, while the commercial proposed price for the Horus device is 2000 USD, ten times more expensive.

Design of Solar Energy Harvesting System to Extend the AT Device Life-Time

Furthermore, a wearable solar energy harvesting system has been developed for power supplying the designed AT device, so allowing it to extend its energy autonomy. In particular, two 6 W flexible mono-crystalline solar panels (model HX160-220P, manufactured by Huaxu Energy Co., Shenzen, China), connected in parallel, have been placed on the back of a jacket by using metallic clip buttons, for easier removal for garment washing. Specifically, they are polyethylene terephthalate (PET) laminated solar cells, featured by 7.2 V open-circuit voltage, 1100 mA short-circuit current, 6 V peak voltage, 1000 mA peak current, 19.5% conversion efficiency, and 160 mm × 220 mm × 2.8 mm dimensions (Figure 4). The solar cells are interfaced with an S18V20F6 buck-boost voltage converter (manufactured by Pololu Co., Las Vegas, NV, USA), featured by a wide input range (from 2.9 to 32 VDC), a fixed 6 VDC output voltage with 4% accuracy, 2 A maximum output current, and typical efficiency between 80% and 90%. Furthermore, the board is equipped with reverse voltage protection (up to 30 V), over-current protection, and over-temperature shutdown circuit. The charge extracted by the solar panels is stored into a 4600 mAh, 6 VDC NiMH battery pack (manufactured by Vapextech UK Ltd., Kent, UK), used to power supply the developed wearable AT device (as shown in Figure 4). Since the NiMH battery pack when fully charged reaches 6 VDC, voltage value incompatible with the power supply range of the Jetson Nano board, a buck converter, based on the MP2315 (manufactured by Monolithic Power Company, Kirkland, WA, USA) synchronous controller, has been employed to provide the 5 VDC stabilized voltage required by the SoM (Figure 4). The MP2315 Integrated Circuit (IC) is featured by a wide input voltage range (from 4.5 V to 24 V), a 3 A maximum load current, and high efficiency (97%).

The battery pack and the electronic section have been placed into pockets realized in the internal part of the garment, whereas the connections have been realized by means of highly flexible cables sewn on the fabric.

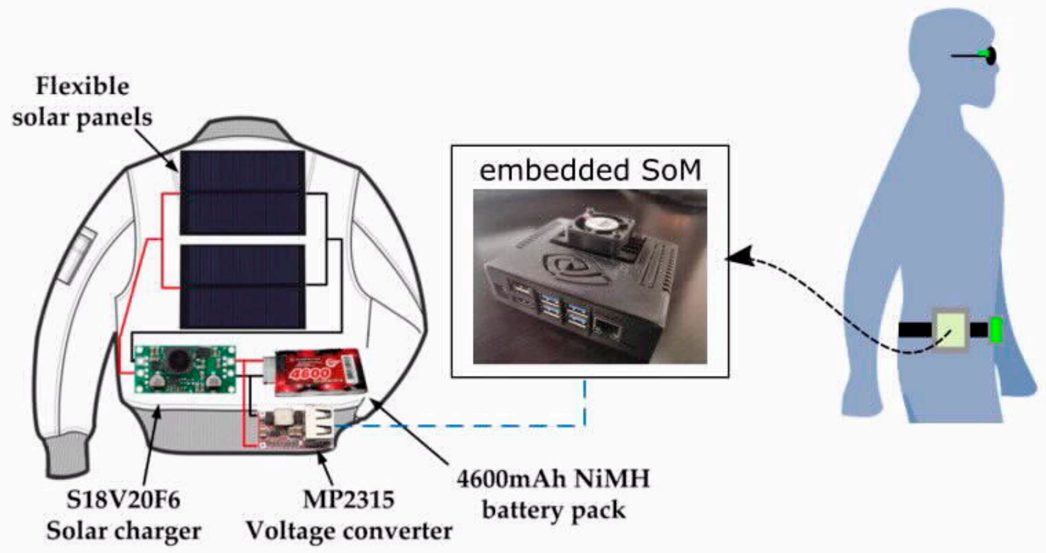


Figure 4. Graphical representation of the wearable solar harvesting system, integrated with the developed wearable AT device.

The solar harvesting system, above described, has to be worn by the user some hours before connecting the wearable AT device, in order to fully charge the 4600 mAh NiMH battery pack; afterward, this last is used to power supply the realized object recognition device, thus ensuring an energy autonomy, in the total absence of luminous sources, comprised between the two limit values below reported, related to the cases in which the Jetson Nano board is continuously configured in mode 0 (MaxN mode, 2A absorbed current) and *mode 1* (5 W mode, 1 A absorbed current), respectively.

$$\text{Energy Autonomy}_{\text{ mode 0}} = \frac{\text{Battery Capacity [mAh]} \times (1\text{–Discharge\_Margin})}{\text{Absorbed current [mA]}} = \frac{4600 \text{ mAh} \times 0.7}{2000 \text{ mA}} = 1.6 \text{ h}. \tag{1}$$

$$\text{Energy Autonomy}_{\text{ mode 1}} = \frac{\text{Battery Capacity [mAh]} \times (1\text{–Discharge\_Margin})}{\text{Absorbed current [mA]}} = \frac{4600 \text{ mAh} \times 0.7}{1000 \text{ mA}} = 3.2 \text{ h}. \tag{2}$$

where the Discharge_Margin is the percentage limit to the discharge of the 4600 mAh NiMH battery pack (typically 30%). However, the obtained values have to be considered as the minimum energy autonomy since the energy contribution provided by the harvesting section during the operation of the wearable AT device will allow to increase further the device's autonomy. The solar contribution depends on the environmental conditions (i.e., source typology, luminous intensity, inclination, and orientation of the solar panels with respect to the light source). However, field tests demonstrate that placing the garment perpendicularly to the sun, with 52.000 lux illuminance and leaving the device stationary, the energy autonomy increases of about a factor 2 (i.e., up to 6.1 h), compared to the total absence of any luminous source (namely, 3.2 h). Furthermore, in the same condition, the energy harvesting section employs about 3.2 h to fully charge the 4600 mAh battery pack, before connecting the developed solar-powered wearable device.

3.2. Software Section

The SoM-based software structure is shown in Figure 5; as highlighted, it consists of three major modules: dataset configuration, object recognition, and object positioning.

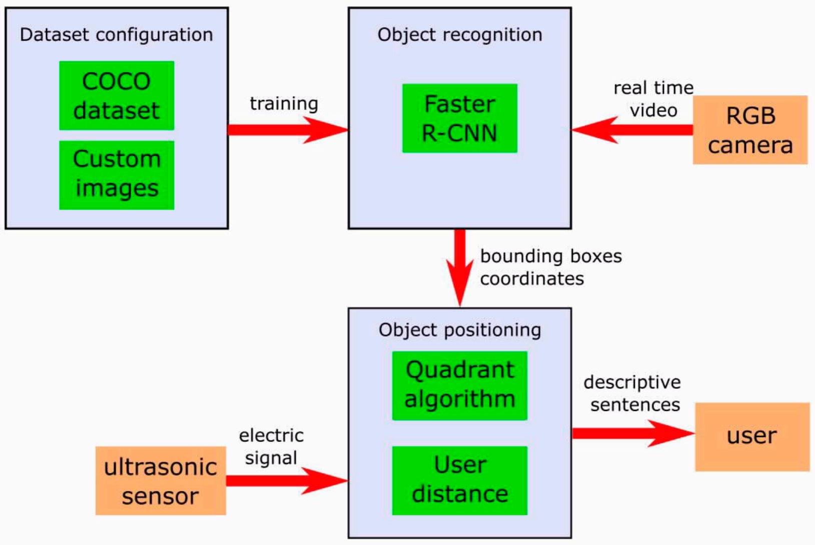


Figure 5. Overview of the software structure.

The dataset configuration module encompasses the Microsoft Common Objects in COntext (COCO) dataset [59], which contains 91 categories of common objects, with 82 categories having more than labeled instances. In total, COCO contains 328,000 images with 2,500,000 labeled instances. A main feature of COCO is that it offers non-canonical views of the objects (for example objects in the background or partially occluded or amid clutter), which improve recognition performance. The module can also manage custom images or images of interest added by the user.

The dataset configuration module is used to train the object recognition module. Based on Faster R-CNN, the former is the backbone of the system, being responsible for detecting and recognizing common objects found in everyday scenes and situations from real-time video. This module outputs the coordinates of the bounding boxes enclosing the object. The positioning module detects the location of the object in the image via a simple Cartesian quadrant algorithm and together with the information coming from the ultrasonic sensor let the user know the presence of the object, its location, and user-distance via audible descriptive sentences.

The SoM runs a Linux operating system (OS); all code was implemented by employing the Python software, running into the Jetson Nano SoM [52,53]. The Jetson Nano Developer Kit was set up according to the official Nvidia installation guide [60]. In this case, the archive JetPack 4.3 and the framework Deepstream are recommended for Tensorflow 1.14.0 according to Nvidia official documentation [61]. Python 3.6 was utilized, being the most recent release compatible with TensorFlow 1.14.0. The essential libraries used to support the training and execution on the Jetson nano SoM are the following: Numpy, Pycocotools, PyCuda, OpenCV, Time, Serial, and MatPlotLib, PIL, all in their latest release.

3.2.1. Object Recognition

As aforementioned, Faster R-CNN is the last development of the R-CNN family architectures. It increases the computational efficiency reducing both the training and testing times and improves object recognition performance. Its architecture is shown in Figure 6.

Faster R-CNN consists of three main modules: a convolutional layer, a region proposal network, and a classification layer. The convolutional layer is the feature extraction stage. It involves of a set of filters that activate when they detect visual features in the input images. such as edges, colors, and specific orientations. This stage outputs feature maps. The region proposal network stage generates locations of possible objects contained in the feature maps. Resulting region proposals are then applied to the feature maps. Finally, the classification layer is used to determine which class the objects found belong to.

