B.TECH. PROJECT ON

SmartGuard-Report and Request System based on Object Detection and Image Comparison

Submitted by
Pratham Jain 2020UCO1698
Ashish Kant 2020UCO1668
Aman Soni 2020UCO1711

Under the guidance of
Mrs. Ritu sibal
(Professor, Department of Computer Engineering)

A Project in partial fulfillment of requirement for the award of
B.TECH. in
Computer Science and Engineering

Department of Computer Engineering,
NETAJI SUBHAS UNIVERSITY OF TECHNOLOGY
DELHI
NEW DELHI-110078

2024

DECLARATION

We hereby declare that the work presented in the report entitled **"Report and Request System based on Object Detection and Image Comparison"** submitted by us in partial fulfillment of the requirements for the degree of Bachelor of Engineering in Computer Engineering at Netaji Subhas University of Technology, New Delhi, is an authentic record of our work carried out under guidance of Professor Ritu Sibal. Due acknowledgments have been given in the report to all material used. This work has not been submitted anywhere else for the reward of any other degree.

Pratham Jain	Ashish Kant	Aman Soni
2020UCO1698	2020UCO1668	2020UCO1711

Computer Engineering Department
Netaji Subhas University of Technology (NSUT)
Azad Hind Fauj Marg
Sector-3, Dwarka, New Delhi
PIN – 110078

ACKNOWLEDGEMENT

We wish to express our profound gratitude and indebtedness to Professor Ritu Sibal, Department of Computer Engineering, Netaji Subhas University of Technology, Delhi for introducing the project topic and for her inspiring guidance, constructive criticism and valuable suggestions throughout the project work. Our sincere thanks and appreciation are also to the staff members and people who helped us during the course of this project.

Pratham Jain	Ashish Kant	Aman Soni
2020UCO1698	2020UCO1668	2020UCO1711

CERTIFICATE

This is to certify that the report entitled **"Report and Request System based on Object Detection and Image Comparison"** being submitted by Pratham Jain,Ashish Kant and Aman Soni to the Department of Computer Engineering, NSUT, for the award of bachelor's degree of engineering, is the record of the bonafide work carried out by them under my supervision and guidance. The results contained in this report have not been submitted either in part or in full to any other university for the award of any degree .

Prof. Ritu Sibal

Department of Computer Engineering

Netaji Subhas University of Technology

Dwarka, New Delhi - 110078

Date: _____

ABSTRACT

A lot of items are reported lost or found on a daily basis in a large organization. With the ever-increasing number of students in an organization such as NSUT, a fully functional system for reporting and requesting such articles is the need of the hour. This project is an attempt to establish such a system for facilitating and automating any such requests in volumes that can be handled intelligently only by automated systems.

With the help of machine learning tools, we decided to train a computer to automatically categorize items and also to return the items to their rightful owner with minimal to zero intervention of man. We established secure accounts for the users of all kinds of the system.

Besides the accuracy and reliability of the system, we focused also on the interface part so that a greater number of students find it encouraging and uplifting to allot time to the system for reporting any item.

The goal is to accumulate the results from various studies in the domain and to create a convolutional neural network to classify the images in the system. The data obtained in the step is further used for allocating items to the actual owner during the comparison of the images. The project also gives useful insights on the models which can be used for classification of images and which one is best suited for any application. Further, it gives the accuracy comparison which can be used to determine the efficiency, cost and computing threshold for any particular situation. The project further dives into methods which can be used for image comparison which is used for matching the owners to their items. The project concludes with an interactive web-based application which applies the rigorous machine learning algorithms in the backend to generate the above mentioned results.

TABLE OF CONTENT

List of Figures

CHAPTER 1- INTRODUCTION

It is quite easy for humans to detect and identify the objects present in an image. Human visualization power is fast and accurate. Humans can perform complex tasks like identifying multiple objects & detect obstacles with their thinking ability. However, it is now possible to train the computers to identify various objects in an image with high accuracy owing to the requirement of faster GPUs, availability of large amounts of data and better algorithms.

Object identification is the branch of computer science which deals with formally classifying the objects into various categories according to the training data. Object identification can be done using various methods like Haar features, Scale Invariant Feature Transform (SIFT), Histogram of Oriented Features (HOG) as well as deep learning models including Convolutional Neural Networks (CNN) and One-Shot Recognition.

In the standard CNN method, input is fed to the number of layers and at the last layer the probabilities of classes are output. For e.g., if we train a model on images of cats and dogs, we can send another image of similar cat or a dog to predict its category. In the case

of one-shot classification, only one training example of a class is used. Model is trained on several instances but the images need to be in a similar domain as the training example. Facial recognition is one such example of one-shot recognition.

The method implemented in the project is the CNN method. It is highly useful in computer vision algorithms where there is a lot of training data and computational power available. In a typical convolutional neural network, the system is fed with a lot of training data of images classified into certain categories. The model is trained on the same and is used as such for object detection.

