

SCENARIO BASED IMAGE GENERATION

Submitted in partial fulfillment of the requirements for
the award of
Bachelor of Engineering degree in Computer Science and Engineering

by

S. SIVA (Reg. No. 39290094)
S. VASANTH (Reg. No. 39290130)



DEPARTMENT OF COMPUTER SCIENCE
SCHOOL OF COMPUTING

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited with Grade "A" by NAAC
JEPPIAAR NAGAR, RAJIV GANDHI SALAI,
CHENNAI – 600 119

MARCH - 2022



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)

Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE

BONAFIDE CERTIFICATE

This is to certify that this project report is the bonafide work of **S. SIVA (Reg. No. 39290094)** and **S. VASANTH (Reg. No. 39290130)** who carried out the project entitled "**Scenario based image generation**" under the supervision
from _____ to _____.

Internal Guide

Dr. KANIPRIYA

Head of the Department

Dr. L. LAKSHMANAN M.E., Ph.D.,

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

I **S. SIVA (Reg. No. 39290094)** and **S. VASANTH (Reg. No. 39290130)** hereby declare that the Project Report entitled “**Scenario based image generation**” is done by me under the guidance of **Mrs. Dr. KANIPRIYA** Department of Computer Science at Sathyabama Institute of Science and Technology is submitted in partial fulfillment of the requirements for the award of Bachelor of Science degree in Computer Science.

DATE:

PLACE: CHENNAI

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr.T.Sasikala M.E., Ph.D, Dean**, School of Computing **Dr. L. Lakshmanan M.E., Ph.D. ,** and **Dr.S.Vigneshwari M.E., Ph.D. Heads** of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide Dr./Mr./ Ms for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

This model works to get the scenario of an image irrespective of what it contains in it. It functions on open cv, NLP, webpage. This system searches the image with the dictionary and give the desired result. The user can get the correct scenario of any kind of image he has. The vision of the computer has been improved magnificently in recent years. By using all the new technologies object detection became very easy and also gives accurate results. The next gen computers are not only detecting objects but also matching the scenarios. The algorithm used to find scenario is ResNet – 50. Image Net is the dataset we used to train this model.

TABLE OF CONTENTS

ABSTRACT	v
LIST OF FIGURES	viii

CHAPTER No.	TITLE	PAGE No.
1.	INTRODUCTION	1
	1.1 INTRODUCTION	1
	1.2 RESEARCH AND SIGNIFICANCE	1
2.	AIM AND SCOPE	2
	2.1 AIM OF PROJECT	2
	<i>2.1.1 OBJECTIVES</i>	2
	2.2 SCOPES	2
	<i>2.2.1 ADVANTAGES</i>	2
3.	SYSTEM DESIGN & METHODOLOGY	3
	3.1 EXISTING SYSTEM	3
	3.2 SOFTWARE REQUIREMENTS	4
	3.3 HARDWARE REQUIREMENTS	4
	3.4 SYSTEM FUNCTIONALITIES	4
	3.5 TRAINING	6
	3.6 REPRESENTATION LEARNING	7

4.	SOFTWARE DEVELOPMENT METHODOLOGY	9
	4.1 SOFTWARE DEVELOPMENT STRATEGIES	10
	4.2 THE STAGE OF SOFTWARE DEVELOPMENT LIFECYCLE	12
	4.2.1 PLANING AND REQUIREMENT ANALYSIS	12
	4.2.2 DEFINING REQUIREMENT	12
	4.2.3 DESIGNING THE PRODUCT ARCHITECTURE	12
	4.2.4 BUILDING OR DEVELOPING THE PRODUCT	12
	4.3.5 TESTING THE PRODUCT	
	4.3.6 DEPLOYMENT IN THE MARKET AND MAINTENANCE	13
		13
5	RESULTS AND DISCUSSION	15
6	CONCLUSION AND FUTURE WORK	21
	REFERENCES	
	APPENDIX	
	A. SOURCE CODE	

LIST OF FIGURES

FIGURE No.	FIGURE NAME	PAGE No.
4.1	WORKING DIAGRAM	14
5.1	HOME PAGE	15
5.2	UPLOADING IMAGE	17
5.3	RESULT 1	18
5.4	SELETING IMAGE	19
5.5	RESULT 2	20

CHAPTER 1

INTRODUCTION

INTRODUCTION

In the 21st century with advancement of the technology is very much helpful for all the people including physically challenged. In ancient times it was too difficult for them to learn study but now with the use of technologies it became very easy for them.

