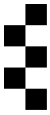

University of Essex

School of Computer Science and Electronic Engineering

MASTER OF SCIENCE IN INTELLIGENT SYSTEMS AND ROBOTICS

Explainable Machine Learning Model: smart glasses for blind people navigation

Carlos Romano Gómez

Supervisor: **Dr. Şefki Kolozali**

August 29, 2023
Colchester, Essex

Abstract

This is abstract text.

Acknowledgements

No need to include, but can if want to.

Contents

List of Figures

CHAPTER 1

Introduction

According to an assessment conducted by the World Health Organization (WHO), an estimated global populace of around 2.2 billion individuals struggles with visual impairments [1]. A considerable amount of this demographic group encounters challenges in navigation, particularly within social indoor settings such as universities, medical facilities, and commercial establishments like supermarkets.
Within these environments, the affected individuals frequently encounter difficulties in acquiring spatial orientation cues and directional information, depleting their ability to determining their whereabouts and optimal routes that will not damage them toward their destinations. This often hampers them and leads to a reluctance to go outdoors, which consequently affects in their social lives and contributing to their isolation. Furthermore, this situation affects people on a large scale, exerting damage on the integral development of young children and compromising the general quality of life experienced by adults. Consequently, along the growth of this high-priority circumstance there has been a noticeable increase in technological innovations. This surge in technological progress has notably generated the conception and construction of novel navigation assistance mechanisms [2] like seen in the recent years. Diverse designs of devices has surfaced with the intent of providing substantive support to visually impaired individuals. Noteworthy among these are devices such as white canes, Global Positioning System (GPS) trackers, smart glasses, and assorted Artificial Intelligence (AI)-driven methodologies [3]. These revolutionary efforts are underpinned by the overall goal of improving the sphere of mobility, fostering a greater sense of autonomy, and fostering a general improvement in the quality of life for people struggling with visual impairments.
Despite their reliability, these devices are not without limitations. For instance, certain methods encounter difficulties in detecting objects, particularly in specific situations, such as head-level objects, a challenge observed in white cane devices [4]. Additionally, it exists a lack of transparency in using deep learning techniques, which poses a significant concern. Users may struggle to comprehend the rationale behind the device's actions, limiting trust and confidence in its use.
Furthermore, developing robust and adaptable algorithms to accommodate various environments requires substantial computational power. Unfortunately, some of the current devices lack the necessary performance and computational power balance, leading to failures when navigating in new environments, restraining them of a better effectiveness and usability.
Addressing these limitations is of much importance to further enhance the functionality and impact of

assistive navigation devices for visually impaired individuals. Ongoing research and advancements in AI, Machine Learning (ML), and computer vision offer new opportunities to develop more transparent, efficient, and user-friendly solutions that can empower and provide them with improved independence and mobility to those with visual impairments.

1.1 Objectives

Hence, the prime endeavor of my research endeavors to present a comprehensive solution in the form of a navigation assistive apparatus tailored to surpass these multifaceted challenges. This apparatus will encompass an intricate system, meticulously orchestrated to involve a ML algorithm. This algorithm will be strategically deployed to effectuate object detection within the immediate surroundings, particularly in scenarios relevant to individuals navigating with visual impairments.
To impart a heightened measure of interpretability, the devised system will utilize tensor decomposition, more precisely, the Canonical Polyadic (CP) decomposition approach, along with tensor regression. This choice is propelled by its potential to facilitate an interpretive dialogue between the user and the model, thereby enhancing the transparency of the decision-making process [5]. Notably, this systematic innovation will rigorously pursue equilibrium between computational power and precision, thereby encouraging cautious compensation. The overall goal is to enable timely and astute object detection without compromising the algorithm's versatility in various contextual scenarios.
In the pursuit of operationalizing the envisioned algorithms, the research is willingly to address the ensuing inquiries:

1. Can a interpretable system perform as well as deep learning methods?
2. Can such system efficiently work on a resource constrained device?
3. Can using smart glasses provide a better user experience for blind people?