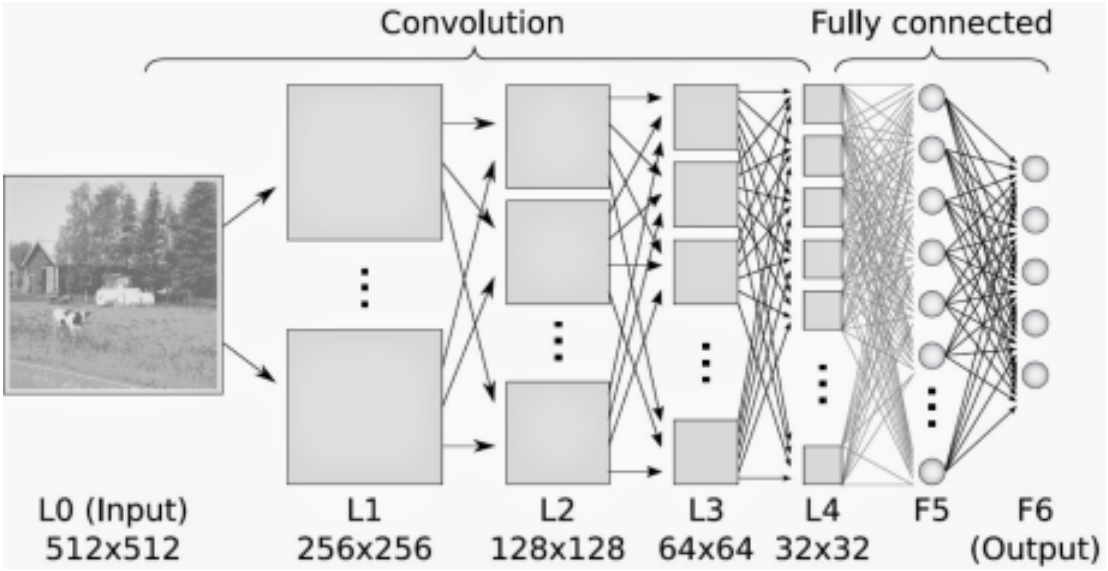

Fig 1.1 CNN Layers for the object detection

It is used to find out the item category from the given image of the item to the system. The trained model uses the image input and formulates a fully connected layer to generate the probability of each class on which the model was already trained. The class with the highest probability is assigned to the object, hence achieving the required task of Object Detection. Various Object detection models were also compared in the project, namely Inception V3, Inception V4, Xception, VGG16, VGG19.

a. Watch

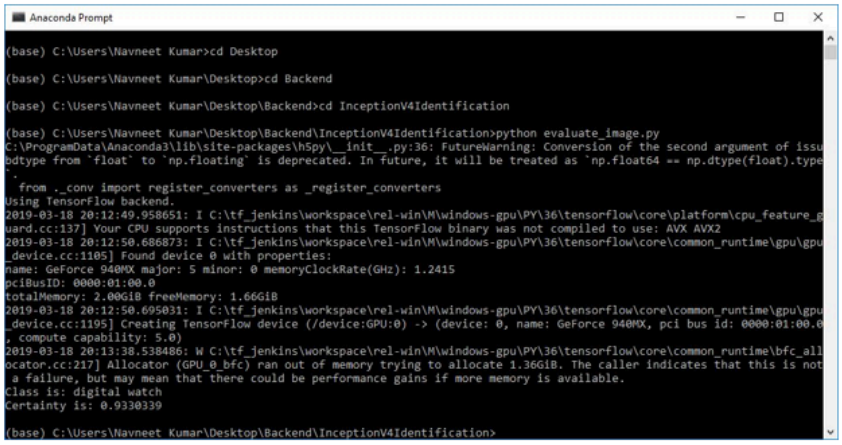

Fig 1.2 Inception V4 model based on ImageNet used for Object Detection

1.1 Image Comparison

Image comparison is the method in computer vision that is used to find the most similar image to the image provided to the system. Humans can identify on the basis of provided images and can find the most similar image to a particular image in a given set of images. Although with computers, it is not so. Human visualization is fast and accurate whereas computers need to be trained to find the most similar image to a particular image.

The methods which are used in the project are either dependent on the feature vectors of a particular image or reducing the image features to produce generalization. The method based on the first approach are:

1. Feature Vectors using Inception V3: In this method, the given image is passed through the Inception V3 pretrained model to find the feature vector of the particular image. The second image is also passed through the same Inception V3 model to generate the feature vector of the second image. The feature vectors thus generated are compared using either Euclidean Distance or Cosine Similarity Metrics.

2. Feature Vectors using VGG19: This method is similar to the one described above. Here the VGG19 pre-trained model is used instead of the Inception V3 model. The image is input to the VGG19 model to produce the feature vectors which are then compared to

another set of feature vectors generated from a different image. Fig 1.3 represents the dissection of the VGG19 model used to generate the feature vectors. In this case, the kNN (k-Nearest Neighbor) method is used to generate the most similar images to a given image. A plot is also generated on the same image.

3. Siamese Network: These are dual connection Convolutional Neural Networks (CNNs) which generate a set of feature vectors which are then combined/merged using contrastive loss function or binary cross entropy to generate a similarity metrics which is used to predict if two images represent the similar object or not. Fig 1.4 represents a sample Siamese Network which is using the Contrastive Loss Function after a fully connected layer to find if two images represent a same object or not.