A **scenario** is a specific possibility. When people are wondering what to do, they think of different scenarios. For example, if a man is walking on a road. By using our webpage user can give any image and he /she can get the exact scenario of that image he gives. That's how this project going to work. This idea will be useful for so many people in so many ways. In this document we are going to see how we are developing this, what are all the technicalities used, what are all the uses of it and so on.

Visually impaired people finding it very difficult to visualize things in all aspects. Even though they have various different methods to learn still they are having very much difficulty in finding the exact scenario of a picture. So this idea will be definitely helpful for them to visualize the scenario they wanted

RESEARCH ANDSIGNIFICANCE

The primary aim of this proposed idea is to find the scenario from the image which user gives. The image can be anything. Our system will give the exact scenario of that image. We are building our dataset in a way that it can give the exact scenario of any image which is shared by the user.

CHAPTER 2

AIM AND SCOPE

AIM OF PROJECT

The primary aim of this proposed system is to find the scenario from the image which user gives. The image can be anything. Our system will give the exact scenario of that image.

OBJECTIVES

The proposed model “Image based scenario generation” is based on the new technologies which makes the whole process efficient. It is going to be very important and profitable for the whole society especially for the visually challenged people and also for the people who loves to surf images. This system will provide more accurate scenario based on the image which user gives. It will be very helpful for the visually impaired to visualize the scenario and thus it will help them to learn more efficiently.

SCOPE

This model will help in creating exact scenario from the image which user gives. This will be helpful for the people who loves to search images especially for the visually limited people. This will surely help in bringing down the barrier between visually impaired and sighted people. That’s how it will work.

ADVANTAGES

- Better results while surfing images.
- Breaching the

CHAPTER 3

SYSTEM DESIGN & METHODOLOGY

EXISTING SYSTEM

In existing system, it aims at generating automated captions by learning the contents of the image. At present images are annotated with human intervention and it becomes nearly impossible task for huge commercial databases. The image database is given as input to a deep neural network (Convolutional Neural Network (CNN)) encoder for generating "thought vector" which extracts the features and nuances out of our image and RNN (Recurrent Neural Network) decoder is used to translate the features and objects given by our image to obtain sequential, meaningful description of the image. In this paper, we systematically analyze different deep neural network-based image caption generation approaches and pretrained models to conclude on the most efficient model with fine-tuning. The analyzed models contain both with and without 'attention' concept to optimize the caption generating ability of the model. All the models are trained on the same dataset for concrete comparison.

Describing the contents of images is a challenging task for machines to achieve. It requires not only accurate recognition of objects and humans, but also their attributes and relationships as well as scene information. It would be even more challenging to extend this process to identify falls and hazardous objects to aid elderly or users in need of care. This research makes initial attempts to deal with the above challenges to produce multi-sentence natural language description of image contents. It employs a local region-based approach to extract regional image details and combines multiple techniques including deep learning and attribute learning through the use of machine learned features to create high level labels that can generate detailed description of real-world images. The system contains the core functions of scene classification, object detection and classification, attribute learning, relationship detection and sentence generation. We have also further extended this process to deal with open-ended fall detection and hazard identification. In comparison to state-of-the-art related research, our system shows superior robustness and flexibility in dealing with test images from new, unrelated domains, which poses great challenges to many existing methods. Our system is evaluated on a subset from Flickr8k

and Pascal VOC 2012 and achieves an impressive average BLEU score of 46 and outperforms related research by a significant margin of 10 BLEU score when evaluated with a small dataset of images containing falls and hazardous objects. It also shows impressive performance when evaluated using a subset of IAPR TC-12 dataset.