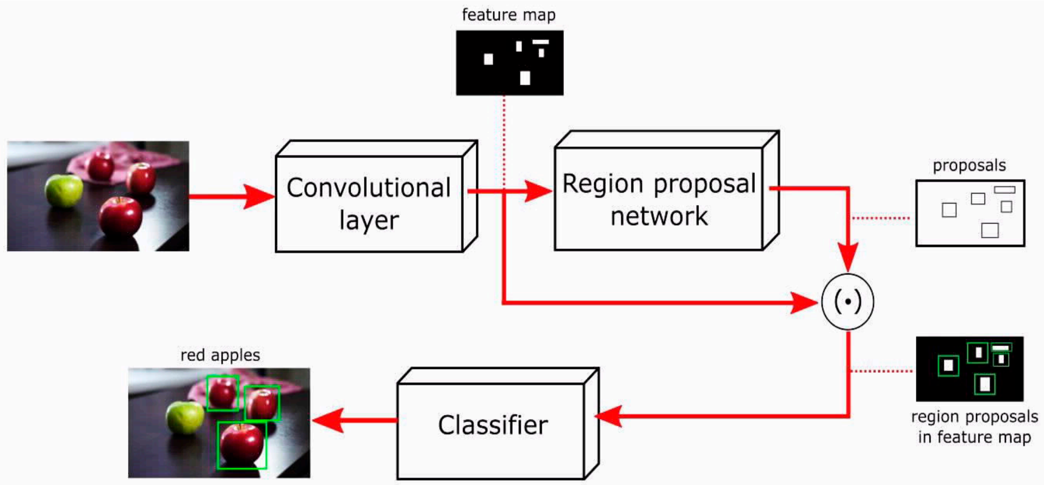


Figure 6. General architecture of the Faster R-CNN.

In this work, the VGG16 architecture was used for the convolutional layer (Figure 7). VGG16 has exhibited the best performance in image recognition tasks [18]. We have previously used it in facial emotion recognition for predicting consumer acceptance of food products with satisfactory results [62]. VGG16 is a 16-layer CNN; it encompasses five filter blocks containing a total of 13 filter layers and five pooling layers. Each of the 13 filter layers includes a rectified linear unit (ReLU) in order to allow a faster and more effective neural network training. Three fully connected layers follow the stack of filter layers to flatten the high-level features in the data. The softmax function is the final layer of the architecture. It maps the non-normalized data to a probability distribution, which can be used as input of the following modules.

The Region Proposal Network module (Figure 8) takes the feature map as input to generate a set of rectangular object proposals (named bounding boxes). To generate them, a mask slides along the feature map. Resulting values are fed to two parallel submodules: a classifier and a regressor. The first determines the probability p_i of a proposal having the target object. The former provides the pixel coordinates (x, y) of the proposal.

Finally, the classification layer is based on the Fast R-CNN architecture (Figure 9). Its input is the feature map with the region proposals. To reduce the amount of data and thus the computation to be performed, a RoI (Region of Interest) pooling layer is first used to down-sample the feature maps. Each down-sampled feature value goes to a fully connected layer (the learning one) to detect the non-linear combinations of these features. As in the previous stage, the resulting values are fed to a classifier and a regressor. The output of this module is the object detection.

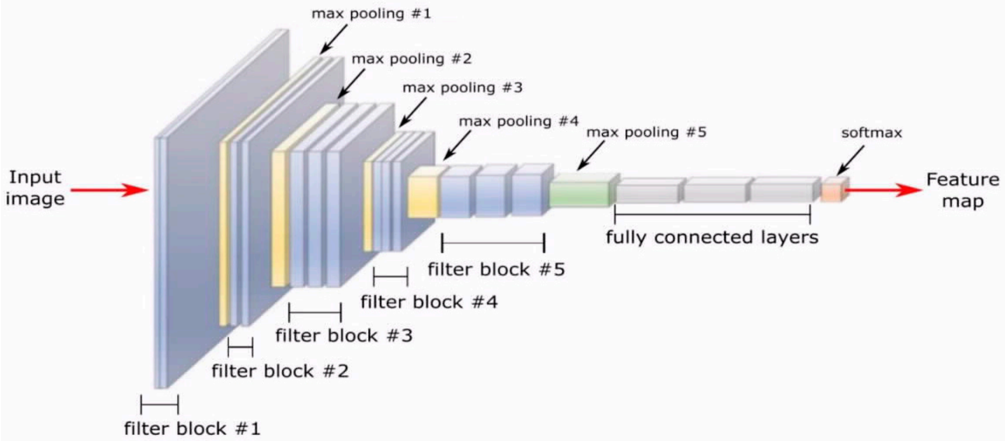


Figure 7. Convolutional layer based on the VGG16 architecture.

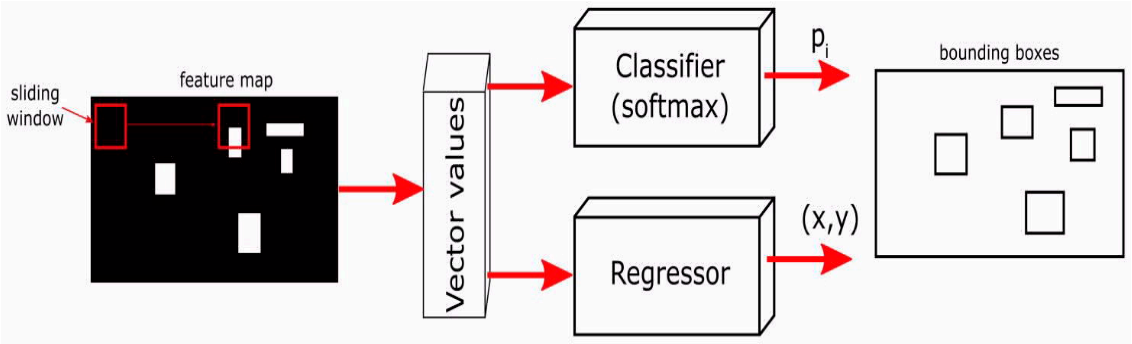


Figure 8. The Region Proposal Network module.

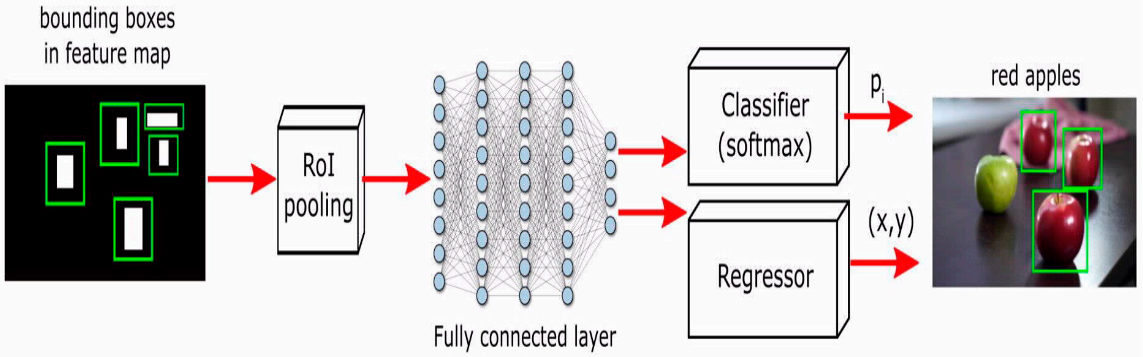
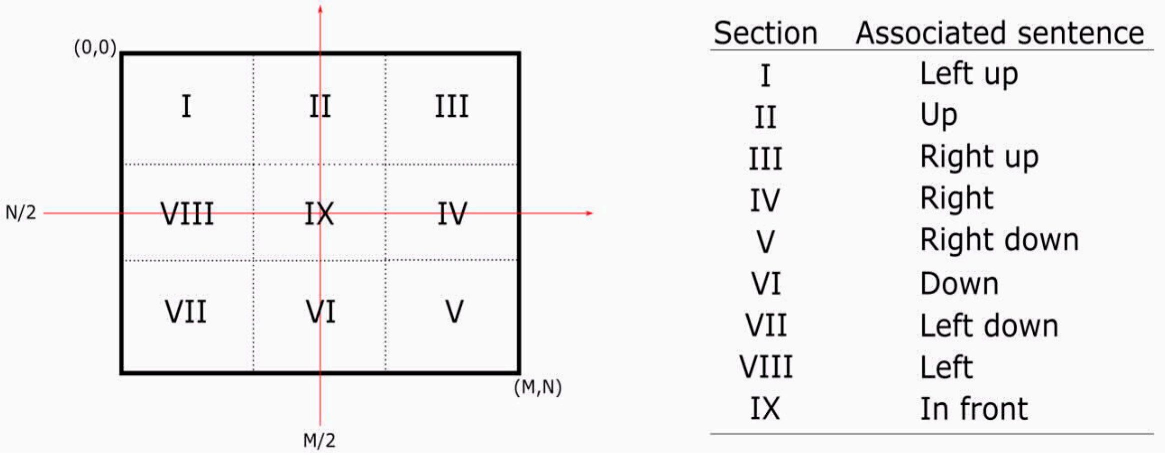


Figure 9. The classification layer based on the Fast R-CNN architecture.

3.2.2. Object Positioning

Once an object has been detected in the FOV of the camera, it is useful to let the user know about its location and user-distance. The software architecture encompasses a simple and fast computing object-positioning algorithm based on Cartesian quadrants. The concept is illustrated in Figure 10. The image is divided into four quadrants; nine sections are defined within the Cartesian space. When an object is detected, its coordinates can be safely positioned in one of these sections and an audible descriptive sentence is associated (example: "power outlet left down", "doorknob in front", "white cane down", etc.). Users can then be aware of the presence and location of an object and decide whether to approach towards it.



Section	Associated sentence
I	Left up
II	Up
III	Right up
IV	Right
V	Right down
VI	Down
VII	Left down
VIII	Left
IX	In front

Figure 10. Object positioning algorithm based on Cartesian quadrants.

There might be cases were objects span over two or more sections rendering ambiguous this positioning approach. To solve these cases, the geometrical center of mass (CoM) of the bounding box enclosing the object was taken as a reliable indicator for deciding on its positioning section.

