

**REQOG - A VISUAL QUESTION ANSWERING INTERFACE  
FOR THE VISUALLY IMPAIRED**

**Muhammed Akram Azarm**

A dissertation submitted in partial fulfilment of the requirement for  
Bachelor of Engineering (Honours) degree in Software Engineering

**Department of Computing  
Informatics Institute of Technology, Sri Lanka  
in collaboration with  
University of Westminster, UK**

**2021**

## **Abstract**

Vision is one of the most important senses in a person and it plays a major role in a person's life. It is important in communication, learning coordination, navigation and most importantly in accessing educational materials. Vision impairment is described as any kind of loss in vision. At least 2.2 billion individuals, according to the World Health Organization, globally are experiencing visual impairment and it spans across all age categories and genders. In this digital era, most of the information available are graphical for instance images on the internet, and maps for navigation. These are not suitable for a person with low to no vision. Accessibility features like screen readers, figure captions, labelled images, tagged social media pictures are not enough for the visually impaired to understand the context of these content.

Visual Question Answering is a problem that requires multi modal capabilities to answer questions based on an image. One of its main use cases is to help the visually impaired. However there has not been any commercial level application built for this purpose.

This research aims at developing a prototype level mobile application which utilizes accessibility features to enable the visually impaired to ask questions based on images captured. This work also attempted to contribute to provide better results for number type answers since the existing works suggested that these types of questions perform poorly in a VQA setting. This work was able to successfully develop a prototype application that addresses the problem in a basic level and got good results for number type answers. This work opens the path to other works like localization.

**Keywords:** Visual Question Answering, Computer Vision, Natural Language Processing.