SOFTWARE REQUIREMENTS

The basic software requirements to run the program are:

- i. Windows 10
- ii. Flask
- iii. HTML, CSS, JavaScript
- iv. Browser e.g., Mozilla Firefox, Safari, chrome, etc.
- v. Torch Version
- vi. TensorFlow
- vii. ResNet
- viii. Python
- viii. Cuda

HARDWARE REQUIREMENTS

The basic hardware required to run the program are:

- i. Hard disk of 20 GB
- ii. System memory (RAM) of 4 GB
- iii. I3 processor-based computer or higher.
- iv. VRAM 10GB
- v. GPU

System Functionalities

- **Python:** Used for Frontend
- **Flask:** Used for Backend
- **HTML, CSS, JavaScript:** Used as Frontend

- **TensorFlow:** Used for training dataset
- **ResNet:** Used for Object detection
- **Cuda:** Used to increase the speed of training
- **VRAM:** Used for storing processes virtually

CLIP is a significant step towards flexible and practical zero-shot computer vision classifiers. As mentioned above, the comparison to Visual N-Grams is meant for contextualizing the performance of CLIP and should not be interpreted as a direct methods comparison between CLIP and Visual N-Grams as many performance relevant differences between the two systems were not controlled for. For instance, we train on a dataset that is 10x larger, use a vision model that requires nearly 100x more compute per prediction, likely used over 1000x their training compute, and use a transformer-based model which did not exist when Visual N-Grams was published. As a closer comparison, we trained a CLIP ResNet-50 on the same YFCC100M dataset that Visual N-Grams was trained on and found it matched their reported ImageNet performance within a V100 GPUday. This baseline was also trained from scratch instead of being initialized from pre-trained ImageNet weights as in Visual N-Grams. CLIP also outperforms Visual N-Grams on the other 2 reported datasets. On Yahoo, CLIP achieves a 95% reduction in the number of errors, and on SUN, CLIP more than doubles the accuracy of Visual N-Grams. To conduct a more comprehensive analysis and stress test, we implement a much larger evaluation suite detailed in Appendix A. In total we expand from the 3 datasets reported in Visual N-Grams to include over 30 datasets and compare to over 50 existing computer vision systems to contextualize results. CLIP is pre-trained to predict if an image and a text snippet are paired together in its dataset. To perform zero-shot classification, we reuse this capability. For each dataset, we use the names of all the classes in the dataset as the set of potential text pairings and predict the most probable (image, text) pair according to CLIP. In a bit more detail, we first compute the feature embedding of the image and the feature embedding of the set of possible texts by their respective encoders. The cosine similarity of these embeddings is then calculated, scaled by a temperature parameter τ , and

normalized into a probability distribution via a SoftMax. Note that this prediction layer is a multinomial logistic regression classifier with L2-normalized inputs, L2-normalized weights, no bias, and temperature scaling. When interpreted this way, the image encoder is the computer vision backbone which computes a feature representation for the image and the text encoder is a hypernetwork (Ha et al., 2016) which generates the weights of a linear classifier based on the text specifying the visual concepts that the classes represent. Lei Ba et al. (2015) first introduced a zero-shot image classifier of this form while the idea of generating a classifier from natural language dates back to at least Elhoseiny et al. (2013). Continuing with this interpretation, every step of CLIP pre-training can be viewed as optimizing the performance of a randomly created proxy to a computer vision dataset which contains 1 example per class and has 32,768 total classes defined via natural language descriptions. For zero-shot evaluation, we cache the zero-shot classifier once it has been computed by the text encoder and reuse it for all subsequent predictions. This allows the cost of generating it to be amortized across all the predictions in a dataset.

