Universidad Autónoma de San Luis Potosí
Facultad de Ingeniería
Centro de Investigación y Estudios de Posgrado

Towards An Intelligent Electric Wheelchair: Computer Vision Module

T E S I S

Que para obtener el grado de:

Maestría en Ingeniería de la Computación

Presenta:
Jesús Gerardo Torres Vega

Asesor:

Dr. Juan Carlos Cuevas Tello

San Luis Potosí, S. L. P. Septiembre de 2022

Abstract

Handicapped people represent fifteen percent of the world population. Autonomy is one of the key things that disabled people seek to have. Intelligent wheelchairs could contribute to people's autonomy by removing assistance needs and performing tasks such as autonomous navigation, real-time object detection and obstacle avoidance. The present work proposes a new intelligent electric wheelchair architecture, and then focuses only on the computer vision module. To test this module, a new dataset was created with twenty different classes, using three different vendor cameras at the same location with the same illumination conditions. We employ YOLOv4, a real-time object detector based on CNNs, installed into an embedded system, the LattePanda Alpha 864s minicomputer. Mean Average Precision was the metric used to evaluate the performance, in terms of localization and classification of objects. The iDS camera, an RGB color format and high-resolution images, demonstrated to have the best performance.

Resumen

La población que exhibe alguna discapacidad representa el quince por ciento de la población mundial. Autonomía es una palabra clave que la populación discapacitada busca tener. Las sillas de ruedas inteligentes pueden contribuir a la autonomía de las personas mediante la reducción de asistencia de terceros y desarrollando actividades como la navegación autónoma, detección de objetos en tiempo real y evasión de obstáculos. El presente trabajo propone una nueva arquitectura para una silla de ruedas inteligente, después se enfoca únicamente en el módulo de visión computacional. Para probar este módulo, un nuevo conjunto de imágenes fue creado donde aparecen veinte clases diferentes, y usando tres cámaras de diferentes marcas en la misma ubicación y condiciones de iluminación. Empleamos YOLO v4, un detector de objetos en tiempo real basado en redes neuronales convolucionales, instalado en una minicomputadora: la LattePanda Alpha 864s. Mean Average Precision fue la métrica utilizada para evaluar el desempeño, en términos de localización y clasificación de los objetos. La cámara iDS, un formato RGB a color y una alta resolución de imágenes demostraron tener el mejor desempeño.

21 de julio de 2022

DR. JUAN CARLOS CUEVAS TELLO
P R E S E N T E.

Por medio de la presente me permito informarle, que en sesión ordinaria del H. Consejo Técnico Consultivo celebrada el día 21 de julio del presente, fue analizada su petición en la cual solicitó autorización para que el **Ing. Jesús Gerardo Torres Vega** de la **Maestría en Ingeniería de la Computación**, se titule mediante la modalidad: **Publicación de Artículo en Congreso Internacional con Arbitraje o en Revista Indizada**, con el articulo denominado: **"Towards An Intelligent Electric Wheelchair: Computer Vision Module".**

Al respecto, me permito informarle que su solicitud fue aprobada de conformidad, de acuerdo a las evidencias que avalan el requerimiento de titulación antes mencionada.

Sin otro particular de momento, le reitero las seguridades de mi atenta y distinguida consideración.

"MODOS ET CUNCTARUM RERUM MENSURAS AUDEBO"

A T E N T A M E N T E

UNIVERSIDAD AUTONOMA DE SAN LUIS POTOSI FACULTAD DE INGENIERIA SECRETARIA

DR. RICARDO ROMERO MÉNDEZ
SECRETARIO DEL CONSEJO.

Copia. H. Consejo Técnico Consultivo.
*etn.

Av. Manuel Nava 8
Zona Universitaria • CP 78290
San Luis Potosí, S.L.P.
tel. (444) 826 2330 al39
fax (444) 826 2336

UASLP
Universidad Autónoma de San Luis Potosí

FACULTAD DE INGENIERÍA

21 de julio de 2022

DR. JUAN CARLOS CUEVAS TELLO
P R E S E N T E.

Por medio de la presente me permito informarle, que en sesión ordinaria del H. Consejo Técnico Consultivo celebrada el día 21 de julio del presente, fue analizada su petición en la cual solicitó autorización para que el **Ing. Jesús Gerardo Torres Vega** de la **Maestría en Ingeniería de la Computación**, se titule mediante la modalidad: **Publicación de Artículo en Congreso Internacional con Arbitraje o en Revista Indizada**, con el articulo denominado: **"Hacia una Silla de Ruedas Eléctrica Inteligente: Módulo de Visión Computacional"**

Al respecto, me permito informarle que su solicitud fue aprobada de conformidad, de acuerdo a las evidencias que avalan el requerimiento de titulación antes mencionada.

Sin otro particular de momento, le reitero las seguridades de mi atenta y distinguida consideración.

"MODOS ET CUNCTARUM RERUM MENSURAS AUDEBO"

A T E N T A M E N T E

UNIVERSIDAD AUTONOMA DE SAN LUIS POTOSI FACULTAD DE INGENIERIA SECRETARIA

DR. RICARDO ROMERO MÉNDEZ
SECRETARIO DEL CONSEJO.

Copia. H. Consejo Técnico Consultivo.
*etn.

100 ANIVERSARIO DE LA AUTONOMÍA UASLP 2023

www.uaslp.mx

Av. Manuel Nava 8
Zona Universitaria • CP 78290
San Luis Potosí, S.L.P.
tel. (444) 826 2330 al39
fax (444) 826 2336

Table of contents

Chapter 1

Paper

This chapter contains the paper presented in the 6th World Conference on Smart Trends in Systems, Security and Sustainability, WorldS4 2022, London, UK. A certificate of conference participation is provided.

