University of Central Punjab

BSCS FINAL PROJECT
Phase 1

REAL-TIME LANDMARK RECOGNITION FOR TOURISM: PILOT PROJECT FOR LAHORE

Project Advisor

DR. ADNAN N. QURESHI

Presented by:
Group ID: F20BS001

Student Reg#	Student Name
L1F17BSCS0328	Waleed Kamran
L1F17BSCS0344	Ali Sehran
L1F17BSCS0499	Muhammad Hammad

Faculty of Information Technology

University of Central Punjab

Software Requirements Specification

Version #1

REAL-TIME LANDMARK RECOGNITION FOR TOURISM: PILOT PROJECT FOR LAHORE

Advisor: Dr. Adnan N. Qureshi

Group: F20BS001

Member Name	Primary Responsibility
Waleed Kamran	Coordinate with “walled city” officials and collecting the dataset.
Ali Sehran	Scheduling plans for data collection, keep track of progress, and collecting the dataset.
Muhammad Hammad	Documenting the group’s working and collecting dataset.

Note: These responsibilities are confined for 1st phase only.

Table of Contents

1. Introduction

1.1 Review of Related Literature

Recently, touristic landmark recognition has attracted attention in the computer vision area. To classify various monuments based on the features of monument images, Saini and Gupta et al. [1] used the dataset of 100 Indian monuments, landmarks and used state-of-the-art DCNN to achieve 92.7% accuracy for a manually acquired dataset of 100 different Indian monuments which have been considered in the dataset for classification. Zheng and Adam et al. [2] shared similar strategies to explore landmarks at a world-scale. They built up a landmark recognition engine consolidating 5,312 landmarks and attractions from 1,259 cities in 144 nations. Among them, most landmarks are situated in Europe and North America and only a few are situated in Asia. The authors assigned this distribution bias to the user community of their data source. Boiarov and Tyantov [3] proposed a metric learning-based approach that efficiently handles a large number of landmarks, their model uses deep neural networks and requires a single pass inference and they developed a method for cleaning landmark data, used hierarchical clustering procedure, and deployed to production at scale and used for recognition on user photos on a cloud application. The built-up system includes 11381 landmarks from 502 cities and 70 countries. To predict Chinese Landmark, Wu and Chen [4] developed a Chinese landmark dataset containing 42,548 images of 987 unique Chinese landmarks. They built their model using CNN and transfer learning techniques and their landmark recognition system is easy to integrate with mobile applications.

Since no such work is directly targeting the landmarks of Pakistan and negligible work has been done for Non-English-speaking areas in the domain of landmark recognition on a global level. The work is done in China and India also targets their own countries and does not cater to the issues of variance in size and illumination issues.

1.2 Problem Statement

As the tourism industry is growing in Pakistan, yet there isn’t any application for this region that can help the tourist to explore without any guide, which implies certain restrictions such as time alignment, budget management, and language barrier. No application is directly targeting landmarks of Pakistan and a negligible amount of work has been done for Non-English-speaking areas. Existing landmark recognition models targets majorly English-speaking areas and missing features such as adding robustness with support of providing variance in light, size, and angle.

1.3 Proposed Solution

Abstract:

Our project will have a huge impact on the revolution of the tourism industry in the sense that tourists will be able to get free of cost guidance and they will be able to know about the landmarks and their brief description (extracted from a source e.g. Wikipedia) by either via text or voice assistant or both in our real-time landmark recognition mobile application. Our landmark recognition model will be recognizing landmarks with high precision at any time of the day and on different scales giving consistent user experience and availability in obvious scenarios. It will be fast enough to work with our real-time application.

Explanation:

This project is targeting foreign and local tourists with little or no idea about landmarks in Lahore ("Walled City"), and for people who are interested in exploring landmarks. We'll be creating a mobile application that will be acting as a real-time landmark recognition guide using deep learning for tourists and eliminating the need for a local and unauthentic guide for foreigners. The absence of large annotated datasets of natural and human-made landmarks and attractions which can be anything: buildings, monuments, cathedrals, parks, waterfalls, castles, etc. makes landmark recognition a valuable, yet challenging task. The tourism industry is huge in Pakistan and our past is enriched with history and culture. Past related work has shown promising results on the global level but it frequently ignores non-English speaking areas, for example, Pakistani landmarks and there is still room available for improvements such as adding robustness to algorithms by including variance in size, angle, and light. Our research will not only help future researchers who want to research in the landmark recognition but can also be beneficial in other image recognition problems. Tests using business landmark recognition APIs additionally demonstrate the incompleteness of such landmark data. We are determined to turn the final landmark recognition model to integrate with our memory-efficient mobile application which can be later converted into API to facilitate tourism-related platform in Pakistan.

Like many state-of-the-art image retrieval methods, our landmark recognition system will be based on a deep neural network. Our framework will consist of modeling and further application deployment. In data pre-processing, landmark images will be collected either by extracting them from various online resources or manually and then the dataset will be cleaned using image clustering and other possible help functions. In data preparation, data cropping and data augmentation may be applied for imbalanced landmark images in the dataset. We may then use up-to-date pre-trained networks, for example, Residual Network (ResNet), to build the recognition model. In this case, we shall have to retrain the model for recognizing new landmark objects. Additional information like geoinformation may be used to narrow the search of similar landmarks which will be in the surroundings of the shooting location. A hierarchical clustering procedure might be used to cater to issues of variance in size and angle. We may consider the use of ONPP [5] and GANs [6] for handling the illumination variation in landmark images which will be improving our system's robustness including variance in light.