Training

We train a series of 5 ResNets and 3 Vision Transformers. For the ResNets we train a ResNet-50, a ResNet-101, and then 3 more which follow Efficient Net-style model scaling and use approximately 4x, 16x, and 64x the compute of a ResNet-50. They are denoted as RN50x4, RN50x16, and RN50x64 respectively. For the Vision Transformers we train a ViT-B/32, a ViT-B/16, and a ViT-L/14. We train all models for 32 epochs. We use the Adam optimizer (Kingma & Ba, 2014) with decoupled weight decay regularization (Loshchilov & Hutter, 2017) applied to all weights that are not gains or biases, and decay the learning rate using a cosine schedule (Loshchilov & Hutter, 2016). Initial hyperparameters were set using a combination of grid searches, random search, and manual tuning on the baseline ResNet50 model when trained for 1 epoch. Hyper-parameters were then adapted heuristically for larger models due to computational constraints. The learnable temperature parameter was initialized to the equivalent of 0.07 from (Wu et al., 2018) and clipped to prevent scaling the logits by more than 100 which we found necessary to prevent training instability. We use a very large minibatch size of 32,768. Mixed-precision (Micikevicius et al., 2017) was used to accelerate training and save memory. To save additional memory, gradient checkpointing (Griewank & Walther, 2000; Chen et al., 2016), half-precision Adam statistics (Dhariwal et al., 2020), and half-precision stochastically rounded text encoder

weights were used. The calculation of embedding similarities was also sharded with individual GPUs computing only the subset of the pairwise similarities necessary for their local batch of embeddings. The largest ResNet model, RN50x64, took 18 days to train on 592 V100 GPUs while the largest Vision Transformer took 12 days on 256 V100 GPUs. For the ViT-L/14 we also pre-train at a higher 336pixel resolution for one additional epoch to boost performance similar to Fires (Touvron et al., 2019). We denote this model as ViT-L/14@336px. Unless otherwise specified, all results reported in this paper as “CLIP” use this model which we found to perform best.

Representation Learning

While we have extensively analyzed the task-learning capabilities of CLIP through zero-shot transfer in the previous section, it is more common to study the representation learning capabilities of a model. There exist many ways to evaluate the quality of representations as well as disagreements over what properties an “ideal” representation should have. Fitting a linear classifier on a representation extracted from the model and measuring its performance on various datasets is a common approach. An alternative is measuring the performance of end-to-end fine-tuning of the model. This increases flexibility, and prior work has convincingly demonstrated that fine-tuning outperforms linear classification on most image classification datasets. While the high performance of fine-tuning motivates its study for practical reasons, we still opt for linear classifier-based evaluation for several reasons. Our work is focused on developing a high-performing task and dataset-agnostic pre-training approach. Fine-tuning, because it adapts representations to each dataset during the fine-tuning phase, can compensate for and potentially mask failures to learn general and robust representations during the pre-training phase. Linear classifiers, because of their limited flexibility, instead highlight these failures and provide clear feedback during development. For CLIP, training supervised linear classifiers has the added benefit of being very similar to the approach used for its zero-shot classifiers which enables extensive comparisons and analysis in Section 3.1. Finally, we aim to compare CLIP to a comprehensive set of existing models across many tasks. Studying 66 different models on 27 different datasets requires tuning 1782 different evaluations. Fine-tuning opens up a much larger design and hyperparameter space, which makes it difficult to fairly evaluate and computationally expensive to compare a diverse set of techniques as discussed in other large scale empirical studies. By comparison, linear classifiers require minimal hyper-parameter tuning and have standardized implementations and evaluation

procedures. Please see Appendix A for further details on evaluation. Figure 10 summarizes our findings. To minimize selection effects that could raise concerns of confirmation or reporting bias, we first study performance on the 12-dataset evaluation suite from Kornblith et al. (2019). While small CLIP models such as a ResNet-50 and ResNet-101 outperform other ResNets trained on ImageNet-1K (BiT-S and the originals), they underperform ResNets trained on ImageNet-21K (BiTM). These small CLIP models also underperform models in the EfficientNet family with similar compute requirements. However, models trained with CLIP scale very well and the largest model we trained (ResNet-50x64) slightly outperforms the best performing existing model (a Noisy Student EfficientNet-L2) on both overall score and compute efficiency. We also find that CLIP vision transformers are about 3x more compute efficient than CLIP ResNets, which allows us to reach higher overall performance within our compute budget. These results qualitatively replicate the findings of Dosovitskiy et al. (2020) which reported that vision transformers are more compute efficient than convnets when trained on sufficiently large datasets. Our best overall model is a ViT-L/14 that is fine-tuned at a higher resolution of 336 pixels on our dataset for 1 additional epoch. This model outperforms the best existing model across this evaluation suite by an average of 2.6%. As Figure 21 qualitatively shows, CLIP models learn a wider set of tasks than has previously been demonstrated in a single computer vision model trained end-to-end from random initialization. These tasks include geo-localization, optical character recognition, facial emotion recognition, and action recognition. None of these tasks are measured in the evaluation suite of Kornblith et al. (2019). This could be argued to be a form of selection bias in Kornblith et al. (2019)’s study towards tasks that overlap with ImageNet. To address this, we also measure performance on a broader 27 dataset evaluation suite. This evaluation suite, detailed in Appendix A includes datasets representing the aforementioned tasks, German Traffic Signs Recognition Benchmark, as well as several other datasets adapted from VTAB.

