IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 1, February 2024

Empowering Independence through Real Time Object Identification and Navigation for People with Disabilities

Tulika Biswas[1], Rounak Kumar[2], Karthik Jain[3], Dr. Nirmala H[4]

Undergraduate Students, Department of Information Science and Engineering[1,2,3]

Professor, Department of Information Science and Engineering[4]

Global Academy of Technology, Bangalore, India

Abstract: *Recent studies in assistive technologies for visually impaired individuals showcase a diverse range of methodologies, algorithms, and implementations aimed at enhancing their independence. A notable focus revolves around leveraging cutting-edge technologies such as YOLO (You Only Look Once), SSD (Single Shot Multibox Detector), and Faster R-CNN (Region-based Convolutional Neural Network) to develop real-time object detection systems and deep learning-based smartphone navigation solutions . One prevalent theme in these advancements is the incorporation of auditory feedback to facilitate enhanced user interaction. This is achieved through sophisticated text-to-speech conversion and the integration of audio cues. The utilization of auditory cues not only aids in real-time awareness of the surroundings but also significantly contributes to the overall user experience . Despite remarkable progress, challenges persist in the realm of assistive technologies for the visually impaired. Issues such as processing speed, the occurrence of false positives and negatives, and the adaptability of these systems to various environmental conditions remain prominent. These challenges underline the need for continued research and development in this field to address existing limitations and refine the effectiveness of these assistive technologies .In essence, this survey provides a comprehensive understanding of the current landscape of assistive technologies for the visually impaired. By identifying both achievements and existing challenges, it serves as a valuable resource for researchers and practitioners, contributing to ongoing advancements that ensure tailored solutions and improved independence for individuals with visual impairments.*

Keywords: Visual Impairment, Assistive Technologies, Object Detection, Smartphone Navigation Deep Learning, Auditory Feedback, Indoor Navigation

I. INTRODUCTION

In recent years, there has been a growing focus on leveraging advanced technologies to address crucial societal issues. This paper delves into the realm of assistive technologies designed for the visually impaired, particularly those utilizing deep learning and artificial intelligence (AI). The exploration is based on a comprehensive survey of 25 relevant research papers, each contributing unique methodologies and insights.The survey encompasses diverse approaches, ranging from smartphone-based navigation assistants to real-time object detection systems and wearable devices. Noteworthy methodologies include the integration of state-of-the-art models like YOLO, SSD, and Faster R-CNN, coupled with innovative features such as text-to-speech conversion, motion tracking, and multi-object detection. Despite remarkable advancements, the surveyed papers acknowledge certain limitations, including challenges in real-world adaptability, processing speed, and occasional false positives/negatives .By synthesizing these research findings, this paper aims to contribute to the ongoing discourse on enhancing the lives of visually impaired individuals through cutting-edge technology. The subsequent sections will delve into the methodologies, limitations, and key insights gleaned from the surveyed papers, offering a comprehensive overview of the current landscape in the field of assistive technologies for the visually impaired.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