1.2 Structure

The next sections of this document will explain the following: In chapter 2, a comprehensive review and analysis of relevant literature and existing works related to assistive navigation devices for visually impaired individuals will be presented and discussed. The aim is to identify the strengths and weaknesses in the current devices, providing a foundation for the proposed methodology. In chapter 3 the proposed methodology for the assistive system will be introduced. This encompasses the design and integration of a interpretable algorithm into user-friendly smart glasses, aimed at aiding visually impaired individuals in navigating their surroundings. Also containing the theoretical and technical aspects of the complex algorithm, incorporating Tensor Decomposition techniques. In chapter 4 the outcomes and findings of the implemented assistive system will be presented. Performance metrics, including accuracy, recall, and precision, will be used to evaluate the effectiveness of the algorithm in object detection and decision-making. The discussion will delve into the implications of the results and provide insights into the device's real-world applicability and potential areas of improvement. Finally, chapter 5 will offer a comprehensive conclusion based on the study's outcomes and analysis, summarizing the main contributions of the research and its significance in the field. Additionally, this section will outline potential future research directions and improvements to further enhance the device's capabilities and impact.

CHAPTER 2

Related Work

The text goes here ...

2.1 Assistive Technologies for Visually Impaired Individuals

Ademas al ser unos lentes se ofrece una libertad de movimiento en las manos que permite al usuario realizar mayores acciones.

Review existing assistive devices such as white canes, GPS trackers, and smart glasses. Discuss the strengths and limitations of these devices in aiding blind people's navigation.

2.2 Machine Learning in Assistive Navigation

Explore previous studies that have applied ML techniques for object detection and navigation assistance. Highlight the role of ML algorithms in enhancing the accuracy and effectiveness of assistive devices.

2.3 Tensor Decomposition in Machine Learning

In order to came up with an innovative algorithm, it is necessary to comprehend the alternatives that exists to the traditional neural netowrks. To this end, an exhaustive exploration of scholarly literature in the domain of neural networks has substantively informed the conception of this labor.

Before continuing, an integral part of this discourse is the definition of the term "tensor", which serves as a fundamental unit of information encompassing complex data constructs. Tensors make it easy to analyze multidimensional data, thus unraveling latent patterns that can evade detection when analyzed within a solitary dimension. For example, this concept finds prominent application in the realm of multimodal data analysis, where tensors, often recognized as matrices, skillfully navigate various dimensions to extract information. Illustratively, in the context of images, tensors effectively encapsulate the trifecta of height, width, and color dimensions, facilitating an understanding of the data set and what it contains.

A Tensor Regression Network (TRN),this type of neural network, in addition to being recent, its construction is similar to the already existing Convolutional Neural Networks (CNN) since they take

advantage of the input of three-dimensional tensors of the convolutional layers of the CNNs, which allows maintaining a multilinear structure but Unlike these, they are not connected to a fully connected layer, but to a tensor contraction layer to preserve the multilinearity and multimodality that the system may have, also reducing the number of necessary parameters.

Tucker decomposition, which has good properties but loses its efficiency in compressing tensors very quickly as the order of tensors increases - An efficient tensor regression for high-dimensional data - Yuefeng Si

Provide an overview of tensor decomposition techniques and their applications in various domains. Discuss the potential benefits of utilizing tensor decomposition for developing the ML model.

CHAPTER 3

Methodology

This document presents a comprehensive proposal describing a revolutionary smart glasses system designed to boost the navigation capabilities of individuals afflicted by visual impairments. The core of this proposal resides in the incorporation of an innovative algorithm synergistically built upon a 3D printed frame, a depth camera and an onboard computational unit. The primary ambition is to generate a sophisticated apparatus equipped with the ability to discern objects within the user's proximate vicinity, subsequently alerting the user concerning the identified object as well as its spatial proximity. The schematic delineating the operational intricacies of this system is depicted in Figure 3.1 and will be explained later in each subsystem's chapters.