CERTIFICATE

THIS IS TO CERTIFY THAT

Jesus Gerardo Torres Vega

HAS DIGITALLY PRESENTED A PAPER TITLED

TOWARDS AN INTELLIGENT ELECTRIC WHEELCHAIR : COMPUTER VISION MODULE

AT THE

6th WORLD CONFERENCE ON

SMART TRENDS IN SYSTEMS, SECURITY AND SUSTAINABILITY

ORGANIZED BY

GLOBAL KNOWLEDGE RESEARCH FOUNDATION

MIKE HINCHEY
PRESIDENT - IFIP

AMIT JOSHI
ORGANIZING SECRETARY

6TH EDITION WORLDS4 2022

NILANJAN DEY
PUBLICATION CHAIR

24-27 AUGUST, 2022
DIGITAL PLATFORM : ZOOM

XIN-SHE YANG
CONFERENCE CHAIR

GR FOUNDATION | Springer | SPRINGER NATURE | GR SCHOLASTIC | INTER VIT | ifip

Towards An Intelligent Electric Wheelchair: Computer Vision Module

Jesus Gerardo Torres-Vega, Juan C. Cuevas-Tello, Cesar Puente, Jose Nunez-Varela, and Carlos Soubervielle-Montalvo

Universidad Autonoma de San Luis Potosi, Facultad de Ingeniería, Av. Dr. Manuel Nava No. 8, Zona Universitaria, San Luis Potosi, SLP, CP 78290 México,
a232559@alumnos.uaslp.mx

Abstract. Handicapped people represent fifteen percent of the world population. Autonomy is one of the key things that disabled people seek to have. Intelligent wheelchairs could contribute to people's autonomy by removing assistance needs and performing tasks such as autonomous navigation, real-time object detection and obstacle avoidance. The present work proposes a new intelligent electric wheelchair architecture, and then focuses only on the computer vision module. To test this module, a new dataset was created with twenty different classes, using three different vendor cameras at the same location with the same illumination conditions. We employ YOLOv4, a real-time object detector based on CNNs, installed into an embedded system, the LattePanda Alpha 864s mini-computer. Mean Average Precision was the metric used to evaluate the performance, in terms of localization and classification of objects. The iDS camera, an RGB color format and high-resolution images, demonstrated to have the best performance.

Keywords: wheelchairs, autonomous driving, Convolutional Neural Networks, YOLOv4, Mean Average Precision, embedded system

1 Introduction

There are more than 1,000 million disabled people in the world, representing the 15% of the world population, from that, almost 190 million need assistance services frequently [1]. According to the 2020 population and housing census, there are around 6,179,890 disabled people living in Mexico, representing 4.9% of the entire population in the country [1]. Of this population, the 48% exhibits difficulties in walking, moving, go up or down stairs, requiring assistance from a third party, a wheelchair, a walker, or any kind of artificial device. The assistance requirements for elderly people or disabled persons have a direct impact into their autonomy, restricting them to perform simple tasks such as moving from a room to another. Mobility limitations are the most frequent disabilities in Mexico [1].

An autonomous driving system, such as an intelligent electric wheelchair, is an alternative for handicapped people, that could contribute to give them back most of their autonomy by allowing them to move with no assistance from other people. Intelligent electric wheelchairs are expected to perform autonomous navigation, following predefined or real-time trajectories, by means of non-mechanical commands from the user. Autonomous navigation requires real-time object recognition to identify obstacles in the wheelchair's trajectory. Also, computer vision could help handicapped persons to find objects such as medicines, medical devices, food or any other important objects in an indoor environment. Some solutions have been proposed using Convolutional Neural Networks (CNNs), user's gaze images, voice commands and environment images as data Input, performing better in outdoor environments, and exploring without taking care of objects around the wheelchair or along the trajectory [2, 5]. These projects have demonstrated that sensing the environment through cameras, CNNs, and object detectors [8], show a better performance in computer vision tasks for object recognition and navigation when automating wheelchairs [2, 4].

This paper proposes a new architecture for an intelligent electric wheelchair that is capable to perform autonomous navigation in an indoor environment, being aware of obstacles around it. Within the architecture, the computer vision module, together with other modules, are in charge of detecting obstacles, navigate without collide and also identify objects of interest to the person. This research focuses on the computer vision module, where a comparison among three different vendor cameras is presented using You Only Look Once v4 (YOLOv4). We use the Mean Average Precision (mAP) metric in order to evaluate the performance of the object detection task in terms of localization and classification.

The paper is organized as follows: Section 2 presents the current state of the art. Section 3 covers the general architecture and the computer vision module. Section 4 presents the YOLOv4 architecture. Section 5 describes the experimental setup. Section 6 contains the results, and Section 7 presents the conclusions and future work.

2 State-of-the-Art for Intelligent Electric Wheelchairs with Computer Vision

There are many approaches for autonomous vehicles using different techniques. This paper involves deep learning techniques for autonomous driving using cameras. For this reason, this work presents only a state-of-the-art study that includes projects related to intelligent electric wheelchairs using deep learning.

Table 1 summarizes the comparison of the related work. Through this table, one can identify several attributes of the proposed solutions: i) if the study captures pictures from the environment (images); ii) the performing environment, i.e. either outdoor or indoor; iii) if the study involves embedded systems because an electric wheelchair is an autonomous vehicle that requires specific HW; iv) if the study contains modules for object recognition or image classification; v) if the study implements obstacle avoidance techniques; vi) if it measures the performance with formal metrics. Finally, we highlight that only one work implements obstacle avoidance through ultrasonic sensors. Few research projects have developed systems capable of performing object recognition in an indoor environment. Also, few of the projects found implement embedded systems (HW) capable of performing all the required tasks. Therefore, the opportunity areas detected are related to indoor environments, embedded system platforms, object recognition and obstacle avoidance.

Table 1. Summary of the state of the art. Intelligent electric wheelchairs with deep learning and computer vision tasks.

Year	Images	Type	HW	Recognition	Obstacle Avoidance	Measures Performance	Ref.
2020	✘	Indoor	✘	✘	✔	✔	[2]
2020	✘	Indoor	✔	✘	✘	✔	[3]
2020	✔	Outdoor	✘	✔	✘	✔	[4]
2020	✘	Outdoor	✔	✘	✘	✘	[6]
2019	✔	Outdoor	✘	✔	✘	✔	[5]
2017	✔	Outdoor	✘	✔	✘	✘	[7]

3 Intelligent Electric Wheelchair Architecture

In Fig. 1 is the proposed architecture for an intelligent electric wheelchair, where all the modules communicate among them. The description of each module is as follows:

