VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"Jnana Sangama", Belagavi , Karnataka, INDIA

A Project Report
on

"LANDMARK RECOGNITION USING CONVOLUTIONAL NEURAL NETWORKS"

Submitted in partial fulfillment of the requirement for the award of the degree of

Bachelor of Engineering
in
Computer Science and Engineering

Submitted By

C.P YASHWANTH	**1GA17CS035**
MADDALI SOWMYA	**1GA17CS080**
RAKSHITHA MURTHY	**1GA17CS118**
SWARAJ PARIDA	**1GA17CS163**

Under the Guidance of

Prof. KAMLESHWAR KUMAR YADAV
Dept. of CSE

Department of Computer Science and Engineering
Accredited by NBA(2019-2022)

GLOBAL ACADEMY OF TECHNOLOGY

Rajarajeshwarinagar, Bengaluru - 560 098

2020 – 2021

GLOBAL ACADEMY OF TECHNOLOGY

Department of Computer Science and Engineering

Accredited by NBA(2019-2022)

CERTIFICATE

Certified that the Project Entitled **"Landmark Recognition using Convolutional Neural Networks"** carried out by **C.P Yashwanth**, bearing **USN 1GA17CS035**, **Maddali Sowmya**, bearing **USN 1GA17CS080**, **Rakshitha Murthy**, bearing **USN 1GA17CS118**, **Swaraj Parida**, bearing **USN 1GA17CS163**, bonafide students of Global Academy of Technology, is in partial fulfillment for the award of the **BACHELOR OF ENGINEERING** in **Computer Science and Engineering** from Visvesvaraya Technological University, Belagavi during the year 2020-2021. It is certified that all the corrections/suggestions indicated for Internal Assessment have been incorporated in the report submitted to the department. The Partial Project report has been approved as it satisfies the academic requirements in respect of the project work prescribed for the said Degree.

Prof. Kamleshwar K Yadav Dept. of CSE GAT, Bengaluru.	Dr. Srikanta Murthy K Professor & HOD Dept. of CSE GAT, Bengaluru.	Dr. N. Rana Pratap Reddy Principal GAT, Bengaluru.

GLOBAL ACADEMY OF TECHNOLOGY

Rajarajeshwarinagar, Bengaluru – 560 098

DECLARATION

We, **C.P Yashwanth**, bearing **USN 1GA17CS035**, **Maddali Sowmya**, bearing **USN 1GA17CS080**, **Rakshitha Murthy**, bearing **USN 1GA17CS118**, **Swaraj Parida**, bearing **USN 1GA17CS163**, students of Seventh Semester B.E, Department of Computer Science and Engineering, Global Academy of Technology, Rajarajeshwarinagar Bengaluru, declare that the Project Work entitled **"Landmark Recognition using Convolutional Neural Networks"** has been carried out by us and submitted in partial fulfillment of the course requirements for the award of degree in **Bachelor of Engineering** in **Computer Science and Engineering** from **Visvesvaraya Technological University, Belagavi** during the academic year **2020-2021**.

C.P YASHWANTH	1GA17CS035
MADDALI SOWMYA	1GA17CS080
RAKSHITHA MURTHY	1GA17CS118
SWARAJ PARIDA	1GA17CS163

Place: Bengaluru

Date: 29/01/2021

ABSTRACT

Landmark recognition can be an important aspect in our day to day lives. Especially with the exponential growth in digital photography, a large set of images are readily available which can be used for processing tasks. With the advancement of computer vision technologies, many image classification approaches have been introduced, where researchers have attempted to extract features out of images in the most optimal and accurate way. When it comes to image processing and classification tasks, Convolutional Neural Networks have gained immense popularity lately, due to its ability to classify large sets of images with high accuracy. This work attempts to apply Convolutional Neural Network models in order to classify landmark images available in the Google Landmark Dataset.

While image retrieval and instance recognition techniques are progressing rapidly, there is a need for challenging datasets to accurately measure their performance – while posing novel challenges that are relevant for practical applications. Google landmarks dataset is a new benchmark for large-scale, fine-grained instance recognition and image retrieval in the domain of human-made and natural landmarks.

With the help of Fast Region-based Convolutional Neural method (Fast R-CNN) we are able to detect objects. Fast R-CNN builds on previous work to efficiently classify object proposals using deep convolutional networks. Fast R-CNN employs several innovations to improve training and testing speed while also increasing detection accuracy.

This project aims to present an efficient end-to-end method to perform instance-level recognition employed to the task of labelling and ranking landmark images. Initially we embed images in a high dimensional feature space using Convolutional Neural Networks (CNNs) trained with an additive angular margin loss and classify images using visual similarity. We then efficiently re-rank predictions and filter noise utilizing similarity to out-of-domain images.

