

MES COLLEGE OF ENGINEERING, KUTTIPPURAM
DEPARTMENT OF COMPUTER APPLICATIONS
20MCA245 – MINI PROJECT

PRO FORMA FOR THE APPROVAL OF THE THIRD SEMESTER MINI PROJECT

(Note: All entries of the pro forma for approval should be filled up with appropriate and complete information. Incomplete Pro forma of approval in any respect will be rejected.)

Mini Project Proposal No : _____
(Filled by the Department)

Academic Year : 2020-2022

Year of Admission : 2020

1. Title of the Project : **AI Optics: Object Recognition And Caption Generation For Blinds Using Deep Learning Methodologies**
2. Name of the Guide : _____ Mr. Vasudevan TV
3. Number of the Student: MES20MCA-2026 _____
4. Student Details (in BLOCK LETTERS)

Name	Roll Number	Signature
1 _____ MOHAMMED AFNAN PP	_____ 26	_____

Date: 01/12/2021

Approval Status : Approved / Not Approved

Signature of
Committee Members

Comments of The Mini Project Guide

Dated Signature

Initial Submission : _____

First Review : _____

Second Review : _____

Comments of The Project Coordinator

Dated Signature

Initial Submission: _____

First Review _____

Second Review _____

Final Comments :

Dated Signature of HOD

AI Optics: Object recognition and caption generation for Blinds using Deep Learning Methodologies

Mohammed Afnan PP

INTRODUCTION & OBJECTIVE

Millions of people around the world face major disability of visual impairment. Vision provides all the information needed for reading, body movement, mobility and its loss can severely affect an individual's professional and social advancement. It was reported by the World Health Organization (WHO) that out of 1.3 billion people that suffer from one or another form of visual impairment, 36 million suffer from complete blindness. Problems are often faced by people with impaired vision or complete blindness once they are out of their familiarized environments. Corporeal development is one of the major issues for the people suffering from impaired vision. They also are unable to recognize an object without physically feeling it and can't savor the beauty of the nature. Many assistive devices have been made commercially available for the visually impaired community of the society to help them read and recognize objects, enhancing their experience. In this paper we propose an end-to-end accessible solution to provide purposeful acknowledgement and guidance using object recognition and caption generation for enabling video to audio aid for the visually impaired community of the society. The aim of purposeful acknowledgement and guidance is to extract the range and direction of the obstacles within a finite and defined free space captured by the camera of the device. The object detection and recognition algorithm will be fed results to the caption generation algorithm for explaining the scene of the surrounding to the visually impaired in audio format

Problem definition and initial requirements

Existing System:

There is no existing system like proposed one. Recently, object detection or image detection has gained a lot of popularity and interest of a large number of people. There are various kind of advances in the domain which aid in detecting natural language generation but are restricted in their outcome. But some of researches are done on this subject. That are given by

- ☐ In a research it initiates with detection and combines them all-together to form a final outcome which consists of detected object and relationships
- ☐ Similarly, in another research that utilized observations to produce a triplet of the image and changed them into text phrases with the help of template. Much greater model based on language parsing have been also utilized.

The above methods have proved themselves useful in various conditions but one issue that remains with them is that they are highly hand designed and rigid when used for text

Proposed System:

With the exponential development in the field of artificial intelligence in recent years, many researchers have focused their attention towards the topic of image caption generation. With this topic being that of arduous task and interest people take it as a challenge to perform to excel in the field of AI. Automatic generation of neutral language descriptions or ‘captions’ according to the composition detected in an image, i.e., scene understanding is the main part of image caption generation which can be achieved by combining both natural language processing along with computer vision. In this paper, we tackle the task of generating captions by using the concepts of Deep Learning. Here we can help a blind by using object recognition and caption generation. Automatic generation of a caption of an image is itself a big hurdle in artificial intelligence that involves connecting computer vision with natural language processing. But solution to this problem could prove to be a better understanding of the outside world for the visually impaired people. The task involves object detection and classification with high accuracy and accessibility with large of flexibility in inputs along with priorities to various situations.