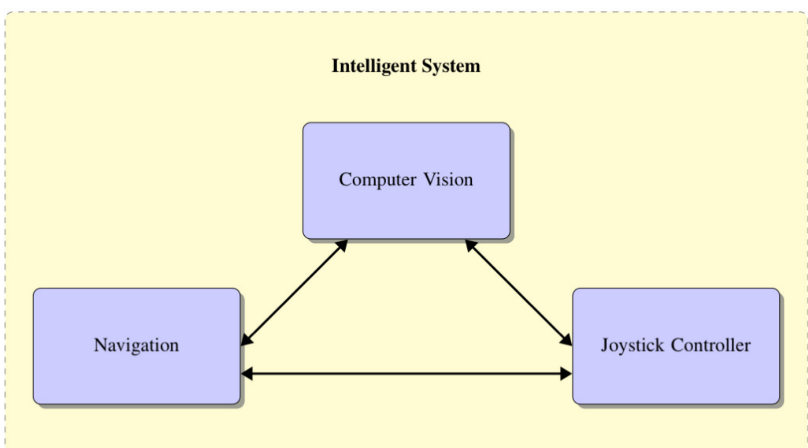


Fig. 1. Architecture of an Intelligent Electric Wheelchair

Intelligent System. This is the main module where all the decisions are taken. The input to the main module depends on the needs of disabled persons: i) a touch screen, ii) voice commands, iii) brain data (i.e. EMOTIV EPOC), iv) web interface, etc. It also receives the information from the navigation and computer vision modules, and it sends the instructions to the joystick controller.
Navigation. This module involves all related to localization of the wheelchair that is origin and destination points. At the first stage, the wheelchair will move in a controlled space. The next stage is that the wheelchair can learn the environment. Undergoing research involves reinforcement learning, so the wheelchair can adapt and create the navigation map by itself. Our goal is to use only cameras as sensors because current technology provides stereo depth, collision avoidance and positional tracking, e.g. Intel RealSense [10].
Joystick Controller. This module means automatizing the joystick by adapting servo motors, so it can be controlled automatically; see [2]. This module is also called the embedded system or hardware system. It receives the movement instructions from the main module. And it also receives information from the navigation and computer vision modules because the navigation map and the obstacle avoidance information are obtained through these modules.
Computer Vision. This module is the main topic of this paper. It requires a camera in order to obtain all the information about the environment. The camera will be located in front of the wheelchair. This module has the restriction that all algorithms must run on an embedded system where the hardware resources are limited; compared to a personal computer or a workstation. Additionally, the image processing must be in real time. The following sections explain the undergoing research, experiments and the future work.

4 Computer Vision Module

The main task of this module is to identify objects in the environment through the images obtained from a camera. Depending of the selected camera, this module can also help to the navigation module if the camera provides stereo vision, where the depth of objects could be obtained.

YOLO is one of the state-of-the-art object detectors based on CNNs, because it is one of the most accurate (Average Precision 10% higher) and fast (FPS 12% higher) [8]. We use the fourth version of YOLO, known as YOLOv4, with the following architecture [8]: i) *Backbone*: Named as CSPDarknet53, is an architecture that contains 29 convolutional layers3×3, a 725×725 receptive field and 27.6 M parameters. This architecture shows a good performance with MS COCO dataset. ii)*Neck*: Spatial Pyramid Pooling network (SPP-net) and Path Aggregation Network (PANet)[8]. iii) *Head*: Same head architecture used in YOLOv3.

5 Experimental Setup

In this section we describe the different cameras under study, the dataset and the metric used to evaluate the performance of our computer vision module.

Cameras Description. There are many factors that have a direct impact on the performance of the object detector, such factors include the image resolution, color formats, environmental conditions. Through this paper we analyze three different cameras: i) *Common Webcam* from the brand Perfect Choice was selected for this study because the low cost, availability and similarity in characteristics to other cameras; ii) *iDS Camera*, Model UI-1240SE-C-HQ developed by IDS Imaging Development Systems with CMOS color sensor and 8-bit depth RGB camera [9]; iii) *Intel RealSense T265 Camera* is based on a fusion of visual and inertial sensors, used in robotics, drones and virtual reality [10]. The tracking camera has two “fisheye” lenses, an IMU (Inertial Measurement Unit) module and an Intel-own ASIC processor for parallel processing.

Embedded System Design Platforms. The wheelchair can be seen as an autonomous vehicle, so embedded system design platforms are needed in the processing stage. None of the projects presented in the state of the art, are capable to perform object recognition in real-time using embedded systems, along with obstacle avoidance and navigation algorithms; see §2. The embedded system used across the literature is a Raspberry Pi [3], others works require the use of computers with more computational resources such as GPU, memory, CPU, etc. The hardware capacity needed to accomplish all the tasks proposed in this paper requires a more powerful embedded system platform than a Raspberry Pi. Therefore, all experiments presented here were developed using a LattePanda Alpha 864s [13].

Dataset. A dataset was created with images commonly found in indoor environments to test the object detector. To achieve reliable results, the images created had to be from a real environment with objects that are part of the pretrained classes in YOLOv4. The dataset consists of three different scenes: office, living room and dining room. First two scenes contain four different images, the third scene contains three. Each image with a different perspective at the same reference point. Since we used three different cameras, this yields a total of 33 images. We classify 20 object classes (backpack, book, bottle, bowl, cellphone, chair, cup, dining table, fork, handbag, keyboard, knife, laptop, mouse, person, potted plant, remote control, sofa, spoon and tv monitor). Fig. 2 shows an example of an office scene, where objects commonly used within an office appear such as a laptop, a cell phone, a keyboard, a mouse, a book, a backpack, a chair and a tv monitor with its remote. For each image it was necessary to generate the bounding box of each object in each image. This is our ground truth dataset, which is necessary for the mAP metric described below.

Mean Average Precision (mAP) is an evaluation metric used for object detection [11]. The object within an image is located and classified by an object detection algorithm. *Precision* is used to measure how accurate the model is, where $Precision = TP/(TP + FP)$. *Recall* is used to measure the sensibility of the model, where $Recall = TP/(TP + FN)$. *Intersection over Union*

a) b) c)

Fig. 2. Example images of the office scene: a) Webcam Perfect Choice, b) iDS, c) Intel RealSense T265. All cameras are located at the same reference point.

(IoU) measures the overlap between the bounding box of the ground truth and the predicted bounding box, where $IoU = AreaofIntersection/AreaofUnion$. If the IoU is equal to one it implies that the predicted and the ground truth bounding boxes perfectly overlap [12]. In object detection, precision and recall are calculated based on an IoU threshold. The next step is to calculate the Average Precision (AP) for each class. Once the AP is calculated for a single class, the calculations are repeated for all classes with the same IoU threshold and confidence threshold, and this experiment was repeated for all cameras. Mean Average Precision (mAP) can be calculated as the average of all the APs.