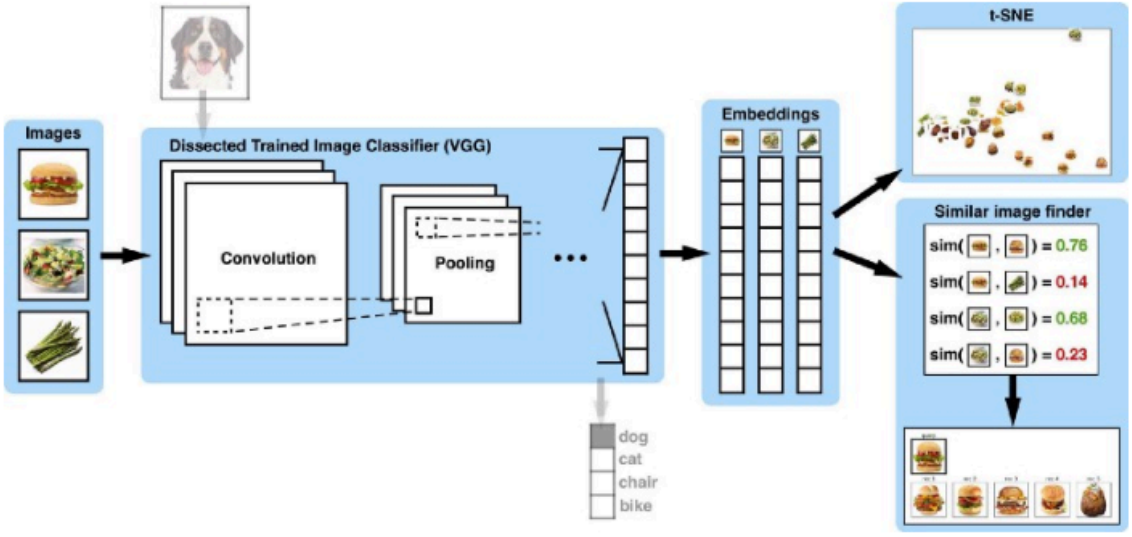

Fig 1.3 Dissecting of VGG19 model which is used to build similar image finder

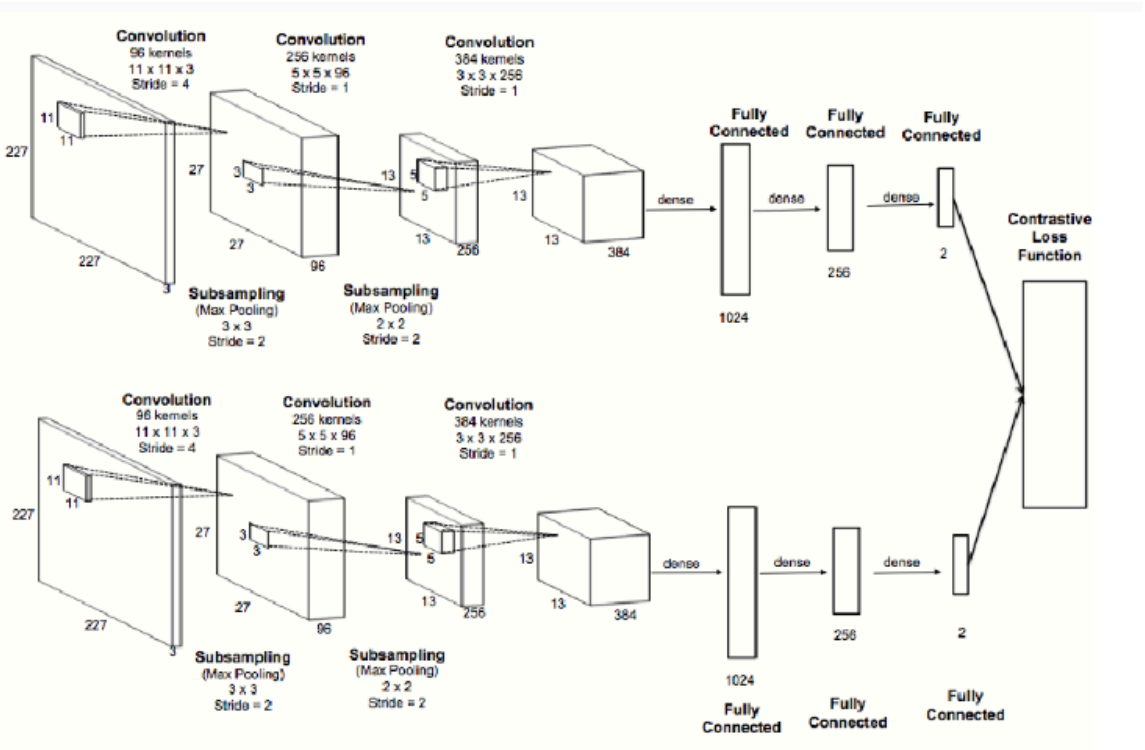

Fig 1.4 Example of Siamese Network using Contrastive Loss Function

The second method which involves reduction of image dimensionality to represent generalization uses the autoencoders to generate the most similar images to a particular input image. The method based on the second approach is described below:

1. Autoencoders for Image Similarity: Autoencoders are a type of neural network and are comprised generally of an encoder and a decoder. The encoder is used to encode the image data and decoder is used to decode the image data so that an almost perfect reconstruction of the given image is possible. The image thus generated is a good representation of the initial image and can certainly be used as a generalization for the object represented in the image. Since the image and the autoencoder neural network now has lower dimensionality, we can train the autoencoder to use the most salient features of the object during training. Therefore, only macroscopic details of the input data are retained where microscopic details are removed for the task. The disadvantage of the method is that the model needs to be trained on new data before it can be used on it

whereas the method described using feature vectors can be used as provided because they are already pre trained on the required categories of objects.

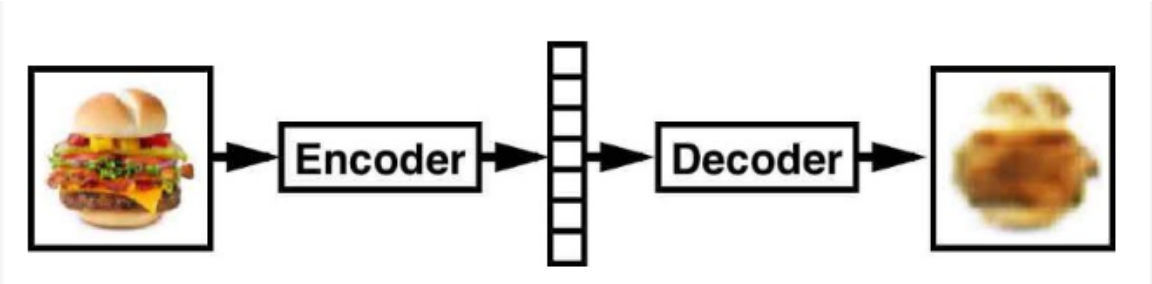

Fig 1.5 Schematic diagram representing the working of autoencoder model

The given Fig 1.5 represents how the autoencoder uses the combination of an encoder and a decoder to generate an almost perfect reconstruction of the image which can then be compared using kNN (k-Nearest Neighbor) for finding the most similar images in the database to the query image.

CHAPTER-2 MOTIVATION

The main motivation to take up such a project was the absolute lack of a functional system present in NSUT. As of now, reporting of lost and found items is mostly done through emails. The details of items are universally available and therefore vulnerable to false claims.

There may be a case that a student is not active on G-Mail and also there are chances of incorrect entitlement by other students. Users can use this system to locate their lost items.

Each day items of high value and low value are lost. Incidents like these can lead to panic and cause students to take the time to retrace their steps or to hope that the item may be found using a typical lost-and-found system. This system can help the students to find their lost item or to report any found item. Reporting of the lost items or found items can also be done by the administrative account in case the student is not interested in manually completing the task. This project offers a great on-ground usability.

A point-wise summary of the points suggested is given below:

1. Large organizations such as colleges tend to have a huge number of articles/items lost

each day with no proper reporting of such items. As a result, the rightful owner of the item may or may not get his item in due time.

2. As the proper address of remote requests can only happen online, we try to bridge the gap between lost & found units using a central directory managed through an administrative account.

3. Machine learning algorithms help greatly in dealing with matching enormous numbers of articles with great accuracy and hence performing return of appropriate items when the number of requests coming to the system are of magnitude of thousands or even more.