To determine the user-object distance, the ultrasonic sensor transmits out an eight-cycle 5 V ultrasonic pulse at 40 kHz and waits for the reflected echo. When an object is detected, the echo signal sets high (5 V) and its width t (in µs) is proportional to the object's distance as depicted in the following Figure 11 and Equation (3).

$$\text{distance [cm]} = \frac{t}{58} \tag{3}$$

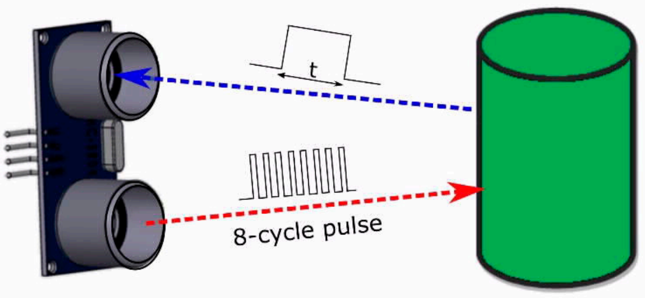


Figure 11. Distance sensing principle with the ultrasonic sensor.

3.2.3. Implementation

Figure 12 shows the use-case diagram for the system summarizing its functionality. The first step involves the camera obtaining a frame. In parallel, the SoM's integrated accelerometer determines if the user is moving. If so, it sets the Jetson nano SoM to normal power consumption mode (namely, 10 W); otherwise, it will remain in the default low power consumption mode (namely, 5 W). Once a frame has been captured, the Faster R-CNN analyses it looking for objects defined in the database. In case one or more detections turn positive, rectangles will be drawn accordingly. Next, the Cartesian quadrant algorithm positions the object(s) in the frame and triggers the ultrasonic sensor to provide the distance value. Finally, the speaker conveys the audible information to the user.

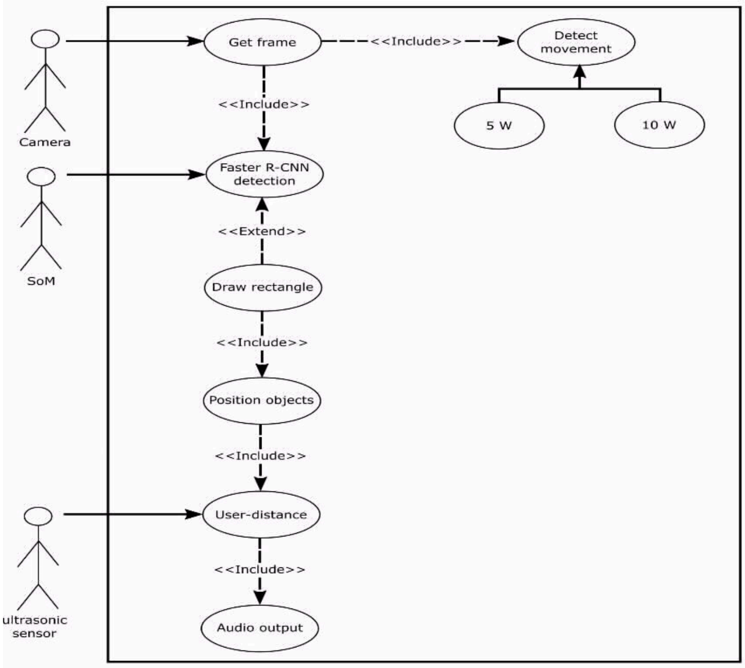


Figure 12. The use-case diagram for the AT device.

Figure 13 shows the activity diagram representing the activity flow of the AT software. In this diagram, ten activities can be identified.

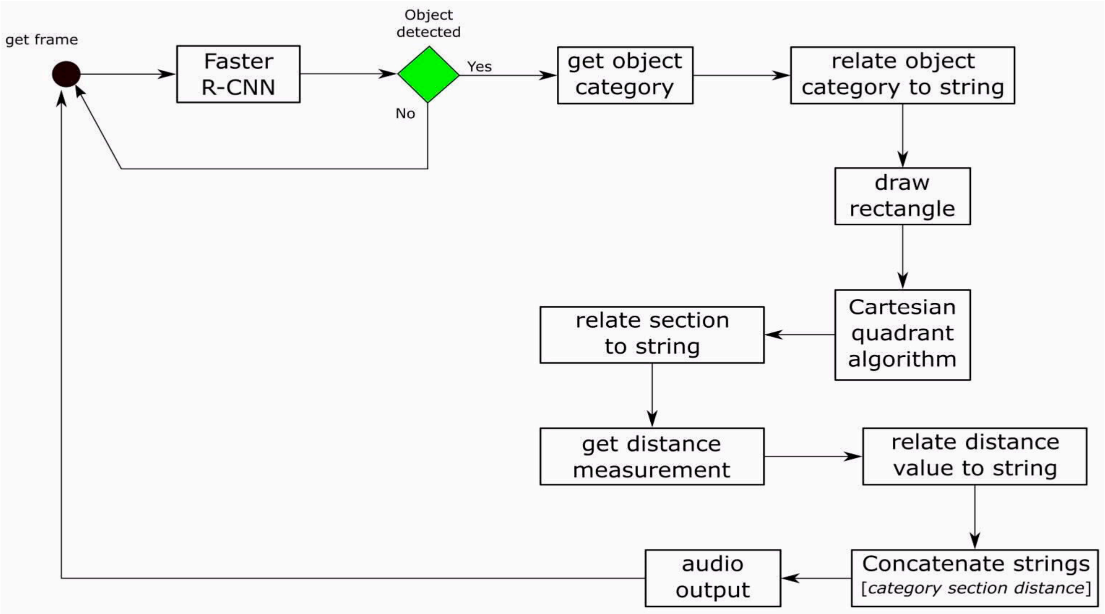

get frame
Faster R-CNN
Object detected
Yes
No
get object category
relate object category to string
draw rectangle
Cartesian quadrant algorithm
relate section to string
get distance measurement
relate distance value to string
Concatenate strings [category section distance]
audio output

Figure 13. The activity diagram for the AT software.

Overall, once a video frame is obtained, it is analyzed by the Faster R-CNN architecture. If no object is found, the system proceeds to analyze the next frame. If an object was found, its object category is determined using the image dataset. A string consisting of the object's name is then associated to it. Next, a rectangle enclosing the object is drawn and the coordinates are analyzed by the Cartesian quadrant algorithm. The resulting section is then related to a string (see Figure 10). The distance measurement is then obtained from the ultrasonic sensor and it is associated to a string containing the distance values in steps of 0.5 m (i.e., 0.5, 1, 1.5, . . . , 4 m). All three strings (object category, section, and distance value) are concatenated into one, building in this way the descriptive sentence. Finally, it is conveyed to the user via the integrated speaker.

4. Experimental Results

4.1. Training of the Designed Object Detection System

As mentioned in Section 3.2, the COCO dataset was used to train the Object detection module. We used a frozen inference graph model, which is an already COCO-trained Faster R-CNN architecture. This model serves as the basis of the dataset configuration module (see Figure 5). Model files can be downloaded from [63] and contain COCO's 91 object categories (as shown in Table 3). Custom objects can also be added to the dataset configuration module. For this project, six extra categories were considered: power outlet, doorknob, switch, white cane, and a group of eight human subjects (which can be easily increased to a few tens depending on the specific application). A total of 600 images (approximately 100 for each object category plus 200 images in total for the 8 human subjects) were incorporated into the general database.

Training of these extra twelve objects was externally conducted using a desktop computer with the following hardware specifications: 8 Gb RAM, AMD Ryzen 5 2500 U processor, and NVIDIA GTX 1050 with 4 GB of VRAM Graphics board. Using this standard computing equipment, a total of 200,000 iterations were performed taking approximately 20 h. The Faster R-CNN training configuration parameters are shown in Table 4, with a detailed description of their meaning. These values are normally used for optimally training this kind of deep learning-based structures [17].

Both COCO and custom files are then uploaded to the SoM module. Files cannot be modified once in the SoM, they are only for execution. This poses an advantage: even if the SoM's battery runs out, trained files remain ready to be used once power is reestablished.

Table 3 summarizes the object detection capabilities of the proposed AT device.

Table 3. Objects detected by the AT solar-powered device: 103 in total.

Frozen Inference Graph Model (COCO Dataset)					Custom Objects
apple	carrot	giraffe	person	suitcase	doorknob
backpack	cat	hair brush	pizza	surfboard	power outlet
banana	cell phone	hair drier	plate	teddy bear	switch
baseball bat	chair	handbag	potted plant	tennis racket	white cane
baseball glove	clock	hat	refrigerator	tie	human subject 1
bear	couch	horse	remote	toaster	human subject 2
bed	cow	hot dog	sandwich	toilet	human subject 3
bench	cup	keyboard	scissors	toothbrush	human subject 4
bicycle	desk	kite	sheep	traffic light	human subject 5
bird	dining table	knife	shoe	train	human subject 6
blender	dog	laptop	sink	truck	human subject 7
boat	donut	microwave	skateboard	TV	human subject 8
book	door	mirror	skis	umbrella	
bottle	elephant	motorcycle	snowboard	vase	
bowl	eye glasses	mouse	spoon	window	
broccoli	fire hydrant	orange	sports ball	wine glass	
bus	fork	oven	stop sign	zebra	
cake	Frisbee	parking meter	street sign	suitcase	
car					

Table 4. Training configuration parameters used for the Faster R-CNN.

Training Parameter	Value	Description
Total trained classes	103	The total number of objects (91 of the COCO dataset plus 12 custom).
Dimensions (pixels)	1024 × 768	Resolution of the images used for the training phase.
Batch size (images)	12	The number of samples processed before the model is updated.
Optimizer	Adaptive Moment Estimation (Adam)	Optimizers are methods used to configure the attributes of the neural network, such as weights and learning rate in order to reduce the losses.
Learning rate (first 60 k mini-batches)	0.0002	The learning rate is a hyper parameter that controls the model's response face to the estimated error each time that the model weights are updated.
Learning rate (60 k–120 k mini-batches)	0.00002	
Weight decay	0.0005	Weight decay allows the neural network to decrease its complexity and to reduce the training time without penalizing the detector's accuracy.
Intersection over Union (IoU)	0.55	IoU is an evaluation metric used to measure the accuracy of an object detector on a particular dataset.
Dropout	FALSE	Dropout is a regularization method that approximates training a large number of neural networks with different architectures in parallel.
Shuffle	TRUE	Shuffle is used to mix up the custom dataset, so it prevents the neural network from memorizing it.
Max detections per class	100	It is the maximum number of detection of one object (class) in one frame.
Max total detections	300	The maximum detections, of any class, allowed in one frame.

