Hearing glasses to visually impaired people prototype

Vinícius Mafra Viti
Universidade Federal do ABC
Brasil

Bruno Augusto Dorta Marques
CMCC - Universidade Federal do ABC
Brasil

***Abstract*—This research aims to study the feasibility of glasses with a camera as a guide for the visually impaired. This idea will be developed to provide these people a way to "see, feel, and experience" the world in a different way. Thus, the development of such technology will be done employing Artificial Intelligence and Computer Vision resources/software. Some attention is given to visually impaired people, although many still experience great difficulty getting around and living alone, directly impacting their quality of life. Hence, implementing the proposed idea is extremely necessary since it will bring a sense of the environment to those who cannot visualize it and suffer daily. Therefore, the research will consist of conducting an improved study on the best methodology and technology to identify objects in front of blind people. The objective is to identify the most relevant objects, providing distance cues to the visually impaired person.**

I. Introduction

There are approximately 36 Million visually impaired persons in the world [1]. While innovation and technology advance more and more in people's lives, there are still challenges for these people in terms of accessibility. The difficulty in locomotion, the perception of where they are, and difficulty in reading visual signals harm and hinder their daily lives but are still very common.

In this sense, it is necessary to create some technology to help in the daily life of these people regarding locomotion and recognition of the environment and their surroundings. In this way, with the aid of some existing computer software, this work aims to develop a type of technology that provides help for these people.

The idea is to use glasses with a camera that will analyze and understand what is being visualized. It is worth pointing out that the idea is to recognize the object in front of the visually impaired person, and from this recognition, transmit the information to the visually impaired person in a friendly and accessible way.

II. Development

The technology developed in this work makes use of methods from computer vision [2], an area of knowledge that is inspired in the human vision to develop computer algorithms that have been revolutionizing science.

The concepts of computer vision work by processing real-world images in a computer, which helps create new technologies through applications such as the detection of visual information in the environment. In this work, the recognition of relevant objects is done through a webcam which transmits an image that is feed to an algorithm implemented using Python language [3], with the help of the OpenCV [4] library and a deep learning detection system powered by the YOLO [5], an object detection algorithm, making this idea possible.

A. Computacional Vision

A use-case study is the starting point for understanding the idea and challenges that the users face on a daily basis. With their feedback, the planning and implementation of the work can be done, providing and prototype based on a practical algorithm that is simple to use. The use-case study will provide information on the most important objects to be tackled by our method and how to present this information in a friendly and useful manner to the user.

The scope of this work starts with the development of the application based on the Python language. The pre-processing, image manipulation, and additional image processing are handled by the OpenCV library that provides ready-to-use algorithms and methods. With the image data formatted in the appropriate shape. Object detection is performed by a pre-trained neural network developed using the YOLO architecture. The YOLO convolutional neural network (CNN) is capable of identifying a group of pre-determined objects. An example of object detection is shown in Fig.1.

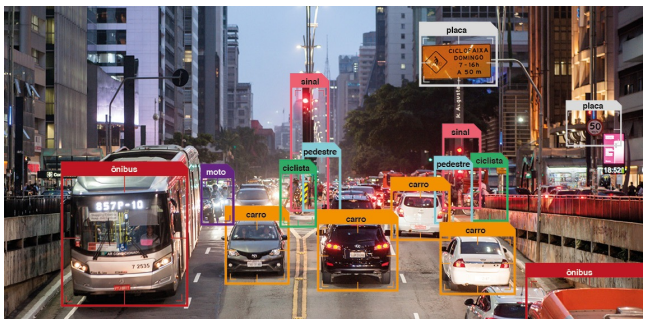

Fig. 1. Object detection example. The bounding box accompanied by its label identifies objects and their spatial position in the image. Source [6]

Initially, we chose to adopt a CNN pre-trained on the COCO [7] dataset for our first prototype. We plan to identify the most important objects and train a neural network on a new dataset that contains only the most relevant objects in the future.

