COMSATS University Islamabad,
Abbottabad Campus

Project Proposal

for

Automatic Image Captioning to Help Blind Peoples of Pakistan

By

Irtisam Ali	**CIIT/SP19-BCS-058/ATD**
Shahid Ali	**CIIT/SP19-BCS-060/ATD**
Sardar Badar Saghir	**CIIT/SP19-BCS-022/ATD**

Supervisor

Dr. Faiz Ali Shah

Bachelor of Science in Computer Science (2019-2023)

No.	Comment	Action

SCOPE DOCUMENT REVSION HISTORY

Supervisor Signature

Date:

Table of Contents

Project Category: (H, C, E)

○ **A-**Desktop Application/Information System ○**B-**Web Application/Web Application based Information System ○**C-** Problem Solving and Artificial Intelligence ○ **D-**Simulation and Modelling ○ **E-** Smartphone Application ○ **F-** Smartphone Game ○ **G-** Networks ○ **H-** Image Processing ○ Other (specify category) _____

Abstract

In Pakistan, people with visual impairment face many hurdles in performing day to day to task. They must have to estimate the state of the object. The issue can be addressed by using a computer vision, but the lack of Urdu images captions dataset makes it much harder. The project aims to manually develop the image dataset with Urdu captioning and then create a model on it. The dataset focuses on some of the common scenarios faced by the visually impaired people. After the dataset is created, VGG16 CNN algorithm is use for generating a valid Urdu caption. The model is used with the android application to develop a computer vision app that can create and read Urdu captions at real time for visually impaired peoples.

Introduction

The people with visual impairment face difficulties in estimating the state of the object. Thus, such a system is needed to that create the caption by analysing the state object in real time. The advancement in computer vision makes it possible to develop software's that can detect the objects in real time. The most of algorithm requires great amount of data for high accuracy result. The datasets that have been created in the past are not particularly focus on the population of Pakistan. The lack of Urdu dataset is one of the core problems for creating the system that can help the visual impaired people. Thus, project design to create images datasets with Urdu captions dealing with common scenario face by visually impaired people and then develop a app that can visualize the states of object in real time by creating caption's using computer vision models. Then user is informed about the caption through TTS engines by converting text into speech.

Problem Statement

According to research in 2017 that about 2.21 million people in Pakistan either are blind or have sever visual impairment [1]. The aim of project is to design a mobile application that help the visually impaired people of Pakistan in their routine tasks. The android app like "Be My Eye" [2], or "Sullivan" [3] try to solve this problem by either hiring volunteer which can be expensive for end user or by training models that do not support Urdu language. Thus, this project aim is to fulfil the needs of impaired and blind users of Pakistan to develop such system. During the development of the system, it is highly likely that we will encounter the technologies that required the skills related to machine learning and computer vision. Thus, it is expected that we will polish our skills related to convolutional neural network (CNN) more specifically about VGG16. In the end of the project, it is anticipated that we will learn technique related to data gathering, feature extraction and training the model on cloud platforms.

Problem Solution for Proposed System

The project is design to create images datasets with Urdu captions dealing with common scenarios face by visually impaired people and then develop an app that can visualize the states of object in real time by creating caption's using computer vision models. Then user is informed about the caption through TTS engines by converting text into speech. It provides the simple way for visually impaired people to interact with their surroundings. The simple UI design with the real time caption with voice, help the visually impaired to identified state of objects near him. The project not only help the blind people, but it will also create the datasets which provide the basis for Urdu image datasets for the blind people which can be used in the future for other problem having similar characteristics.

Related System Analysis/Literature Review

Many systems have try to solve the problem of similar nature but they have some of the limitation that is trying to solve by this project

1. Be My Eye

The system is developed to support visually impaired people. The premises behind the app is to provide the platform where call can be arranged between visually impaired and volunteers who want to assist [4]. The only issue with this app is that vastly dependent upon availability of volunteers thus it is highly likely that volunteer will be unavailable when user is looking for the assistant [5].

2. Sullivan+

The app is available in both android and IOS users that can help in visual details. The premise of app to recognize the face, text, image description and colour [6]. The issue is that app has complex UI that might be useful for that low vision person but cannot work with blind persons. It lacks the voice feature thus not useful in real life for scenarios for blind peoples. Another issue is that it lacks the support of the Urdu language thus not good for illiterate peoples of Pakistan.

3. Lookout Assisted Vision

This app can recognize simple document in English and try to identify plants. But this app is not useful for common use as it does not deal with the common scenarios like identifying the objects and the situation context. Thus, issue has been tried to encounter in our project.

Table 1 Related System Analysis with proposed project solution

Application Name	Weakness	Proposed Project Solution
Be My Eye	The issue with this app is that vastly dependent upon availability of volunteers and has no advantage has been taken of computer vision.	Introduce computer vision for explaining scenario to eliminate the need of volunteers.
Sullivan+	The issue is that app has complex User Interface for blind persons. It lacks the voice feature. Another issue is that it lacks the support of the Urdu language.	Our aim is to create user friendly by that can be used by blind people's. The Urdu language will be there for Pakistan Peoples.

