Environment Model Configuration from Low-quality Videos

López De Luise Daniela
CAETI – Universidad Abierta Interamericana – Facultad de Tecnología Informática
Av. Montes de Oca 745, Ciudad de Buenos Aires, Argentina.

CI2S Labs
Pringles 50, Ciudad de Buenos Aires, Argentina.

Instituto de Investigaciones Científicas (IDIC), Universidad de la Cuenca del Plata (UCP), Facultad de Ingeniería, Tecnología y Arquitectura, Formosa, Corrientes, Argentina.
daniela_ldl@ieee.org

Park Jin Sung
CI2S Labs
Pringles 50, Ciudad de Buenos Aires, Argentina.
zeroalpha2000@gmail.com

Hoferek Silvia
Instituto de Investigaciones Científicas (IDIC), Universidad de la Cuenca del Plata (UCP), Facultad de Ingeniería, Tecnología y Arquitectura, Formosa, Corrientes, Argentina.

Universidad Siglo 21, Decanato de Ciencias Aplicadas, Argentina
srhoferek@gmail.com

Avila Lautaro Nicolás
Instituto de Investigaciones Científicas (IDIC), Universidad de la Cuenca del Plata (UCP), Facultad de Ingeniería, Tecnología y Arquitectura, Formosa, Corrientes, Argentina..
lautyx027@gmail.com

Benitez Micaela Antonella
Instituto de Investigaciones Científicas (IDIC), Universidad de la Cuenca del Plata (UCP), Facultad de Ingeniería, Tecnología y Arquitectura, Formosa, Corrientes, Argentina.
benitezmicaelaantonela@gmail.com

Bordon Sbardella Felix Raul
Instituto de Investigaciones Científicas (IDIC), Universidad de la Cuenca del Plata (UCP), Facultad de Ingeniería, Tecnología y Arquitectura, Formosa, Corrientes, Argentina.
raulbordon250@gmail.com

Fantín Rodrigo Iván
Instituto de Investigaciones Científicas (IDIC), Universidad de la Cuenca del Plata (UCP), Facultad de Ingeniería, Tecnología y Arquitectura, Formosa, Corrientes, Argentina.
rodrignr4@gmail.com

Machado Gastón Emmanuel
Instituto de Investigaciones Científicas (IDIC), Universidad de la Cuenca del Plata (UCP), Facultad de Ingeniería, Tecnología y Arquitectura, Formosa, Corrientes, Argentina.
gastonmachado44@gmail.com

Mencia Aramis Oscar
Instituto de Investigaciones Científicas (IDIC), Universidad de la Cuenca del Plata (UCP), Facultad de Ingeniería, Tecnología y Arquitectura, Formosa, Corrientes, Argentina.
aramismencia@gmail.com

Ríos Anahí Ailén
Instituto de Investigaciones Científicas (IDIC), Universidad de la Cuenca del Plata (UCP), Facultad de Ingeniería, Tecnología y Arquitectura, Formosa, Corrientes, Argentina..
anahiriosailen@gmail.com

Rios Emiliano Luis
Instituto de Investigaciones Científicas (IDIC), Universidad de la Cuenca del Plata (UCP), Facultad de Ingeniería, Tecnología y Arquitectura, Formosa, Corrientes, Argentina.
mcwiths@gmail.com

Riveros Nahuel Edgardo
Instituto de Investigaciones Científicas (IDIC), Universidad de la Cuenca del Plata (UCP), Facultad de Ingeniería, Tecnología y Arquitectura, Formosa, Corrientes, Argentina..
Nahuel42425@gmail.com

***Abstract*—This article aims to describe main findings on a prototype for assisting blind people. To improve its functioning the main approach is to build a model dynamically using Intelligent System and Machine Learning. After several partial models the prototype is able to detect and recognize the outline of a user environment, specifically to determine the spatial organization of multiple objects. This paper encompasses a comprehensive set of activities aimed at evaluating and enhancing a system with efficient metrics for feature assessments upon, video , image segmentation, and data mining on the fly. Additionally, this work covers automatic image tagging, and a set of risk rules. It also evaluates and depicts specific techniques and approaches to be applied to create models with high pattern-detection efficiency. The algorithm used is required to be light and quick, in order to be used in standard cell phones to assist blind people and provide meaningful information to the user. As part of the current paper a small statistical analysis is also performed.**

Keywords—Blind people assistance, Video processing, Object detection, Data Mining, Environment configuration.

I. INTRODUCTION

The most challenging activity for blind people is to walk outdoors independently of any support. HOLOTECH [1] aims to provide guidance in this context by simply using a standard cell phone. It outperforms many of current proposals [2][3][4] as traditional tools are the cane, trained animals, and assistants. World statistics indicate that individuals with visual impairments grow year by year [5] [6]. Exposure to nature has benefits for people's mental and physical health. One way to achieve this is by ubiquitous and mobile technologies. However, existing research in this area is primarily focused on people without visual impairments and is not inclusive of blind and partially sighted individuals.

Outdoor experiences in the natural environment for these people present specific needs and barriers that could be addressed by technology. According to [8] they can be classified into three main concerns: independence, knowledge of the environment, and sensory experiences.

For most people who are blind, exploring an unknown environment can be unpleasant, uncomfortable, and unsafe. Authors in [9] explore an adaptation of the use of virtual reality as a learning and rehabilitation tool for people with disabilities, based on the hypothesis that the supply of appropriate perceptual and conceptual information through compensatory sensorial channels may assist people who are blind with anticipatory exploration. His work aims to allow the user to explore a virtual environment with two main goals: evaluation of different modalities (haptic and audio) and navigation tools, and evaluation of spatial cognitive mapping employed by people who are blind. Preliminary results indicate that comprehensive cognitive maps can be built by exploring the virtual environment.This study's results indicate that prototypes and research in this way are very promising.