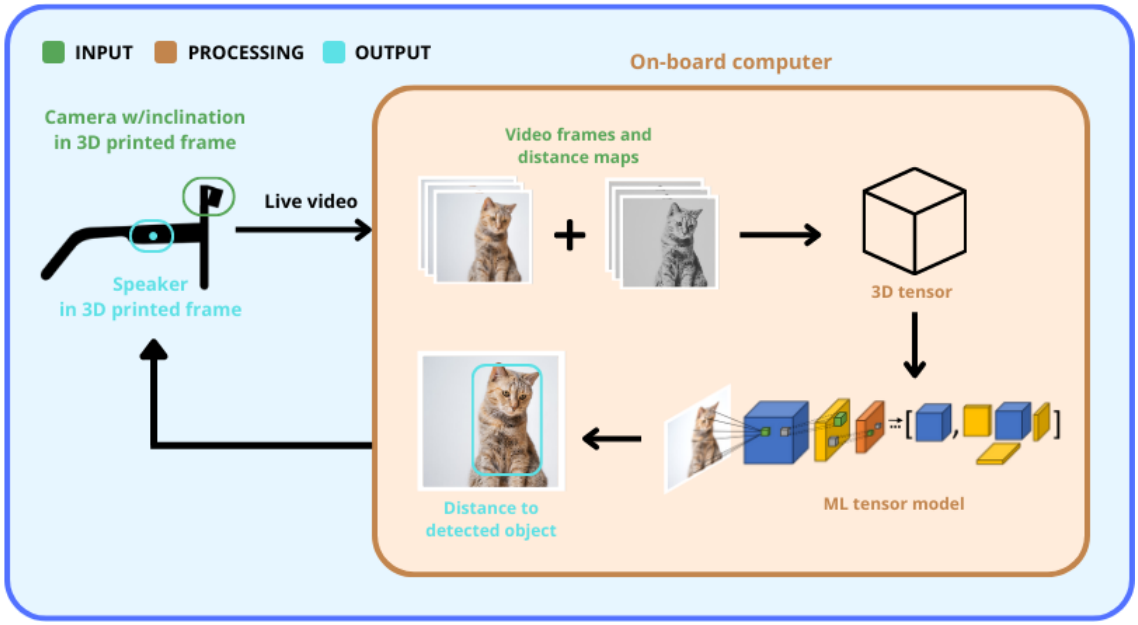

Figure 3.1: System architecture. An image its capured by the camera, passed as frames to the ML tensor model and onced processed, feedback is sent to the user.

In essence, the envisioned system encompasses two vital phases that function in unison to grant its operational efficacy. The software component, constituting the ML algorithm, stands by an explicable algorithm that ingeniously strikes a harmonious equilibrium between computational efficiency and power.

This algorithm assumes a crucial role in orchestrating multifarious tasks spanning from data collection and model training to real-time user feedback, which are the system's core functionalities. This facet prioritizes a proper balance between computational robustness, precision, and system performance. The hardware facet, on the other hand, revolves around the fabrication of the glasses frame tailored to accommodate visually impaired users, alongside the provisioning of a portable element for the onboard computational entity. This facet, resonating with user-centric considerations, underscores the importance of user comfort and operational convenience in the overall design.

3.1 Smart Glasses

In order to provide the proposed system with the necessary portability and the perfect integration for users, it is necessary to adopt an innovative strategy. Central to this strategy is the integration of a meticulously designed 3D printed glasses frame, which assumes the role of containing the camera apparatus for detecting objects within the wearer's immediate view. Importantly, the inherent versatility of this design is highlighted by the provision of strategically allocated spaces, prepared to accommodate potential sensor integration's, ensuring a forward-thinking approach to technology augmentation.
Of great significance is the configuration of the camera's mounting. Positioned at the forefront of the frame, this apparatus assumes an inclination of a predetermined angle. This particular configuration, denoted by the angle θ, serves as a critical determinant in the camera's viewing spectrum. By properly tilting the camera, the resulting panoramic visual expanse encompasses the frontal domain, amplifying the detection range manifold. This efficacious design empowers the camera to discern objects occupying varying altitudes, including those at head level and ground level. The illustrative elucidation of this design elucidation is graphically depicted in the Figure 3.2.