4.2. Object Recognition

One of the advantages of using deep learning-based structures is that they work in a wide range of conditions: illumination level, indoor/outdoor environments, different view angles, partial occlusions, and different object properties (dimensions, colors, textures, etc.), among others. Therefore, there is no need to have a controlled setup to test the system and obtain good results.

Figures 14 and 15 show some results on a single object and multiple-objects recognition performed by the proposed approach, respectively. In this experimental work, both indoor and outdoor scenes exhibiting different levels of illumination, object view angles, object partial occlusions, and objects properties have been considered. Note that object recognition for both COCO and custom objects is performed with high accuracy and robustness.

One of the most difficult tasks VI people face is the identification of people. The inability to recognize known individuals limit their social interactions and might put them at risk in particular situations, such as being disoriented on the street or expecting someone to meet in public spaces [64].

Among the custom objects included in the system's database, a group of eight human subjects were included. They gave explicit written consent to use their faces for all the purposes of this research according to the Universidad Panamericana ethics guidelines.

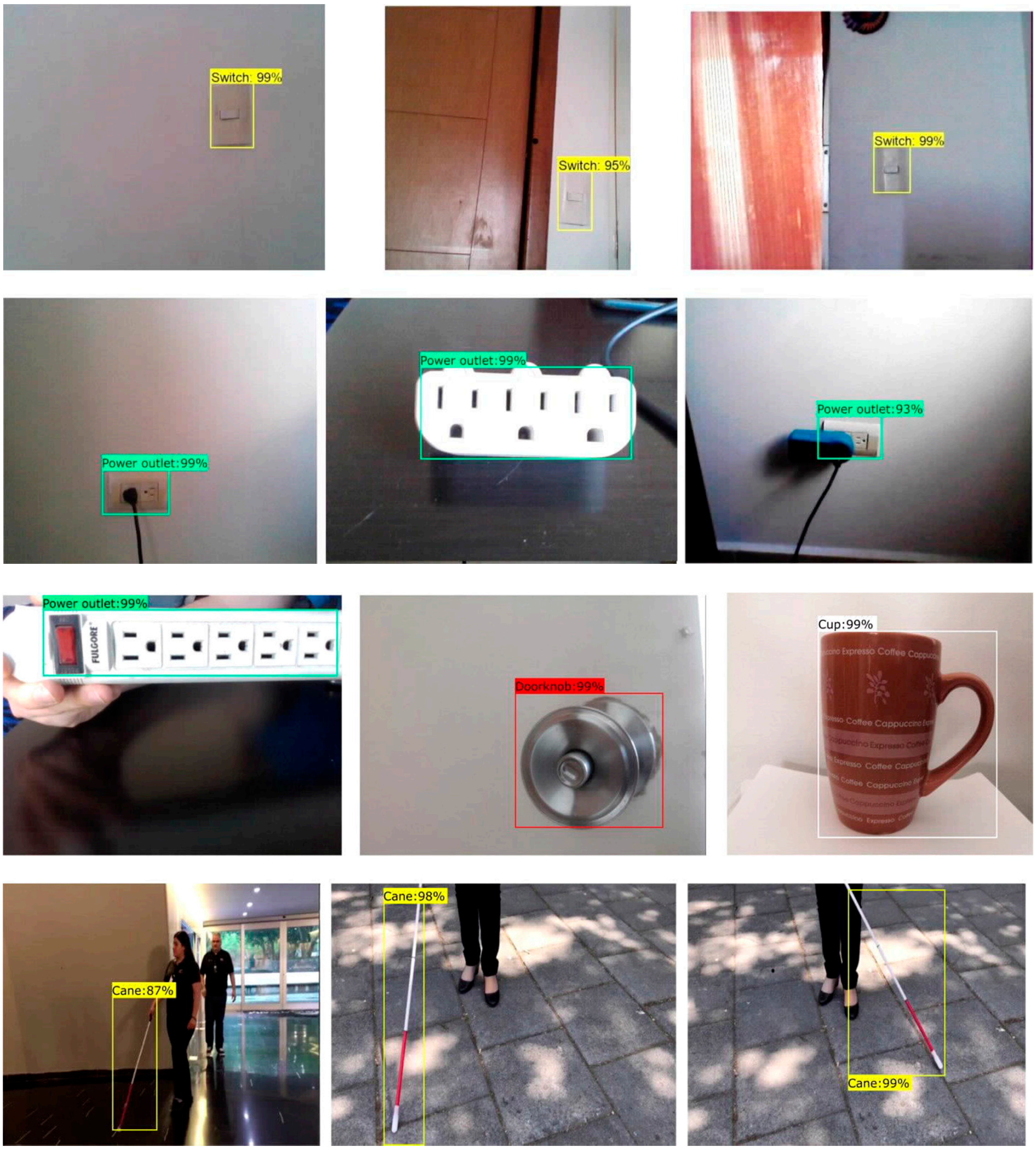


Figure 14. Selected examples of single object detection with the developed AT wearable device. Custom objects: switch, power outlet, door-knob, and white cane. COCO objects: cup.

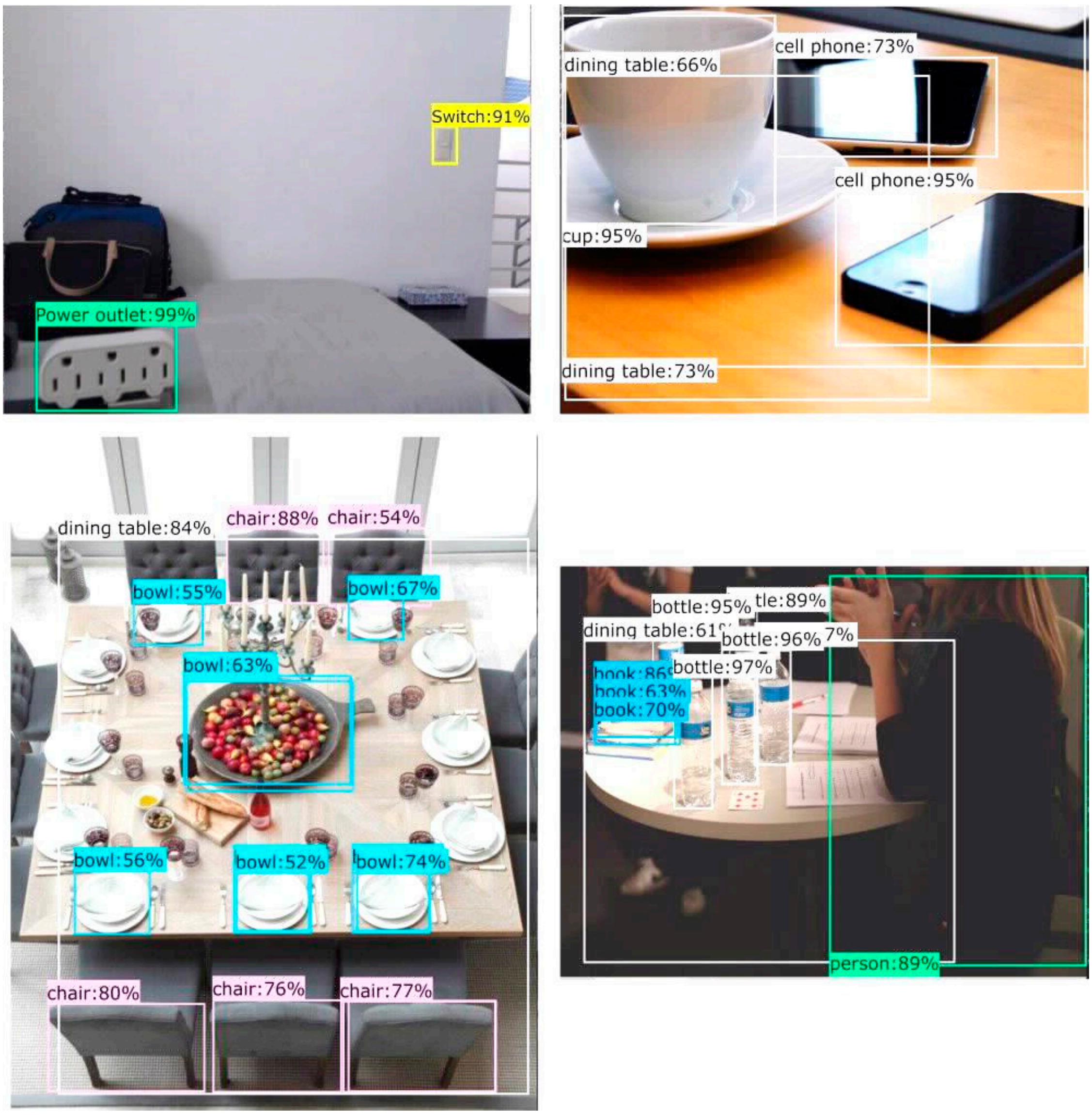


Figure 15. Selected examples of multiple object detection. Custom objects: switch and power outlet. COCO objects: cup, cell phone, dining table, chair, bowl, book, bottle, and person.

Figure 16 shows the system's performance for detecting persons in a scene and further matching their faces with a specific category. Note in the upper figure, that the system is capable of detecting (general) people in the scene. In the lower figures, the system still recognizes the human subjects but it further specifies the names of recognized human faces being included in the database (namely, Bernardo and Claudio, respectively, the human subject 2 and 4 in Table 3). Such a feature is intended to enhance VI people participation in society and to contribute to the general awareness of their context and surroundings.

