SECURITY ASSISTANCE FOR VISUALLY CHALLENGED USING MACHINE LEARNING

PROJECT SYNOPSIS

OF MAJOR PROJECT

BACHELOR OF TECHNOLOGY

Computer Science and Engineering

SUBMITTED BY--

ATHARVA NAMDEO (2100290100037)

DEVESH KUMAR (2100290100053)

PARTH SHARMA (2100290130119)

August 2023

Project Guide: Dr. Sushil Kumar

KIET Group of Institutions, Delhi-NCR, Ghaziabad (UP) Department of Computer Science and Engineering

TABLE OF CONTENTS

INTRODUCTION

As our world grows more interconnected and technology becomes more and more integrated into our daily lives, we must work towards accessibility and inclusivity for all people, no matter what their skills. The community of visually impaired people stands out among individuals who encounter particular difficulties on a daily basis as a group in need of specialised solutions to improve their safety and security. The emergence of Machine Learning (ML) and its notable progress in multiple fields have made it feasible to leverage technology to develop cutting-edge security support systems tailored to the needs of individuals with visual impairments.

This topic, "Security Assistance for Visually Challenged Using Machine Learning," tackles the urgent need to give those who are blind or visually impaired the instruments and resources they need to safely and confidently traverse the environment. It attempts to investigate how ML might be used to create innovative solutions that surpass conventional approaches and provide visually impaired people with a more independent and safe way of living.

The convergence of security and assistive technology for the visually impaired is a complex area that includes navigation, item identification, facial recognition, and general situational awareness. The field of machine learning, especially in the areas of computer vision, natural language processing, and wearable technology, holds great promise for transforming the way people with visual impairments perceive and engage with their surroundings.

This introduction lays the groundwork for an extensive investigation into the ways in which the community of visually impaired people might benefit from the use of machine learning to increase their security, independence, and self-sufficiency. It explores the creative advancements and solutions that can significantly improve the lives of those with visual impairments and eventually lead to a more just and inclusive society.

OBJECTIVES

Three Major Objectives of our project are:-

1. To develop a new ML model to aid the visually impaired.

2. To develop a machine learning model for Image to Audio.

3. To produce effective and precise audio for given input images.

LITERATURE REVIEW

1- Show and Tell: A Neural Image Caption Generator

By Vinyals et al. (2015) presented a revolutionary method for captioning images. A notable innovation in computer vision and natural language processing, it merged Convolutional Neural Networks (CNNs) with Recurrent Neural Networks (RNNs) to generate high-quality captions for photographs. This paper is a cornerstone in the field, with its significance extending to applications such as content recommendation systems and accessibility tools.

2- An attention-based paradigm for picture captioning and visual question answering is presented in "Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering" by Anderson et al. (2018). Pre-extracting picture features is part of the "bottom-up" method, whereas the "top-down" attention mechanism aids in producing more pertinent and contextually rich descriptions. This research is a key contribution to multimodal deep learning as it greatly enhanced the quality of question answering and picture captioning in visual applications.

3- Tacotron: Towards End-to-End Speech Synthesis

In the field of text-to-speech (TTS) synthesis, Wang et al.'s 2017 research paper "Tacotron: Towards End-to-End Speech Synthesis" is a groundbreaking work. It presents the Tacotron model, which is intended for speech synthesis from beginning to end. One of this paper's main achievements is the development of a revolutionary sequence-to-sequence architecture with attention mechanisms for text-to-speech (TTS), which enables speech to be generated directly from text input without the need for intermediary elements like phonemes. The Tacotron model made synthesised speech far more expressive and natural-sounding. This study established the groundwork for future developments in the field of TTS technology and represented a significant advancement in the field.

4- Parallel WaveGAN: A fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram

A noteworthy advancement in the field of text-to-speech synthesis is "Parallel WaveGAN: A fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram" by Yamamoto et al. (2020). This work presents a sophisticated method for effectively producing speech waveforms of excellent quality. The model delivers impressive results in terms of speed and audio quality by integrating multi-resolution spectrogram approaches with generative adversarial networks (GANs). Waveform generating technology introduced in the paper has shown to be innovative, leading to faster and more efficient text-to-speech systems. These systems find applications in a variety of fields, such as audiobook narration and virtual assistants.

5- ImageNet Classification with Deep Convolutional Neural Networks

The paper "ImageNet Classification with Deep Convolutional Neural Networks" is a landmark in the field of deep learning and computer vision. It presents a pioneering model called AlexNet, which revolutionized image classification tasks by demonstrating the effectiveness of deep convolutional neural networks (CNNs).

FEASIBILITY STUDY

Feasibility:

Given current developments in computer vision and machine learning technology, the possibility of this project appears encouraging. The project intends to create a system that can help people who are blind or visually impaired navigate their environment securely by utilizing these technologies. The availability of reasonably priced gear, including good quality cameras and the potential for crowdsourced data collecting, promotes feasibility. However, it's crucial to take into account issues with consumer acceptance, data privacy concerns, and the accuracy and reliability of object recognition algorithms.