CHAPTER 4

SOFTWARE DEVELOPMENT METHODOLOGY

The creation of computer software is a unique activity. Unlike the design of other products, such as building a house or writing a book, small mistakes may only have a minor impact. However, computer software is very integrated so small mistakes can have a cascading impact on many functions in the software package. Also, software source code is so readily modified that there can be a lot of pressure to get major features working first, then go back to the same sections of code to fix the behavior for unusual circumstances. Because of these factors, the testing and debugging cycle is key to getting reliable, robust, bug free software. Software testing is generally broken into two broad categories. Unit tests are tests of an individual section of code, as independently from the rest of the software as possible. These sections of code may be termed methods, functions, subroutines, objects, or procedures depending upon the programming language. Functional tests are tests of the entire piece of software. These tests may be designed to test various aspects of the software including; correct behavior, security, performance, robustness, handling heavy loads, standards compliance, and usability. Occasionally, the same set of tests may be done multiple times to check for compatibility with various operating systems, web browsers, or other interdependent components. Many of the currently popular software development methodologies focus on test driven development (TDD). This is a process in which unit tests are designed before or while creating the software to independently ensure that each section of code is working correctly. Key tests may even be specified in advance in the design specification documents. This cuts down on the amount of time that will be spent on testing and debugging later. This is key to being able to frequently release new versions of the software with some level of confidence in how well they will work.

Developing high-quality and reliable software is a challenging task that demands a comprehensive and structured approach. Here comes to help the concept of the software development life cycle (SDLC).

Software Development Strategies

Here we will describe from a rather high altitude the various development methods and processes employed for software today. We focus on designing, creating, and maintaining large-scale enterprise application software, whether developed by vendors or in-house development teams. The creation and use of one-off and simple interface programs is no challenge. Developing huge operating systems such as Microsoft XP with millions of lines of code (LOC), or large, complex systems such as the FAAs Enroute System, bring very special problems of their own and are beyond the scope of this book. This is not to say that the methodology we propose for robust software architecture is not applicable; rather, we will not consider their applications here. The time-honored enterprise software development process generally follows these steps :

Specification or functional design, done by system analysts in consort with the potential end users of the software to determine why to do this, what the application will do, and for whom it will do it.

Architecture or technical design, done by system designers as the way to achieve the goals of the functional design using the computer systems available, or to be acquired, in the context of the enterprise as it now operates. This is how the system will function.

Programming or implementation, done by computer programmers together with the system designers. Testing of new systems (or regression testing of modified systems) to ensure that the goals of the functional design and technical design are met.

Documentation of the system, both intrinsically for its future maintainers, and extrinsically for its future users. For large systems this step may involve end-user training as well.

Maintenance of the application system over its typical five-year life cycle, employing the design document now recrafted as the Technical Specification or System Maintenance Document.

Any organization that will develop software is faced with a difficult choice of choosing the right software development method. Whereas the software development methods used, play a significant role in the overall software development process. Software development methods are needed so that the software development process can be systematic so that it is not only completed within the right time frame but also must have good quality. There are various methods of software development in System Development Lyfe Cycle (SDLC).

Each SDLC method provides a general guiding line about different software development and has different characteristics. Each method of software development has its drawbacks and advantages so that the selection of software development methods should be compatible with the capacity of the software developed. This paper will compare three different software development methods: V-Shaped Model, Parallel Development Model, and Iterative Model with the aim of providing an understanding of software developers to choose the right method.