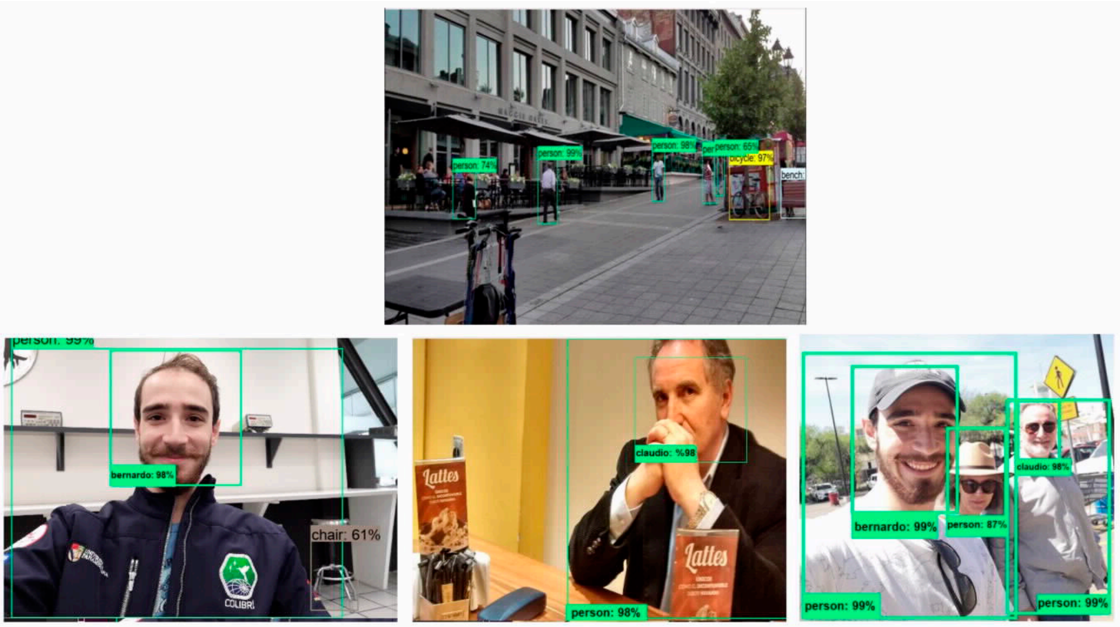


Figure 16. Selected examples of human subject detection: known and unknown individuals.

Note that the bounding boxes enclosing the recognized objects include a confidence value, i.e., the probability (in %) that the bounding box contains the target object. Figure 17 shows an example of decreasing confidence values of an object due to its progressive occlusion (from left to right in Figure 17). Note in the central image that the finger inside the bounding box area decreases the confidence value from 93% to 87%. Just as with human vision, there is a limit at which the AT device is capable of detecting an object.

As aforementioned, deep learning object recognition is robust to many scene parameters, one of them is the object view angle, particularly relevant because the camera is attached to the user glasses. As the head moves, regardless of the user being static or moving, the objects captured in the frame may not be aligned with ground's baseline. Figure 18 shows a set of examples of objects misaligned with the ground's baseline. Note that recognition is independent from the camera view angle.

Real-time video processing usually leads to some false-positive results appearing only in one or two frames during a window of time. The main cause of false positives is illumination variation [65]. To filter uncertain cases and thus reduce false positives, we established a confidence threshold of 55%. Therefore, objects below this probability value will not be taken into consideration [66,67].

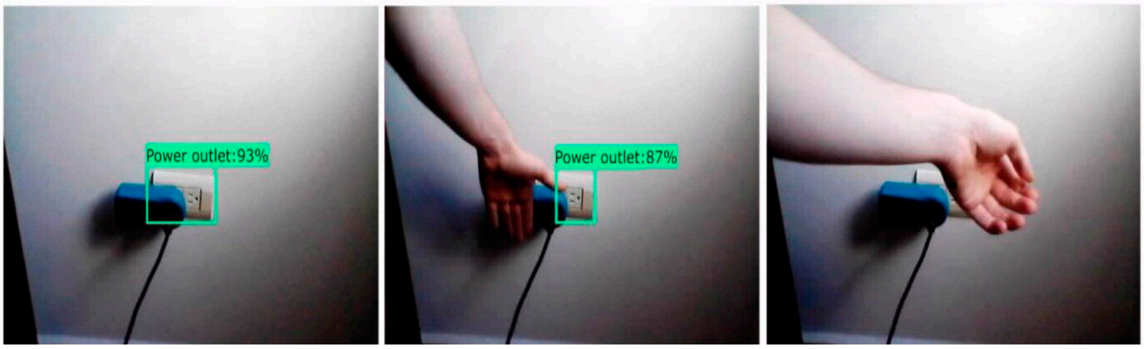


Figure 17. Example of a progressive confidence value decrease due to the object occlusion (from left to right).

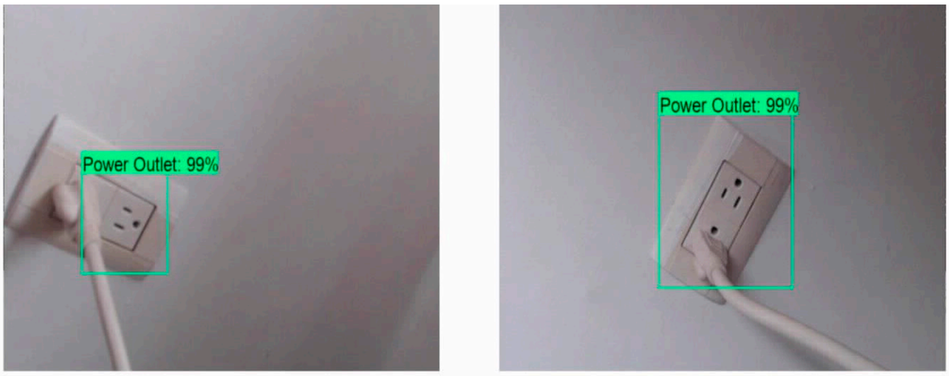
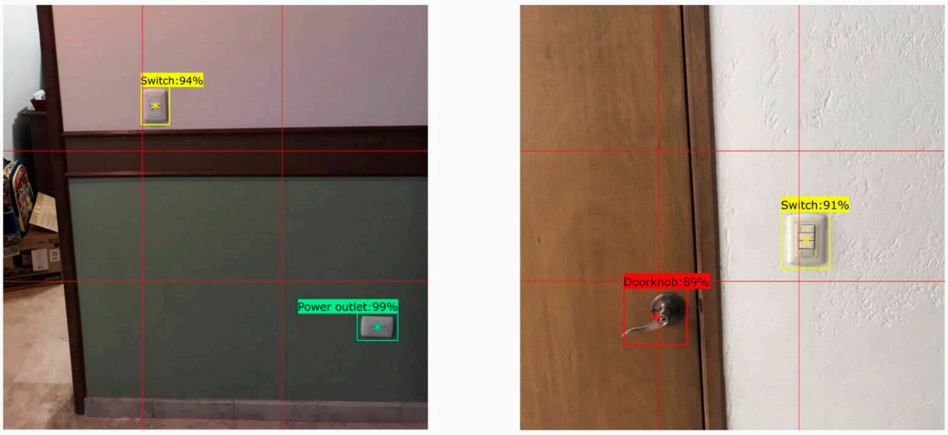


Figure 18. Example of recognition performance with different object view angles.

4.3. *Object Positioning*

Figure 19 shows two representative results of the object-positioning module. For the left Figure 19, the Cartesian quadrant algorithm locates the switch in Section II, while the power outlet in Section V (as represented in Figure 10). The ultrasonic sensor gives a lecture of 2 m to the wall containing both. For the right Figure 19, the switch is located in Section IV while the doorknob in Section VII at a distance of 1.5 m. Note the simplicity and effectiveness of the developed algorithm.



Switch **Up**, Power outlet **Right down, 2 meters**

Switch **Right**, Doorknob **Left down, 1.5 meters**

Figure 19. Example of object positioning based on Cartesian quadrants and ultrasonic sensing. The symbol * represents the CoM of the bounding box. It is taken as a decision factor and is especially useful in cases where objects span over more than one section.

Nevertheless, the ultrasonic sensing approach has its limitations specially when the objects found in the scene are not located at the same user-distance. Recall that the emitted acoustic pulse is a beam with just a 15° measuring angle. Therefore, the delivered distance is that of the object directly facing the ultrasonic sensor; the following Figure 20 shows this situation with the two recognized objects located at different distances from the user. As the user is facing the wall, the ultrasonic sensor will deliver the distance to the switch missing that of the power outlet.

To solve situations like this, we have envisaged to use two more ultrasonic sensors for a total of three to comprehensively cover the camera's FOV.

As the user freely moves his/her head or approaches the object, the item's relative position might change in the frame thus changing the section assigned by the Cartesian quadrant algorithm. This effect does not pose any problem since the audible descriptive sentences update accordingly warning the user. It is important to note that this module does not intend to provide precise positioning information that could enable an accurate grabbing of objects but to offer users a general overview of where objects are located taking as reference their gaze (the camera on the glasses moves according to the user gaze).

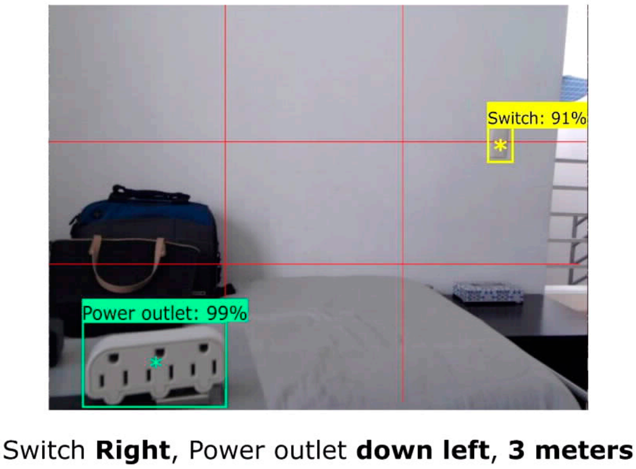


Figure 20. Example of ultrasonic sensing for objects located at different user-distances. The delivered distance value is that of the wall containing the switch.

4.4. Performance Metrics

In this sub-section, we detail the following performance metrics of interest for our AT device; namely, for overall object detection, the mean Average Precision (mAP), and the mean test-time speed. For object classification: the loss functions. Finally, for the dataset module: the confidence value according to the images used for training.

To correctly determine the mAP and mean test-time speed (i.e., the computing time) for the software execution (object recognition and object positioning), the SoM-based AT device was subjected to operation in different operating conditions. A total of 27 different objects (the twelve custom and 21 COCO objects) were captured by the camera in different views, angles, and distances (so far up to 4 m away from the user), illumination levels, as well as in different environmental conditions (outside and inside a house). An ensemble of 8100 images were processed in real-time. Figure 21 shows a boxplot analysis summarizing the obtained results. The mAP can be estimated at 86% (Figure 21a), while the mean test-time speed at 215 ms (which corresponds to an update rate for the position information equals to 4.65 Hz), specifically for the operating mode (i.e., *mode 0*) with higher clock frequency (Figure 21b). These numbers allow us to confirm that an accurate and reliable real-time object recognition can be performed with the proposed approach.