Since blindness can be caused by things like genetics, infection, disease or injury the help these individuals may highly differ in every case. As a case, someone with 20/200 vision sees an object from 20 feet that a person with 20/20 vision is able to see from 200 feet.Then, environmental Challenges for People who are completely blind or have impaired vision usually depends on its origin and age. In spite of that, they have a difficult time navigating outside the spaces that they're accustomed to. In fact, physical movement is one of the biggest challenges for blind people, since it requires a special explanation of World Access for the Blind [10]. One of the most compelling difficulties with different types of blindness [7] is the loss of awareness of the environment, exposing them to a high risk of accidents. Based on the grade of blindness and the individual's age, the risk factor varies for older people. Falls are more prevalent in this group and can have a greater impact, potentially leading to more significant complications. Despite this, there are studies dedicated to preventing such issues [11], but they require time and effort. In addition to the issues under investigation, new complications associated with the global pandemic have emerged [12]. Since most visually impaired individuals are not independent, the necessary social distancing measures during the pandemic have introduced complications, significantly impacting their daily lives. While the pandemic may be coming to an end, there are still existing limitations in the tools available to aid blind individuals with mobility. In addition to the numerous advantages provided by the traditional white cane [13], there are also several disadvantages. Addressing these limitations and disadvantages is both an issue and a necessity for the blind community.

Recognizing the necessity for further assistance to enhance the autonomy and independence of individuals with visual impairments, we are utilizing current technology and machine learning to fulfill this need.

The project HOLOTECH aims to provide information and complement the lost awareness in an alternate way: with a simple language based on sounds. The first steps of the prototype are to get video, detect certain objects by pattern matching, analyze their location, velocity, and direction, and finally generate an audible alarm with certain features mapping main information about surrounding obstacles.

This work focuses on the tuning of pattern-matching steps, which is relevant to increase fidelity and precision in real-time. The approach is to train a Machine Learning model to detect a specific type of object and to use it for recognizing obstacles. The current study involves methods to guarantee the accuracy of the detection and improve coding performance to be lightweight enough to be used on a portable device. To adjust the accuracy, newly created models are compared to legacy pre-trained models that come with OpenCV[14]. New models correspond to objects that are to be detected but do not exist in the database of pre-trained models. All of them are obstacles of interest found with the interaction with potential users contacted as part of the collaboration of the Circle of the Non-Sighted (CINOVI). They determined a specific list of potential risks and objects that need to be identified. Note that the interaction with real users helps to debunk misconceptions, such as the need to detect people in the environment, since they can easily perform this. It is interesting to mention that the objects that constitute the most frequent problems can be handled with a cane. But those requiring more complex detection, are typically at a high of one meter or more from floor level or pits that are at a strange angle and can't be detected with that device.

Several previous results from the project with members of the Argentine Library for the Blind (BAC), the CAETI center of UAI, and the current team in conjunction with CINOVI, are in [1][15][16]. The information derived from the models implemented in the prototype can detect any type of object with different degrees of confidence. This article presents the approach of multiple model synchronization intended to be handled by an expert system.

The background of the current proposal relies on multiple advances in the field of video processing. Among others, slicing, segmentation, and Machine Learning (ML) have been used by López De Luise et. al. [17] for 3D inferred Scenes using 2D images (Fig. 1), and mood inference, where an AI artist changes a piece or human art to incorporate the inferred observer internal status choosing among intelligent "brushes", and then updates the picture in real-time (Fig. 2). Other authors performed smart coloring from grayscale images [18], endoscopic 3D reconstruction [19], image restoration [20], deep image inference [21], video compressing [22], and other extensive applications [23] of computational intelligence on video processing.

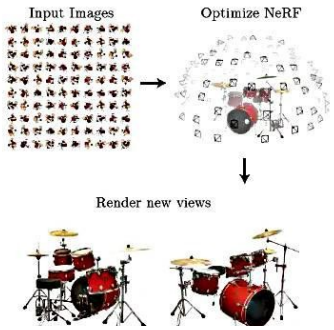


Fig. 1. 2D figures (left) and resulting 3D inferred figures (right)

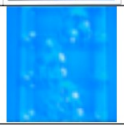
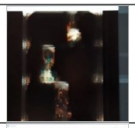
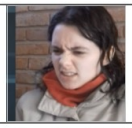
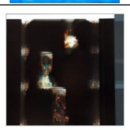
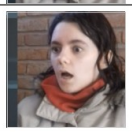
Mood	Table Column Head		
	Previous status	Final status	Expression
Neutral	NO		
Sad			
Horror			
Surprise			

Fig. 2. Table of Lumiere productions (an AI artist) using mood inference

The rest of this text covers the case study, statistical analysis, and conclusions derived from training of the models for the prototype. It is important to highlight that identifying the objects and analyzing them is key for the progress of the prototype as the main goal is to provide precise information to lower potential risks of collision during indoor and outdoor displacement for individuals with visual impairments.