4. Users will be able to upload pictures of lost items and can also provide descriptions of the items. The application will pair lost and found items if they match.

5. A simple and user-friendly Web-based UI helps the system to be acknowledged by greater public and thus the reach of the system is improved.

6. All of the lost item's information will be maintained in real-time so the user can get up-to-date information about the items lost.

Fig 1.6 Motivation elements for taking up the project

Chapter-3 literature review

In the paper [1], Karen et al. investigated the effect on accuracy of large-scale image recognition based on the convolutional network depth. The main constraint that they imposed was reducing the convolution filters to 3*3 so that they could perform thorough evaluation of networks on increasing depths. The process showed significant improvements by increasing the depths up to 16-19 weight layers. Their findings were based on the ImageNet 2014 submission, where they excelled in localization and classifications tracks. They also proved accuracy of models to other datasets and hence achieved generalization. The research work was also available publicly to facilitate further research on computer vision deep visual representation.

In the paper [2], Francois et al. interprets inception modules as intermediate step in-between regular convolution and depth wise differentiable convolution operation. Depth wise separable convolution operation can thus be introduced as Inception module in form of maximally large number of towers. Inception modules replaced by depth wise separable convolutions is form adapted in the Xception model. It slightly out performed Inception V3. The performance gain is more due to efficient manner of usage of model parameters rather than numbers of parameters as both of them have the same number of parameters.

In the paper [3], regarding Inception V3 model, Christian et al. explored some ways to scale up networks in such a way that directs the usage of the added calculus in an accurate way to get maximum efficiency and maintaining the time constraints. It was used first in 2012 in the ImageNet Large Scale Visual Representation Competition. Factorized convolutions and aggressive regularization were used for increasing computational efficiency. The accuracy although was good and the model had top-1 error of 21.2% and top-5 error of 5.6%. It uses less than 25 million parameters for object

detection. When used with 4 model ensemble and multi-crop evaluation, it reported 3.5% top-5 error and 17.3% top-1 error thus improving efficiency.

In the article [4], Anson et al. describes a method used to build a similar image finder by dissecting the VGG19 model and using the feature vectors from an image database to search for the most similar images. The model works on the principle of transfer learning i.e., “low level features are transferable”. Trained model such as VGG is used and with its last layers removed, the dissected version of the model can be used to generate a set of low dimensional feature vectors. The features vectors available can be compared using cosine similarity or Euclidean distance to make similarity predictions.

In the article [5], Anson et al. explained how autoencoders can be used to find similar images. He used the method in unlabeled image dataset. He explained what is an encoder and more concisely focused on usefulness of autoencoders. It was also made clear that autoencoders generate an almost perfect reconstruction to produce a generalized image of lower dimensionality which is used for image similarity comparison. After passing the images through the encoder, he used k-Nearest Neighbor (kNN) algorithm to find out the 5 most similar image to the given image. The performance of the algorithm was not perfect, although it provided useful insights to the process of unsupervised image similarity.

In the paper [6], Gregory et al. used the Siamese neural network of One-Shot image recognition. In cases where only minimal data is available, process of learning features of an object becomes computationally expensive and practically unusable. To counter the same, One-Shot recognition method is devised which is used to make correct predictions based on only single example of any new class. In the paper, they explored Siamese neural networks which have inherent property of ranking similarity between inputs. After tuning the network, they capitalized on discriminative features to explore the predictive power, not just to new sample data, but also to the data belonging to entirely new classes. Using convolutional architecture, they were able to achieve better performance which exceeded the performance of other deep learning models.

Chapter-4 Problem Statement

4.1 PROBLEM STATEMENT

Many organizations, including businesses, educational institutions, and public facilities, encounter challenges in managing lost and found items efficiently. The current methods often involve manual recording and storage of lost items, leading to inefficiencies, misplacements, and difficulties in reuniting lost items with their owners. This results in frustration for both individuals who have lost items and staff responsible for managing the lost and found process.

Key Challenges:

1. Inefficient Reporting: Employees or visitors often face difficulties in reporting lost items promptly, leading to delays in initiating the search and retrieval process.

2. Manual Recording: The traditional approach of manually recording lost items on paper or spreadsheets is prone to errors, inconsistencies, and data loss.

3. Limited Visibility: Without a centralized system, tracking the status of lost items and matching them with found items becomes challenging, reducing the chances of successful recovery.

4. Lack of Communication: Inadequate communication channels between staff and individuals reporting lost items hinder effective coordination and updates regarding the search process.

5. Security Concerns: Without proper security measures, there's a risk of theft or mishandling of valuable lost items stored in the organization's premises.

6. Ownership Verification: Verifying the ownership of found items and ensuring rightful return to their owners can be a cumbersome and time-consuming process without proper documentation and authentication procedures.

7. Unclaimed Items Management: Managing unclaimed items effectively becomes a challenge, requiring clear policies and procedures for disposal or donation while respecting privacy and legal considerations.

4.2 OBJECTIVES

To develop a comprehensive Lost and Found Management System that addresses the aforementioned challenges by providing a streamlined, automated, and secure process for reporting, recording, tracking, and retrieving lost items within the organization. Proposed Solution Features:

1. Online Reporting Portal: A user-friendly platform for employees, visitors, or customers to report lost items conveniently and promptly.

2. Centralized Database: A centralized database to record detailed information about lost and found items, including descriptions, locations, dates, and images.

3. Search and Matching Algorithms: Advanced search algorithms to match lost and found items efficiently, increasing the likelihood of successful recovery.

4. Automated Notifications: Automated notifications to inform stakeholders about reported lost items, updates on search progress, and item retrieval.

6. Authentication Mechanisms: Robust authentication mechanisms to verify the ownership of found items, ensuring their rightful return to the owners.

Therefore we will aim:

- To study the various machine learning algorithms that can be used for Object Identification.
- To study Convolutional Neural Network (CNN).
- To study the Inception V4 model for Object Detection.
- To study the various machine learning algorithms that can be used for Image Comparison
- To study Image Comparison Using model VGG-19.
- To compare various Image Comparison models on the basis of Accuracy.

- To be able to feed into the database i.e MONGODB various images of objects which will be later used as datasets for classification and comparison.
- To connect the software to a database and show the details of the objects stored.