ACKNOWLEDGEMENT

The satisfaction and the euphoria that accompany the successful completion of any task would be incomplete without the mention of the people who made it possible. The constant guidance of these persons and encouragement provide, crowned our efforts with success and glory. Although it is not possible to thank all the members who helped for the completion of the phase - 1 of project individually, we take this opportunity to express our gratitude to one and all.

We are grateful to management and our institute **GLOBAL ACADEMY OF TECHNOLOGY** with its very ideals and inspiration for having provided us with the facilities, which made this, phase - 1 project a success.

We express our sincere gratitude to **Dr. N. Rana Pratap Reddy,** Principal, Global Academy of Technology for the support and encouragement.

We wish to place on record, our grateful thanks to **Dr. Srikanta Murthy K**, HOD, Department of CSE , Global Academy of Technology, for the constant encouragement provided to us.

We are indebted with a deep sense of gratitude for the constant inspiration, encouragement, timely guidance and valid suggestion given to us by our guide **Prof. Kamleshwar Kumar Yadav,** Assistant Professor, Department of CSE, Global Academy of Technology.

We are thankful to all the staff members of the department for providing relevant information and helped in different capacities in carrying out this phase -1 project.

Last, but not least, we owe our debts to our parents, friends and also those who directly or indirectly have helped us to make the phase - 1 project work a success.

C.P YASHWANTH	**1GA17CS035**
MADDALI SOWMYA	**1GA17CS080**
RAKSHITHA MURTHY	**1GA17CS118**
SWARAJ PARIDA	**1GA17CS163**

TABLE OF CONTENTS

LIST OF FIGURES

GLOSSARY

ANN	Artificial Neural Network
CNN	Convolutional Neural Network
CV	Computer Vision
DeLF	Deep Local Features
DFD	Data Flow Diagram
DNN	Deep Neural Networks
FC	Fully Connected
FCN	Fully Convolutional Network
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
PCA	Principal Component Analysis
RANSAC	Random Sample Consensus
VGG	Visual Graphics Geometry

Chapter 1

INTRODUCTION

1.1 Definitions

Computer vision (CV) is a field of Machine Learning that deals with how computers can gain high-level understanding from digital images or videos. Some of its tasks include methods for acquiring, processing, analyzing and understanding digital images, and extraction of high-dimensional data from the real world in order to produce numerical or symbolic information.

Advancements in the field of Deep Learning and improved hardware performance have accelerated the advancements in Computer vision and related fields. Applications of Computer Vision range from tasks like Automatic Inspection, Medicine, Autonomous Systems, Military, etc.

Artificial Neural Networks (ANN) is a series of algorithms that endeavours to recognize underlying relationships in a set of data through a process that mimics the operations of human brain. It works by adjusting "weights" during training, and predicts the output based on these adjusted weights.

Convolution Neural Networks (CNN) is a class of Deep Neural Networks (DNN), most often used to analyzing visual imagery. They have applications in image and video recognition, recommender systems, image classification, medical image analysis, natural language processing, and financial time series.

Has variety of applications which include

- image-to-image translation,
- video generation,
- voice generation,
- Image enhancement, etc.

1.2 Project report outline

As depicted by the title "Landmark Recognition using Convolutional Neural Networks", landmarks are instance specific. Object recognition techniques are still primitive in nature when it comes to identifying instances of any object class as there is huge spectrum of features. Further improvement is still ongoing with respect to instance recognition.

This project aims at recognizing landmarks from an image, as in tackling the problem of recognizing an instance.

This report covers the introduction of the project, the system requirements, hardware and software requirements, functional and non-functional requirements, the literature survey and the model architectures used.

Chapter 1 : In this chapter topics like computer vision, neural networks and convolutional neural networks are introduced. Various applications of the same are also mentioned.

Chapter 2 : In this chapter thorough System Study of the Literature survey is conducted and proposed work towards the project is mentioned. Further the scope of the project and its purpose is given.

Chapter 3 : In this chapter System requirements like functional and non-functional, software and hardware requirements are mentioned.

Chapter 4 : In this chapter System Design and architectures are discussed. Data Flow diagrams and Use Case diagram is also shown. Further architecture of models are presented as well. Finally, module is split into sub-modules are the same are briefed.