The modal basically focuses on the three important criteria:- first being on the generation of complete natural language sentences, second being on making the generated sentence semantically and grammatically correct and third making the caption consistent with the image.

BASIC FUNCTIONALITIES OF PROJECT

This Object recognition and caption generation for Blinds using Deep Learning Methodologies shall provide the following type’s easy-to-use, interactive, and intuitive interfaces.

- This app shall provide with login to access their specified account using a username and unique password
- During login process the app will verify the specific user account
- App contains four sections Admin, Blind, Volunteer/care taker
- Admin should contain the following functional requirements.
 - It should capture the following data
 - ❖ Username
 - ❖ Password

User Modules

There are three types of modules. Admin, Blind, Volunteer/care taker. Each of them have distinct login section each of them can login their account section by conforming their unique.

1. Admin

Admin can login the system using username and password. Admin can control the overall system and should have the functionality to monitor overall process.

Admin can control the overall workflow.

- Login
- View Volunteers
- View Blind Persons
- Object Management
- View Complaint and sent reply

2. Volunteer/ care taker

Care taker can login the app using his/her unique username and password and they can do the functionalities that are given below

- ☐ Signup
- ☐ Login
- ☐ View profile, edit , change password
- ☐ Blind person management(add,edit,view,delete)
- ☐ View emergency help
- ☐ Sent message to blind
- ☐ Sent complaints and view reply
- ☐ View location of blind person
- ☐ Video Calling

3. Blind person

Blind person can login the app using his/her unique username and password and they can do the functionalities that are given below

- Authentication check using imei number
- Object detection and voice out
- ☐ Emergency command invocation using voice (sms,web)
- ☐ Location updation using gps
- Video Calling

Functional Modules

CNN

In deep learning, a **convolutional neural network** (CNN, or **ConvNet**) is a class of artificial neural network, most commonly applied to analyse visual imagery. They are also known as **shift invariant** or **space invariant artificial neural networks** (SIANN), based on the shared-weight architecture of the convolution kernels or filters that slide along input features and provide translation equivariant responses known as feature maps. Counter-intuitively, most convolutional neural networks are only equivariant, as opposed to invariant, to translation. They have applications in image and video recognition, recommender systems, image classification, image segmentation, medical image analysis, natural language processing, brain-computer interfaces, and financial time series.

CNNs are regularized versions of multilayer perceptrons. Multilayer perceptrons usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. The "full connectivity" of these networks make them prone to overfitting data. Typical ways of regularization, or preventing overfitting, include: penalizing parameters during training (such as weight decay) or trimming connectivity (skipped connections, dropout, etc.) CNNs take a different approach towards regularization: they take advantage of the hierarchical pattern in data and assemble patterns of increasing complexity using smaller and simpler patterns embossed in their filters. Therefore, on a scale of connectivity and complexity, CNNs are on the lower extreme.

Convolutional networks were inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. Individual cortical neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. The receptive fields of different neurons partially overlap such that they cover the entire visual field.

CNNs use relatively little pre-processing compared to other image classification algorithms. This means that the network learns to optimize the filters (or kernels) through automated learning, whereas in traditional algorithms these filters are hand-engineered. This independence from prior knowledge and human intervention in feature extraction is a major advantage

HARDWARE AND SOFTWARE REQUIREMENT

This specifies the hardware and the support software required to carry out the development.

➤ **Hardware Requirements**

The selection of hardware is very important in the existence and proper working of any software. Then selection hardware, the size and capacity requirements are also important.

- ☐ Processor : 64 bit
- ☐ RAM : Min 3 GB
- ☐ Hard Disk : 10 GB

➤ **Software Requirements**

One of the most difficult task is selecting software for the system, once the system requirements is found out then we have to determine whether a particular software package fits for those system requirements. The application requirement:

- ☐ OPERATING SYSTEM: WINDOWS 10
 - ☐ FRONT END: HTML, CSS, JAVASCRIPT
 - ☐ BACK END: Mysql
 - ☐ IDE: JetBrains Pycharm, Android studio
 - ☐ TECHNOLOGY USED: PYTHON, JAVA
 - FRAME WORK USED: Flask
-