The input of our algorithm is an image provided by a camera located in the glass of the user. We feed the frames captured by this camera to the python algorithm that should

output the detected objects and their spatial position. With the provided information, we are able to provide a piece of feedback information to the users about the environment in their surrounding area.

B. Object Detection

As shown in Fig.1, it is possible to notice several objects detected, which leads to some questions about how the programmed software can recognize each object? What differentiates them?

The computer vision method is based on a neural network that is capable of distinguishing each object based on its details and a kind of pre-defined notion that is learned through a training process in which we provide a dataset containing the objects and their position information. The training process is performed a single time. After this process is over, the algorithm is capable of recognizing the objects that were present in the image based on the provided dataset.

It is worth pointing out that these requirements for the application of the project will be refined in order to recognize objects closer to the visually impaired person and not unrelated/unimportant objects.

C. Convolutional Neural Network

A convolutional neural network [8] is a network that uses filters to identify what is being processed. The core idea of this type of technology is to process a set of filters on a given image, refining it to a certain level, creating a type of output data that allows the image to be represented.

This principle consists principally of three types of layers that serve to do this idea of content filtering, called: convolutional layer, pool layer, and fully connected layer. The extraction of this data is allowed by some weights. Such weights end up contributing to the definition of the image so that, in an overview, the filters applied to the images end up generating a certain final data that results in a vector containing information that aids in recognition of the image. This idea of convolutional layers aims to overcome an intense process of identifying and engineering features that help in decoding and analyze the images for detection, classification, and regression tasks.

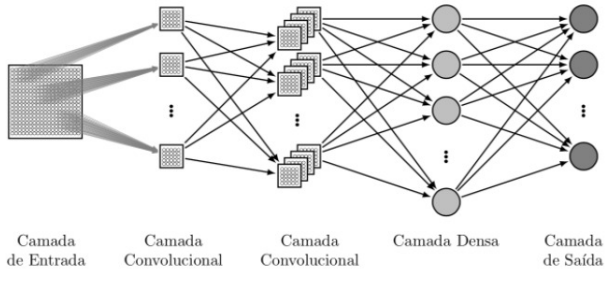

Fig. 2. EXAMPLE OF A CONVOLUTIONAL NEURAL NETWORK. Source [9]

D. Hardware Resources

With the software portion of the method worked out, we now have the question of defining the hardware technology that will be employed in this project. In computational terms, we choose to use a 'mini-computer' , Jetson Xavier (Fig.3), which is a deep learning specific hardware optimized for neural network inference that will be used to run all of the developed algorithm implementations.

For image capture, a high-quality webcam will be used to get every possible detail. Finally, to consolidate all these tools, we have the frame of a pair of glasses with a certain lens that will not affect the technical aspects. The frame should be robust enough to secure the webcam while allowing the user to be comfortable while using our prototype.

Fig. 3. JETSON XAVIER

The choice of each of these elements was made with the needs of the project in mind, where the glasses were taken into consideration due to the fact that they are in an excellent location that allows a large field of view for the accessory to identify as many objects as possible and thus avoid any kind of problem during detection.

This technology has the opportunity for drastic changes in the lives of several visually impaired people. Many people suffer nowadays in several aspects related to mobility because, besides the difficulty of having an idea of what is in front of them, there are still other issues, both social and infrastructural, that hinder the locomotion of these people.

This work aims to prove the viability of this concept by using well-established methods of computer vision. This work aims to be a technological innovation by providing an application that affects blind people, contributing to a new lifestyle with more safety and mobility.

III. Analyse and Tests

A user study will be performed with the objective of knowing and ascertaining the needs of a visually impaired person.

The user study will be held with a group of blind people, which will provide information about the main problems faced in their daily lives, thus helping in the best development of the technology in question. In addition, a code analysis will be done to adapt our methodology to their needs.