II. Methodology and Materials Selection

Due to the context for using the proposed solution, the concept for the prototype's architecture involves utilizing compact, lightweight hardware, such as a smartphone in conjunction with an Arduino Nano. This combination enables the model to incorporate hardware that is typically not found in a smartphone, such as an ultrasound sensor. The purpose of this architecture is to serve as an initial prototype, facilitating our exploration of the proposed solution. However, a crucial aspect of the prototype is to have hardware capable of capturing information, recognizing, and executing various routines and actions. Therefore, the minimum requirement for the architecture is the ability to capture video and process it using a trained model for recognition. The idea to recognize objects is to predict and filter behavior, a chair or desk are unmovable objects by themselves so the only way for a impaired person to collide is by his own movement, the same can be applied to other objects like trees. For that reason it is imperative to understand that even if the requirement is to be detected in real-time the process to recognize can take some reasonable time. For that reason the architecture proposed is centered using a smartphone. Nevertheless for real-time cases that need fast detecting reactions we are incorporating ultrasonic sensors that will allow us to know when a sudden object appears in the trajectory and in this case the need to know what object is in the path doesn't matter, only the direction where it came from, the proposed architecture also aims to facilitate communication with impaired individuals through various methods, such as using sound or vibration in extreme cases. The article focuses on detecting and recognizing the object from a certain distance: how precise are the models and with the architecture used as a base.

Considering the hardware restrictions, there is a strategic limitation on the model to be designed for identifying any object of interest. In previous research, statistics show this limitation imposed on images used to train the model with quality and resolution around 320x240 dpi to 720x480 dpi.

In order to test and train the model, there is a pre-processing to improve the boundary of any targeted object, a tool named Labelme [24] is used. It performs automatic graphic image annotations. This enables data for detecting the desired object using polygons and tags to classify it automatically. These metadata are stored in a JSON file. There is an extra step to adapt the JSON formatting to YOLO [25], a conversion required before the training [26]. The Neural Network tool used here for object detection is YOLO for its unique approach that ensures quickness. There are other popular tools like Faster R-CNN [27] and SSD [28], but YOLO is simpler to use and more popular, with better community support.

This project employs PyTorch [29] to implement the object detection model based on YOLO [25]. PyTorch offers tools for defining neural network architectures, optimizing models, computing gradients automatically, enabling distributed training, and more. Additionally,

PyTorch is compatible with execution on either GPU or CPU, providing a valuable option for users without access to a CUDA-enabled graphics card. Moreover, Ultralytics [30] was utilized for training the neural network, and Pillow [31].

As a first step to the multi-model management of the prototype, the models are for a limited diversity of objects. They correspond to those that currently exist in the pre-trained model. This way it is possible to compare their performance with the newly created models with HOLOTECH.

The comparison of both models is a preliminary step to evaluate the degree of precision increment in the processing with models improved with ML. Despite that, it is important to remark that for a tiny subset of targets, the pre-trained models will be used. The option will depend on the performance of the newly created models. In some cases, pre-trained models are not loaded to reduce the overloading of information, reducing the number of patterns and therefore minimizing the firing rate of filtering data tasks.

The selected targets to work in the current article are “chairs”, “desks”, “doors”, and “persons”. Test conditions for all of the objects are mostly found indoors and outdoors. As an intermediate procedure, there is an additional validation of the need to use a model based on the classification group. This is because in previous research special cases emerged where the shapes of different obstacles induce mistakes in the patterns. As an example, some chairs and desks look very similar depending on the perspective.

Finally, two special targets are doors and persons. The first is because they do not present any pre-trained model in the tool. The last is because they will be detected and filtered out. Blind people do not need any alarm disturbing every time a person approaches the individual, as they don't represent any collision risk.

II. Architecture Schema

With the ongoing shift in technology there are 2 architecture approaches, the first one is creating a standalone python application that has the models included inside. The second one is a client-server architecture where the client sends the data feed to the server and the server is in charge of processing and applying the model.

Each of the architectures has their pros and cons, for the first architecture we found that is hard to create the binary for the differents hardware (Android - Iphone), more so that both hardware theirs software requires that the binary is made of a specific programming language and neither of those are adapted easily to use A.I models. Furthermore in case that needs to be provided and updated it would require re-building the application. However once created the advantage is that it would allow it to respond faster and it would not be dependent on external resources.

On the other hand, the second architecture would be the other way around, because it is separate in 2 client and server, the advantage would be that there won’t be a compatibility issue in the client side, furthermore the models would be on server, this would allow any update that requires a change on the model won't impact the client in a bad way. Contrary to the first architecture, this architecture is easier to scale in the future. Even so the disadvantages are that it would be dependent that the hardware has an internet connection, even more the server needs to be hosted on a server that runs 24hs, more users would require a more sizable server. All of this would require a significant amount of cost.

For the purpose of this research, we moved forward to implement both alternatives. To understand where would be the point that neither of the alternatives would be not more viable in the future. Currently both are valid alternatives.

III. Prototype Performance and Test Cases

The main goal of this article is to present the first stage of the tuning process for the HOLOTECH prototype. The threshold for efficiency metrics has been defined to ensure a precision rate with minimum quality in real-life functioning. From the set of pre-trained models, a lower boundary of the acceptance interval is defined. Note that every performance in the current context corresponds to tests running against a database of video clips recorded according to the list of selected targets. As mentioned previously, the preference choices refer to those obstacles declared as of interest due to their potential risk during the indoor and/ or outdoor displacement. The list is the result of interactions with volunteers in collaboration with the CINOVI, and consists of both indoor and outdoor obstacles and/or events.