Lookout Assisted Vision	App is not useful for common scenarios due to limited capability of the object detection	Create general purpose data set with support of common household scenario, currency, vehicle, and common sign recognition

Advantages/Benefits of Proposed System

The system offers certain benefit's that include

1. It covers the more general scenario as VizWiz dataset is use for reference.
2. It supports Urdu Language.
3. The real time Urdu captioning creates the opportunity to target the untouched market.
4. The application is saleable which can be extended to further languages.
5. Integrated Pakistani currency identification for impaired people of Pakistan is unique to this application.

Scope

The main function of app is to detect and create Urdu caption for Vehicles, Furniture, Appliances, Persons, Bench's, Street Sign's, Stop Sign's, Kitchen object's and Currency of Pakistan. The model dataset is limited to above scenarios. The captions of image will be converted to voice with the help of TTS engines. By studying datasets like VizWiz it is estimated that number of images require for developing the project could be between 20-40k.

The project does not cover all scenario face by the visually impaired person due to constrains of time and resources. The object caption will be generated when user capture the image. It will be quite difficult to generate real time captions due technical issues. The software is limited to identifying Pakistan currency, Vehicle, Person, Street signs, and Electronics. The activities it can detect are whether Door is open, Refrigerator is open, Someone is eating Food, Someone is cooking Food, Someone is ironing the clothes, Someone is working on a laptop, Someone is Sleeping, Someone is playing Cricket, Someone is doing Gym and someone is using Phone.

The project is mainly dealing with both machine learning and data science aspects. The project will be in a pipeline instead of sequential with data collection, data pre-processing, feature extraction, machine learning and deployment can be run in parallel. The most appropriate software development for our project is agile methodology.

Modules

The project can divide into Data Collection, Data Prepossessing, Features Extraction, Model Training, Model Testing, Model Deployment and Android App Development.

1. Module 1: Data Collection

The raw data is collected by capturing manual images or extracting relevant images form internet. Beside that relevant image will also be extracted from publicly available datasets.

2. Module 2: Data Prepossessing

The raw data is than process into relevant categories, and image that is not relevant to require data set will be remove and then image will be label with relevant captions. The number captions can varies depending upon the image.

3. Module 3: Features Extraction

Features from images are extracted using convolutional neural network (CNN), So that we can use those features to train the model.

4. Module 4: Model Training

The algorithm like vgg16 using will be use training the model from features extracted during the features extraction process.

5. Module 5: Model Testing

The testing data is used for testing the model. In this process model is tested with both deployed and development conditions. If the accuracy is less than the require amount than model will be train again.

6. Module 6: Model Deployment

The model will be deployed with tflite on cross-platforms. The train model with extension of .tflite will be embedded with the application.

7. Module 7: Cross-platform App Development

The project will be package into cross platform app that has basic permission of internet and camera. The app will utilize the model and then create the captions based on model prediction and finally TTS engine convert into Urdu voice.

8. Module 8: Software Testing

Final module will be the testing of both app and model from different user to get the unbiased result. And in this phase bugs that will be found will be fixed. And performance of will be optimize in the end of testing phase.

System Limitations/Constraints

The app is limited to data set used is created by group thus it cannot cover every aspect of visual impaired people. Train model accuracy is uncertain. The model is limited to detecting activities

like are whether Door is open, Refrigerator is open, Someone is eating Food, Someone is cooking Food, Someone is ironing the clothes, Someone is working on a laptop, Someone is Sleeping, Someone is playing Cricket, Someone is doing Gym and someone is using Phone. The software is limited identifying object like Pakistan currency, Vehicle, Person, Street signs, and Electronics

Software Process Methodology

The Procedural methodology is adopted for training model and object-oriented approach for developing the application. Agile model is selected for the project. As individual interaction is more important for us than the process. Flexibility and adoption to change make it more relevant for our projects.

Tools and Technologies

Table 2Tools and Technologies for Proposed Project

	Tools	**Version**	**Rationale**
Tools And Technologies	Google Collab		Online Editor
	Android Studio	2022	IDE
	Figma		Design Work
	MS Word	2015	Documentation
	MS Power Point	2015	Presentation
	Pencil	2.0.5	Mockups Creation
	Technology	**Version**	**Rationale**
	Python	6.0	Programming language
	Flutter	2	Programming language
	Tensor Flow	2.x	Library

Project Stakeholders and Roles

Table 3Project Stakeholders for Proposed Project

Project Sponsor	COMSATS University, Islamabad, Abbottabad
Stakeholder	The stack holder in our project • Sardar Badar Saghir • Shahid Ali • Irtisam Ali • Project Supervisor Name: Dr Faiz Ali Shah • Final Year Project Committee: Evaluation of project