1.4 Problem Scope

We care about this issue because it will help foreign and local tourists, travelers, and people who are interested in exploring landmarks and attractions. There is no work published for Pakistani landmarks with high precision and which caters to issues like the variance in size, angle, and light, and there is an absence of landmark datasets. There is no free of cost and authentic guidance available for the tourists. Tourism in Pakistan is evolving with time and tourists often face problems of recognizing landmarks and also getting authentic information about them. Tourists face other privacy problems as well while hiring local guides for information regarding landmarks.

1.5 Challenges

- Implementation of complex deep learning model which will be catering the issues of variance in size, angle, and illumination for all famous landmark recognition of Lahore
- Manual collection of the dataset by site visiting of all famous landmarks of Lahore ("Walled City") in the current pandemic situation
- Development of dataset using different techniques including web scrapping, unsupervised machine learning, and augmentation
- Integration of the developed model with the mobile application

1.6 Knowledge Areas Required

- Data Mining
- Computer Vision
- Research
- Deep Learning
- Mobile Application Development

1.7 Completeness Criteria

S.No.	**Criteria**	**Weightage %**
1	Dataset Collection	20
2	Preprocessing and Model Training	30
3	Model Optimization	20
4	Mobile Application Development	20
5	Mobile App Publish	10

1.8 Research Outcomes/Nature of End Product

The nature of our end product will be in two forms: first will be the research work done which will be first of its kind for Pakistan and achieve better performance than its global competitors in terms of variance in size, angle, and light, lastly, a mobile application that will be the end product of our research which will be recognizing landmarks in a real-time environment and acting as a guide for tourists and people who want to explore the landmarks of Lahore.

1.9 Learning Outcomes

- Having a better command in PYTHON and having a good understanding of MATLAB as well
- Understanding computer vision
- Understanding an important category of image recognition models and their representations in form of pre-trained models
- Understanding a research paper in image recognition or an algorithm related to ours in the paper
- Learning integration of deep learning models with the mobile application
- Becoming mature in the mobile application development area

2. Background Study and Literature Survey

Problem domain:

This project is targeting the tourism industry, we care about this issue because it will help foreign and local tourists, travelers, and people who are interested in exploring landmarks and attractions. There is no work published for Pakistani landmarks with high precision and which caters to issues like, the variance in size and light, and there is an absence of landmark datasets. It will have a huge impact on the revolution of the tourism industry in the sense that tourists will be able to get free of cost guidance and they will be able to know about the landmarks and their brief description (extracted from a source e.g. Wikipedia) by voice assistant or text in our real-time landmark recognition mobile application. Existing landmark recognition models work well in the specific day time, angles, and from specific distance causing the problem of limitation in time, distance, and angle. Our landmark recognition model will be recognizing landmarks with high precision at any day time and different scales giving consistent user experience and availability in obvious scenarios. It will be fast enough to work with our real-time application.

Literature Survey:

During our literature review, we went through several research papers and found a few of them to be relatable to what we are doing. Although those research papers proposed some promising methods yet those methods were not the answer to our problem. Few pieces of research did help us in finding appropriate direction, for such research paper's abstract are as follow:

- **"Improving Landmark Recognition using Saliency detection and Feature classification":**

Landmark recognition is one of the wanted but difficult kinds of classification in the field of computer vision. By now there exist many models to classify buildings and monuments within images, while now advanced machine learning algorithms allow people to focus on the classification of each building or monument. This paper proposes a model for Classifying Indian Landmarks. The model discussed in this paper provides a robust classification by a combination of Graph-Based Visual Saliency(GBVS) along with feature-based classification algorithms such as kNN and Random Forest.

- **"Tour the World: Building a web-scale landmark recognition engine"**

As we already have discussed in our SRS, a negligible amount of work has been done for non-English-speaking areas, this paper also highlights the same issue. This paper states that landmark recognition is not easy when implemented on a world-scale as there is no available list for landmarks for the whole world, due to which collecting data for world scale and making the final model accurate is a big problem. This paper uses data available on the internet while using advanced object recognition and clustering techniques. The model used pruning images using image matching and clustering techniques. The final result includes 5312 landmarks from 1259 cities in 144 countries. This experiment indicates that the model can deliver satisfactory recognition performance with high efficiency.

- **"Large Scale Landmark Recognition via Deep Metric Learning"**

This paper has used the landmark image dataset available at Mail.ru. For that dataset, this paper has presented a landmark recognition model. This model recognizes famous places, buildings, monuments, and other landmarks in user photos. The main challenge is to differentiate what object is a landmark and what is not a landmark. There are no specific criteria to differentiate. Also, there is no appropriate sized dataset available to train a recognition model. This paper proposed a metric learning-based approach which deals with existing challenges along with handling a large number of landmarks. The model uses DNN (deep neural network) and requires a single pass inference that makes it fast to use in production. This paper also describes an algorithm for cleaning landmark dataset. This paper describes the basic components of neural network architecture, the learning strategy, and the features of the metric learning approach. This paper also includes the results for the distribution of photos with and without landmarks from a user collection and also compared their method with others to evaluate. The described system has been deployed as a part of a photo recognition solution at Cloud Mail.ru, which is the photo-sharing and storage service at Mail.ru Group.