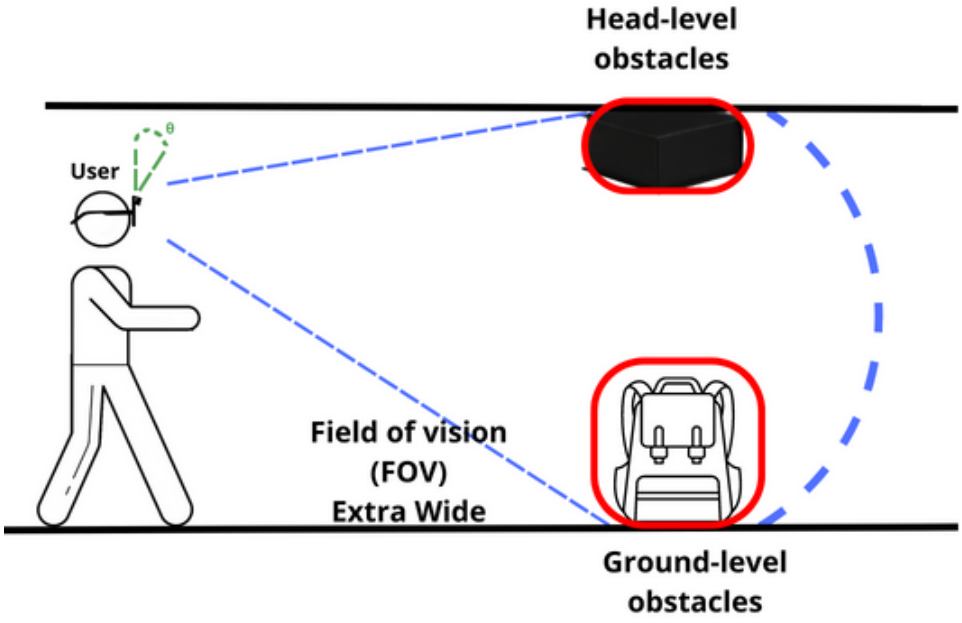

Figure 3.2: A side view of the system. The system will be able to detect objects in a head level and ground level, due to an inclination of the camera in the frame.

In this project, the ZED Mini camera, a product of the StereoLabs company, was employed. This camera system encompasses two distinct cameras, each characterized by unique functionalities. The

first camera claims a resolution of up to 2.2K and exhibits a variable frame rate of up to 100 Frames Per Second (FPS) depending on the chosen resolution. The second camera, in addition to its imaging capabilities, possesses the capacity to perform depth measurements at a range of up to 15 meters [6]. The carefull selection of this device arise from its multifaceted attributes, including its compact form factor. The intrinsic versatility of the ZED Mini camera is a defining factor in its selection. This versatility is exemplified by its efficacy in diverse data acquisition scenarios. The camera accommodates data acquisition via recording, facilitating the accumulation of training data, as well as real-time data acquisition via streaming, thereby catering to continuous user engagement. Moreover, the camera's inherent adaptability extends to the domain of image quality. The ability to manipulate image quality, whilst influencing the rapidity of user response, stand out the dynamic responsiveness to this system.

It should be noted that the camera is of greater importance since it also has auxiliary capabilities. Among these features are motion sensors, that is, sensors such as gyroscope and accelerometer to detect the movement of the camera, and also has pre-installed object detection models. While these ancillary features remain secondary to the current project goals, their potential addition holds promise for subsequent iterations of this technological innovation.

Similarly, to achieve the required computational prowess while preserving the portability of the device, the adoption of an on-board computer or carrier board was considered appropriate. Here, an onboard computer refers to a processing power unit embedded within a carrier circuit board along with all the essential laptop-like components, including RAM, cooling mechanisms, and add-on components. Notably absent are peripheral devices like monitors, mice, and keyboards. But in return, a carrier board can encompass ports for USB or Ethernet connectivity and come with an operating system pre-installed, facilitating optimal functionality.

In light of this, the Jetson Orin Nano on-board computer by Nvidia surged as the chosen solution. This compact computing device, ideal for the creation of entry-level AI-powered robotics, intelligent drones, and smart cameras [7] akin to the one envisaged within this document, was deemed optimal. The selection of the Jetson Orin Nano was primarily driven by Nvidia's offerings in terms of video components, capable of expeditiously processing multimedia data. Furthermore, this diminutive computing marvel aligns seamlessly with the implementation of AI algorithms, offering an interface for swift and efficient integration. Given these considerations, in the midst of the large number of specifications offered by both devices, this project has opted for the adoption of LOSSLESS image compression for the camera, specifically at a scale of 2:1. This choice emanates from the innate properties of the specific Nvidia Jetson model used in this context, due to this particular model lacks a hardware accelerator, which prevents the adoption of more aggressive image compression techniques that can yield higher compression rates. Consequently, a setting of 60 frames per second (FPS) along with a high-definition (HD) resolution of 720 was identified as the fastest configuration, although the resulting operating speed measures approximately 10 FPS. This noticeably leisurely pace, however, it also means that the model will have the ability to encapsulate substantial information within each frame, a critical attribute that enables efficient user alerts.