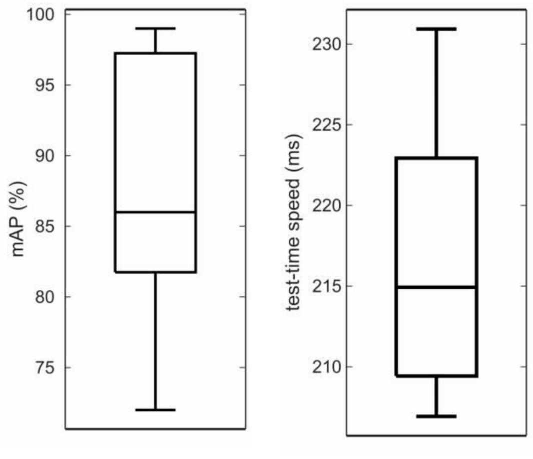
To verify that the system produces stable and consistent results over time, a test-retest reliability analysis was performed by repeating the same task a week later, in the same environmental and operating conditions. For this purpose, the ensemble of 8100 images were stored and fed again to the SoM-based processing unit. Both mAP and mean-test time were measured.

The Pearson correlation results indicate a reliability coefficient of r = 0.947 and r = 0.917 for mAP and mean-test time, respectively, thus confirming a high positive association between the scores. It can be therefore concluded that the AT designed system for VI people produces consistent results.

Figure 22 shows a set of loss functions representing the inaccuracy of predictions during the classification process of the classification layer (Figure 9). Figure 22a shows the loss function for the initial classification, i.e., object recognition alone. Note that inaccuracy rapidly decreases upon the number of iterations. The 200,000 iterations performed for the six custom images (Section 4.1) allow us to expect an inaccuracy of just 2.5%; regarding this, referring to Figure 22a, note that for the value of 200 K iterations (X axis), we obtained an inaccuracy value of 0.025, thus 2.5%. For greater clarity with reference to Figure 22, the inaccuracy value (i.e., "*Loss Function*") comes from 1 (namely, 100%) during the first iterations but rapidly decreases, as the network is learning, as the number of iterations increases (X axis).

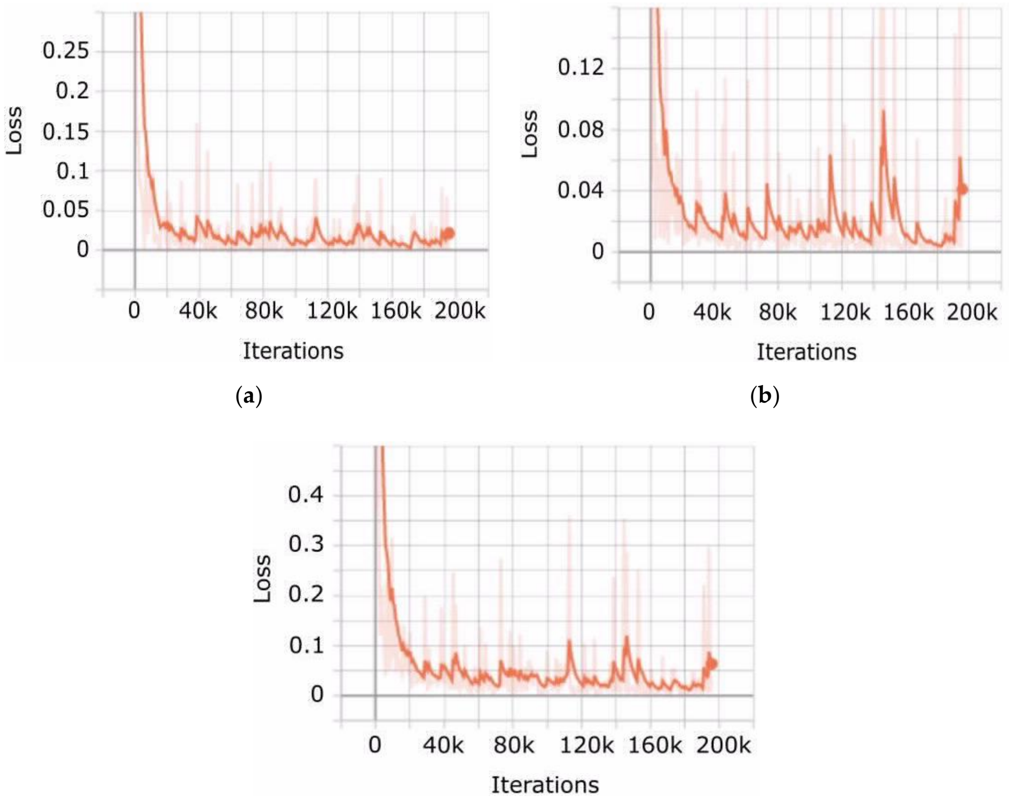
Similarly, Figure 22b shows the loss function for localization, i.e., the bounding boxes enclosing the objects. A 4% inaccuracy in bounding boxes tracing can be expected. Figure 22c shows the loss function for the whole classification layer, i.e., object recognition and localization. A 6% error is observed over the number of iterations. Note that these loss functions guarantee accurate and reliable object detection.

Figure 23 shows the confidence value evolution as a function of the number of images used for training in the dataset configuration module. A logarithm behavior is observed. Note that 12 images already guarantee the 55% detection threshold and can be used for fast training. The 100 images used for training each custom objects achieve an accuracy of 89%.



(**a**) (**b**)

Figure 21. Boxplot analysis for (**a**) the mAP and (**b**) the mean test-time speed.



(**a**) (**b**)

(**c**)

Figure 22. Loss functions (%) for the classification layer as function of the number of iterations: (**a**) object classification, (**b**) localization, and (**c**) overall layer classification.

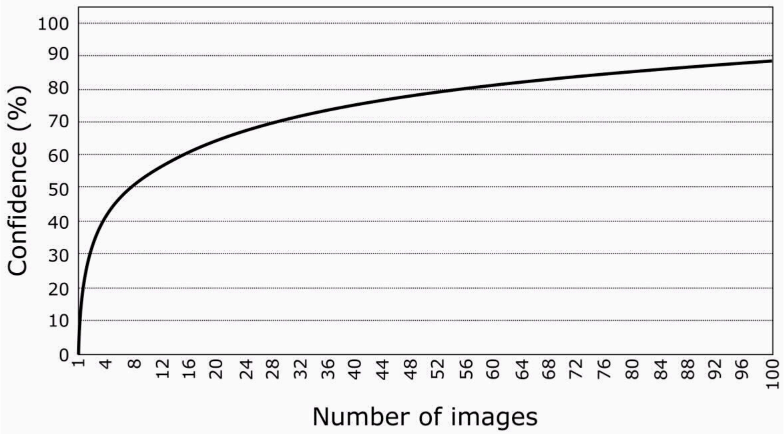


Figure 23. The confidence value as a function of images used for training one custom object.

5. Results Discussion

Experimental results of the proposed system confirm an accurate real-time object recognition with an 86% mean average precision (mAP) and 215 ms mean computing time in the case of *mode 0* operation modality with high-frequency clock (value that increases up to 360 ms in case of low-speed *mode 1* enabling). The overall object classification loss function shows that the training process was performed correctly and that classification inaccuracies can be expected in the order of 6%.

A direct comparison with systems already reported in the literature targeting the assistance of the visually impaired is not evident since most of them miss to report their mAPs [29,30,32]. Other object recognition works [44,68,69] for different applications report comparable mAPs between 70% and 90% exploring other deep learning approaches, such as SSD (Single Shot MultiBox Detector), YOLO (You Only Look Once), and R-FCN (Region-based Fully Convolutional Networks) with the use of other image databases for training, such as the PASCAL VOC (Visual Object Classes), SUN (Scene UNderstanding), and the KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute) one.

Nevertheless, our mAP agrees with those reported in [70,71]. In the first, an 81.09% mAP was obtained for moving vehicle detection. In the second scientific work, the authors report an 86.67% mAP for object detection in sports images. Both explore the same couple of this work: Faster R-CNN and the COCO dataset; however, their computing platforms are not a compact solution as the Jetson Nano embedded device employed in this work. The study of the confidence value evidently confirms that the more images involved in the network training, the better. For the AT device network architecture, full trainings guaranteeing 89% recognition rates can be achieved with 100 images per object while fast training conducted with only 12 images can ensure the minimum fixed confidence threshold of 55%.

Future work will explore other computer vision techniques, such as multi-scale analysis, wavelets, image segmentation, multi-resolution, and pyramids, to improve even further the current mAP. These methods will certainly require other types of hardware to perform real-time detection. A comprehensive assessment involving mAP, computing time, hardware's wearability, and cost will be performed to determine the best option.

The proposed system is ready for user evaluation. Current work focuses on the design of the user tests. Their objective will be to assess how subjects make use of the information provided by the AT device. In particular, we seek to determine the appropriate audible feedback refresh rate, i.e., how often it is pertinent to convey the descriptive sentences to the user. Our previous works with AT for visually impaired people [5,7,10] has given interesting insights on how information must be conveyed to the users: it must be simple and not continuous, because it will surely overwhelm and cognitively fatigue users, which might finally reject the device regardless of its technological advancements and benefits. In addition, we will investigate the pertinence of the object-positioning module. We are interested in determining whether the 2D space general information available from images plus the distance to the object obtained from the ultrasonic sensor are sufficiently useful for users to interact with the environment. The possibility of incorporating two more ultrasonic sensors will be studied to cover a wider range similar to that of the camera.