- **"Chinese Landmark Recognition"**

Landmark recognition is a useful but difficult task because there is no annotated dataset having a large number of landmarks. The work done before related to this does not cover non-English speaking areas, Therefore, this model is designed for Chinese Landmarks. The commercially available landmark recognition APIs also lack such landmark data. This project consists of a dataset having Chinese landmarks and attractions which are human-made and as well as natural landmarks. The dataset contains 42,548 images having 987 unique landmarks. The landmark recognition model designed in this paper is based on CNN (Convolution Neural Network) and transfer learning techniques. Lastly, an application is built for iOS based on this recognition model. It allows users to do landmark prediction on images taken by camera and images from the device gallery.

Conclusion:

Recently, touristic landmark recognition has attracted attention in the computer vision area. To classify various monuments based on the features of monument images, Saini and Gupta et al. [1] used the dataset of 100 Indian monuments, landmarks and used state-of-the-art DCNN to achieve 92.7% accuracy for a manually acquired dataset of 100 different Indian monuments which have been considered in the dataset for classification. Zheng and Adam et al. [2] shared similar strategies to explore landmarks at a world-scale. They built up a landmark recognition engine consolidating 5,312 landmarks and attractions from 1,259 cities in 144 nations. Among them, most landmarks are situated in Europe and North America and only a few are situated in Asia. The authors assigned this distribution bias to the user community of their data source. Boiarov and Tyantov [3] proposed a metric learning-based approach that efficiently handles a large number of landmarks, their model uses deep neural networks and requires a single pass inference and they developed a method for cleaning landmark data, used hierarchical clustering procedure, and deployed to production at scale and used for recognition on user photos on a cloud application. The built-up system includes 11381 landmarks from 502 cities and 70 countries. To predict Chinese Landmark, Wu and Chen [4] developed a Chinese landmark dataset containing 42,548 images of 987 unique Chinese landmarks. They built their model using CNN and transfer learning techniques and their landmark recognition system is easy to integrate with mobile applications.

Since no such work is directly targeting the landmarks of Pakistan and negligible work has been done for Non-English-speaking areas in the domain of landmark recognition on a global level. The work is done in China and India also targets their own countries and does not cater to the issues of variance in size and illumination issues

3.

3.1 Proposed Solution

Abstract:

This project is being created for foreign and local tourists with little or no idea about landmarks in Lahore, and for people who are interested in exploring landmarks. We'll be creating a mobile application that will be acting as a real-time landmark recognition guide using deep learning for tourists and eliminating the need for a local and unauthentic guide for foreigners. The absence of large annotated datasets of natural and human-made landmarks and attractions which can be anything: buildings, monuments, cathedrals, parks, waterfalls, castles, etc. makes landmark recognition a valuable, yet challenging task. The tourism industry is huge in Pakistan and our past is enriched with history and culture. Past related work has shown promising results on the global level but it frequently ignores non-English speaking areas, for example, Pakistani landmarks and there is still room available for improvements such as adding robustness to algorithms by including variance in size, angle, and light. Our research will not only help future researchers who want to research in the landmark recognition area in Pakistan but can benefit researches for other image recognition problems. Tests using business landmark recognition APIs additionally demonstrate the incompleteness of such landmark data. We are determined to turn the final landmark recognition model to integrate with our memory-efficient mobile application which can be later converted into API to facilitate tourism-related platform in Pakistan.