Some of the elements included in the priority list are:

- Cars parked on sidewalks.
- Cars parked in unauthorized spaces.
- Motorcycles parked in unauthorized locations or moving irregularly.
- Bicycles parked in unauthorized locations or moving irregularly.
- Open gates or gates opening onto the street.
- Plant branches or any object protruding the path of travel, at a knee height and above.
- Any public municipal ramp.
- Devices installed on walls, such as air conditioners at non-recommended heights, especially when protruding the pathway.
- Potholes or breaks in the ground, with or without fencing.
- Potentially dangerous moving obstacles such as strollers, various vehicles, cyclists, etc.

The list continues and might be extended after the initial deployment of the actual prototype after tuning the performance within the acceptance level interval.

IV. Model Training and Results

This section explains the tests performed for making the pattern recognition neural model with YOLO and depicts some of the essential steps to fit the performance in YOLO to build a better model. The activity is performed on two versions of the model evaluation.

A. *First version: reduced set*

The training aims to build a model that outperforms the one included by default in Labelme. The process includes a validating step, the accuracy level assessment, and a final comparison between the new model. Special attention deserves the fact that the training database of videos used for the legacy neural network in the platform is not available to end users. Therefore a new training set with similar obstacles has to be generated and labeled in advance. The object classes are named here as ModelGroup, and each of them has an independent model specifically trained for the subset of specific objects in it. Considering the reduced list covered in the current test, the training encompassed four models, each utilizing a set of 150 images. Every image corresponding to any of the target objects has been conveniently labeled using Labelme.

In order to set the necessary information from an image, the tool Labelme is used to streamline the process. The tagging is carried out manually to ensure that obstacles are accurately identified within the image. Fig. 3 shows a screenshot of the process.

Fig. 3. Tagging process using Labelme

As can be seen from the picture, Labelme lets the user define metadata without the need to use complex image processing concepts to filter unwanted information and to retrieve the required data. Once an image is tagged Labelme provides the image context characteristic, and features of the tagged object in a JSON format (Fig. 4).


```
File  Edit  Format  View  Help
{
  "version": "5.3.1",
  "flags": {},
  "shapes": [
    {
      "label": "Mesa",
      "points": [
        [
          25.307881773399018,
          93.04187192118228
        ],
        [
          24.568965517241395,
          97.22906403940887
        ],
        [
          28.509852216748783,
          99.692118226601
        ],
        [
```


Fig. 4. Example features automatically generated in JSON format

The JSON file requires some extra processing to transform it into the required YOLO format, as the set is then fed to YOLO. This step is performed with a tool called Labelme2yolo which generates the structure for model training. Additionally, the tool facilitates the segmentation of images into test, train, and validation subsets, and if necessary, it can apply pre-processing techniques to convert the images as in Fig. 5.

1) *Dimension issue*

In order to enhance image processing performance, resizing images to standard sizes is implemented: small (100 dpi or 320x240), medium (200 dpi or 500x300), and large (300 dpi or 720x480). Despite potential minor distortion of object shapes by approximately 1% or less during rescaling, this procedure centers and refines images by eliminating noise during processing and model development.

To accomplish this resizing process, we devised a compact system capable of obtaining images with their respective names and specified formats. The Python Imaging Library (PIL) was utilized as the tool for resizing. For each size category, a precondition was established: if the image size exceeded a certain threshold, it would be resized to fit within the designated size for that category, and the updated file would be saved with the same name and extension.

Fig. 5. Example of a pre-processing step

2) *Model*

The Model train-set set comprises a set of chair, desk, and door pictures. The premise is to apply the pattern matching to similar objects, as mentioned previously, to determine the confusion date of the model, evaluate the confidence level of the object identification, and limit the failure rate.

The model training sessions consist of 100 epochs using the yolov8m-seg model type from YOLO. Training is performed in two independent batches. Due to the use of a Mac Pro workstation, the training process took several hours to complete. It is important to note that this is not a problem since the model is not expected to be retrained during its application in the field. Nevertheless, it is possible to reduce the training time by utilizing an NVIDIA graphics card in a Linux environment. The configuration used to train the models is the initial step in understanding the response of the training and fine-tuning it if necessary to improve the precision of the model.

To emphasize the object in the individual models, an additional "background" tag is added to indicate when an object other than the expected object was selected. This was done in an attempt to create false positives and have a clear understanding of the background noise (see an example of the problem in Fig. 6).

Fig. 6. Example of tagging with included background

3) Results

It is interesting to note that the ModelGroup, managed to find the object's category but failed to correctly identify the specific object when it was trained with objects sharing similar characteristics, such as a desk and a chair. The problem is shown in Fig. 7.

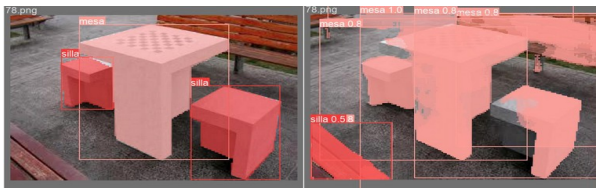


Fig. 7. Comparison of the tagged objects and predicted outcome

The precision of some identified objects is high enough for the model to be confident that the object is a desk, with a score of 0.8. However, in some cases, the model still finds objects incorrectly and determines a wrong classification with a high confidence score. The problem arises just in cases where the out-shape is similar between two or more patterns (as in the case of chair and table in Fig. 7, where the model ends up identifying everything as a desk). One solution for this particular case was to add a boundary restriction to emphasize that a chair needs to be within a certain dimension, and the same for the desk. However, this solves specific cases like the confusion between chairs and desks. As more diverse objects are to be added to the model training, it will not be ideal and the restrictions will become more complex.