II. LITERATURE STUDY

In [1],It addresses the challenges faced by visually challenged individuals, emphasizing the importance of real-time object detection for their mobility and safety. It explores existing works using various algorithms like YOLO, Faster R-CNN, and MobileNet SSD. While acknowledging advancements, the review highlights the need for a system compatible with low-computation devices like smartphones, providing a foundation for the proposed research agenda. [2] It introduces the application of deep learning, specifically YOLOv2, for real-time object detection on mobile platforms in the context of Industry 4.0. It highlights the significance of smart and autonomous mobile platforms in optimizing production processes. The review emphasizes YOLOv2's advantages, such as its speed, precision, andability to detect a wide range of object classes, making it an optimal choice for integration into mobile robots. [3] discusses the evolution of navigation systems for visually impaired individuals, focusing on smartphone-based solutions. Highlighted are obstacle recognition techniques using Faster R-CNN and YOLO algorithms, a proposed system with stable and fast modes for varied scenarios, and its superior capabilities over traditional methods. User feedback supports the system's effectiveness, with acknowledgment of support from the Ministry of Science and Technology in Taiwan. In [4], The YOLO (You Only Look Once) approach revolutionizes object detection by framing it as a regression problem. In contrast to traditional methods, YOLO employs a single neural network for end-to-end optimization, achieving real-time processing speeds and outperforming existing detectors. It addresses challenges from deformable parts models, R-CNN, and other fast detectors, presenting a unified, efficient model for general-purpose object detection with broad applicability. Acknowledgments include support from ONR, NSF, and The Allen Distinguished Investigator Award. [5] The introduction of YOLOv4_Resnet101 for VIPs attains a remarkable 96.34% accuracy in obstacle detection, integrating text-to-speech for auditory feedback and incorporating image preprocessing. The customizable system strikes a balance between speed and accuracy, yet lacks explicit discussion on development challenges. Continuous refinement is emphasized for practical applications, with future endeavors focusing on exploring aural and touch-based feedback avenues.[6] The proposed system offers a simple, affordable, and user-friendly object detection solution for the visually impaired, leveraging a smartphone and the YOLO V3 algorithm. Emphasizing real-time detection, YOLO V3 enhances processing speed crucial for user interaction. The literature review highlights the system's reliance on accurate and efficient audio communication to convey real-time object information to users.[7] the author developed a women's safety system incorporating GPS and fingerprint technology to address the rising crimes against women. This portable device, activated by fingerprint scanning, sends the user's location to authorized personnel and predefined contacts via SMS during emergencies.[8] The literature highlights a real-time object identification system for the visually impaired, employing Raspberry Pi and YOLO algorithm. Incorporating Pyttx3 for audio feedback, the outdoor-focused system relies on YOLO accuracy, addresses dynamic challenges, and assumes focal length for effective machine learning-based object recognition.[9] The literature reflects on a CNN-based SSD model elevating object detection for the visually impaired, surpassing Fast R-CNN in mapping and speed. Challenges include distant object recognition, necessitating extensive training. Recommendations for future enhancements involve exploring object tracking techniques. The study underscores advancements in computer vision technologies for aiding individuals with visual impairments and identifies areas for further research to optimize object detection systems.[10] , The literature highlights a novel mobile-based navigation system employing deep learning algorithms for object recognition and distance estimation via smartphone cameras. Challenges include limited recognition of predefined objects, emphasizing the necessity for augmented datasets. Addressing accurate depth estimation remains a challenge, underscoring the significance of advancements in deep learning for improving mobile navigation systems.

[11] The study presents a novel approach to real-time hand gesture recognition for human-machine interaction, employing a Faster R-CNN Inception V2 model. The research addresses challenges in unconstrained conditions, utilizing a custom dataset for training. Evaluation includes diverse optimization algorithms, emphasizing the superiority of the ADAM optimizer. The proposed model demonstrates promising accuracy and efficiency, contributing to advancements in pattern recognition-based gesture recognition systems.[12] The paper proposes a lightweight hand gesture recognition model based on YOLOv3 and DarkNet-53 convolutional neural networks, emphasizing real-time detection without additional image preprocessing. The model exhibits remarkable accuracy, reaching 97.68%, even in complex environments and low-resolution settings. Comparative analysis with SSD and VGG16 demonstrates the superiority of the proposed model, achieving an accuracy range of 82-85%. The study underscores the potential

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 1, February 2024

applications in communication aids for the deaf, gesture-based signaling systems, and other domains, highlighting the need for YOLO-labeled datasets to further advance research in hand gesture recognition. Future endeavors aim to enhance the model's capabilities for detecting multiple gestures simultaneously and exploring hybrid methods for broader applications in smart mobile applications and robotics.[13] This paper addresses the challenging task of real-time hand gesture recognition under unconstrained conditions using a Faster R-CNN Inception V2 model. The proposed model, trained on a custom dataset captured in diverse environments, demonstrates superior performance, particularly with the Adam optimizer. Achieving an average precision of 0.991 and efficient prediction times, the study emphasizes the significance of Convolutional Neural Network architectures for accurate and timely object detection in dynamic settings.[14] This study introduces a robust approach for hand gesture recognition using the Faster-RCNN deep learning algorithm with a five-layer neural network. Leveraging a benchmark database with diverse hand gestures, the proposed method demonstrates exceptional accuracy (up to 99.2%) in complex environments. The work emphasizes the significance of efficient hand segmentation, employing skin color segmentation in the preprocessing stage, and highlights the Faster-RCNN's effectiveness for real-time human-computer interaction applications.