Software development methods are needed so that the software development process can be systematic so that it is not only completed within the right time frame but also must have good quality. Sommerville [10] argues that the software development method is a representation of the software process. Each development method represents a software development process from a particular point of view. The software development process model differs from the software development methodology, a method controls through a stage (data specification, allocation of needs, control, etc.), while the model has an interest in providing guidance in a certain order (several steps) in carrying out the tasks that need to be done to complete a project software.

A major goal of modern software development is to follow organized and disciplined practices throughout the system specification and construction process, rather than to take an ad hoc approach to programming. The key idea is to develop and use a software development methodology, a process for the creation of software. A methodology combines technical methods with management procedures for software development in a synergistic fashion, and automated tools in a development support system provide additional assistance. The underlying philosophy is that use of a methodology can improve many aspects of the entire software development process, including a better fit to user requirements, fewer errors in the resulting system, better documentation throughout the entire process, and significantly reduced costs for system evolution.

The Stages of Software Development Lifecycle (SDLC)

Stage 1: Planning and Requirement Analysis: Requirement analysis is the most important and fundamental stage in SDLC. It is performed by the senior members of the team with inputs from the customer, the sales department, market surveys and domain experts in the industry. This information is then used to plan the basic project approach and to conduct product feasibility study in the economical, operational, and technical areas. Planning for the quality assurance requirements and identification of the risks associated with the project is also done in the planning stage. The outcome of the technical feasibility study is to define the various technical approaches that can be followed to implement the project successfully with minimum risks.

Stage 2: Defining Requirements: Once the requirement analysis is done the next step is to clearly define and document the product requirements and get them approved from the customer or the market analysts. This is done through SRS (Software Requirement Specification) document which consists of all the product requirements to be designed and developed during the project life cycle.

Stage 3: Designing the product architecture: SRS is the reference for product architects to come out with the best architecture for the product to be developed. Based on the requirements specified in SRS, usually more than one design approach for the product architecture is proposed and documented in a DDS - Design Document Specification. This DDS is reviewed by all the important stakeholders and based on various parameters as risk assessment, product robustness, design modularity, budget and time constraints, the best design approach is selected for the product. A design approach clearly defines all the architectural modules of the product along with its communication and data flow representation with the external and third party modules (if any). The internal design of all the modules of the proposed architecture should be clearly defined with the minutest of the details in DDS.

Stage 4: Building or Developing the Product: In this stage of SDLC the actual development starts and the product is built. The programming code is generated as per DDS during this stage. If the design is performed in a detailed and organized manner, code generation can be accomplished without much hassle. Developers have to follow the coding guidelines defined by their organization and programming tools like compilers,

interpreters, debuggers etc are used to generate the code. Different high level programming languages such as C, C++, Pascal, Java, and PHP are used for coding. The programming language is chosen with respect to the type of software being developed.

Stage 5: Testing the Product: This stage is usually a subset of all the stages as in the modern SDLC models, the testing activities are mostly involved in all the stages of SDLC. However, this stage refers to the testing only stage of the product where products defects are reported, tracked, fixed and retested, until the product reaches the quality standards defined in the SRS.

Stage 6: Deployment in the Market and Maintenance: Once the product is tested and ready to be deployed it is released formally in the appropriate market. Sometime product deployment happens in stages as per the organizations. business strategy. The product may first be released in a limited segment and tested in the real business environment (UAT- User acceptance testing).