Chapter 2

REVIEW OF LITERATURE

2.1 System Study

Our world consists of breath-taking landmarks A landmark is a recognizable natural or artificial feature used for navigation, a feature that stands out from its near environment and is often visible from long distances. Landmarks are usually classified as either natural landmarks or man-made landmarks, both are originally used to support navigation on finding directions. A variant is seamark or daymark, a structure usually built intentionally to aid sailors navigating featureless coasts. Natural landmarks can be characteristic features, such as mountains or plateaus. Examples of natural landmarks are Table Mountain in South Africa, Mount Ararat in Turkey, Uluru in Australia, Mount Fuji in Japan and Grand Canyon in the United States. Trees might also serve as local landmarks, such as jubilee oaks or conifers In modern sense, landmarks are usually referred to as monuments or prominent distinctive buildings, used as the symbol of a certain area, city, or nation. Some examples include the Statue of Unity in Narmada, the White House in Washington, D.C., the Statue of Liberty in New York City, the Eiffel Tower are often very tall and visible from many miles around, thus often serve as built landmarks. Also town hall towers and verifies often have a landmark character.

Many researchers used various supervised and unsupervised leaning algorithms like nearest neighbours, neural networks image processing ,open cv for recognition of these landmarks. But recognising these landmarks still improving.

Our proposed project uses different techniques like VGG.16, DeLF models by setting and tuning the hyper-parameters to get good accuracy in classifying and predicting the recognition of the landmarks.

2.2 Proposed Work

In this project, we have tried different approaches like indexing, retrieval, geometric verification using different models and predicting all of these modules. All the models make use of approaches like extract features, and perform predictions in their unique way, which ultimately leads to the recognition of the landmark present in the image, thus letting us overcome almost all challenges which we might encounter.

2.3 Scope of the project

Landmark recognition, the problem comes from a famous Kaggle competition, the Google Landmark Recognition Challenge. Training set contains **over 1.2 million images** spread across **14,951 classes** of landmarks**,** varying from one to thousands of images per class. This problem of extreme classification is something that is very prevalent in the data science community today with the advancement of deep learning.

2.4 Purpose of the project

Landmark recognition can help in recognising the labels of those places where the name of the place or name of the landmark is unknown. It can predict labels directly from the image pixels which help users to understand and organize their photos.

Chapter 3

SYSTEM REQUIREMENT SPECIFICATION

3.1 Functional Requirements

- This project is required to execute the given model and project the expected outputs. The ultimate goal of this project is to make the model to detect the landmark present in the image and output the correct name of the landmark.
- In the training of the model the model should report the accuracy of the output.
- The tool should be enabling the user to upload the image from various formats.
- The tool should be enabling the user to upload the image of any size.
- The tool should be providing the user with the feature where the user can run the software thus expecting the best result all the time.

3.2 Non Functional Requirements

- The developed model is scalable, the model is able to detect the landmark present in the image given the features of the landmark.
- The user should satisfy all the hardware and software requirements, only then the model is going to support your actions.
- The model is expected to produce the results within no time (sometimes a few seconds)
- The tool should never fail in the middle of the operation.
- The response time of the tool should be acceptable.

3.3 Hardware Requirements

- Intel i3 6^{th} gen and above
- 4 GB ram (min)
- 2 GB graphics card
- 100 GB HDD (min)

3.4 Software Requirements

- Python 2.0 and above
- TensorFlow 1.4 and above
- Anaconda 3
- Any web browser : Chrome, Firefox, Safari
- Operating System : Windows or MacOS or Linux

Chapter 4

SYSTEM DESIGN

4.1 Design Overview

System Design is the process of defining architecture, components, and modules Interfaces and data for a system to satisfy requirements. System design could see it as the application of system theory to product development. There is some overlap with the description of system analysis, system architecture and system engineering.

If the broader topic of "product development" blends the perspective of marketing, design, and manufacture into a single approach to product development. System design is therefore the process of defining and developing systems to satisfy specified requirements of the user object oriented analysis and design methods are becoming the most widely used methods for computer system design. The UML has become the standard language in object oriented analysis and design. It is widely used for modelling software systems and is increasingly used for high designing non-software systems and organizations.

System design is one of the most important phases of software development process. The purpose of the design is to plan the solution of the problem specified by the requirement documentation. In other words the first step in the solution to the problem is the design of the project.

The design will continue the specification of all the modules, their interaction with other modules and the desired output from each module. The output of the design process is a description of the software architecture as shown in Figure 4.1.