III. ARCHITECTURE

1. Input Source: The system takes input from a live camera feed, capturing real-time visual information of the user's surroundings.
2. Object Detection and Classification: Utilizes YOLOv2 (You Only Look Once) and Faster R-CNN (Convolutional Neural Network) for object detection and classification in the camera feed.
3. Real-Time Scene Understanding: Combines YOLOv2's speed and accuracy with Faster R-CNN's image recognition to provide a detailed scene description, including object colors and room layouts.
4. Gesture-to-Speech Conversion: Implements gesture recognition to convert user gestures into speech for effective communication.
5. Text-to-Speech Conversion: Integrates text-to-speech functionality to verbally communicate written content, enhancing communication capabilities.
6. Image-to-Text/Speech: Provides the capability to convert visual information into text or speech, offering a comprehensive understanding of the surroundings.
7. SSD Integration: Incorporates Single Shot Multibox Detector (SSD) to further enhance real-time object detection, ensuring speed and accuracy, crucial for dynamic environments.
8. Natural Language Processing (NLP): Utilizes NLP to convert spoken language into text and vice versa, facilitating communication for individuals who are deaf and blind.
9. Privacy Safeguards: Implements robust measures to prioritize user privacy, ensuring compliance with ethical considerations and building user trust.
10. Continuous Improvement Mechanism: Includes a continuous improvement mechanism to adapt to evolving technologies, ensuring ongoing relevance and addressing emerging needs of individuals with disabilities.

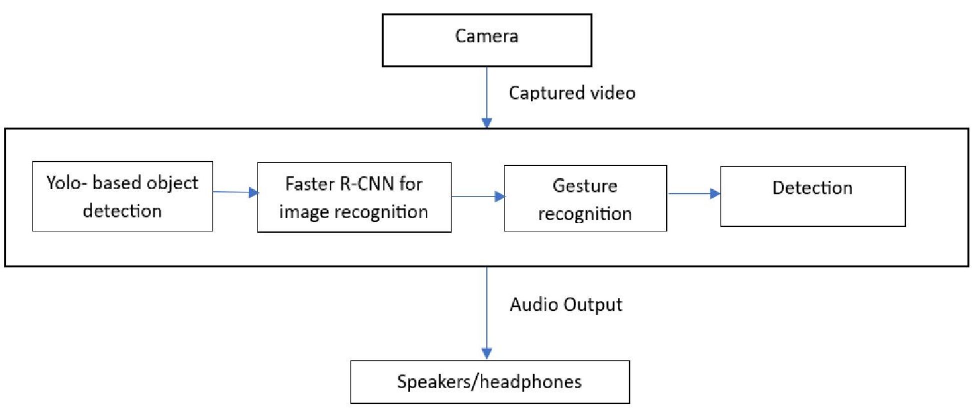

Fig. System Architecture

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 1, February 2024

The system's operation commences with the acquisition of a live camera feed, facilitated either through a device-connected camera module or an integrated camera within a smartphone. This real-time video stream undergoes intricate processing using the YOLO (You Only Look Once) deep-learning algorithm, enabling the system to perform instantaneous object detection and classification. The objects identified, along with their respective classifications, are meticulously extracted, laying the groundwork for subsequent in-depth analysis. To fortify the system's image recognition capabilities, a Faster R-CNN (Region-based Convolutional Neural Network) is seamlessly integrated. This component goes beyond mere object identification by extracting additional features, such as colors, shapes, or spatial layouts, providing a rich and nuanced understanding of the environment. For an added layer of refinement in real-time object detection, the architecture optionally incorporates the Single Shot Multibox Detector (SSD).