It seems like the models trained individually with the background did not end up being as precise as expected, as shown in Fig. 8.

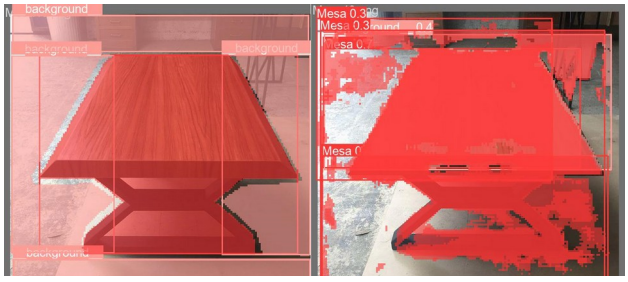


Fig. 8. Predicted output (right) using dedicated models

This time, the precision of the predicted output was not perfect. As can be seen in the picture, the model ended up finding multiple readings regarding the objects identified, with a confidence range between 0.3 to 0.7, and identifying an object as a background too. The primary problem with the model appears to be the various interpretations of the recognized objects, with confidence scores ranging from 0.3 to 0.7. Although the identification of the background as an object can be eliminated, the consistent trend of readings falling within this confidence range for the predicted objects is worrisome. This issue renders the models unsuitable for use, as it could potentially result in hazardous situations for individuals.

To better understand the usability of the models, a study was conducted using the output from the training. This data can be used to statistically understand the models and fine-tune them to make them more precise. The outputs are presented in Fig. 9 and Fig. 10.

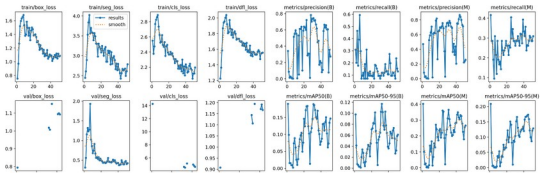


Fig. 9. Results of the group training model

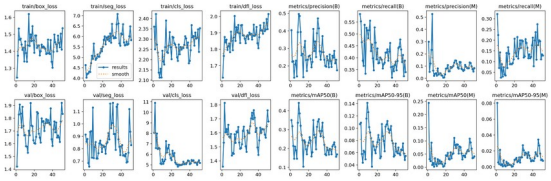


Fig. 10. Results of the desk training model

The addition of the background tagging for the individual trained model was to reduce the background noise found in the analysis of the confusion matrix trained on the ModelGroup model Fig. 11.

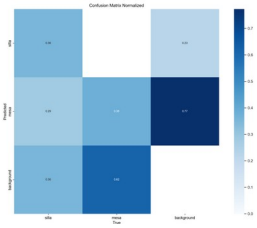


Fig. 11. Confusion matrix for ModelGroup

Although the data used to train the model did not contain a background tag, the model predicted the desk object as background in about 77% of the cases. In order to reduce the rate of mispredictions, the background was provided during training for the other models.

The inclusion of the background tag did not result in the expected enhancement in the individual models. This caused the background to be identified as a separate object, as shown in Fig. 12. This indicates that additional modifications or fine-tuning may be required to resolve this problem. Although the misprediction rate was lower than

the other models, there were still too many issues with the outcome.

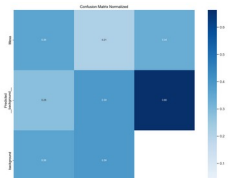


Fig. 12. Confusion matrix of Desk Model

The predicted outcome for the desk object was similar for both models, with a prediction of 38% and 36%, respectively.

B. Second version: optimized set

The training process was conducted similarly to the previous version, but with the distinction that this time, the dataset consisted of 38 curated images primarily focusing on two categories: people and chairs. Given the nature of the objects of interest (people and chairs), a new phase was introduced to identify these objects. In this process, additional images were gathered to augment the dataset for testing purposes. The primary goal is to evaluate the effectiveness of the newly developed code in detecting these specific objects.

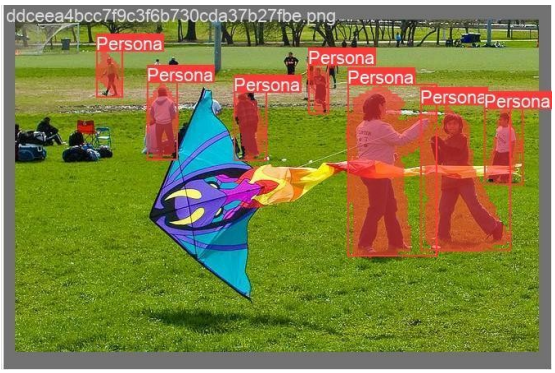


Fig. 13. Tagging of persons

1) Model

The new model is named ModelGroup, focused on detecting people and chairs as previously mentioned. This model was trained for 30 epochs in 2 batches using YOLO. The training was conducted on a workstation running Windows OS, utilizing CPU processing which led to longer training times (Fig. 14).

2) Results

Upon evaluating the results obtained from ModelGroup, we observed an improvement in detecting people compared to previous training iterations. However, the model still struggles to detect multiple individuals accurately. Specifically, in one instance depicted in Figure 15, half of the group of people was not tagged. The precision of object tagging in these results achieved a confidence level of 0.6 (see fig 16 – 18).