4.2 System Architecture

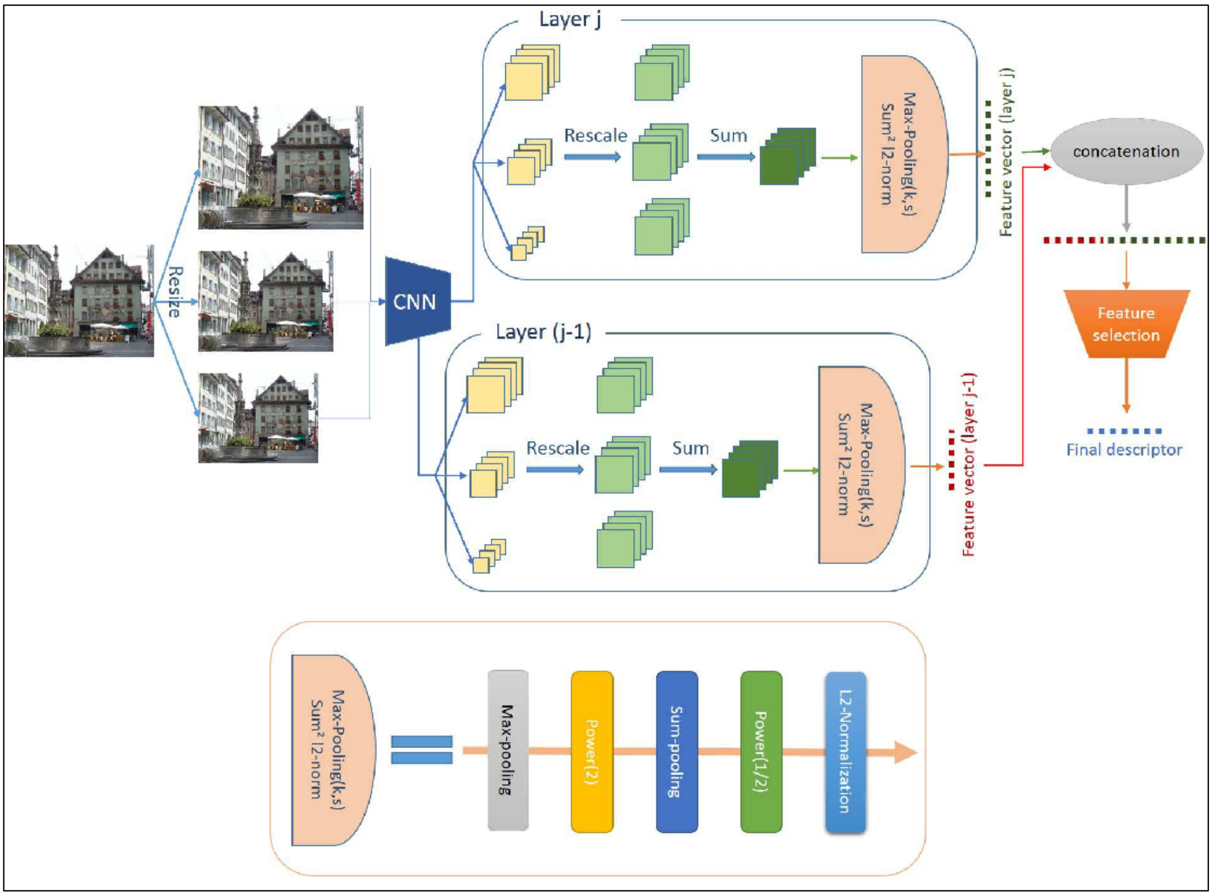

Figure 4.1 : System Architecture

The figure 4.1 shows a general diagram describing the activities performed by this project.

4.3 Data Flow Diagrams

A data flow diagram is the graphical representation of the flow of data through an information system. Data flow diagram is very useful in understanding a system and can be efficiently used during analysis.

A data flow diagram shows the flow of data through the system. It views a system as a function that transforms the inputs into desired outputs. Any complex systems will not perform this transformation in a single step and data will typically undergo a series of transformations before it becomes the output as shown in Figure 4.2, 4.3, and 4.4.

With a data flow diagram, users are able to visualize how the system will operate that the system will accomplish and how the system will be implemented. Old system data flow diagrams can be drawn up and compared with a new data flow diagram to draw comparisons to implement a more efficient system.

DFDs can be used to provide the end user with a physical idea of where the data they input, ultimately as an effect upon the structure of the whole system.