A pivotal facet of user interaction involves a gesture recognition module, which interprets user gestures with precision and translates them into meaningful commands. Recognized gestures have the potential to trigger specific actions within the system or prompt the generation of corresponding speech output. In parallel, real-time auditory announcements are dynamically generated based on the amalgamation of identified objects, interpreted gestures, or user-initiated commands. These auditory cues provide immediate and relevant insights into the user's surroundings, significantly enhancing user confidence and ensuring a heightened sense of safety. This comprehensive and integrated approach underscores the system's commitment to delivering a sophisticated and user-centric experience in object identification and environmental awareness.

IV. METHODOLOGY

The methodology for implementing the project "Empowering Independence Through Real-Time Object Navigation for People with Disabilities" encompasses a meticulously designed approach to seamlessly integrate cutting-edge technologies while placing a paramount focus on user-centricity, ethical considerations, and robust privacy safeguards. To commence, the system initiates the process by acquiring input from a live camera feed, ensuring the continuous capture of the user's surroundings in real-time. The core of the object detection and classification process lies in the utilization of state-of-the-art technologies, namely YOLOv2 (You Only Look Once) and Faster R-CNN (Region Convolutional Neural Network). This combination facilitates the accurate identification and classification of objects present in the captured images. The integration of Single Shot Multibox Detector (SSD) further amplifies the system's capabilities, ensuring swift and precise detection of objects, especially in dynamic and rapidly changing environments. Real-time scene understanding forms a pivotal aspect of the methodology, achieved through the synergistic amalgamation of YOLOv2's speed and Faster R-CNN's image recognition prowess. This comprehensive scene analysis extends beyond mere object identification, encompassing intricate details such as object colors and the spatial layout of the surroundings. The integration of gesture recognition technology is another critical element, enabling the conversion of user gestures into speech. This innovative feature not only facilitates effective communication but also enhances the overall user experience. The project takes a stride further by incorporating text-to-speech functionality. This addition empowers the system to articulate written content audibly, catering to users who may face challenges in reading text. Furthermore, the implementation of image-to-text and speech capabilities ensures a nuanced understanding of visual information, offering a more comprehensive and inclusive experience. Natural Language Processing (NLP) is seamlessly woven into the fabric of the project, playing a pivotal role in converting spoken language into text and vice versa. This inclusion addresses the specific needs of individuals who are deaf and blind, fostering a more inclusive and accessible communication environment. Privacy safeguards are a non-negotiable aspect of the methodology, with robust measures implemented to protect user privacy and adhere to stringent ethical standards. The system is designed to prioritize user confidentiality and data security, fostering a sense of trust and reliability. In an ever-evolving technological landscape, the project embraces a continuous improvement mechanism. This iterative process ensures adaptability to emerging technologies, incorporation of user feedback, and a commitment to delivering a solution that remains not only relevant but also highly effective. Through this holistic methodology, the project endeavors to empower independence by providing individuals with disabilities a comprehensive, real-time understanding of their surroundings while upholding the principles of user privacy, ethical considerations, and inclusive design.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

V. CONCLUSION

In summary, the surveyed papers showcase a collective effort to address challenges faced by visually impaired individuals and enhance their safety using advanced technologies. The implementation of deep learning, computer vision, and AI in navigation and object detection systems reflects significant strides toward inclusivity and independence. For the visually impaired, smartphone-based navigation assistants and object recognition systems demonstrate progress but face challenges in processing speed and adaptability. Ongoing refinement is crucial for optimal performance. In the context of women's safety, the integration of biometric verification, IoT, and mobile applications highlights a commitment to leveraging technology for personal security. Emphasis on real-time location sharing and immediate assistance reflects a holistic approach to empowering. Despite unique contributions, common limitations include dataset constraints and the need for ongoing improvements. The surveyed literature forms a foundation for future endeavors in assistive technology and safety applications, paving the way for more responsive systems addressing diverse user needs and societal challenges.