Fig. 14. Tagging of persons

Fig. 15. Comparison between tagged and predicted persons

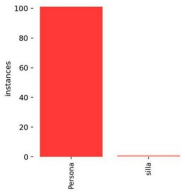


Fig. 16. Bar prediction rate for tagging

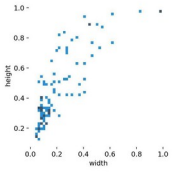


Fig. 17. Plot prediction hit for tagging

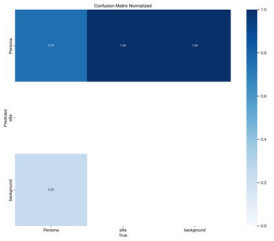


Fig. 18. Confusion matrix

Based on figures 16, 17, and 18, it can be seen an improvement in the results obtained compared to the previous instance.

C. Third version: additional objects' set

For the third iteration, we followed a similar procedure as before but made a modification to the dataset by including an additional tagged object, specifically tables.

The objects of interest, now totaling three for this case and trained simultaneously, presented certain challenges and difficulties in identifying and tagging each element. The goal here is to assess the effectiveness of training with a group of three different object types.

The JSON file requires some extra processing to transform it into the required YOLO format, as the set is then fed to YOLO.

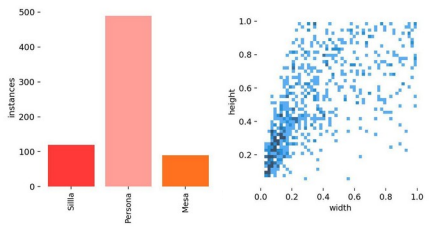


Fig. 19. Hit rate of the extended tagged objects

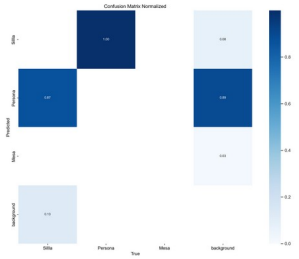


Fig. 20. Cohen metric of the extended tagged objects

D. *Fourth version: extended set*

In this latest test, the samples are only images tagged as containing persons and supplemented the dataset with additional images, bringing the total to 271. Since its focus is on a single object in the scenario, the model finds it easier to detect a target. The aim of this version is to focus solely on a single object and to observe if this approach yields better results and detection rate. The training is kept with 100 epochs with 6 batches using YOLO, following the same configuration as in previous models to facilitate comparison.

With an increased number of images and tagged objects to refine the neural network, we have achieved the best results thus far, with responses and detections approaching perfection, accurately tagging the full body of individuals. However, one issue we encountered was with objects or people situated too far away or overlapped with other elements, as illustrated in the figures 21 and 22.

Despite these complications, it can be stated that with a higher frame rate, results with less noise can be obtained, thus allowing better control of the situation with this group of individuals and having a certainty rate close to 90/95%.

V. Conclusions

The analysis of the models for identifying different targets show similar predictions accurate no matter the type of detection. Also, the extra work required to filter noise, does not provide significant improvement in the hit rate. As it is just for objects that represent low risk for blind people it can be avoided. Regarding the splitting of a global model into class models, tests show that it is necessary but shall not be trained for every object but for predetermined classes or ontological groups in order to improve precision. As shown in tests, there are different ways to implement this. Among others creating models without background tagging, improving the input data set, adding more false positives, and fine-tuning the model.

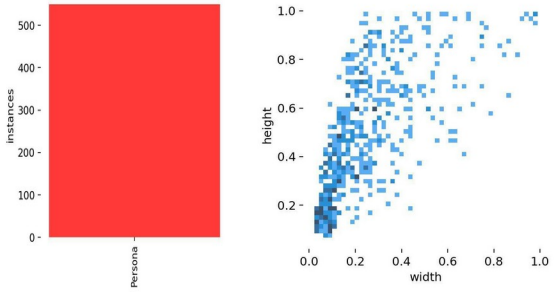


Fig. 21. Rate plots of the extended set

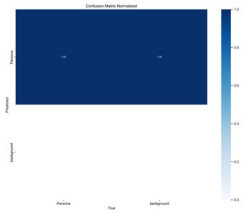


Fig. 22. Cohen metric of the extended set

At the current state of this research it is not possible to reach the same level of precision as the pre-trained commercial models in certain specific cases presented here. Therefore it requires extra work. Regarding the architecture implemented in the prototype, there also remains a tuning for selecting a proper aggregation of objects.

The evaluation of the models for identifying target objects revealed that whether using a group of objects or analyzing individual objects, the predictive outcomes are largely comparable. Nevertheless, superior performance is observed with individual object analysis, leading to enhanced detection, reduced noise, and a notable decrease in false positives.

VI. Risk Rules

For the application of models and to allow a fast response of the system, we set a couple of rules where depending on coordinates of the object identified the system would proceed to further track or ignore it. The concept is that there would be no point to follow up an object that would not present any foreseeable damage to the user. For that there would be some constants and variables that we would set depending on the user and after the detection apply the different equations, as illustrated in figures 23, that would allow us to determine the risk the object would incur to the user. We used in the equations 2 sequence of detection to identify the velocity, dispersion and direction of the object.