Since we use the YOLOv4 [10] of-the-shelf convolutional neural network for part of our computer vision solution, we will make performance tests regarding inference time, object detection accuracy, and usability of this algorithm. The main objective of those tests is to identify flaws in the algorithm in order to adapt and improve the CNN model for our use case. For example, it is important to detect true-negative inference results in order to ensure that our system is reliable for the real-use scenario.

IV. Future Works

As future work, we plan to improve the developed application to infer the distance of objects by employing depth-based estimations. With depth information, the application will be able to provide more accurate and valuable information to the users.

We also aim to research different architectures, to improve and optimize the inference time by reducing computational cost and allowing a smooth experience for the users. In addition, seeking to establish a method for that does not require a high-end GPU [11] for training purposes, we plan to study different models, such as YOLOv4-tiny, which focus on a CNN with few layers in order to improve the training and inference time.

As previously discussed in relation to neural networks, our initial work focus on the COCO dataset, a standard dataset for object detection. Although it contains a series of objects, we plan to augment this dataset by providing new samples that are relevant to our application.

V. Conclusions

A headset glasses aiming to help in the mobility and environment of visually impaired people is a suitable application of the existing computer vision technologies.

Our work purpose is to detect objects in the field of view of the user that may affect the trajectory of the visually impaired. While off-the-shelf solutions can perform the detection of many objects, this work aims to refine current solutions to identify and detect specifically the most relevant objects that aid the visually impaired in their daily lives. We focus on the accuracy of this recognition and the inference time. We plan to produce the first prototype with spatial information and improve this prototype in the future by providing depth cues to the estimation model. The addition of depth cues would allow the inference of distance and the ability to provide some kind of importance sampling of objects.

The development of our application and prototype is still in a working-in-progress phase. The application development is still in the early phases, but we are already obtaining few encouraging results. Although the tests are still in the planning phase, this is an important step to validate our approach. In addition, the study of some detection issues is still necessary. We aim to provide a precise classification to improve the usability of our method

References

[1] Fabio. (2020, apr) Estatísticas sobre deficiência visual no brasil e no mundo. [Online]. Available: https://louisbraille.org.br/portal/2020/04/13/estatisticas-sobre-deficiencia-visual-no-brasil-e-no-mundo/

[2] K. Engenharia. (2021, jun) VisÃo computacional. [Online]. Available: https://kotengenharia.com.br/visao-computacional-o-que-e/

[3] U. Roveda. (2020, oct) O que é python, para que serve e por que aprender? [Online]. Available: https://kenzie.com.br/blog/o-que-e-python/

[4] C. Technologies. (2020, jan) Opencv: Uma breve introdução à visão computacional com python. [Online]. Available: http://blog.cedrotech.com/opencv-uma-breve-introducao-visao-computacional-com-python

[5] F. Kühne. (2018, jun) Yolo: um sistema para detecção de classes de objetos em tempo real. [Online]. Available: https://www.profissionaisti.com.br/yolo-um-sistema-para-deteccao-de-classes-de-objetos-em-tempo-real/

[6] R. de Oliveira Andrade. (2019). [Online]. Available: https://revistapesquisa.fapesp.br/as-maquinas-que-tudo-veem/

[7] L. Tsung-Yi, "Microsoft coco: Common objects in context," in *European conference on computer vision*, 2014.

[8] A. M. P. C. e. C. N. V. Ana Caroline Gomes Vargas, "Um estudo sobre redes neurais convolucionais e sua aplicação em detecção o de pedestres," *SIBGRAPY16*, 2016.

[9] R. Sakurai. (2017). [Online]. Available: http://www.sakurai.dev.br/cnn-mapreduce/

[10] A. B. Chien-Yao Wang and H.-Y. M. Liao, "Scaled-yolov4: Scaling cross stage partial network," *COMPUTER VISION FUNDATION*, 2021.

[11] TecMundo. (2012, jun) O que é gpu? [Online]. Available: https://www.tecmundo.com.br/hardware/1127-o-que-e-gpu-.htm