6 Results

Experiments are performed for different IoU thresholds, so we calculate mAP for $0.5, 0.55, 0.65, \ldots, 0.95$ values. Fig. 3 shows the results, where iDS camera has the best performance with mAP@0.5:0.05:0.95 = 0.70138, right away the common webcam with a value of 0.6515 and finally the Intel RealSense T265 with 0.48146. iDS camera captures higher resolution and brighter images in comparison with the common webcam. The resolution allows to detect more objects with a higher confidence and get less False Positives. Also, the CMOS color sensor with 8 bit-depth in the iDS camera represents an important feature to capture a high-quality image which can be better processed by the object detector. A color image is another important factor.

Prediction Time. Table 2 shows the average time of each camera for the prediction analysis, where the best results are in bold. As the iDS camera showed better results with mAP@0.5:0.05:0.95 = 0.70138, the average processing time is significantly higher compared with the others, such as Intel RealSense T265 camera with a low resolution and the lowest prediction average time. If the implementation of the object detector requires a higher mAP value, regardless of processing time, a color camera with a higher resolution is the best option. Nevertheless, if the processing time is an important feature of the implementation, a hardware architecture equipped with a GPU and a camera with high resolution may bring better results.

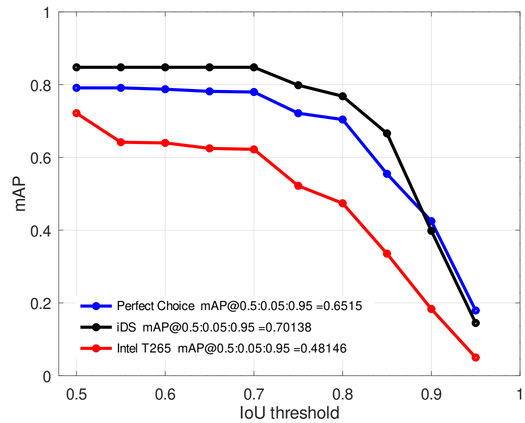


Fig. 3. Comparison of performance for three cameras analyzed. Each point is the mAP calculated at a specific IoU threshold

Table 2. Prediction average time and mAP for each camera.

Camera	**Average time (sec)**	**mAP@0.5:0.05:0.95**
Perfect Choice	13.264867	0.65150
iDS	18.685641	**0.70138**
Intel RealSense	**12.489101**	0.48146

7 Conclusions and Future Work

We introduced a new architecture towards an intelligent electric wheelchair, and we presented the results obtained for the computer vision module. Also, a new dataset was generated to measure the performance of three different cameras. Besides the images captured with cameras, it was also necessary to generate a ground truth dataset containing the bounding boxes of all objects within the images. The YOLOv4 object detector was used in this module, which is based on CNNs, and running on an embedded system platform (LattePanda). The mAP metric was used, where the iDS camera shows the best performance with mAP@0.5:0.05:0.95 = 0.70138, see Fig. 3. The iDS camera suggests that taking RGB and high-resolution images are the best characteristics to use with the object detector. On the other hand, Table 2 shows that iDS camera has the worst prediction average time. This research is not only intended to help handicapped people to retrieve part of their autonomy, but also it can be used for any autonomous systems guided by cameras. As further research, these are the research directions: i) re-train the CNN inside YOLOv4 to recognize objects that currently cannot identify; ii) real-time processing, use of different embedded system platforms such as Jetson Nano, which is another minicomputer from NVIDIA with GPU [14]; iii) exploring new technologies, including cloud computing, distributed systems and parallel computing, in order to achieve a better prediction time for the computer vision module and a remote-controlled wheelchair.

References

1. INEGI, "Población. Discapacidad." http://cuentame.inegi.org.mx/poblacion/discapacidad.aspx?tema=P (accessed Apr. 14, 2021).
2. Dahmani, M., Chowdhury2, M. E. H., Khandakar, A., Rahman, T., Al-Jayyousi, K., Hefny, A., Kiranyaz, S.: An Intelligent and Low-Cost Eye-Tracking System for Motorized Wheelchair Control. In: Sensors, vol. 20, num. 14, 2020 doi.org/10.3390/s20143936
3. Anam, S.K., Sujanarko, B.: Design of electrical wheelchair navigation for disabled patient using convolutional neural networks on Raspberry Pi 3. In: AIP Conference Proceedings, vol. 2278, no. 1, p. 020048, 2020. doi.org/10.1063/5.0014513.
4. Kai, T., Lu, H., Kamiya, T.: Object Recognition from Spherical Camera Images Based on YOLOv3. In: 2020 20th International Conference on Control, Automation and Systems (ICCAS), pp. 419–422. doi.org/10.23919/ICCAS50221.2020.9268308.
5. Sakai, Y., Nakayama, Y., Lu, Y., Li, Y., Kim, H.: Recognition of Surrounding Environment for Electric Wheelchair Based on WideSeg. In: 19th International Conference on Control, Automation and Systems (ICCAS), Oct. 2019, pp. 816–820. doi.org/10.23919/ICCAS47443.2019.8971608
6. Rakhmatulin, I., Duchowski, A.T.: Deep Neural Networks for Low-Cost Eye Tracking. In: Procedia Computer Science, vol. 176, pp. 685–694, 2020. doi.org/10.1016/j.procs.2020.09.041.
7. Nakayama, Y., Lu, H., Tan, J.K., Kim, H.: Environment recognition for navigation of autonomous wheelchair from a video image. In: 2017 17th International Conference on Control, Automation and Systems (ICCAS), 2017, pp. 1439–1443. doi.org/10.23919/ICCAS.2017.8204217.
8. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: YOLOv4: Optimal Speed and Accuracy of Object Detection. In: arXiv.org, Cornell University, 2020. http://arxiv.org/abs/2004.10934
9. iDS camera: UI-1240SE. https://es.ids-imaging.com/store/ui-1240se.html
10. Intel RealSense Tracking Camera Datasheet. https://www.intelrealsense.com.
11. Yohanandan, S.: mAP (mean Average Precision) might confuse you!. In: Towards Data Science, 2020. https://towardsdatascience.com/map-mean-average-precision-might-confuse-you-5956f1bfa9e2.
12. Khandelwal, R.,: Evaluating performance of an object detection model. In: Towards Data Science, 2020. https://towardsdatascience.com/evaluating-performance-of-an-object-detection-model-137a349c517b
13. LattePanda Alpha 864s https://www.lattepanda.com/products/lattepandaalpha-864s.html
14. Jetson Nano Developer Kit. https://developer.nvidia.com/embedded/jetson-nano-developer-kit

Chapter 2

Presentation

This chapter contains the slides of the paper presentation at the conference. Some slides were added after the previous exam to improve the presentation.