4.3.1 Data Flow Diagram – Level 0

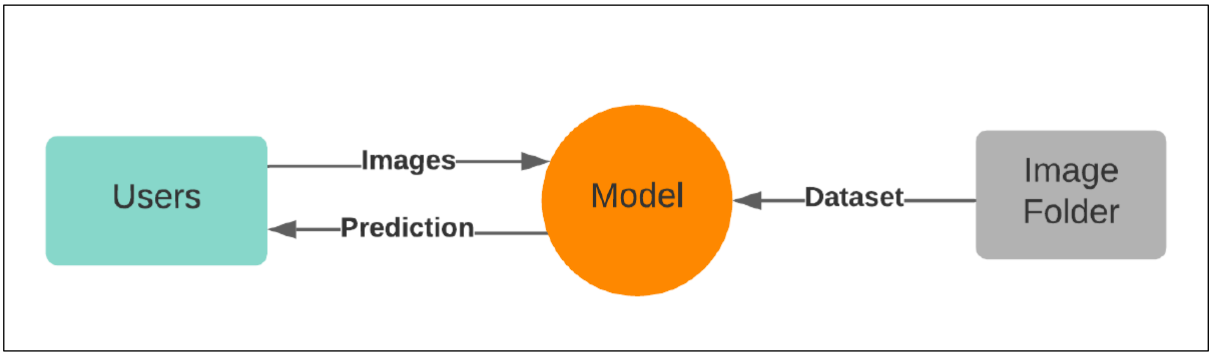

Figure 4.2 – Data Flow Diagram Level 0

4.3.2 Data Flow Diagram – Level 1

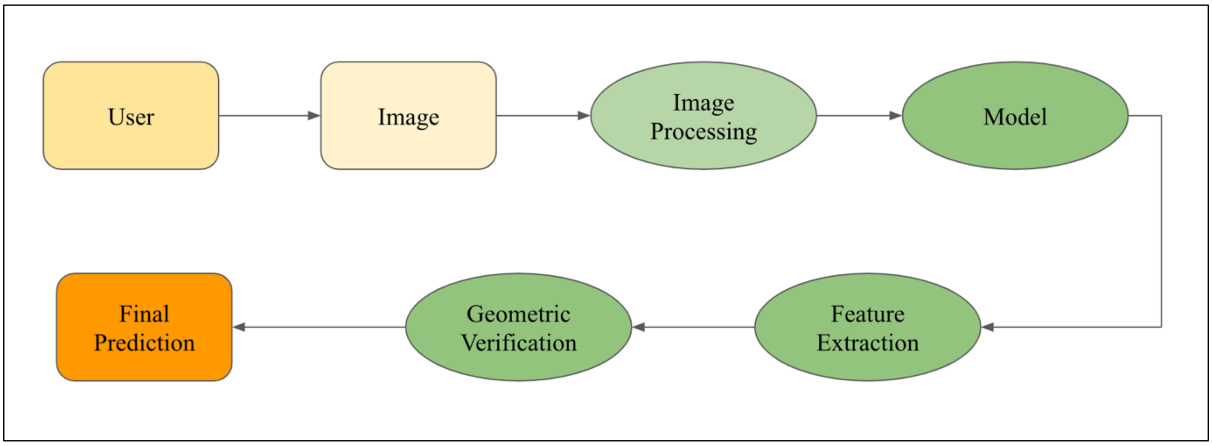

Figure 4.3 – Data Flow Diagram Level 1

4.3.3 Data Flow Diagram – Level 2

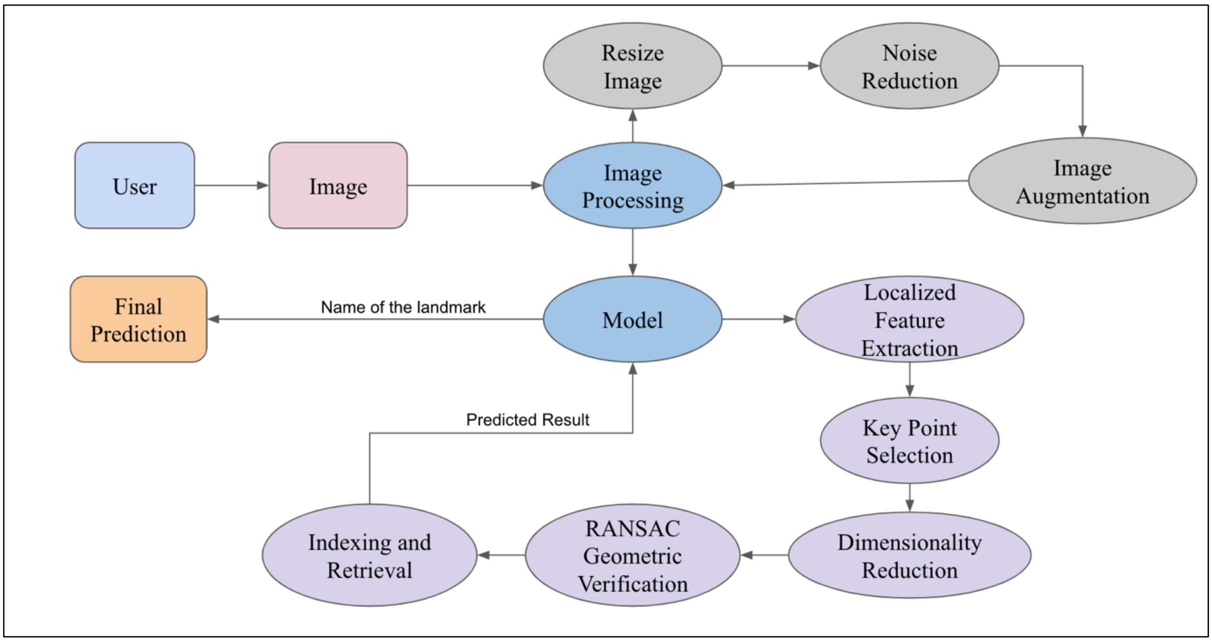

Figure 4.4 – Data Flow Diagram Level 2

4.4 Use Case Diagram

The external objects that interact directly with the systems are called the **actors.** Actors include humans, external devices and other software systems. The important thing about actors is that they are not under control of the application. In this project, user of the system is the actor.