CHAPTER-5 METHODOLOGY

5.1 Structure of website & Flow of control

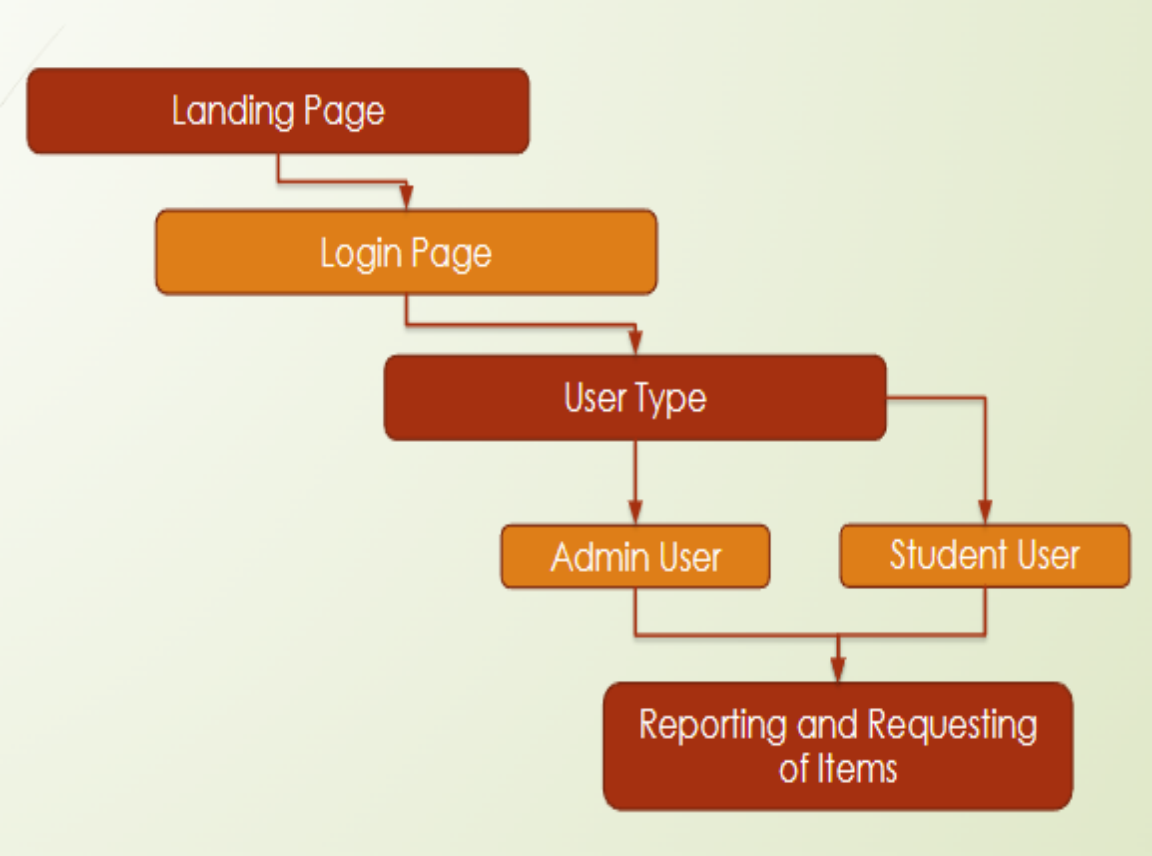

Fig- 1.7

5.2 Report and Request System

Report and Request System is a fully functional web-based application which is used to provide an interactive user interface (UI) for the students as well as admin members. It is

used to report lost items as well as found items through the administration account or student account. The main difference between the two types of accounts supported in the system is that the administration accounts have all the information regarding the items already present in the system and history of the items which have been reported in the system. The students can report any lost item or the found item in the system but they cannot view or change the other lost or found requests present in the system.

The various pages in the system are as follows:

1. Landing Page: The landing page is an animation of 5 different images which are depicted in the form of a transition.

Fig 1.8 Landing page in Report and Request System

2. Login Page: The page is used to ask the current user if he/she wants to log in as an admin or a student.

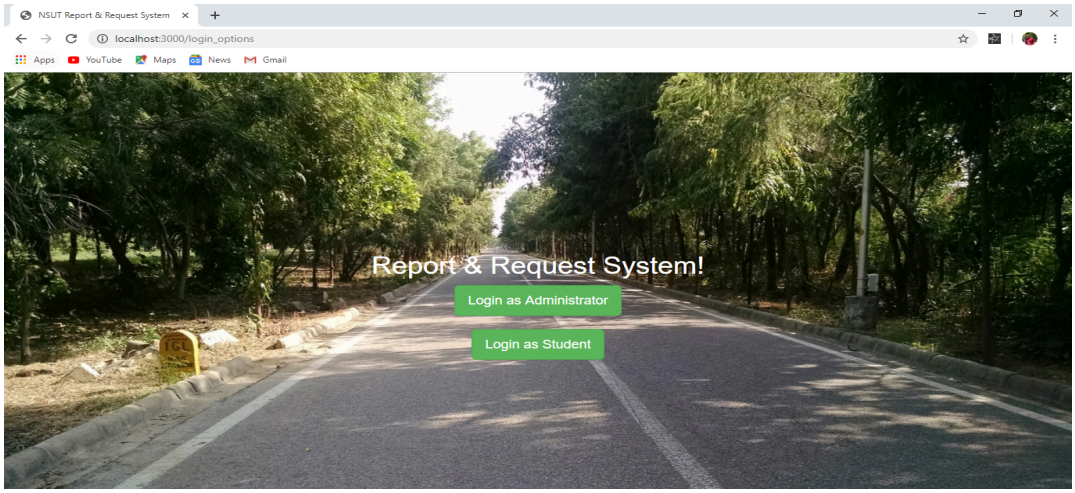

Fig 1.9 Login page in Report and Request System

3. Admin Login Page

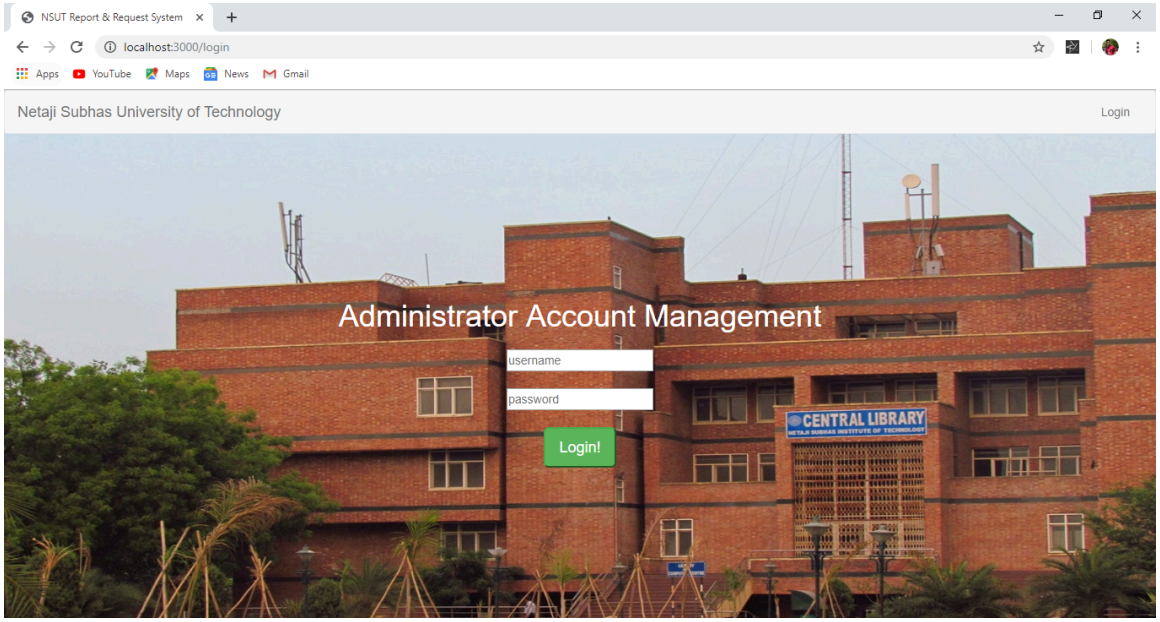

Fig 1.10 Login page for Admin User

4. Student Login Page

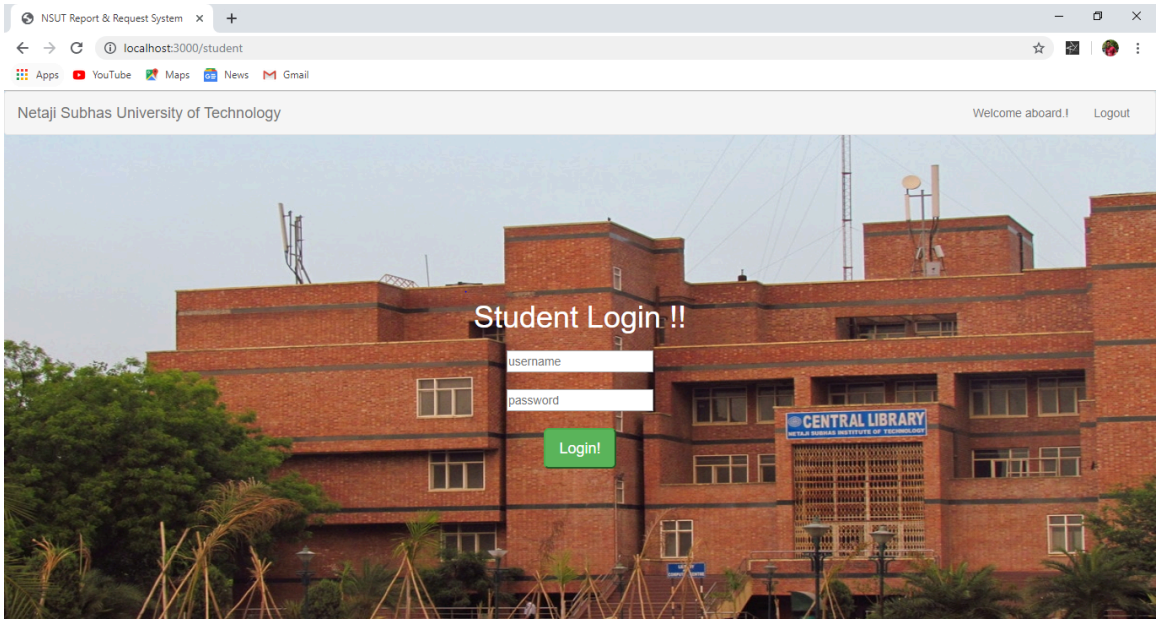

Fig 1.11 Login page for Student User

There are various cases that may arise during the reporting of items in the system:

Case 1. Reporting a lost item in the system through an admin account: It is the simple case where the admin account is logged in and is used to report a lost item. The student in this case approaches the admin and the student's lost request is logged in the system with the necessary details and the audit history is augmented to include the current lost request.

The initial state of the system is an empty web page with no records of any data in the system. Fig 1.10 represents the state of the system in such a case. Fig 1.11 represents the addition of details regarding the object reported in the system. Fig 1.12 represents the state of the system after the addition of the object.

The item added after the request can also be viewed in more detail. The reported item in the system also supports editing of the details. Comments can be added on individual

items reported in the system by members logged in or by other admin members. Comments are recognized by admin name and can be edited only by the same person who has put the comment.