UASLP Universidad Autónoma de San Luis Potosí | Facultad de Ingeniería | Centro de Investigación y Estudios de Posgrado

Towards An Intelligent Electric Wheelchair: Computer Vision Module

Examen de Grado

Ing. Jesús Gerardo TORRES VEGA

Asesor: Dr. Juan Carlos CUEVAS TELLO

26th September 2022

1

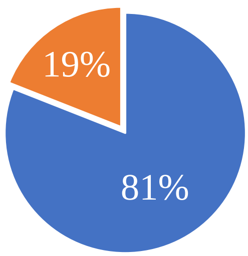
Agenda

2

Disability around the World

+1,000 million of people

Disabled World Population

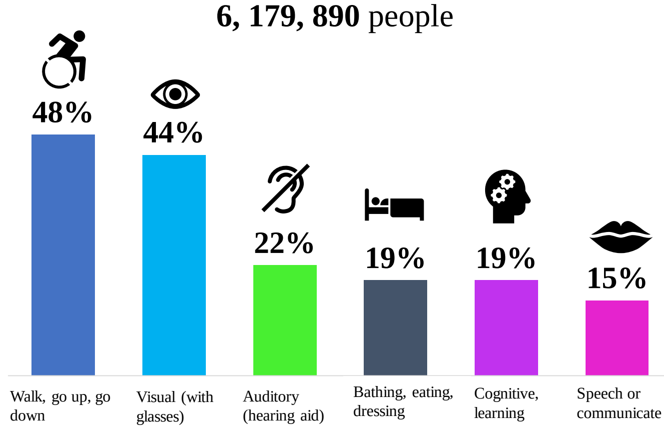


Other disabilities   Mobility disabilities

Source: INEGI, 2020 [1]

Disability in Mexico

6, 179, 890 people

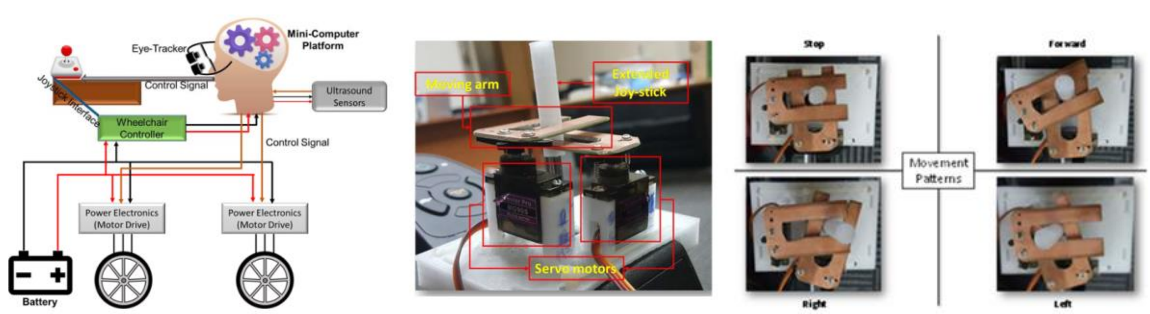


Source: INEGI. Censo de Población y Vivienda 2020 [1]

3

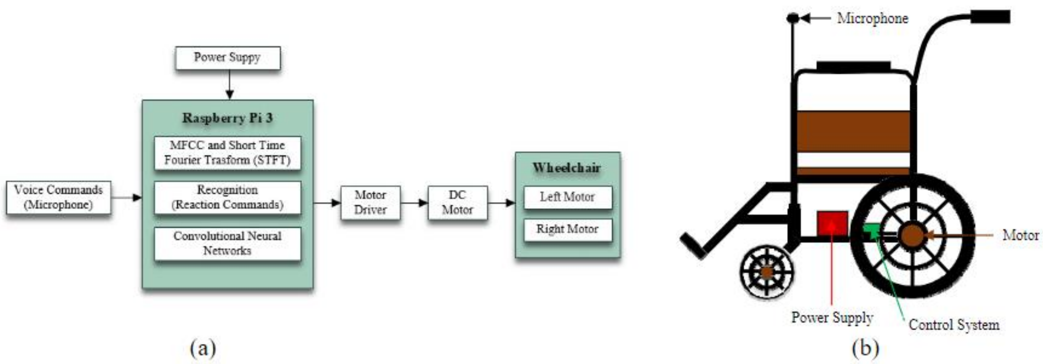
State of the art in Intelligent Electric Wheelchairs

Approach: Intelligent Electric Wheelchairs using Deep Learning Techniques



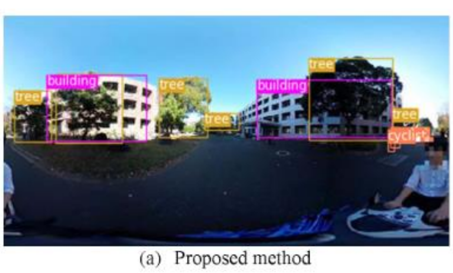
Block representation of the proposed systems for pupil based intelligent eye-tracking motorized wheelchair control [2].

4

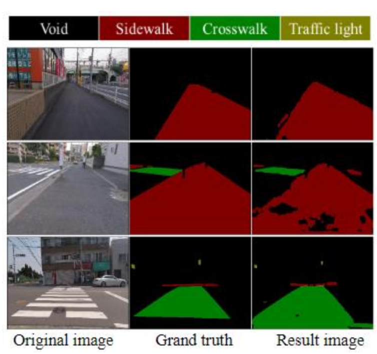


(a) Block of Wheelchair Navigation System, (b) Design of Electrical Wheelchair Navigation [3].