To find use cases, for each actor, list the fundamentally different ways in which the actor uses the system. Each of these ways is a use case as shown in Figure 4.5.

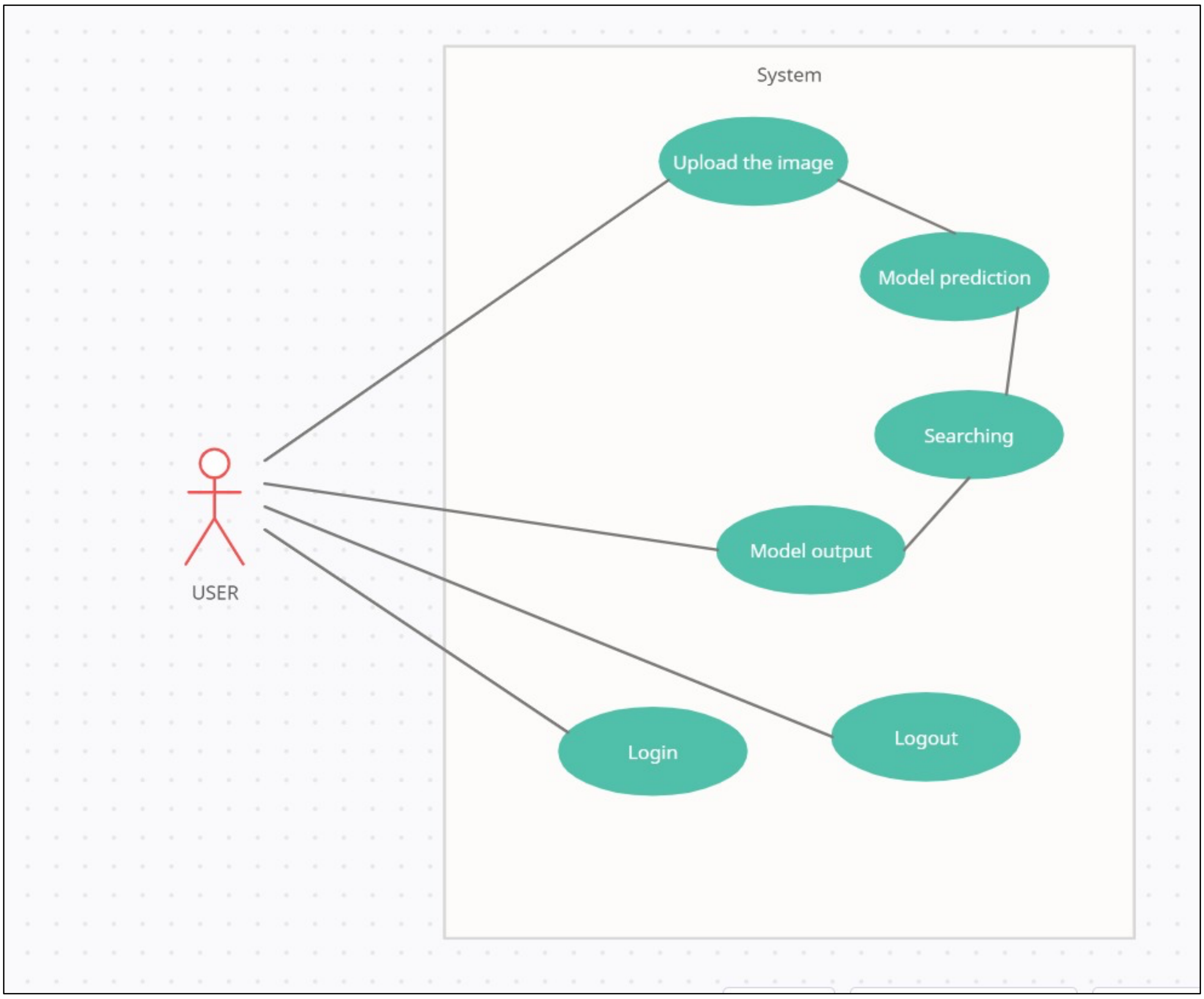

Figure 4.5 – Use Case Diagram

4.5 CNN Architecture Diagrams

CNN architecture is inspired by the organization and functionality of the visual cortex and designed to mimic the connectivity pattern of neurons within the human brain. The neurons within a CNN are split into a three-dimensional structure, with each set of neurons analyzing a small region or feature of the image.