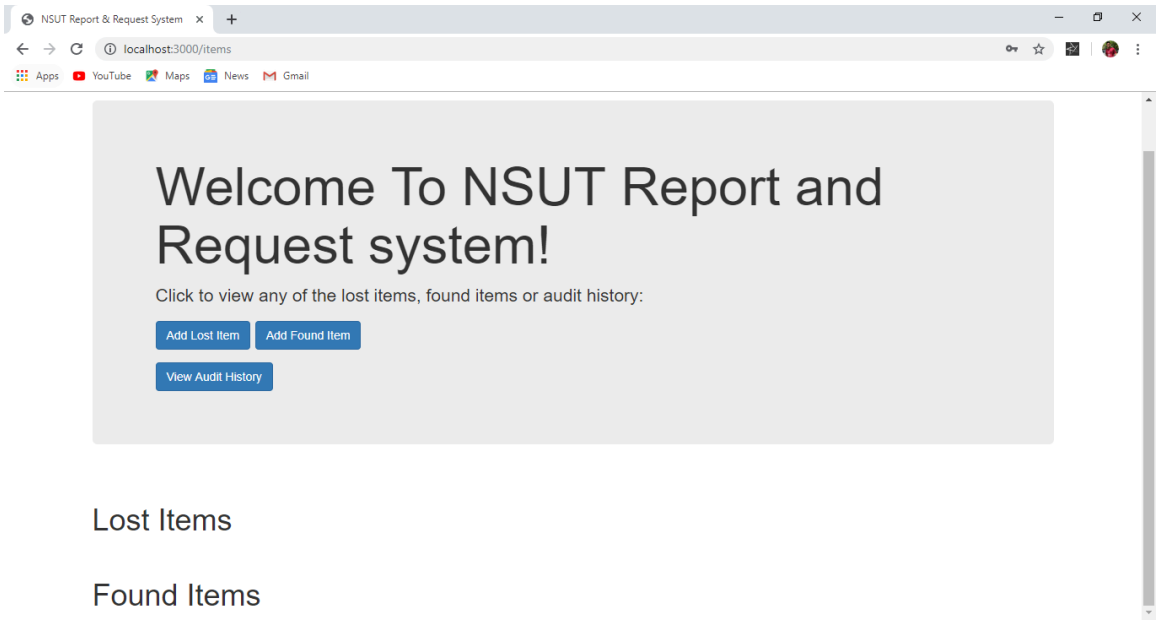

Fig 1.12 Initial state of system logged in by Admin User

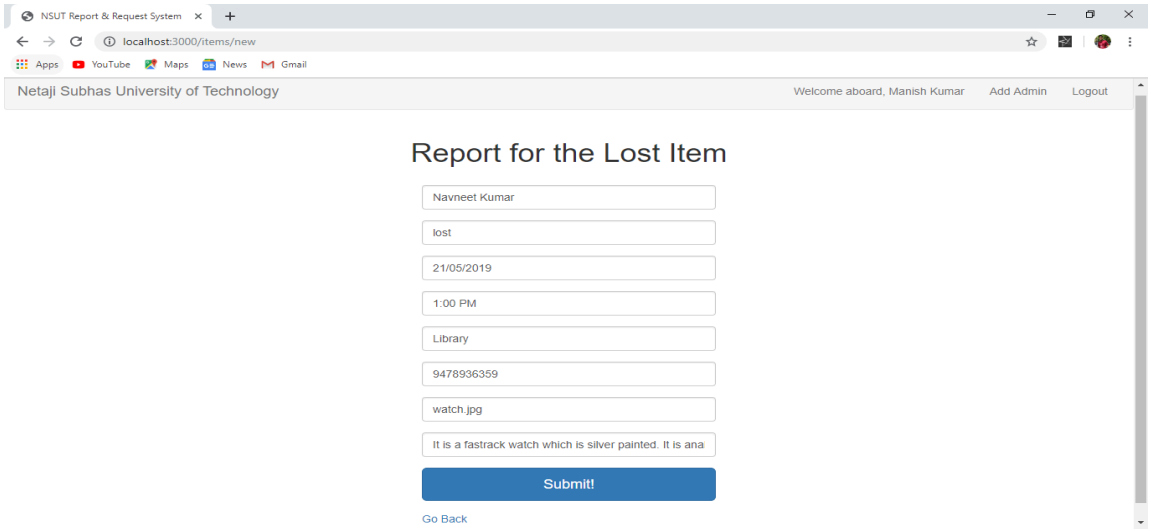

Fig 1.13 Addition of details regarding the lost item

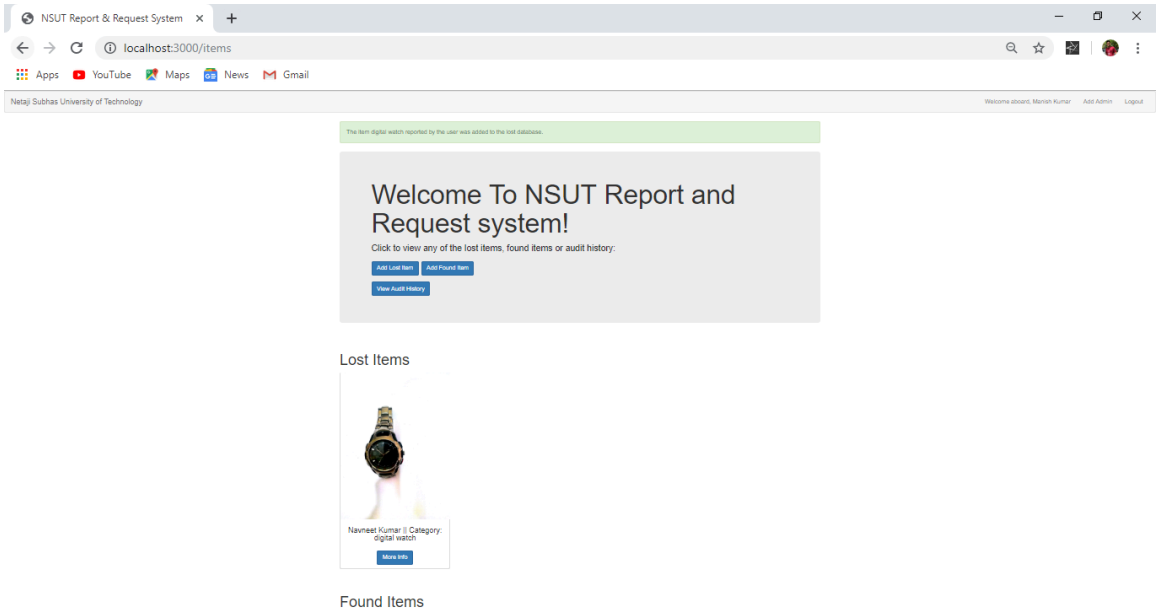

Fig 1.14 State of the system after the addition of watch reported in the system

Case 2. Reporting a found item in the system through an admin account: It is the simple case where the admin account is logged in and is used to report a found item. The student in this case approaches the admin and the student's found request is logged in the system with the necessary details and the audit history is augmented to include the current lost request.