5



Experimental results [4]



Results of dataset from electric wheelchair [5]

Some images of datasets. The aerial image [7]

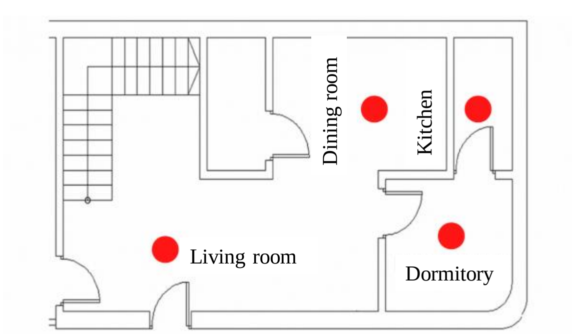
6

Approach: Intelligent Electric Wheelchairs using Deep Learning Techniques

Year	Images	Type	HW	Recognition	Obstacle Avoidance	Measures Performance	Ref.
2020	✘	Indoor	✘	✘	✔	✔	[2]
2020	✘	Indoor	✔	✘	✘	✔	[3]
2020	✔	Outdoor	✘	✔	✘	✔	[4]
2020	✘	Outdoor	✔	✘	✘	✘	[6]
2019	✔	Outdoor	✘	✔	✘	✔	[5]
2017	✔	Outdoor	✘	✔	✘	✘	[7]
2021	✔	Indoor	✔	✔	✔	✔	This

✓ Implements ✘ Does not implement

Solution Proposed



To develop an autonomous driving system:

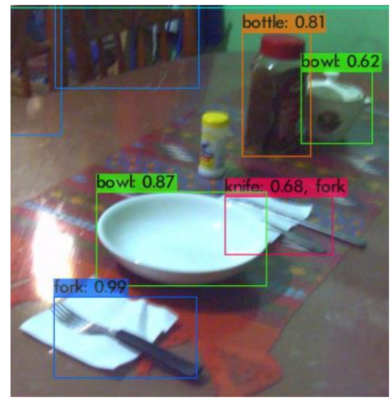
- *Intelligent Electric Wheelchair*

Benefits for the user:

✓ Give back most of their autonomy
✓ Reduce assistance needs
✓ Perform in an indoor environment

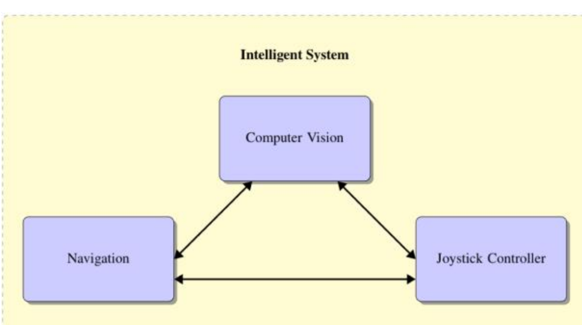
Intelligent Electric Wheelchair

- Perform autonomous navigation without collide
- Be aware of obstacles
- Modify trajectory in real time to avoid obstacles
- Detect objects of interest for the user



9

Disabled Population | State of the Art | Architecture Proposed | Computer Vision Module | mAP | Results | Conclusions



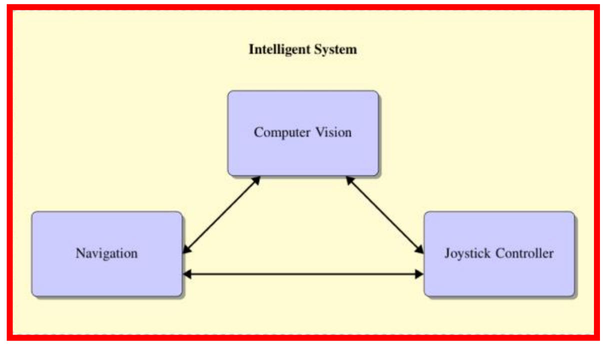
Composed by four major modules:

1. *Intelligent System*
2. *Navigation*
3. *Joystick Controller*
4. *Computer Vision*

10

Intelligent System

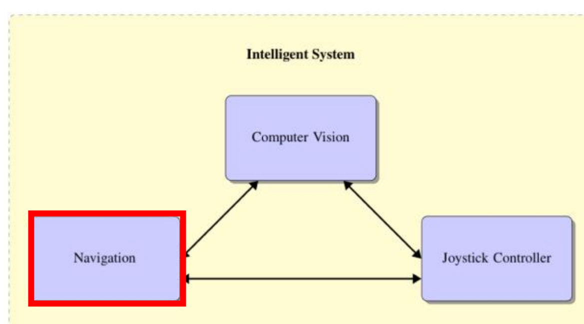
- Platform (OS) of the architecture
- User interface
- Controls data flow
- Manages instructions



11

Navigation module

- Localization of the wheelchair
- Defines trajectory
- Obstacle avoidance technique
- Learn from the environment.

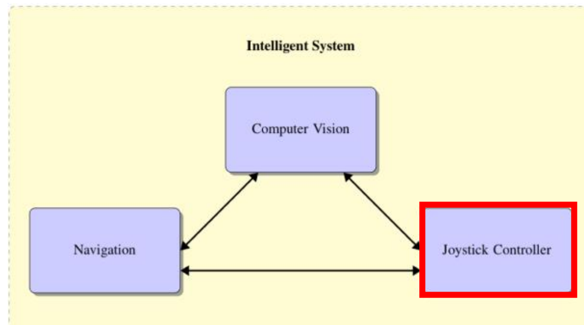


12

Joystick-controller

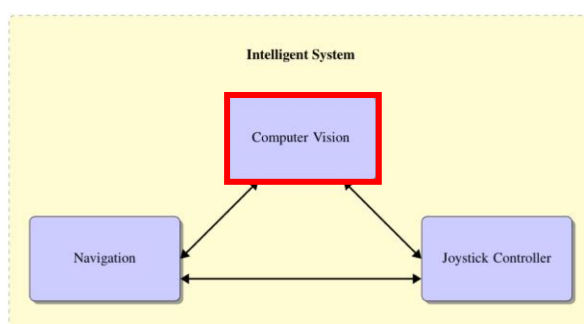
Also known as hardware system

- Perform autonomous movements of the electric wheelchair
- Adaptable to any model
- Avoid alter the architecture

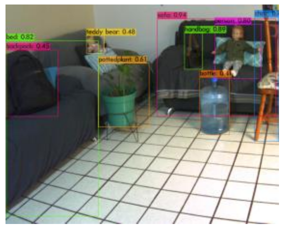


Computer Vision module

- Acquire images
- Pre-processing techniques to adapt images
- Identify and classify objects in real-time
- Use stereo vision cameras for depth of objects calculations



Camera



- Identify and classify objects in the environment
- The object detector YOLO v4 locates and classifies objects.
- Usage of stereo vision cameras can calculate depth of objects to help the navigation module.

Challenges:

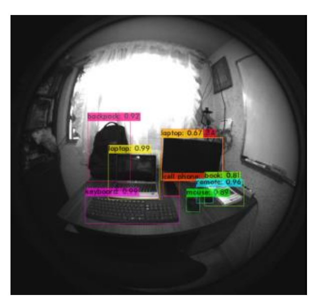
- Which are the features of a camera that achieve better prediction results.
- Use an embedded system to store the architecture.