Condition	Getting Close	Getting Away	In place
Object displacement in the same direction with possibility of collision -MC / Desplazamiento de un objeto en la misma direccion con la posibilidad de Colision -MC	IF(dy > 30 && vy >= iv && y2 - y1 > 0 && ABS(x2-x1)<3) TRUE	ELSE FALSE	IF(dy == 30 && vy >= 0 && ABS(y2 - y1) < 5 && ABS(x2 - x1) < 5) TRUE
Object displacement in the same direction without possibility of collision -MS / Desplazamiento de un objeto en la misma direccion sin posibilidad de Colision -MS	TRUE	IF(dy > 30 && vy >= iv && y2 - y1 < 0 && ABS(Y2-Y1)>5 && ABS(x2 - x1) < 5) TRUE	TRUE
Object displacement in opposite direction with possibility of collision -OC / Desplazamiento de un objeto en una direccion opuesta con posibilidad de Colision -OC	IF(dy<30 && vy >= iv && y2-y1 < 0 && ABS(x2-x1)<3) TRUE	IF(dy>30 && vy >= iv && y2-y1 > 0 && ABS(x2-x1) < 5) TRUE	IF(dy == 30 && vy >= 0 && ABS(y2 - y1) < 5 && ABS(x2 - x1) < 5) TRUE
Object displacement in opposite direction without possibility of collision -OS / Desplazamiento de un objeto en una direccion opuesta sin posibilidad de colision -OS	IF(dy>30 && vy >= iv && y2-y1 < 0 && ABS(x2-x1)<3) TRUE	IF(dy<30 && vy >= iv && y2-y1 > 30 && ABS(x2-x1)<5) TRUE	IF(dy == 30 && vy >= 0 && ABS(x2 - x1) < 5 && ABS(y2 - y1) < 5) true
Object approximation (from right side) with possibility of collision -LDC / Objeto aproximandose (por el lado derecho) con posibilidad de colision -LDC	IF((dy<30 | dx<10) && vx > iv && vy = iv && x2 > 0 && ABS(y2-y1)<3) true	IF((dy>30 | dx>10) && vx < iv && vy < iv && x2 > 0 && ABS(y2-y1)<3) true	IF(vx == 30 && vy >= 0 && ABS(y2 - y1) < 5 && ABS(x2 - x1) < 5) true
Object approximation (from right side) without possibility of collision -LDS / Objeto aproximandose (por el lado derecho) sin posibilidad de colision -LDC	IF((dy>30 | dx>10) && vx > iv && vy = iv && x2 > 0 && ABS(y2-y1)<3) true	IF((dy<30 | dx<10) && vx < iv && vy < iv && x2 > 0 && ABS(y2-y1)<3) true	IF(vx == 0 && vy == 0 && ABS(y2 - y1) < 5 && ABS(x2 - x1) < 5) true
Object approximation (from left side) with possibility of collision - LIC / Objeto aproximandose (por el lado izquierdo) con posibilidad de colision -LIC	IF((dy<30 | dx>-10) && vx > iv && vy = iv && x2< 0 && ABS(y2-y1)<3) true	IF((dy>30 | dx<-10) && vx < iv && vy < iv && x2< 0 && ABS(y2-y1)<3) true	IF(vx == 0 && vy == 0 && ABS(y2 - y1) < 5 && ABS(x2 - x1) < 5) true
Object approximation (from left side) without possibility of collision - LIS / Objeto aproximandose (por el lado izquierdo) son posibilidad de colision -LIS	IF((dy>30 | dx<-10) && vx > iv && vy = iv && x2< 0 && ABS(y2-y1)<3) true	IF((dy<30 | dx>-10) && vx < iv && vy < iv && x2< 0 && ABS(y2-y1)<3) true	IF(vx == 0 && vy == 0 && ABS(y2 - y1) < 5 && ABS(x2 - x1) < 5) true

Fig. 23. Some of the equations to determine the risk

VII. Future Work

This paper aims to enhance the existing models in the project improving precision, and considering a wider range of objects for detection. The resulting models are to be with similar performance as any pre-trained commercial model. The next step is to feed results to an Expert System to interpret the context and use the correct model in a sequence.

Further investigation is required to assess how these models perform with further targets. The creation of this system will derive in an automatic interpretation of the environment using additional technologies like a GPS, or movement sensor, to gain extra knowledge of the individual's indoor or outdoor environment, to analyze the existence of any risks in its close or far away surroundings.

Depending on the architecture selected some of the ideas would provide a more robust challenge to be applied, however that would allow us to determine different alternatives or external resources that would help us improve the prototype.

The improved prototype will aim to be adequate for a real-time response and validation of the context where objects move and eventually produce alarm signals by means of a sound language shared with users for seamless communication between the prototype and its user.

Incorporating a GPS signal could substantially enhance contextual awareness by identifying whether an individual is indoors or outdoors, and it could leverage specific types of models for this purpose.

References

[1] Park, J.S., De Luise, D.L., Hemanth, D.J., Pérez, J. (2018). Environment Description for Blind People. In: Balas, V., Jain, L., Balas, M. (eds) Soft Computing Applications. SOFA 2016. Advances in Intelligent Systems and Computing, vol 633. Springer, Cham. https://doi.org/10.1007/978-3-319-62521-8_30

[2] Bryant Penrose, R. (2023). Anticipating Potential Barriers for Students With Visual Impairments When Using a Web-Based Instructional Platform. Journal of Visual Impairment & Blindness. Tools of the Blind and Visually Impaired. Volume 117 Issue 5

[3] Curing Retinal Blindness Foundation (2023). Tools of the Blind and Visually Impaired. https://www.crb1.org/for-families/resources/tools

[4] Blasch, B. B., Long, R. G., and Griffin, Shirley N. Results of a National Survey of Electronic Travel Aid Use. Journal of Visual Impairment and Blindness, November, 1989, v. 33, n 9, pp 449-453.