4.5.1 Visual Graphics Geometry (VGG.16)

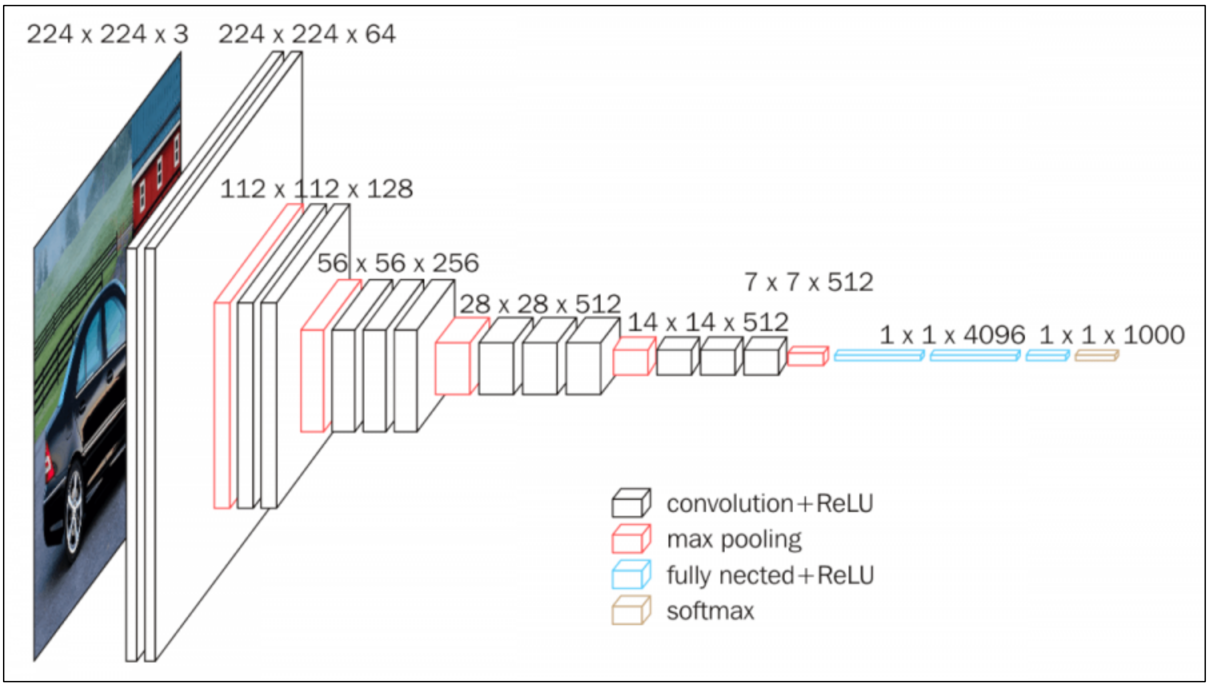

Figure 4.6 – VGG.16 Architecture

Figure 4.6 shows the VGG-16 Architecture. The input to cov1 layer is of fixed size 224 x 224 RGB image. The image is passed through a stack of convolutional (conv.) layers, where the filters were used with a very small receptive field: 3×3 (which is the smallest size to capture the notion of left/right, up/down, centre). In one of the configurations, it also utilizes 1×1 convolution filters, which can be seen as a linear transformation of the input channels (followed by non-linearity). The convolution stride is fixed to 1 pixel; the spatial padding of conv. layer input is such that the spatial resolution is preserved after convolution, i.e. the padding is 1-pixel for 3×3 conv. layers. Spatial pooling is carried out by five max-pooling layers, which follow some of the conv. layers (not all the conv. layers are followed by max-pooling). Max-pooling is performed over a 2×2 pixel window, with stride 2.

Three Fully-Connected (FC) layers follow a stack of convolutional layers (which has a different depth in different architectures): the first two have 4096 channels each, the third performs 1000-way ILSVRC classification and thus contains 1000 channels (one for each class). The final layer is the soft-max layer. The configuration of the fully connected layers is the same in all networks.