The initial state of the system is an empty web page with no records of any data in the system. Final state has the record of the current request.

The item added after the request can also be viewed in more detail. The reported item in the system also supports editing of the details. Comments can be added on individual items reported in the system by members logged in or by other admin members. Comments are recognized by admin name and can be edited only by the same person who has put the comment.

Case 3. Reporting a lost item in the system through student account: It is the case where a student who has logged in, reports that he/she has lost a particular item. He/she also uploads the image of the object from the client side which is uploaded to the lost folder in the system database. Now two cases arise in the same case:

1. The object is not available in found log

In this case the object is added to the lost database and is displayed with its details in the report and request website.

2. The object is available in found log

In this case the object from the found log is deleted and is returned to the current user. The current user is given the number to call from which he can receive his particular item.

Case 4. Reporting a found item in system through student account: It is the case where a student who has logged in, reports that he/she has found a particular item. He/she also uploads the image of the object from the client side which is uploaded to the found folder in the system database. Now two cases arise in the same case:

1. The object is not available in lost log

In this case the object is added to the found database and is displayed with its details in the report and request website.

2. The object is available in lost log

In this case the object from the lost log is deleted and is returned to the user who lost the item. The current user is given the number to call from which he/she can return the particular item.

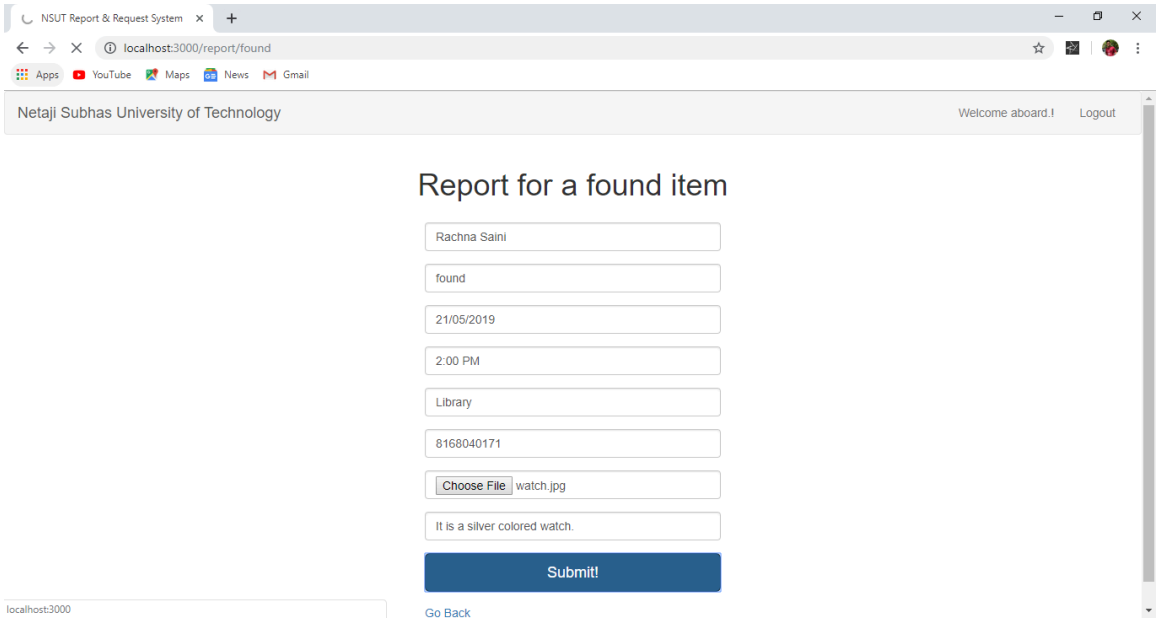

Fig 1.15 Found request by a Student User

WORK DONE TILL DATE

1. We have made the front-end of the project
2. We have studied and researched various methodologies by which we can train our own model
3. We have selected the API which we will use in case we are not able to train a model of suitable accuracy.

FUTURE WORK TO BE DONE

1. Front-end has to be integrated with the back-end.
2. Training of models has to be done.
3. For bonus, we will also try to make an app.

REFERENCES

[1] Branislav Holländer. “Siamese Networks: Algorithm, Applications And PyTorch Implementation”, Internet: https://becominghuman.ai/siamese-networks-algorithm-applications-and-pytorch-implementation-4ffa3304c18, [Sep 24, 2018]

[2] Faizan Shaikh. “Deep Learning in the Trenches: Understanding Inception Network from Scratch”, Internet: https://www.analyticsvidhya.com/blog/2018/10/understanding-inception-network-from-scratch/, [Oct 18, 2018]

[3] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Alex Alemi. “Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning”, Internet: https://arxiv.org/abs/1602.07261, [Feb 23, 2016]

[4] SH Tsang. “Review: Inception-v3—1st Runner Up (Image Classification) in ILSVRC 2015”, Internet: https://medium.com/@sh.tsang/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c, [Sep 10, 2018]

[5] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens. “Rethinking the Inception Architecture for Computer Vision”, [May 23, 2015]

[6] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, Zbigniew Wojna. “Rethinking the Inception Architecture for Computer Vision**”,** [Dec 2, 2015]

[7] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Alexander A. Alemi. “Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning” Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)

[8] Karen Simonyan, Andrew Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition”, [Sept 4, 2014]

[9] Francois Chollet. “Xception: Deep Learning with Depthwise Separable Convolutions”, [Oct 7, 2016]

[10] Christian Szegedy, Vincent Vanhoucke, Sergey loffe, Jonathan Shlens, Zbigniew Wojna. “Rethinking the Inception Architecture for Computer Vision”, [Dec 2, 2015]

[11] Anson Wong. “Building a Similar Images Finder without any training!” Internet: https://towardsdatascience.com/building-a-similar-images-finder-without-any-training-f69c0db900b5, [Dec 16, 2017]

[12] Anson Wong. “Finding Similar Images using Autoencoders” Internet: https://towardsdatascience.com/find-similar-images-using-autoencoders-315f374029ea, [Aug 13, 2017]

[13] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov. “Siamese Neural Network for One-shot Image Recognition”, [Dec 5, 2016]