ARCHITECTURAL DESIGN

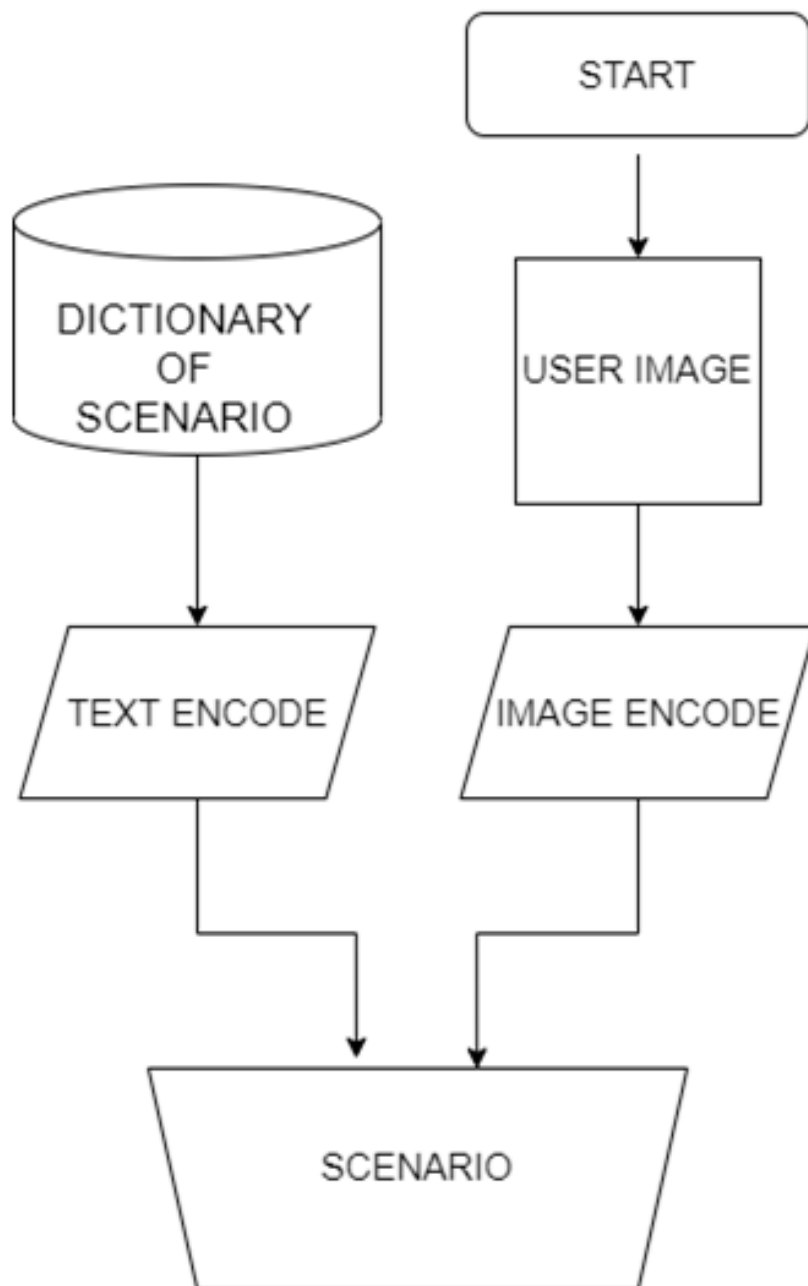


Fig 4.1: working diagram

CHAPTER 5

RESULTS AND DISCUSSION

As a result of this system getting introduced in our defence system, it will act like a bridge to cover the gap of communication between the police and the victim. The police will be able to understand the victim better which will further result in taking quick actions. The victim will not get harassed and will be delivered with justice in time. It will make the world a better place to live in peace and togetherness.