First challenge: Comparison of cameras under study with the dataset created

iDS

Common Webcam

Intel RealSense



A dataset with 33 images where 20 object classes were used in the analysis.

Specifications	iDS	Webcam	Intel RealSense
Model	UI-1240SE-C-HQ	Perfect Choice 1080p	Intel RealSense T265
Resolution	1280 x 1024	1080 x 1850	1280 x 980
Image format	RGB	RGB	Gray
Configuration complexity	Medium	Easy	High
Cost	High	Low	High
Additional accessories	Extra lens	NA	NA
Additional characteristics	CMOS color sensor 8-bit depth	-	Wide eye-fish lens, IMU, Stereo vision

17

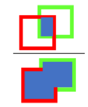
Second challenge: Embedded System Design Platform

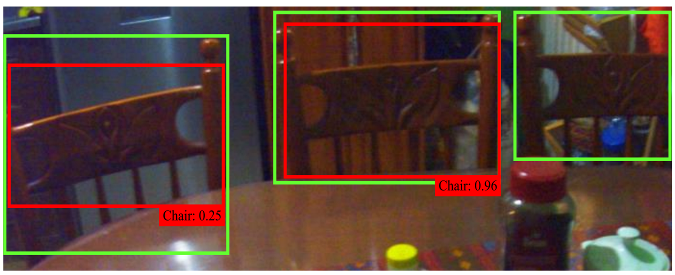
LattePanda Alpha 864S

18

Mean Average Precision (mAP)

mAP is an evaluation metric used for object detection in terms of location and classification

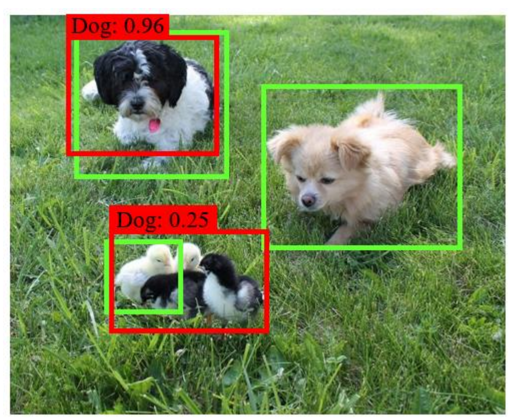
IoU = Area of Intersection / Area of Union = 



The **confidence value** is calculated by the object detector

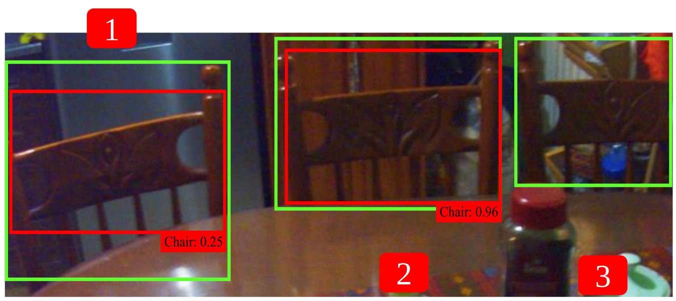
Object detector to classify dogs

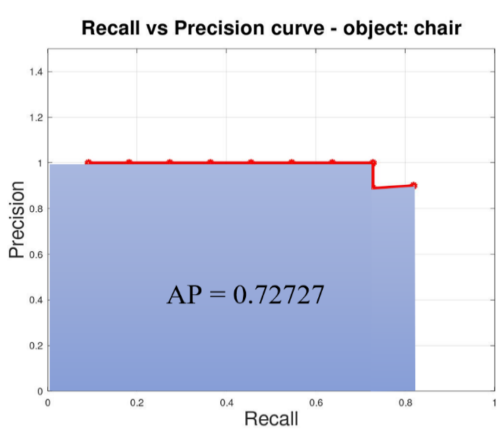
Actual	Prediction	Result
Dog	Dog	True Positive
Bird	Dog	False Positive
Dog	-	False Negative
Bird	-	True Negative



IoU threshold of 0.5 and confidence value of 0.75

1) False Positive
2) True Positive
3) False Negative



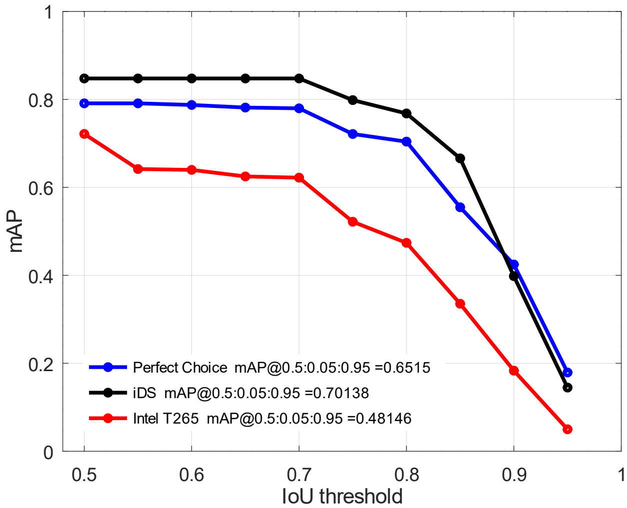


- *Precision*: How accurate is the model

$$Precision = \frac{TP}{TP + FP}$$

- *Recall*: Sensibility of the model

$$Recall = \frac{TP}{TP + FN}$$



Important features:

- ✓ High resolution images
- ✓ Color format
- ✓ CMOS color sensor

Prediction Time

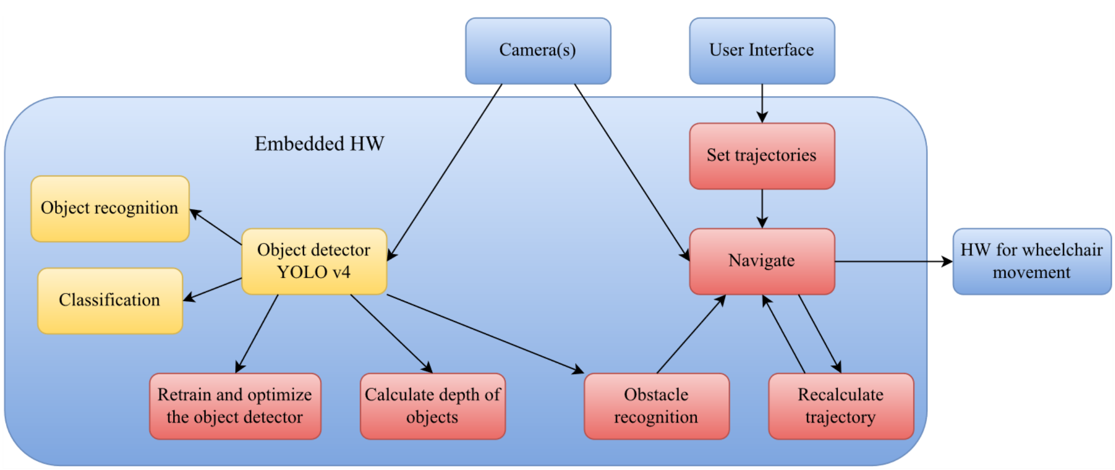
A high-resolution image cause a large processing time.