In Table 5, the comparison between the proposed AT solar-powered wearable device with other visual recognition systems, previously presented in Section 2, is reported; specifically, the main features considered for the comparison are the mean Average Processing (mAP), the mean computing time, the complexity of the proposed solution, both taking into account the computational load required for object detection and the resources for training the recognition algorithm, and finally the cost of the proposed solution. In particular, our proposed solution obtains an optimum mAP (i.e., 86%) and processing time (i.e., 215 ms) compared to other similar systems requiring high-performance and high-frequency processors hardly applicable for a wearable system. For instance, the system in [26] can support a 30 frame/s frame rate only for a high-performance computer with 2.4 GHz Intel Core i7 processor and 32 GB RAM; these conditions make the integration of the system into a portable and wearable device very difficult, involving high implementing costs.

Table 5. Comparison between the proposed object detection system and other similar works reported in the scientific literature.

Solution	mAP [%]	Mean Computing Time [ms]	Complexity	Cost
L.B. Neto [24]	94	2.4	High	High
S. Lu [38]	88	22	High	High
U. Malūkas [40]	96	105	High	High
K. Jayakanth [41]	100	451	Medium-High	High
Our system	86	215	Low	Low

6. Conclusions

This paper has presented an innovative and fully operational solar-powered wearable AT device to assist VI people in finding daily used objects in the nearby space. All the aspects of the new system have been analyzed: (i) a deep and extensive state-of-the-art analysis with a comparison with similar devices and techniques which allowed to appreciate the accuracy and cost-effectiveness of the proposed solution, (ii) performances of adopted deep learning technique (Faster R-CNN), (iii) possibility of extending the system energy autonomy by using a wearable solar energy harvesting system for power supplying the designed AT device, (iv) the AT device performances related to single and multiple objects recognition, so demonstrating that the implemented system is robust to many scene parameters and may detect with high accuracy the real-time images of an extended library.

The system is composed of a miniature low-cost camera, a system on module (SoM) embedded board, and an ultrasonic sensor. The camera is placed on the user's eyeglasses and captures real-time video of the surrounding environment. The SoM can be attached at the waist level as a belt and processes the images sent by the employed low-cost camera. The ultrasonic sensor is placed on the same belt at the buckle level to provide the distance to an object. The battery pack and the electronic section of the wearable solar energy harvesting system, useful for increasing the autonomy of the designed AT device, have been placed into pockets realized in the internal part of the jacket.

The algorithms running on the SoM are capable of detecting objects found in everyday scenes and positioning them in the Cartesian space through a specific introduced methodology that divides the analyzed image into nine quadrants. The feedback to the user consists of audible descriptive sentences involving the object(s) detected, its (their) position(s) in the field of view of the camera, and its (their) user-distance.

The software's architecture exhibits three main modules: dataset configuration, object detection, and object positioning modules. The dataset configuration module contains previously trained Faster R-CNNs encompassing the 91 objects of the COCO dataset and it is also capable of hosting Faster R-CNN trained with custom images. The object detection module consists of a Faster R-CNN processing the real-time video coming from the camera. It is responsible for detecting and recognizing the objects for the VI. Finally, the object-positioning module uses a Cartesian quadrant algorithm to position the objects found and determine their distance, and conveys the appropriate audible messages to the user. All major components of the architecture have been thoughtfully described.

The final results obtained with the proposed solar-powered device and presented in the manuscript are very encouraging; in fact, with complete training done on 100 images, it is possible to reach 89% recognition accuracy, while with reduced training done on only 12 images, it is possible, however, to reach a recognition accuracy value as high as 55%. Moreover, the developed recognition system is equipped with a triaxial MEMS accelerometer, able to detect the user condition (stationary or moving), to dynamically adapt the SoM power consumption (low-power *mode 1* or high-power *mode 0*, respectively). In this way, a reduction of the system energy consumption is obtained leaving unaltered the functionality; after this power consumption analysis and optimization, the on-field tests have demonstrated that the designed energy harvesting section, based on wearable flexible solar panels, significantly increases the device energy autonomy up to 6.1 h (outdoor scenario).

Finally, our vision is not to implement just a laboratory prototype. We envisage a technological transfer that can assist the VI population, also considering the proposed device economy (total cost of only 200 USD), long energy autonomy, and ease of wearing. Future work will include the deployment of a website where the already trained Faster R-CNN files can be downloaded by users. Different categories, such as home appliances, office equipment, cleaning products, and clothes, among many others, will be ready for purchase and upload to the AT device.

Author Contributions: Conceptualization, R.V. and P.V.; methodology, B.C. and R.V.; software, B.C.; validation, B.C.; formal analysis, C.D.-V.-S.; resources, R.d.F. and N.I.G.; data curation, P.V.; writing—original draft preparation, R.V. and P.V.; writing—review and editing, R.d.F., N.I.G. and P.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

1. World Health Organization. Fact Sheet on Blindness and Vision Impairment (October 2019). Available online: https://www.who.int/en/news-room/fact-sheets/detail/blindness-and-visual-impairment (accessed on 2 October 2020).
2. Bourne, R.R.; Flaxman, S.R.; Braithwaite, T.; Cicinelli, M.V.; Das, A.; Jonas, J.B.; Keeffe, J.; Kempen, J.H.; Leasher, J.; Limburg, H.; et al. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: A systematic review and meta-analysis. *Lancet Glob. Health* **2017**, *5*, 888–897. [CrossRef]
3. Velazquez, R. Wearable Assistive Devices for the Blind. In *Wearable and Autonomous Biomedical Devices and Systems for Smart Environment: Issues and Characterization*; Lay-Ekuakille, A., Mukhopadhyay, S.C., Eds.; LNEE, 75; Springer: Berlin/Heidelberg, Germany, 2010; pp. 331–349, ISBN 978-3-642-15687-8.
4. National Academies of Sciences, Engineering, and Medicine. *Making Eye Health a Population Health Imperative: Vision for Tomorrow*; The National Academies Press: Washington, DC, USA, 2016.
5. Velazquez, R.; Fontaine, E.; Pissaloux, E. Coding the Environment in Tactile Maps for Real-Time Guidance of the Visually Impaired. In Proceedings of the IEEE International Symposium on MicroNanoMechanical and Human Science, Nagoya, Japan, 5–8 November 2006; pp. 1–6.
6. Bologna, G.; Deville, B.; Diego-Gomez, J.; Pun, T. Toward Local and Global Perception Modules for Vision Substitution. *Neurocomputing* **2017**, *74*, 1182–1190. [CrossRef]
7. Velazquez, R.; Pissaloux, E.; Rodrigo, P.; Carrasco, M.; Giannoccaro, N.I.; Lay-Ekuakille, A. An Outdoor Navigation System for Blind Pedestrians Using GPS and Tactile-Foot Feedback. *Appl. Sci.* **2018**, *8*, 578. [CrossRef]
8. Real, S.; Araujo, A. Navigation Systems for the Blind and Visually Impaired: Past Work, Challenges, and Open Problems. *Sensors* **2019**, *19*, 3404. [CrossRef]
9. Bhowmick, A.; Hazarika, S.M. An insight into assistive technology for the visually impairedand blind people: State-of-the-art and future trends. *J. Multimodal User Interfaces* **2017**, *11*, 1–24. [CrossRef]

10. Velazquez, R.; Hernandez, H.; Preza, E. A Portable Piezoelectric Tactile Terminal for Braille Readers. *Appl. Bionics Biomech.* **2012**, *9*, 45–60. [CrossRef]
11. Neto, R.; Fonseca, N. Camera Reading for Blind People. *Procedia Technol.* **2014**, *16*, 1200–1209. [CrossRef]
12. Oproescu, M.; Iana, G.; Bizon, N.; Novac, O.C.; Novac, M.C. Software and Hardware Solutions for Using the Keyboards by Blind People. In Proceedings of the International Conference on Engineering of Modern Electric Systems, Oradea, Romania, 13–14 June 2019; pp. 25–28.
13. Watanabe, T.; Kaga, H.; Shinkai, S. Comparison of Onscreen Text Entry Methods when Using a Screen Reader. *IEICE Trans. Inf. Syst.* **2018**, *101*, 455–461. [CrossRef]
14. Rawat, W.; Wang, Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Comput.* **2017**, *29*, 2352–2449. [CrossRef]
15. Jmour, N.; Zayen, S.; Abdelkrim, A. Convolutional Neural Networks for Image Classification. In Proceedings of the International Conference on Advanced Systems and Electric Technologies, Hammamet, Tunisia, 22–25 March 2018; pp. 397–402.
16. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
17. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
18. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
19. Meshram, V.V.; Patil, K.; Meshram, V.A.; Shu, F.C. An Astute Assistive Device for Mobility and Object Recognition for Visually Impaired People. *IEEE Trans. Hum.-Mach. Syst.* **2019**, *49*, 449–460. [CrossRef]
20. Krishnan, A.; Deepakraj, G.; Nishanth, N.; Anandkumar, K.M. Autonomous walking stick for the blind using echolocation and image processing. In Proceedings of the 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), Noida, India, 14–17 December 2016; pp. 13–16.
21. Cardin, S.; Thalmann, D.; Vexo, F. A wearable system for mobility improvement of visually impaired people. *Vis. Comput.* **2007**, *23*, 109–118. [CrossRef]
22. Chen, S.; Yao, D.; Cao, H.; Shen, C. A Novel Approach to Wearable Image Recognition Systems to Aid Visually Impaired People. *Appl. Sci.* **2019**, *9*, 3350. [CrossRef]
23. Li, B.; Zhang, X.; Munoz, J.P.; Xiao, J.; Rong, X.; Tian, Y. Assisting blind people to avoid obstacles: An wearable obstacle stereo feedback system based on 3D detection. In Proceedings of the 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO), Zhuhai, China, 6–9 December 2015; pp. 2307–2311.
24. Neto, L.B.; Grijalva, F.; Maike, V.R.M.L.; Martini, L.C.; Florencio, D.; Baranauskas, M.C.C.; Rocha, A.; Goldenstein, S. A Kinect-Based Wearable Face Recognition System to Aid Visually Impaired Users. *IEEE Trans. Hum.-Mach. Syst.* **2017**, *47*, 52–64. [CrossRef]
25. Katzschmann, R.K.; Araki, B.; Rus, D. Safe Local Navigation for Visually Impaired Users With a Time-of-Flight and Haptic Feedback Device. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 583–593. [CrossRef]
26. Chen, T.; Ravindranath, L.; Deng, S.; Bahl, P.; Balakrishnan, H. Glimpse: Continuous, Real-Time Object Recognition on Mobile Devices. In Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems, New York, NY, USA, 1–4 November 2015; pp. 155–168.
27. Viola, P.; Jones, M.J. Robust Real-Time Face Detection. *Int. J. Comput. Vis.* **2004**, *57*, 137–154. [CrossRef]
28. Jauregi, E.; Lazkano, E.; Sierra, B. Approaches to Door Identification for Robot Navigation. In *Mobile Robots Navigation*; Barrera, A., Ed.; InTech: London, UK, 2010; ISBN 978-953-307-076-6.
29. Niu, L.; Qian, C.; Rizzo, J.-R.; Hudson, T.; Li, Z.; Enright, S.; Sperling, E.; Conti, K.; Wong, E.; Fang, Y. A Wearable Assistive Technology for the Visually Impaired with Door Knob Detection and Real-Time Feedback for Hand-to-Handle Manipulation. In Proceedings of the IEEE International Conference on Computer Vision Workshops, Venice, Italy, 22–29 October 2017; pp. 1500–1508.
30. Panchal, A.; Varde, S.; Panse, M. Character Detection and Recognition System for Visually Impaired People. In Proceedings of the IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology, Bangalore, India, 20–21 May 2016; pp. 1492–1496.
31. Jabnoun, H.; Benzarti, F.; Amiri, H. Object Recognition for Blind People based on Features Extraction. In Proceedings of the International Image Processing, Applications and Systems Conference, Sfax, Tunisia, 5–7 November 2014; pp. 1–6.