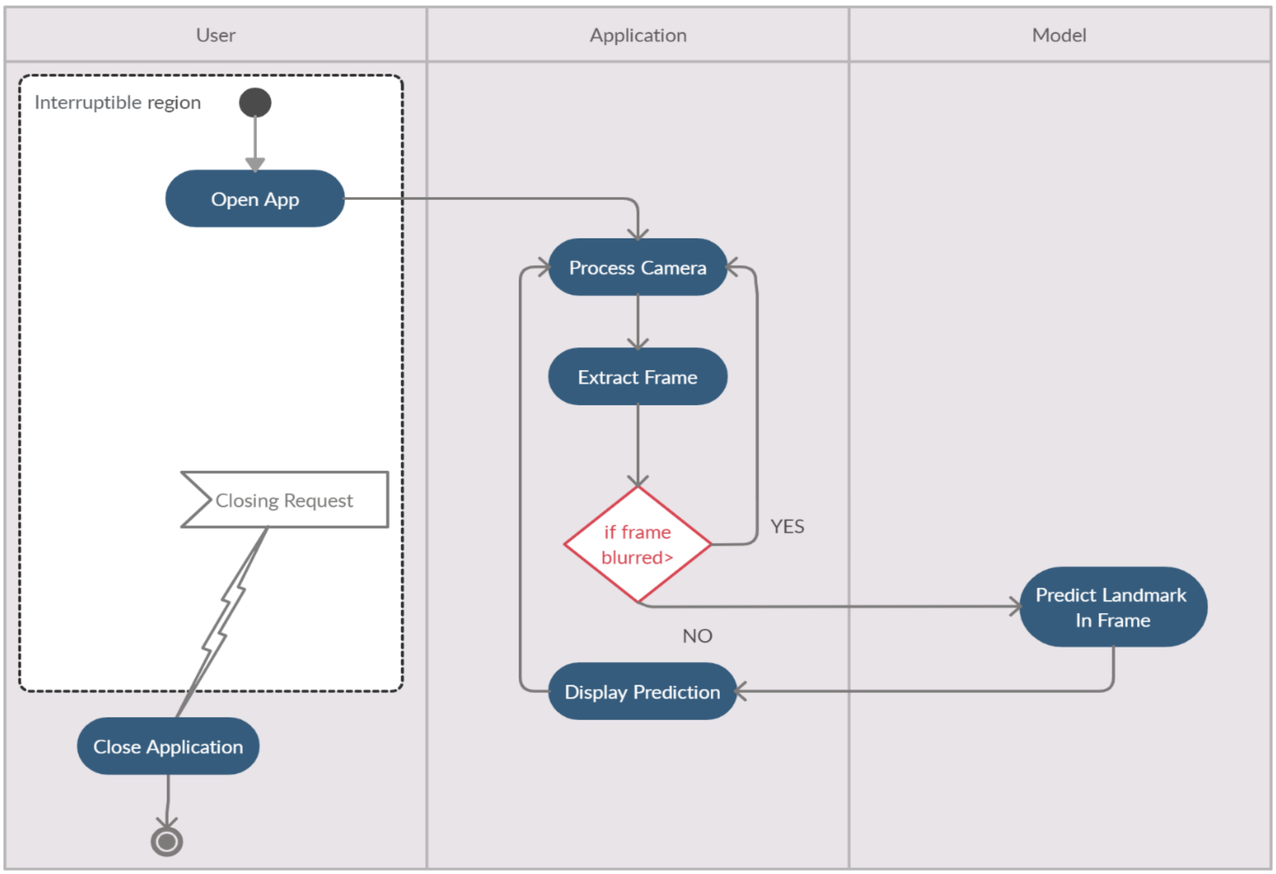
Features (Mobile Application):

1) Scan Landmark:
    a. Will access the device camera
    b. On behalf of deviation will extract a frame

2) Send Frame to model to predict the extracted frame

3) Display returned prediction at bottom of the application.

Activity Diagram:



3.2 User Classes and Characteristics

User Classes:

- **Tourist**
  - criteria: Should have internet
  - Hardware requirements: a smartphone with a camera
  - Software requirements: android
  - Min technical expertise: novice

3.3 Operating Environment

Hardware platform:
- *Mobile device*

Operating system:
- *Android (version 8.0+)*

3.4 Design and Implementation Constraints

Constraints:
- *The system must use python-3 for implementation of the model training*
- *The system must use Java-15 language for the android app development*
- *The system should use the android studio as a developmental tool for app development*
- *Throughout the project, the spiral model will be used as software architecture*
- *The system must be able to operate a camera and a recognition model*
- *For initial working "Google Collab" will be used as a development environment.*

3.5 Assumptions and Dependencies

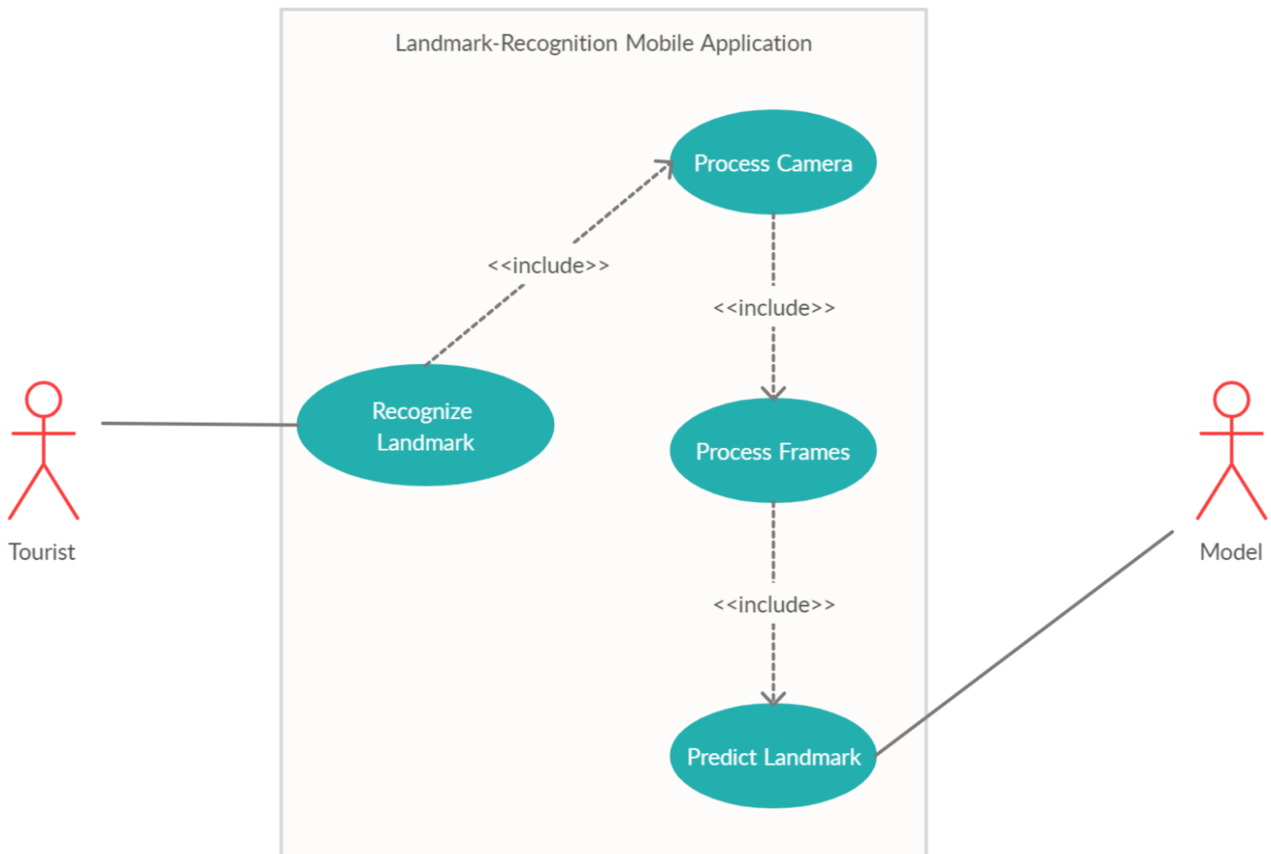
Assumptions:

- *User will be having obvious lighting (threshold illumination factor) and angle conditions to detect a landmark*
- *User will be having enough hardware requirements (RAM) to run the real-time mobile app*
- *User will be having sufficient internet bandwidths to recognize landmarks using our model*

Dependencies:

- *Python libraries such as Keras, Sklearn, and some others*
- *Pre-trained models such as Resnet, VGG may be used*

4. Functional Requirements



Description: There are two general use cases in the scenario above, one will be recognizing landmarks in which our application will extract a frame from the user's device camera and pass to the model, while the second use case will be acted upon when the frame is passes to model and model will predict the label of the frame.

4.1 Use-Case 1

Identifier	UC-1	
Purpose	To recognize a Landmark using Application	
Priority	High	
Pre-conditions	The user has opened the app and facing the device's camera towards a landmark.	
Post-conditions	The extracted frame is transferred to the model for further processing.	
Typical Course of Action		
S#	**Actor Action**	**System Response**
1	Tourist is recognizing a landmark in a real-time environment.	
2		The system processes the camera and extracts the frame to predict the landmark using the deep learning model.
Alternate Course of Action		
S#	**Actor Action**	**System Response**
1	Tourist is recognizing a landmark in a real-time environment with no connectivity.	
2		The system will alert the user to connect to the internet via a dialog box.

Table 1: UC-1

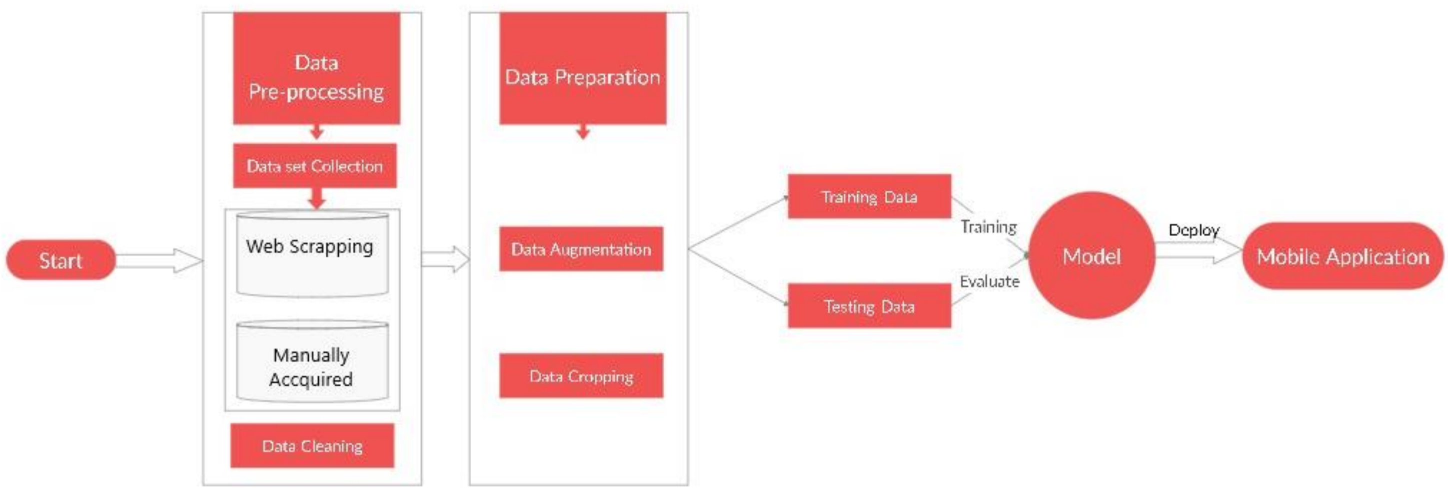
4.2 Use-Case 2

Identifier	UC-2	
Purpose	To predict a Landmark	
Priority	High	
Pre-conditions	The frame has been extracted and passed to the model.	
Post-conditions	Landmark is detected and tourists will be informed about the landmark name and some description via text or speech assistant.	
Typical Course of Action		
S#	**Actor Action**	**System Response**
1	The model predicts the given frame by the tourist.	
2		The system presents the predicted landmark name, some description via text or speed assistant.
Alternate Course of Action		
S#	**Actor Action**	**System Response**
1	Model is not able to predict the given frame by the tourist because the prediction accuracy does not meet the threshold.	
2		The system does not show the prediction and continue checking with the incoming frames.