Although the above deliberations, fundamentally address the viability of implementing an algorithm within resource-constrained device, to address the user comfort, the merits of harnessing 3D printing technology have been studied. This strategic choice arises from multifaceted considerations. Notably, the economic advantages inherent in 3D printing, coupled with its inherent adaptability to personalized customization, render it an optimal avenue for both prototyping and potentially for the eventual definitive model. Furthermore, the innate modular nature of 3D printing facilitates easy incorporation of modular

components, thereby harboring latent potential for future enhancements. This portends the prospect of seamlessly integrating adjunct accessories, including microphones, to enhance the object detection capabilities, as previously elucidated in this exposition. Additionally, the incorporation of 3D printing technology has led to the creation of a protective casing, cleverly designed to house the on-board computer and accompanying power supply. This strategic innovation is of great importance as it fulfills the dual purpose of improving security and guaranteeing the longevity of the system, increasing its durability and resistance.

3.2 Tensor Decomposition Algorithm

The fundamental idea underlying the formulation of this algorithm resides in the pursuit of innovative methodologies to develop a novel problem-solving prototype, characterized by an inherent capacity for a custom-made explanation to user comprehension. Particularly, the conception that emerged is stated upon the combination of factor decomposition and a neural network grounded in tensor regression. Notably, this union is presented as a novel approach, prepared to offer a greater degree of versatility in the assimilation of information within neural networks. This greater versatility is underscored by the inherent structural variability of tensors, spanning a wide spectrum of dimensional manifestations.
As highlighted above, the training phase of the algorithm was done offline, using recorded videos obtained from ZED's camera. However, during the testing phase, a real-time streaming approach was adopted. This strategic difference in data acquisition methodologies facilitated the creation of distinct codes capable of discerning various attributes, such as color images, spatial depth representation, and individual depth values. This heterogeneous information formed the basis on which the resulting tensors were built, invoking a composite combination of two dimensions and two joined matrices. This way of obtaining precision facilitated the analysis of simultaneous data, leaving aside concerns related to the scarcity of data within the tensors, as well as their sparseness.
This creation ends in the formulation of a height tensor, denoted as H, which naturally corresponds to the vertical dimension of the image. Within the scope of this experiment, where a resolution of 720 pixels was maintained, the height tensor, H, is inherently held at 720 pixels. Whereas, the width tensor, W, emerges as a composite construction, divided to encompass two discrete domains. The first part of W corresponds to the values that encapsulate the color image, while the second facet correlates with the values that delimit the representation of spatial depth. Taken together, this composite-width tensor holds a total of 2,560 distinct values, encapsulating the information inherent in the algorithm's data construct. This structural framework finds a visual appearance in Figure 3.3.

Operational within the algorithmic framework, the architecture constitutes a tensor regression neural network. Its preliminary structure involves convolutional layers, designed to extract salient features related to each image, and unlike conventional processes, the back layer abandons the standard flattened configuration. Instead, it adopts a new tensor contraction layer, naturally calibrated to preserve multilinear attributes related to the tensor data. This unique architectural choice encourages the preservation of the multilinearity of the tensor and avoiding the possible sparcity and loss of information. This preservation can also be effected through tensor decomposition, thus effectively smoothing the number of tensor ranks, an approach supported by its ability to improve computational efficiency while preserving essential nuances in the data. Towards the end of the traversed path of the network, the result is manifested as a product that involves the resulting tensor, synthesized through the algorithmic process, together with a low

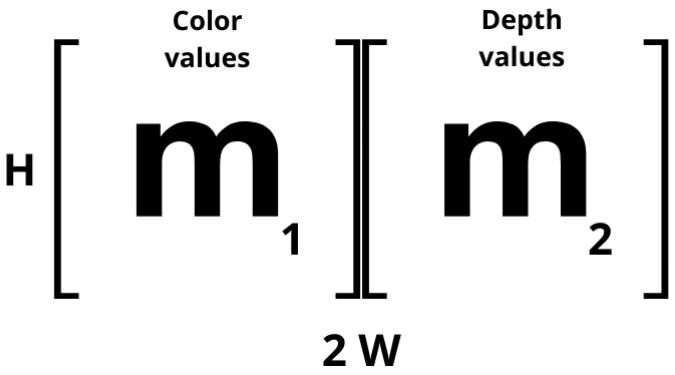

Figure 3.3: A visual representation of the input tensor.

rank tensor, which contains the corresponding weighting coefficients [8]. This product, similar to what happens in a conventional neural network, is subsequently subjected to the backpropagation process, a major component for solving the optimization of neural networks. This procedure, which is run iteratively over several epochs, leads to refinement and alignment of the network parameters, ultimately culminating in the achievement of optimal performance.

