Smart Assistance for Blind: Real-Time Image Captionng using Computer Vision

Mrs. N. Sree Divya [1], Avusula Bhavana[2], Vanathadupula Ushasri[3]

Assistant Professor, Department of IT[1]

B. Tech Student, Department of IT[2,3]

Mahatma Gandhi Institute of Technology, Hyderabad, India

Abstract: *Recent advancements in image captioning technology have significantly improved the lives of people with visual impairments, promoting social inclusivity. Using computer vision and natural language processing, images become more accessible and understandable through textual descriptions. Notable progress has been made in developing photo captioning systems specifically for visually impaired users. However, challenges remain, such as ensuring the accuracy of automated captions and managing images with multiple objects or scenes. This study introduces a pioneering architecture for real-time image captioning based on a VGG16-LSTM deep learning model, supported by computer vision. The system has been built and implemented on a Raspberry Pi 4B single-board computer with GPU capabilities. This setup enables the automatic generation of suitable captions for images taken in real-time with a camera module, making it a convenient and portable solution for visually impaired individuals. The performance of the VGG16-LSTM model is assessed through extensive tests involving both sighted and visually impaired participants in various environments. The results reveal that the proposed system functions effectively, producing accurate and contextually relevant real-time captions. User feedback indicates a notable enhancement in understanding visual content, thereby aiding the mobility and interaction of visually impaired individuals within their surroundings. Multiple datasets were utilized, including Flick8k, Flickr30k, VizWiz captioning, and a custom dataset, for the training, validation, and testing of the model*

Keywords: image captioning technology, visual impairments, social inclusivity, computer vision, natural language processing (NLP), textual descriptions, photo captioning, accuracy, real-time image captioning, VGG16-LSTM deep learning model, portable solutions, automatic generation, extensive testing, contextually relevant captions