[5] WEBAIM (2021) Screen Reader User Survey #9 Results. Web accessibility in mind. Institute for Disability Research. Utah State University. Last updated: Jun 30, 2021. https://webaim.org/projects/screenreadersurvey9/

[6] The Lancet Global Health Commission on Global Eye Health: vision beyond 2020. Crossref DOI link: https://doi.org/10.1016/S2214-109X(20)30488-5

[7] Cleveland Clinic (2022) *Blindness.* https://my.clevelandclinic.org/health/diseases/24446-blindness

[8] Understanding Experiences of Blind Individuals in Outdoor Nature. M. Bandukda · A. Singh · N. Bianchi-Berthouze · C. Holloway. DOI: 10.1145/3290607.3313008. Conference: ACM CHI'19 · 2019

[9] A Virtual Environment for People Who Are Blind – A Usability Study. O. Lahav, D W Schloerb, S Kumar, M A Srinivasan. J Assist Technol. 2012; 6(1). doi:10.1108/17549451211214346. 2016

[10] Challenges That Blind People Face. Written by Kate Beck 18 December, 2018. HealthFully (https://healthfully.com/).Leaf Group Ltd.

[11] Insight (Lawrence). Author manuscript; available in PMC 2018 Dec 28. Published in final edited form as: Insight (Lawrence). 2011 Spring; 4(2): 83–91.

[12] Front Psychol. 2022; 13: 897098. Published online 2022 Oct 28. doi: 10.3389/fpsyg.2022.897098

[13] Rasouli Kahaki, Z., Karimi, M., Taherian, M. et al. (2023) Development and validation of a white cane use perceived advantages and disadvantages (WCPAD) questionnaire. BMC Psychol 11, 253. https://doi.org/10.1186/s40359-023-01282-4

[14] Holzer, R. (2019) OpenCV tutorial Documentation. Release 2019. pp 125

[15] Park, N. et al. (2021). Multi-neural Networks Object Identification. In: Balas, V., Jain, L., Balas, M., Shahbazova, S. (eds) Soft Computing Applications. SOFA 2018. Advances in Intelligent Systems and Computing, vol 1222. Springer, Cham. https://doi.org/10.1007/978-3-030-52190-5_13

[16] López De Luise, D., Park Jin , S., Hoferek , S., Avila Lautaro, N., Benitez Micaela, A., Bordon Sbardella, F. R., Fantín, R. I., Machado, G. E., Mencia Aramis, O., Ríos, A. A., Luis, E. L., & Riveros, N. E. (2023). Detección Automática de Objetos como asistencia a Personas Invidentes. Revista Abierta De Informática Aplicada, 7(1), 37–50. https://doi.org/10.59471/raia202356

[17] Furundarena, F., López De Luise, D., Veiga, M. (2022) Computational Creativity through AI modeling. CASE 2022

[18] Komatsu, T., Saito, T. (2006). Color Transformation and Interpolation for Direct Color Imaging with a Color Filter Array. International Conference on Image Processing, pp. 3301-3304, doi: 10.1109/ICIP.2006.312878

[19] Imtiaz, M. S., Wahid, K. A. (2014) Image enhancement and space-variant color reproduction method for endoscopic images using adaptive sigmoid function. In 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 3905-3908, doi: 10.1109/EMBC.2014.6944477

[20] Wen, Y. W., Ng, M. K., Huang, Y. M. (2008) Efficient Total Variation Minimization Methods for Color Image Restoration. In IEEE Transactions on Image Processing, vol. 17, no. 11, pp. 2081-2088, doi: 10.1109/TIP.2008.2003406

[21] Kuo, T., Hsieh, C., Lo, Y. (2013) Depth map estimation from a single video sequence. In IEEE International Symposium on Consumer Electronics, pp. 103-104, doi: 10.1109/ISCE.2013.6570130

[22] Yakubenko, M. A., Gashnikov, M. V. (2023) Entropy Modeling in Video Compression Based on Machine Learning. In IX International Conference on Information Technology and Nanotechnology (ITNT), Samara, Russian Federation, pp. 1-4, doi: 10.1109/ITNT57377.2023.10139143

[23] De Siva, N. H. T. M., Rupasingha, R. A. H. M. (2023) Classifying YouTube Videos Based on Their Quality: A Comparative Study of Seven Machine Learning Algorithms. In IEEE 17th International Conference on Industrial and Information Systems (ICIIS), Peradeniya, Sri Lanka, pp. 251-256, doi: 10.1109/ICIIS58898.2023.10253580

[24] Russell, B. C., Torralba, A., Murphy, K. P., Freeman, W. T. (2005). LabelMe: a database and web-based tool for image annotation. MIT AI LAB MEMO AIM-2005-025, SEPTEMBER, 2005

[25] Upulie, H. D. I., Kuganandamurthy, L. (2021). Real-Time Object Detection Using YOLO: A Review. DOI: 10.13140/RG.2.2.24367.66723

[26] labelme2yolo 0.1.3 (2023) Project Description. October 2023 release. https://pypi.org/project/labelme2yolo

[27] Ren, S., He, K., Girshick, R., Sun, J. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. Cornell University. Doi: https://arxiv.org/abs/1506.01497

[28] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., Berg, A.C. (2016). SSD: Single Shot MultiBox Detector. Doi: https://arxiv.org/abs/1512.02325

[29] Pytorch Foundation, (2016), PyTorch, https://pytorch.org/

[30] Ultralytics (2023), https://github.com/ultralytics/ultralytics

[31] *Jeffrey A. Clark. Pillow 10.3.0 (2024). https://pillow.readthedocs.io/en/stable/*