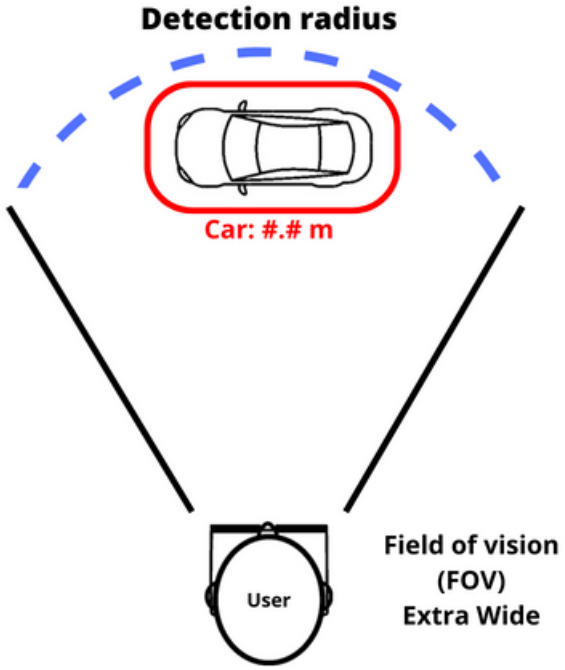

Figure 3.4: A top view of the system. The system will be able to detect objects in front of the user in a wide view angle and give feedback about the distance of the detected objects.

CHAPTER 4

Results and Discussion

The text goes here ...

4.1 Your first section of the first main chapter

... goes here.

4.2 Your second section of the first main chapter

... goes here.

CHAPTER 5

Conclusions

FURTHER WORK IF NOT DONE IN THIS
TENSOR REGRESSION LAYER + DROPOUT FOR MAKING THE MODEL A LITTLE BIT MORE ACCURATE AND MUCH MORE ROBUST TO NOISE, THAT NORMALLY WILL LEAD TO A MISCLASSIFICATION, THING THAT WE CAN NOT PERMIT WHILE PEOPLE'S SECURITY IS INVOLVED.

TAKE ADVANTAGE OF THE OTHER TOOLS THAT ZED CAMERA OFFERS, LIKE SPACIAL MAPPING TO MAP BLINDS HOUSES AND GUIDE THEM WITH EASE AND/OR POSITIONAL TRACKING TO REMEMBER WHICH PATH TAKE TO RETURN.

INTEGRATE OTHER SENSORS LIKE THE MICROPHONES TO A BETTER DETECTION OF WHERE THE OBJECTS ARE

APPENDIX A

A Long Proof

Mathematics is added using dollar signs for in-line math, i.e. $x^2 + y^2 = z^2$, or by using open-bracket close-bracket for a displayed equation.

$$c^2 = a^2 + b^2 - 2ab\cos\theta.$$

Example A.1. This is an example.

Lemma A.2. *This is a lemma.*

Definition A.3. In 1950, Alan Turing published an article in *Mind* titled "Computing Machinery and Intelligence" where he considered the question "Can machines think?". This is known as **Turing's Test**.

Remark A.4. This is a very important remark.

APPENDIX B

Another Appendix

Text goes here

Bibliography

[1] WHO, "Blindness and vision impairment." https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment, October 2022.

[2] A. Bhowmick and S. M. Hazarika, "An insight into assistive technology for the visually impaired and blind people: State-of-the-art and future trends - journal on multimodal user interfaces," 1 2017.

[3] M. P. de Freitas, V. A. Piai, R. H. Farias, A. M. R. Fernandes, A. G. de Moraes Rossetto, and V. R. Q. Leithardt, "Artificial intelligence of things applied to assistive technology: A systematic literature review," *Sensors (Basel)*, vol. 22, p. 8531, Nov. 2022.

[4] R. Manduchi and S. H. Kurniawan, "Mobility-related accidents experienced by people with visual impairment," *American Economic Review Journal*, vol. 4, no. 2, 2011.

[5] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine learning interpretability: A survey on methods and metrics," *Electronics*, vol. 8, no. 8, 2019.

[6] S. Labs, "Zed mini camera and sdk overview." https://cdn2.stereolabs.com/assets/datasheets/zed-mini-camera-datasheet.pdf, 2018.

[7] Nvidia, "Jetson orin nano developer kit carrier board." https://bit.ly/3DWidrQ, 2023.

[8] J. Kossaifi, Z. C. Lipton, A. Kolbeinsson, and A. Khanna, "Tensor regression networks," *Journal of Machine Learning Research*, 2020.