Table 2: UC-2

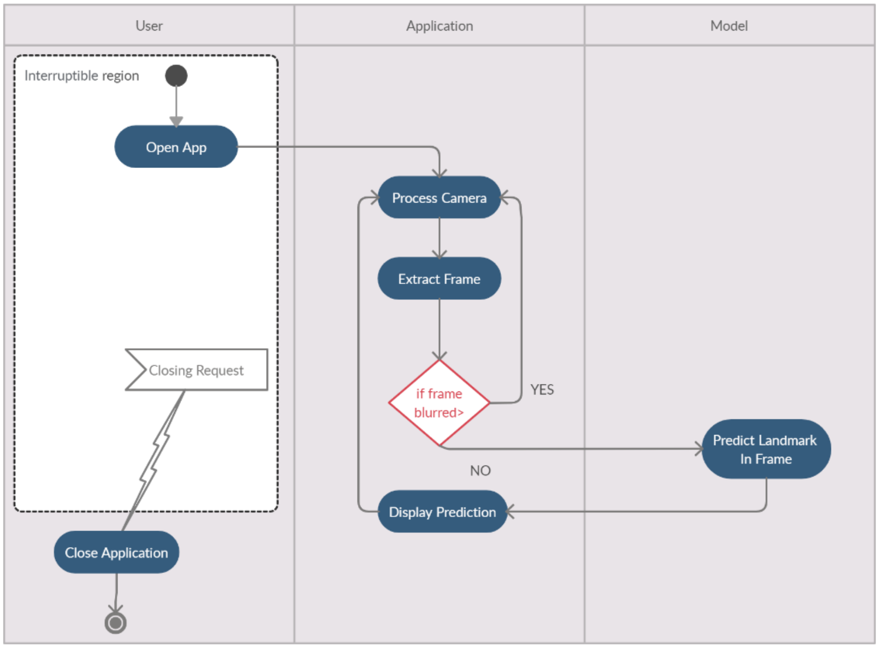
4.3 Proposed Workflow

Block Diagram:

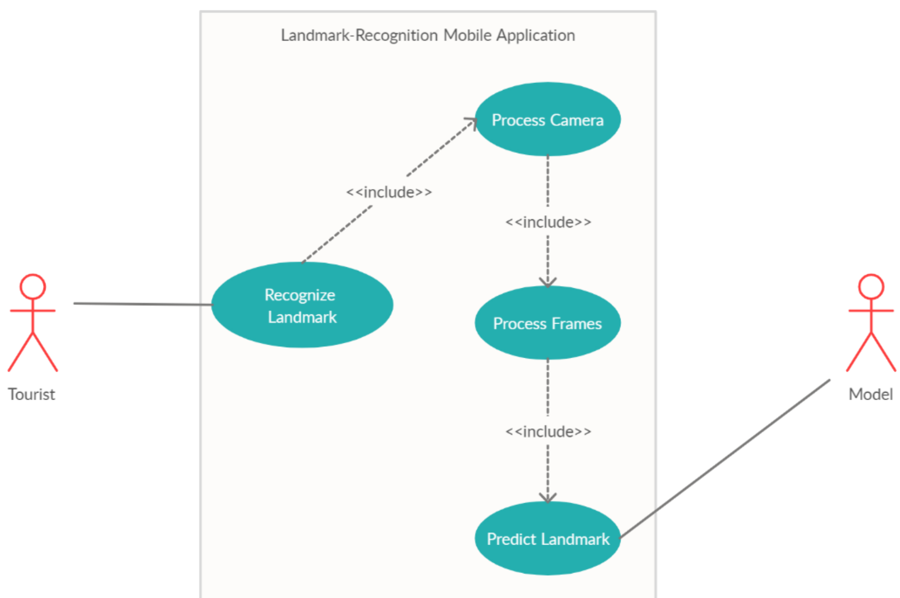


4.4 Analysis and Modeling of Requirements

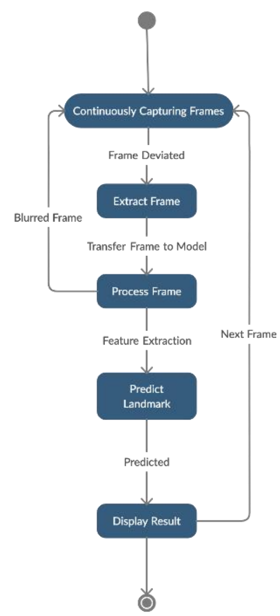
i) Activity Diagram:



ii) Use Case:



iii) State Diagram:



5. Nonfunctional Requirements

5.1 Target Performance

Our landmark recognition model will be recognizing landmarks with high precision at any time of the day and on different scales giving consistent user experience and availability in obvious scenarios. It will be fast enough to work with our real-time application. Our desired accuracy from the model is around 80% along with a response time of under 1 second (irrespective of network delay).