Team Members Individual Tasks/Work Division

Table 4Team Member Work Division for Proposed Project

Student Name	Student Registration Number	Responsibility/ Modules
Sardar Badar Saghir	SP19-BCS-022	Module 1, module 2 and module 3 and module 5. Also Responsible for Gantt Chart
Shahid Ali	SP19-BCS-060	He is responsible for module 1, module 2, and module 6 and module 8. Also responsible for UML Diagram's
Irtisam Ali	SP19-BCS-058	Module 1, Module 2, Module 4 and Module 7. Also, responsible making Mock-up's diagrams

Data Gathering Approach

Data is collected by Interview, Questionnaire, scrapping from online resources and by modifying data set resources.

Concepts

- ## Machine Learning

The project is based upon the machine learning concepts. Thus, we will have to deal lot concept of machine learning on developing the model.

- ## Computer Vision

Our Machine Learning require the knowledge of computer vision domain, as our datasets are base upon the images.

- **Data Science Methodology**

As dataset are built from scratch and thus it requires work, it is required for us to apply various data science to create the quality of dataset.

- **Cross platform development**

Final Product will in the form of both IOS And Android thus it will help us in learning cross platform development

Gantt chart

Automatic Image Captioning to Help Blind People of Pakistan					
TASK NAME	**STATUS**	**DURATION**	**START DATE**	**FINISH DATE**	**STATUS**
Project Selection	Complete	3 days	Mon 14/03/22	Wed 16/03/22	Not Started
Documentation	In Progress	150 days	Thu 17/03/22	Wed 12/10/22	In Progress
Proposal and Feasibility Report (10%)	Complete	8 days	Thu 17/03/22	Mon 28/03/22	Complete
SRS	In Progress	10 days	Tue 29/03/22	Mon 11/04/22	
Identify Currency	Not Started	15 days	Tue 12/04/22	Mon 02/05/22	
UseCase Diagram	Not Started	10 days	Tue 03/05/22	Mon 16/05/22	
Class Diagram	Not Started	10 days	Tue 17/05/22	Mon 30/05/22	
Sequence Diagram	Not Started	10 days	Tue 31/05/22	Mon 13/06/22	
identify 5 Objects	Not Started	10 days	Tue 14/06/22	Mon 27/06/22	
30% Presentation	Not Started	11 days	Tue 28/06/22	Tue 12/07/22	
Data Flow Diagram	Not Started	8 days	Wed 13/07/22	Fri 22/07/22	
Block Diagram	Not Started	8 days	Mon 25/07/22	Wed 03/08/22	
Activity /State Machine Diagram	Not Started	7 days	Thu 04/08/22	Fri 12/08/22	
Analyze Module	Not Started	15 days	Mon 15/08/22	Fri 02/09/22	
Integration of Model in Application	Not Started	15 days	Mon 05/09/22	Fri 23/09/22	
identify 5 Remaining Objects	Not Started	10 days	Mon 26/09/22	Fri 07/10/22	
60% Presentation	Not Started	12 days	Mon 10/10/22	Tue 25/10/22	
Identify 10 Activities	Not Started	7 days	Wed 26/10/22	Thu 03/11/22	
Package Diagram	Not Started	7 days	Fri 04/11/22	Mon 14/11/22	
Deployment Diagram	Not Started	10 days	Tue 15/11/22	Mon 28/11/22	
Thesis	Not Started	5 days	Tue 29/11/22	Mon 05/12/22	
100% Presentation	Not Started	2 days	Mon 05/12/22	Wed 07/12/22	

Conclusion

Project is to create a solution that can help the blind people of Pakistan in daily life. In future this dataset can be extended, thus making it possible to improve application accuracy and deal the scenario. The project is one of its kind that is dealing with the Urdu speaking population of Pakistan and addressing problem of visual impairment.

References

[1] "As of 2017, out of 207.7 million people in Pakistan, an estimated 1.12 million (95% Uncertainty Interval [UI] 1.07–1.19)."

[2] "Be My Eyes - See the world together." https://www.bemyeyes.com/ (accessed Mar. 16, 2022).

[3] "Sullivan plus for visually impaired and low vision | Blind Help Project." Accessed: Mar. 16, 2022. [Online]. Available: https://blindhelp.net/software/sullivan-plus-visually-impaired-and-low-vision

[4] "Our story." https://www.bemyeyes.com/about (accessed Mar. 16, 2022).

[5] "Our story." https://www.bemyeyes.com/about (accessed Mar. 16, 2022).

[6] "Sullivan+ (blind, low vision) - Apps on Google Play." https://play.google.com/store/apps/details?id=tuat.kr.sullivan&hl=en&gl=PK (accessed Mar. 16, 2022).

Plagiarism Report