Need for the Project:

By offering a technology-driven solution to improve their safety and independence, our project addresses a fundamental need for the community of people who are blind or visually impaired. People who are visually impaired frequently struggle to recognize and navigate their surroundings, and moreover, they are not even safe at their homes which puts them at risk for mishaps and dangers. With the aid of machine learning and computer vision, this project aims to alert them and meet their needs by assisting them in real-time object detection and identification, reducing their reliance on outside assistance, and improving their overall quality of life.

Significance of the Project:

Our project is significant because it has the potential to significantly improve the safety, mobility, and autonomy of visually impaired people. The project will use machine learning and computer vision to enable real-time item recognition and

hazard detection, enabling the visually impaired to navigate their surroundings more independently and with more confidence. This technology not only helps them live more independently but also promotes their well-being at their home, making the world more fair and open to all.

METHODOLOGY

Project Initiation: Defining the project's goals, deliverables, and scope. Creation of a multidisciplinary team with experience in project management, computer vision, and machine learning

Requirement Analysis: Gather detailed requirements and identify challenges in the current document approval processes.

Design Phase: Development of a machine learning model, implementing various image recognition algorithms.

Technology Selection: Choosing the most appropriate algorithms and datasets for the training and testing of the model.

Development: Implementing the model and user-friendly interactions, with functionalities like danger alert, recognition, and image-to-audio generation.

Testing and Quality Assurance: Extensive testing of the model's attributes and capabilities. Usability problems and performance bottlenecks are found and fixed.

TIMELINE

1. **5th Semester**: By the end of the 5th semester, the project should have achieved the following milestones:

- Completion of the project proposal and initial research.
- Selection and customization of machine learning models for object detection and navigation.
- Development of a prototype for object recognition.
- Basic integration of voice commands and audio feedback.

2. **6th Semester**: By the end of the 6th semester, the project should reach the following stages:
    - Refinement and optimization of machine learning algorithms for improved accuracy.
    - Development of machine learning model
    - Testing and validation of the system with a focus group of visually impaired individuals.

3. **7th Semester**: The 7th semester should mark the finalization and completion of the project:
    - Full system integration for real-time data updates.
    - Implementation of customization features and accessibility standards.
    - Extensive user testing and feedback collection, leading to system improvements.

- Preparation for project documentation and presentation, potentially involving further refinements based on user input.

FACILITIES REQUIRED

Real-time Image Processing: To process images in real-time, a powerful onboard computer or smartphone with machine learning capabilities is necessary. This device processes the video feed from the cameras and runs machine learning models for object recognition.

Audio Feedback: Convert the visual information into audio feedback. Text-to-speech (TTS) technology can describe the objects, their locations, and provide guidance using aural cues.

Machine Learning Training: Collect and label datasets of various objects and scenarios for training the machine learning models. Continuous training and fine-tuning of the models are essential for improving accuracy.

User-Friendly Interface: Develop an intuitive and user-friendly interface that can be easily understood and operated by individuals with visual impairments.

Accessibility Standards: Adhere to accessibility standards and regulations to ensure that the system is designed to meet the needs of visually challenged users.

Testing and Feedback: Continuously test the system with visually challenged individuals, gather feedback, and make improvements based on their input.

EXPECTED OUTCOMES

The expected outcome of implementing security assistance for visually impaired individuals using machine learning is an innovative system that significantly enhances their safety and independence. This system will provide real-time object detection, navigation guidance, and audio feedback through wearable devices. Users can interact with the system via voice commands, receive tactile cues, and access cloud-based data for up-to-date information. By adhering to accessibility standards and continuously gathering user feedback, the outcome aims to create a user-friendly, customizable, and privacy-conscious solution that empowers visually

impaired individuals to navigate their environment confidently, mitigating obstacles and enhancing their overall quality of life.

REFERENCES

[1] Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2015.7298935
10.1109/cvpr.2015.7298935

[2] Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., & Zhang, L. (2018). Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. doi:10.1109/cvpr.2018.00636
10.1109/cvpr.2018.00636

[3] Yamamoto, R., Song, E., & Kim, J.-M. (2020). Parallel Wavegan: A Fast Waveform Generation Model Based on Generative Adversarial Networks with Multi-Resolution Spectrogram. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). doi:10.1109/icassp40776.2020.9053795
10.1109/icassp40776.2020.9053795

[4] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84–90. doi:10.1145/3065386
10.1145/3065386

[5] Y. Wang, R. J. Skerry-Ryan, D. Stanton, Y. Wu, R. J. Weiss, N. Jaitly, Z. Yang, Y. Xiao, S. Bengio, and R. Clark, "Tacotron: Towards End-to-End Speech Synthesis," in Proceedings of the International Conference on Machine Learning (ICML), 2017.