4.5.2 Deep Local Features (DeLF)

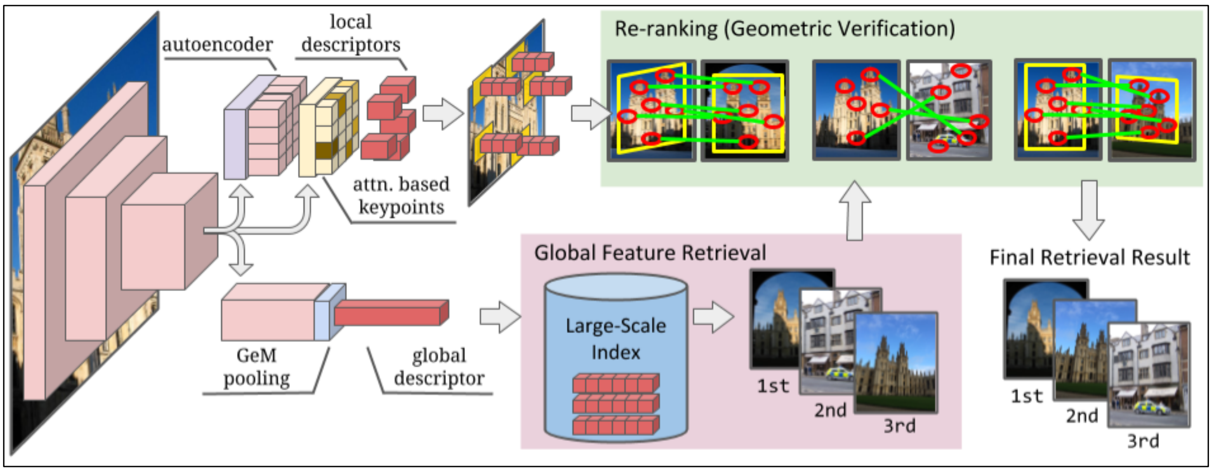

Figure 4.7 – DeLF architecture

DeLF Architecture shown in Figure 4.7 can be decomposed into four main blocks:

(i) dense localized feature extraction,

(ii) keypoint selection,

(iii) dimensionality reduction and

(iv) indexing and retrieval.

(i) We extract dense features from an image by applying a fully convolutional network (FCN), which is constructed by using the feature extraction layers of a CNN trained with a classification loss. We employ an FCN taken from the ResNet50 model.

(ii) Instead of using densely extracted features directly for image retrieval, we design a technique to effectively select a subset of the features. Since a substantial part of the densely extracted features are irrelevant to our recognition task and likely to add clutter, distracting the retrieval process, keypoint selection is important for both accuracy and computational efficiency of retrieval systems.

(iii) We reduce the dimensionality of selected features to obtain improved retrieval accuracy, as common practice. First, the selected features are L2 normalized, and their dimensionality is reduced to 40 by PCA, which presents a good trade-off between compactness and discriminativeness. Finally, the features once again undergo L2 normalization.

(iv) We extract feature descriptors from query and database images, where a predefined number of local features with the highest attention scores per image are selected. Our image retrieval system is based on nearest neighbor search. Finally, we perform geometric verification using RANSAC and employ the number of inliers as the score for retrieved images.

4.6 Modules

4.6.1 Image acquisition

The image is received, and converted into a manageable entity. The user provides the image, which is used to detect if any landmark is present.

4.6.2 Image Pre-processing

The aim of pre-processing is to supress the unwanted distortions or enhance required image features. The images are resized, noise is reduced, and augmented.

4.6.3 Feature Extraction

Feature extraction is a very crucial role in any machine learning task. It is considered as a process of dimensionality reduction where very relevant features are considered for training.

4.6.4 Geometric Verification

Feature descriptors are extracted from query and database images, where a predefined number of local features with the highest attention scores per image are selected. This image retrieval system is based on nearest neighbor search. Finally, geometric verification is performed using RANSAC and employ the number of inliers as the score for retrieved images.

4.6.5 Final Predictions

Predictions from the model is obtained. The predictions will contain the index of the landmark which is present in the trained dataset. From the index the name of the landmark is obtained.

CONCLUSION

The present implemented work, we have proved that a localized feature extraction and geometric verification approach on the landmark image works exceptionally well. Following is our implemented approach with all the pre-processing steps. Currently we are trying to implement different models in the image processing domain. We will try with every possible model to increase the accuracy and efficiency of the project.

BIBLIOGRAPHY

[1] Hyeonwoo Noh, Andre Araujo, Jack Sim, Tobias Weyand, Bohyung Han. "A Large-Scale Image Retrieval With Attentive Deep Local Features". IEEE, 2017, pp. 3456-3465

[2] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik. "Rich
Feature Hierarchies for Accurate Object Detection and Semantic Segmentation tech report (v5)". IEEE, 2014

[3] Ross Girshick. "Fast R-CNN". IEE, 2015, pp. 1440 - 1448

[4] Tobias Weyand, André Araujo, Bingyi Cao, Jack Sim. "A Large-Scale Benchmark for Instance-Level Recognition and Retrieval". IEEE, 2020, pp. 2575-2584

[5] Christof Henkel, Philipp Singer. "Supporting Large-scale Image Recognition with out-of-domain Samples". eprint arXiv:2010.01650, 2020

[6] SeungKee Jeon. "1st Place Solution to Google Landmark Retrieval 2020". eprint arXiv:2009.05132, 2020