VI. FUTURE WORK

Achieving unparalleled accuracy and versatility in object identification systems is made possible through the revolutionary hybrid model that integrates YOLO, SSD, and Faster R-CNN. This innovative approach harnesses the real-time capabilities of YOLO, the precision of Faster R-CNN, and the balancing features of SSD, amalgamating the strengths of each model. This amalgamation ensures optimal performance of the object detection system across various scenarios, delivering precise results for diverse use cases. An integral aspect of this system is the incorporation of natural language processing and improved text-to-speech for context-aware auditory cues, enhancing both accessibility and user interaction. This advanced auditory feedback not only elevates the user experience but also fosters inclusivity, making the system more user-friendly.The real-world adaptability testing phase involves subjecting the system to rigorous. trials in diverse settings, a crucial step in its development. This testing methodology aims to enhance the system's adaptability and ensure its reliable operation in various scenarios. By exposing the system to a spectrum of situations, developers can identify and address potential issues, thereby refining and optimizing its overall functionality. Efficiency is further boosted through dataset expansion and bias reduction. Enlarging datasets to encompass a broader range of objects enhances the system's coverage and object recognition capabilities. Simultaneously, reducing biases in the dataset promotes demographic inclusivity, ensuring effective functionality across diverse user profiles and backgrounds. The integration of touch-based feedback introduces a tactile dimension to the user experience, exploring haptic cues to improve spatial information and enable tactile navigation. This not only enhances the system's flexibility but also adds an extra layer of interaction for all users, with particular benefits for those with visual impairments. Ensuring the system's adaptability to low-power devices is crucial for widespread adoption. Optimizing deep learning models for efficiency on low-power devices extends the system's usability to a broader range of hardware, promoting accessibility and usability across various platforms. Continuous model refinement is an iterative process driven by user feedback, addressing false positives and negatives, optimizing models, and enhancing overall adaptability. User input keeps the system at the forefront of object detection technology, allowing it to evolve to meet changing needs.The integration of wearable devices, such as smart glasses or vests, elevates the user experience by providing a hands-free and immersive engagement. This not only enhances user convenience but also opens up possibilities for applications in navigation and augmented reality. Multimodal approaches that integrate touch-based, audio-based, and visual feedback result in a comprehensive user experience. This extensive feedback system caters to a wide range of user preferences and accessibility requirements, making the system suitable for a broader audience .Improvements in security and privacy are imperative in the current digital landscape. Implementing modern encryption and secure data processing safeguards user information, alleviating concerns about data security and privacy. This commitment to protecting user information enhances the object detection system's credibility and fosters widespread acceptance. In conclusion, the integration of YOLO, SSD, and Faster R-CNN, along with innovative auditory feedback, adaptability testing, dataset expansion, touch-based feedback, low-power device optimization, continuous model refinement, wearable device integration, multimodal approaches, and privacy/security enhancements, positions this object detection system at the forefront of technological innovation. This holistic and user-centric approach ensures not only accuracy and versatility but also accessibility, security, and adaptability in diverse real-world scenarios.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, February 2024

REFERENCES

[1]. Real-Time Object Detection And Identification For Visually Challenged People Using Mobile Platform Neeraj Joshi, Shubham Maurya, Sarika Jain National Institute of Technology, Kurukshetra https://www.semanticscholar.org/paper/Real-Time-Object-Detection-And-Identification-For-Joshi-Maurya/a0029bacef20e2adbc6a69855ce0ac5c67f6be18

[2]. Deep Learning for Real-Time Capable Object Detection and Localization on Mobile Platforms F. Particke1, , R. Kolbenschlag1 , M. Hiller1 , L. Patiño-Studencki1 and J. Thielecke1 1 Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Information Technologies, Erlangen, Gersmanyhttps://www.researchgate.net/publication/320898401_Deep_Learning_for_RealTime_Capable_Object_Detection_and_Localization_on_Mobile_Platforms

[3]. You Only Look Once: Unified, Real-Time Object Detection Joseph Redmon , Santosh Divvala∗†, Ross Girshick , Ali Farhadi∗† University of Washington∗ , Allen Institute for AI† , Facebook AI Research .https://arxiv.org/abs/1506.02640