5.2 Additional Software Quality Attributes

- Robustness: The system should work under various environmental conditions (*400 lux+*).
- Usability: The system should be having a simple interface so that novice user can use (*1-Option*).
- Availability: The system will be available for as long as the device is connected to the internet.
- Reliability: The system should have high accuracy while recognizing the landmark (*Acc: 80%+*).

6. Other Requirements

-For our initial phase, we are working on "Google Collab" which doesn't allow us to store data more than certain limits (Size and Time), therefore we are bound to use Kaggle's database for our dataset and have integrated it with Collab using Kaggle's API.

-While training our final model we will be requiring GPU (*1650Ti-4Gb*) to compute weights.

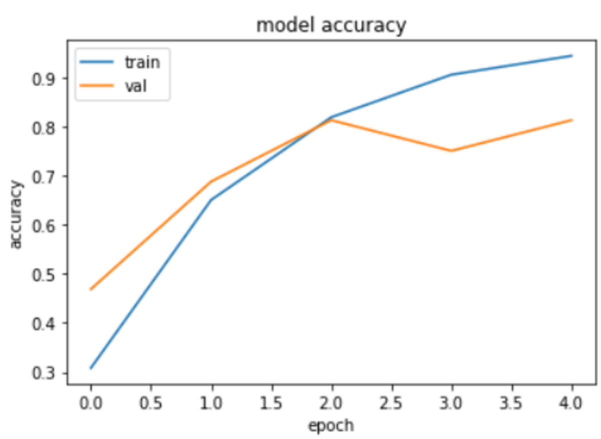
7. Initial Results

As per the plan for phase 1, we were supposed to collect data, either by web scrapping or by manual mean, for that we performed an experiment in which we collected a small dataset for 11 landmarks with 30-50 avg per landmark using web scrapping and used pre-trained model including ResNet-50, VGG-16, VGG-19 and custom CNN with 3 layers and 7 layers.

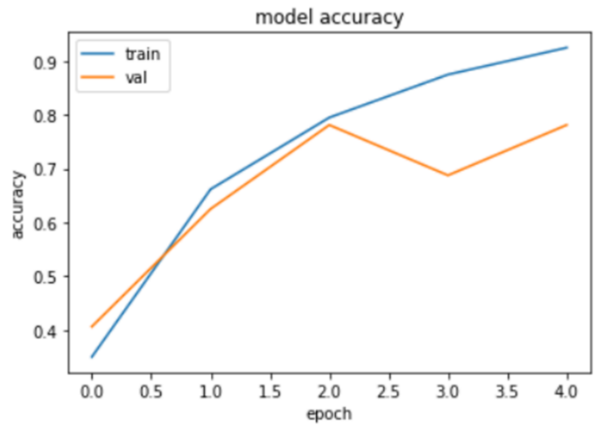
Accuracy we were able to achieve are as follow:

-> VGG-16= *75-80*%
-> VGG-19= *65-75*%
-> Resnet-50= *85-90*%
-> CNN 3-Layered (Grayscale)=*30-34*%
-> CNN 7-Layered (RGB)= *41*%

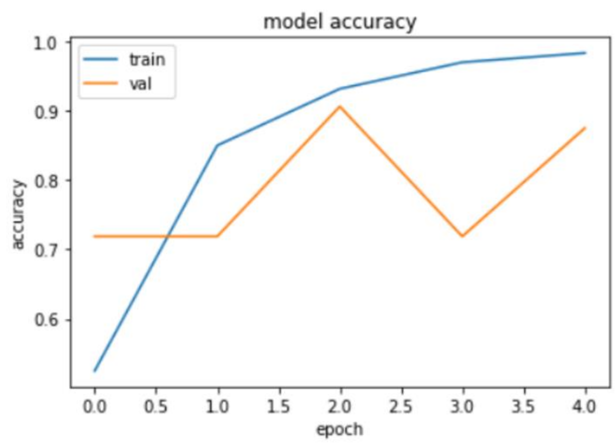
VGG-16:



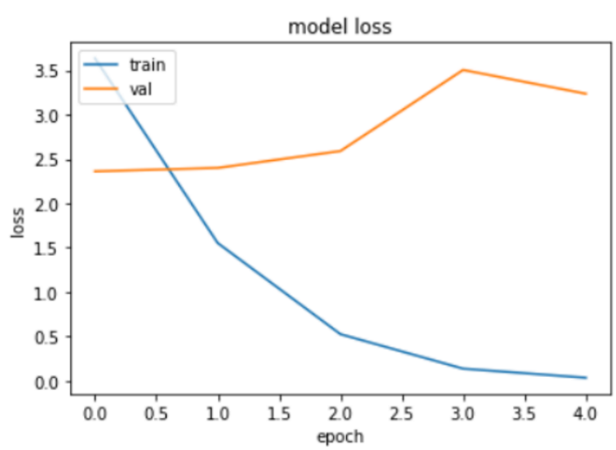
VGG-19:



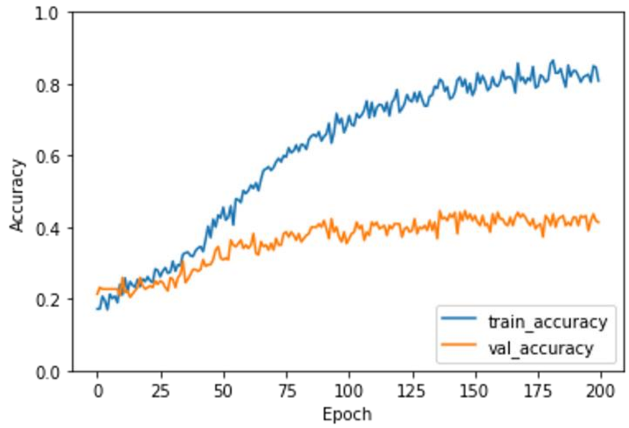
RESNET-50;



3-Layered:



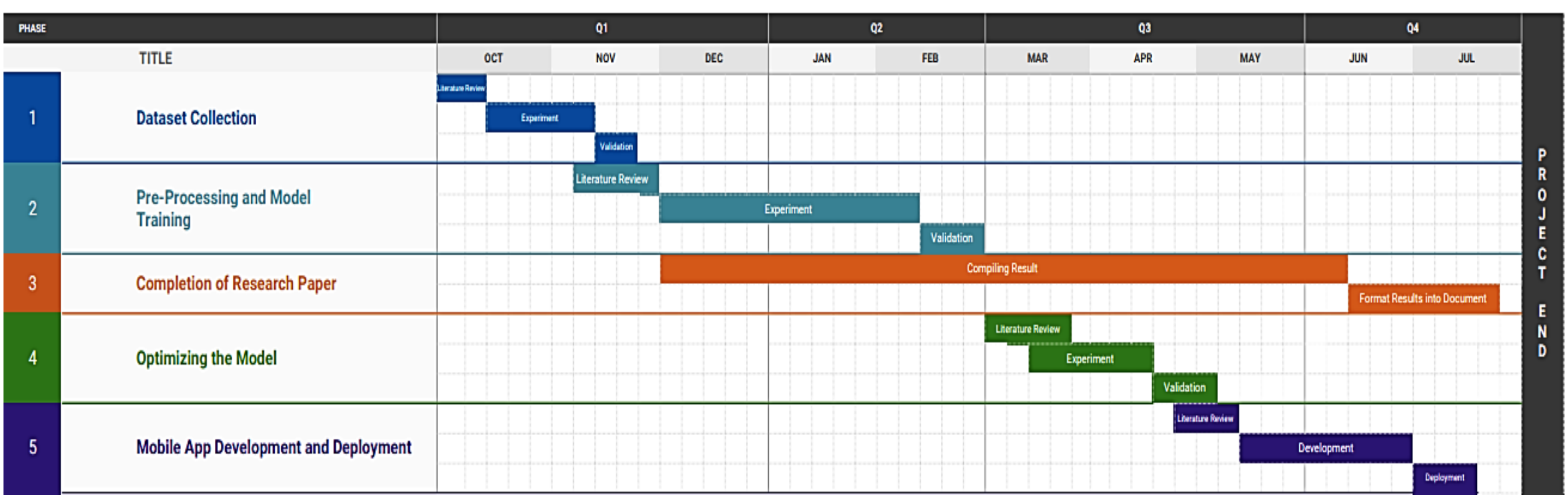
7-Layered:



As for custom layered CNN accuracies are low while in the case of pre-trained models, results were biased stating overfitting, pointed out that web scraping isn't appropriate. We then collaborated with "Walled City Lahore Authority", and were able to gather 30,000+ images for 20+ landmarks. The current dataset is sized around 25 Gb, which can be augmented and can be increased.

8. Revised Project Plan

<u>Project's Gantt chart using Microsoft Office timeline:</u>



9. References

REFERENCES

[1] A. Saini, T. Gupta, R. Kumar, A. K. Gupta, M. Panwar and A. Mittal, "Image based Indian monument recognition using convoluted neural networks," 2017 International Conference on Big Data, IoT and Data Science (BID), Pune, 2017, pp. 138-142.

[2] Y. Zheng *et al*., "Tour the world: Building a web-scale landmark recognition engine," *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, 2009, pp. 1085-1092.

[3] A. Boiarov and E. Tyantov, "Large Scale Landmark Recognition via Deep Metric Learning", *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, New York, USA, 2019, pp. 169-178.

[4] H. Wu and M. Chen, "Chinese Landmark Recognition," *2020 International Conference on Computing, Networking and Communications (ICNC)*, Big Island, HI, USA, 2020, pp. 24-28.

[5] P. Koringa, S. Mitra and V. Asari, "Handling Illumination Variation: A Challenge for Face Recognition", *Advances in Intelligent Systems and Computing*, pp. 273-283, 2016.

[6] D. Sakkos, E. Ho and H. Shum, "Illumination-Aware Multi-Task GANs for Foreground Segmentation", *IEEE Access*, vol. 7, pp. 10976-10986, 2019.

Appendix A: Glossary

<Define all the terms necessary to properly interpret the SRS, including acronyms and abbreviations. You may wish to build a separate glossary that spans multiple projects or the entire organization, and just include terms specific to a single project in each SRS.>

Appendix B: IV & V Report

(Independent verification & validation)
IV & V Resource

Name Signature

S#	Defect Description	Origin Stage	Status	Fix Time	
				Hours	Minutes
1					
2					
3					
…					

Table 2: List of non-trivial defects