Camera	Average time in seconds	mAP@0.5:0.05:0.95
Perfect Choice	13.264867	0.65150
iDS	18.685641	**0.70138**
Intel RealSense	**12.489101**	0.48146

Prediction average time and mAP for each camera.

- It was generated a new dataset with 33 images using three different cameras
- Analysis of cameras using the mAP metric, YOLO v4 object detector, and a ground truth dataset
- Running YOLO v4 on Embedded System: LattePanda
- The best camera: iDS with RGB and high-resolution images, but also exhibits a poor prediction time

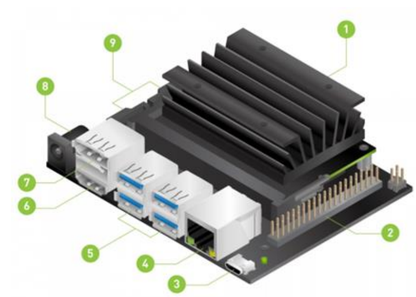
Disabled Population | State of the Art | Architecture Proposed | Computer Vision Module | mAP | Results | Conclusions



Future Work

- Retrain the object detector using images with objects commonly found in indoor environments

- To improve prediction time:
  - I. Use Jetson Nano to train and test the model
  - II. Explore new technologies:
    - Cloud computing
    - Distributed systems



27

Q&A

Towards An Intelligent Electric Wheelchair: Computer Vision Module

Examen de Grado

Ing. Jesús Gerardo TORRES VEGA

26th September 2022

28

[1] INEGI, "Población. Discapacidad." http://cuentame.inegi.org.mx/poblacion/discapacidad.aspx?tema=P (accessed Apr. 14, 2021).

[2] M. Dahmani et al., "An Intelligent and Low-Cost Eye-Tracking System for Motorized Wheelchair Control," Sensors, vol. 20, no. 14, Art. no. 14, Jan. 2020, doi: 10.3390/s20143936.

[3] Sutikno, K. Anam, and B. Sujanarko, "Design of electrical wheelchair navigation for disabled patient using convolutional neural networks on Raspberry Pi 3," AIP Conference Proceedings, vol. 2278, no. 1, p. 020048, Oct. 2020, doi: 10.1063/5.0014513.

[4] T. Kai, H. Lu, and T. Kamiya, "Object Recognition from Spherical Camera Images Based on YOLOv3," in 2020 20th International Conference on Control, Automation and Systems (ICCAS), Oct. 2020, pp. 419–422. doi: 10.23919/ICCAS50221.2020.9268308.

[5] Y. Sakai, Y. Nakayama, H. Lu, Y. Li, and H. Kim, "Recognition of Surrounding Environment for Electric Wheelchair Based on WideSeg," in 2019 19th International Conference on Control, Automation and Systems (ICCAS), Oct. 2019, pp. 816–820. doi: 10.23919/ICCAS47443.2019.8971608.

[6] I. Rakhmatulin and A. T. Duchowski, "Deep Neural Networks for Low-Cost Eye Tracking," Procedia Computer Science, vol. 176, pp. 685–694, Jan. 2020, doi: 10.1016/j.procs.2020.09.041.

[7] Y. Nakayama, H. Lu, J. K. Tan, and H. Kim, "Environment recognition for navigation of autonomous wheelchair from a video image," in 2017 17th International Conference on Control, Automation and Systems (ICCAS), Oct. 2017, pp. 1439–1443. doi: 10.23919/ICCAS.2017.8204217.

Appendix A

Objective of the Intelligent Electric Wheelchair, Contributions and Future Work of the paper

This appendix describes the general and specific objectives of the Intelligent Electric Wheelchair. Then, lists the main contributions of the paper and finally presents the future work.

A.1 General Objective of the Intelligent Electric Wheelchair

Develop an intelligent electric wheelchair capable of navigate autonomously and being aware of the environment and obstacles present along the trajectory.

A.2 Specific Objectives of the Intelligent Electric Wheelchair

1. Set the trajectory to reach a point specified by the user.
2. Navigate without collide and following the trajectory.
3. Detect obstacles along the movement of the electric wheelchair.
4. Modify dynamically the trajectory to avoid obstacles.
5. Identify objects of interest for the user such as medications, personal stuff, etc.

A.3 General Objective of the Analysis for the Computer Vision Module

Perform an analysis to find the characteristics of a camera that makes the object detector YOLO v4 perform the best in localization and classification of objects present in an image.

A.4 Specific Objectives of the Analysis for the Computer Vision Module

1. Compare three vendor cameras with different characteristics.
2. Generate a dataset with images using the cameras positioning at the same reference point, same illumination conditions and number of objects present in each image.
3. The analysis should consider objects present in indoor environments.
4. Generate the ground truth of each image in the dataset.
5. Compare the performance of each camera through an evaluation metric.

A.5 Contributions of the paper

a) Set the initial architecture of the intelligent electric wheelchair
b) Generate a dataset of 33 images from three different vendor cameras to perform a comparison analysis
c) Demonstrate the usage of a minicomputer LattePanda Alpha 864s to install and set up the YOLO v4 object detector.
d) Generate the ground truth of the dataset needed in the analysis
e) Develop a script code to calculate the mean average precision (mAP) using the ground truth dataset and the prediction results from YOLO v4 available in txt files.
f) Find the best characteristics for YOLO v4: RGB image format, high resolution images and a CMOS color sensor available in the iDS camera

A.6 Future work

Following the same line of the computer vision module few approaches are proposed:

1. Retrain the object detector YOLO v4 using images with objects commonly found in indoor environments.
2. Use Jetson Nano minicomputer from NVIDIA to set up YOLO v4, train the object detector and test the model.
3. Perform a new analysis using Jetson Nano minicomputer to improve the prediction time through its GPU integrated.
4. Explore new technologies such as cloud computing and distributed systems to set up the object detector to improve prediction time near real-time.