[4]. Enhancing Object Detection for VIPs Using YOLOv4_Resnet101 and Text-to-Speech Conversion Model Tahani Jaser Alahmadi 1, , Atta Ur Rahman 2 , Hend Khalid Alkahtani 1 and Hisham Kholidy 3 Submission received: 7 July 2023 / Revised: 17 July 2023 / Accepted: 26 July 2023 / Published: 2 August 2023 https://www.mdpi.com/2414-4088/7/8/77

[5]. Object Detection System for Visually Impaired Persons Using Smartphone D. Ravi Kumar, Hiren Kumar Thakkar, Suresh Merugu, Vinit Kumar Gunjan, and Suneet K. Gupta https://www.researchgate.net/publication/356066523_Object_Detection_System_for_Visually_Impaired_Persons_Using_Smartphone

[6]. DeepNAVI: A deep learning based smartphone navigation assistant for people with visual impairments Bineeth Kuriakose, Raju Shrestha, Frode Eika Sandnes Oslo, Norway https://www.sciencedirect.com/science/article/pii/S0957417422017432

[7]. REAL-TIME DYNAMIC OBSTACLE DETECTION FOR VISUALLY IMPAIRED PERSONS Tejas Hari Aher*1, Govind Karvande*2, Tejas Uttam Aher*3, Amey Jadhav*4, Prof. Dr. S.V. Gumaste*5 *1,2,3,4,5Department Of Information Technology, MET Institute Of Engineering Nashik, India. https://www.irjmets.com/uploadedfiles/paper/issue_5_may_2022/24169/final/fin_irjmets1653465179.pdf

[8]. Convolutional Neural Network for Object Detection System for Blind People Y.C. Wong, J.A. Lai, S.S.S. Ranjit, A.R. Syafeeza,N.A.Hamidhttps://www.researchgate.net/publication/333507222_Convolutional_Neural_Network_for_Object_Detection_System_for_Blind_People

[9]. Outdoor Navigation for Visually Impaired based on Deep Learning Saleh Shadi M.Sc., Saleh Hadi PhD, Mohammad Amin Nazari M.Sc, Wolfram Hardt prof. https://www.researchgate.net/publication/337362792_Outdoor_Navigation_for_Visually_Impaired_based_on_Deep_Learning

[10]. Real-Time Hand Gesture Recognition Based on Deep Learning YOLOv3 Model Abdullah Mujahid 1 , Mazhar Javed Awan 2 , Awais Yasin 3 , Mazin Abed Mohammed 4 , Robertas Damaševiˇcius 5,* , Rytis Maskeliunas ¯ 6 and Karrar Hameed Abdulkareem 7 https://www.mdpi.com/2076-3417/11/9/4164

[11]. Hand Gesture Recognition Using Faster R-CNN Inception V2 Model Rubin Bose S ,Sathiesh Kumar V Department of Electronics Engineering MIT Campus /Anna University Chennai Tamil Nadu , India. https://dl.acm.org/doi/abs/10.1145/3352593.3352613

[12]. Hand Gesture Recognition Based on Faster-RCNN Deep Learning Xiaoguang Yu1, Yafei Yuan2,https://www.semanticscholar.org/paper/Hand-Gesture-Recognition-Based-on-Faster-RCNN-Deep-Yu-Yuan/3654e2d07f410d153eb4cf878798024b98ec6b71

[13]. Pattern Recognition based Hand Gesture Recognition model Using Faster R-CNN Inception V2 Model ThitupathiJangapally, Dr. TryambakHiwarkar, Bhopal, M.P, India. https://www.researchgate.net/publication/344157098_Pattern_Recognition_based_Hand_Gesture_Recognition_model_Using_Faster_R-CNN_Inception_V2_Model

[14]. Simple Smartphone-Based Guiding System for Visually Impaired People Bor-Shing Lin 1 , Cheng-Che Lee 1 and Pei-Ying Chiang 2, Department of Computer Science and Information Engineering, National Taipei University,Taiwan.(PDF) Simple Smartphone-Based Guiding System for Visually Impaired People (researchgate.net)