32. Ciobanu, A.; Morar, A.; Moldoveanu, F.; Petrescu, L.; Ferche, O.; Moldoveanu, A. Real-Time Indoor Staircase Detection on Mobile Devices. In Proceedings of the International Conference on Control Systems and Computer Science, Bucharest, Romania, 29–31 May 2017; pp. 287–293.
33. Nascimento, J.C.; Marques, J.S. Performance Evaluation of Object Detection Algorithms for Video Surveillance. *IEEE Trans. Multimed.* **2016**, *8*, 761–774. [CrossRef]
34. Hernandez, A.C.; Gómez, C.; Crespo, J.; Barber, R. Object Detection Applied to Indoor Environments for Mobile Robot Navigation. *Sensors* **2016**, *16*, 1180. [CrossRef]
35. Li, Z.; Dong, M.; Wen, S.; Hu, X.; Zhou, P.; Zeng, Z. CLU-CNNs: Object Detection for Medical Images. *Neurocomputing* **2019**, *350*, 53–59. [CrossRef]
36. Baeg, S.; Park, J.; Koh, J.; Park, K.; Baeg, M. An Object Recognition System for a Smart Home Environment on the Basis of Color and Texture Descriptors. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, CA, USA, 29 October–2 November 2007; pp. 901–906.
37. Paletta, L.; Fritz, G.; Seifert, C.; Luley, P.; Almer, A. Visual Object Recognition for Mobile Tourist Information Systems. In Proceedings of the SPIE 5684, Multimedia on Mobile Devices, San Jose, CA, USA, 14 March 2005; pp. 190–197.
38. Lu, S.; Wang, B.; Wang, H.; Chen, L.; Linjian, M.; Zhang, X. A real-time object detection algorithm for video. *Comput. Electr. Eng.* **2019**, *77*, 398–408. [CrossRef]
39. Trabelsi, R.; Jabri, I.; Melgani, F.; Smach, F.; Conci, N.; Bouallegue, A. Indoor object recognition in RGBD images with complex-valued neural networks for visually-impaired people. *Neurocomputing* **2019**, *330*, 94–103. [CrossRef]
40. Malūkas, U.; Maskeliūnas, R.; Damaševičius, R.; Woźniak, M. Real Time Path Finding for Assisted Living Using Deep Learning. *J. Univers. Comput. Sci.* **2017**, *24*, 475–486. [CrossRef]
41. Jayakanth, K. Comparative Analysis of Texture Features and Deep Learning Method for Real-time Indoor Object Recognition. In Proceedings of the 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 17–19 July 2019; pp. 1676–1682.
42. Jabnoun, H.; Benzarti, F.; Morain-Nicolier, F.; Amiri, H. Video-based assistive aid for blind people using object recognition in dissimilar frames. *Int. J. Adv. Intell. Parad.* **2019**, *14*, 122–139. [CrossRef]
43. Kulikajevas, A.; Maskeliūnas, R.; Damaševičius, R.; Ho, E.S.L. 3D Object Reconstruction from Imperfect Depth Data Using Extended YOLOv3 Network. *Sensors* **2020**, *20*, 2025. [CrossRef] [PubMed]
44. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2016**, arXiv:1506.02640.
45. Cho, M.; Chung, T.; Lee, H.; Lee, S. N-RPN: Hard Example Learning for Region Proposal Networks. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 3955–3959.
46. Silberman, N.; Hoiem, D.; Kohli, P.; Fergus, R. Indoor Segmentation and Support Inference from RGBD Images. In Proceedings of the Computer Vision—ECCV 2012, Florence, Italy, 7–13 October 2012; Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 746–760.
47. Lai, K.; Bo, L.; Ren, X.; Fox, D. A large-scale hierarchical multi-view RGB-D object dataset. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 1817–1824.
48. ImageNet. Available online: http://www.image-net.org/ (accessed on 2 October 2020).
49. Bashiri, F.S.; LaRose, E.; Peissig, P.; Tafti, A.P. MCIndoor20000: A Fully-Labeled Image Dataset to Advance Indoor Objects Detection. *Data Brief* **2018**, *17*, 71–75. [CrossRef]
50. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
51. Velazquez, R.; Varona, J.; Rodrigo, P.; Haro, E.; Acevedo, M. Design and Evaluation of an Eye Disease Simulator. *IEEE Latin Am. Trans.* **2015**, *13*, 2734–2741. [CrossRef]
52. Meet Jetson, the Platform for AI at the Edge. Available online: https://developer.nvidia.com/embedded-computing (accessed on 2 October 2020).
53. How to Configure Youur NVIDIA Jetson Nano for Computer Vision and Deep Learning. Available online: https://www.pyimagesearch.com/2020/03/25/how-to-configure-your-nvidia-jetson-nano-for-computer-vision-and-deep-learning (accessed on 2 October 2020).
54. NVIDIA Co. Jetson Partner Supported Cameras. 2020. Available online: https://developer.nvidia.com/embedded/jetson-partner-supported-cameras (accessed on 2 October 2020).

55. NVIDIA Co. Jetson Partner Hardware Products. 2020. Available online: https://developer.nvidia.com/embedded/community/jetson-partner-products (accessed on 2 October 2020).
56. NVIDIA Jetson Linux Developer Guide: Clock Frequency and Power Management. Available online: https://docs.nvidia.com/jetson/l4t/index.html#page/Tegra%2520Linux%2520Driver%2520Package%2520Development%2520Guide%2Fclock_power_setup.html%23 (accessed on 2 October 2020).
57. Özdemir, A.T. An Analysis on Sensor Locations of the Human Body for Wearable Fall Detection Devices: Principles and Practice. *Sensors* **2016**, *16*, 1161. [CrossRef]
58. This Powerful Wearable Is a Life-Changer for the Blind. Available online: https://blogs.nvidia.com/blog/2016/10/27/wearable-device-for-blind-visually-impaired/ (accessed on 2 October 2020).
59. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In *Computer Vision—ECCV 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; LNCS, 8693; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755. ISBN 978-3-319-10601-4.
60. Getting Started with Jetson Nano Developer Kit. Available online: https://developer.nvidia.com/embedded/learn/get-started-jetson-nano-devkit (accessed on 2 October 2020).
61. Installing TensorFlow For Jetson Platform. Available online: https://docs.nvidia.com/deeplearning/frameworks/install-tf-jetson-platform/index.htm (accessed on 2 October 2020).
62. Alvarez-Pato, V.M.; Sanchez, C.N.; Dominguez-Soberanes, J.; Mendoza-Perez, D.E.; Velazquez, R. A Multisensor Data Fusion Approach for Predicting Consumer Acceptance of Food Products. *Foods* **2020**, *9*, 774. [CrossRef]
63. TensorFlow 1 Detection Model Zoo. Collection of Pre-trained Detection Models. Available online: https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf1_detection_zoo.md (accessed on 2 October 2020).
64. Pissaloux, E.; Velazquez, R. *Mobility of Visually Impaired People: Fundamentals and ICT Assistive Technologies*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2018.
65. Pissaloux, E.; Maybank, S.; Velazquez, R. On Image Matching and Feature Tracking for Embedded Systems: A State of the Art. In *Advances in Heuristic Signal Processing and Applications*; Chatterjee, A., Nobahari, H., Siarry, P., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 357–380, ISBN 978-3-642-37879-9.
66. Visconti, P.; de Fazio, R.; Costantini, P.; Miccoli, S.; Cafagna, D. Innovative complete solution for health safety of children unintentionally forgotten in a car: A smart Arduino-based system with user app for remote control. *IET Sci. Meas. Technol.* **2020**, *14*, 665–675. [CrossRef]
67. Visconti, P.; de Fazio, R.; Costantini, P.; Miccoli, S.; Cafagna, D. Arduino-based solution for in-car-abandoned infants' controlling remotely managed by smartphone application. *J. Commun. Softw. Syst.* **2019**, *15*, 89–100.
68. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A. SSD: Single Shot MultiBox Detector. In *Lecture Notes in Computer Science*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; LNCS 9905; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
69. Xue, Y.; Li, Y. A fast detection method via region-based fully convolutional neural networks for shield tunnel lining defects. *Comput. -Aided Civ. Infrastruct. Eng.* **2018**, *33*, 638–654. [CrossRef]
70. Wang, H.; Yu, Y.; Cai, Y.; Chen, X.; Chen, L.; Liu, Q. A comparative study of state-of-the-art deep learning algorithms for vehicle detection. *IEEE Intell. Transp. Syst. Mag.* **2019**, *11*, 82–95. [CrossRef]
71. Intellica Co. A Comparative Study of Custom Object Detection Algorithms. 2019. Available online: https://medium.com/@Intellica.AI/a-comparative-study-of-custom-object-detection-algorithms-9e7ddf6e765e (accessed on 2 October 2020).

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