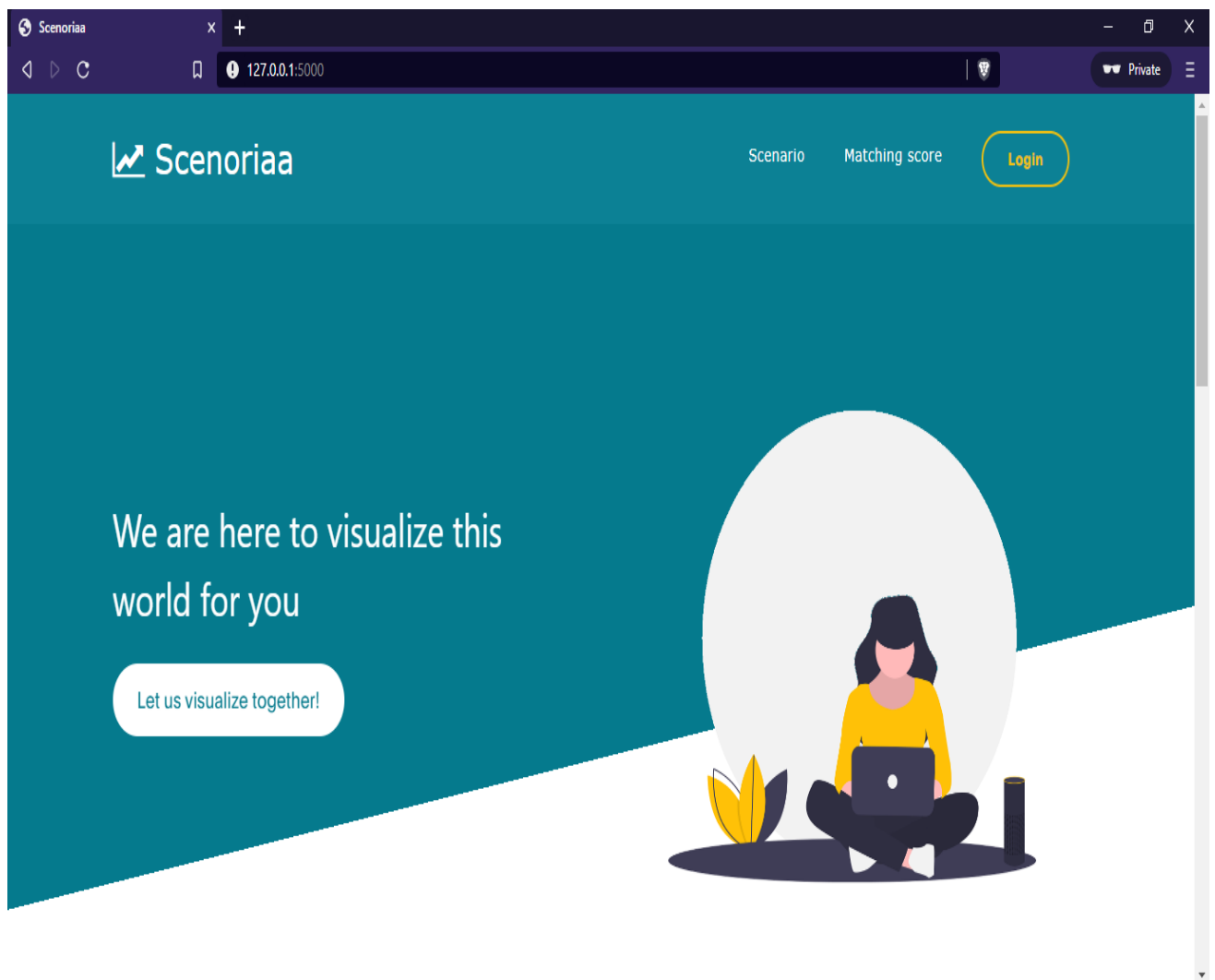


Fig:5.1 Home Page

The Figure 5.1 is the home page of the project. This is the basic outlook of our website. A home page is generally the primary web page a visitor navigating to a website from a search engine will see, and it may also serve as a landing page to attract visitors. The home page is used to facilitate navigation to other pages on the site by providing links to prioritized and recent articles and pages, and possibly a search box. For example, a news website may present headlines and first paragraphs of top stories, with links to full articles. Meanwhile, other websites use the home page to attract users to create an account. Once they are logged in, the home page may be redirected to their profile page. This may in turn be referred to as the "personal home page".

A website may have multiple home pages. Wikipedia, for example, has a home page at wikipedia.org, as well as language-specific home pages, such as en.wikipedia.org and de.wikipedia.org. The Uniform Resource Locator (URL) of a home page is most often the base-level domain name.

In the past, common home page filenames were `index.html` or `default.html`.⁶ Another convention was that if a home page was not created for a website, web servers would default to display a list of files located in the website's root directory. These conventions are largely obsolete on the modern Web. [citation needed]

When a web browser is launched, it may automatically open at least one web page, sometimes referred to as a "start page". Start pages can be a website or a special browser page, such as thumbnails of frequently visited websites. Moreover, there is a niche market of websites intended to be used solely as start pages.

The figure 5.1 is the basic outlook of the website. There user can upload any image and get the exact scenario of that image and also user can compare both the image and the scenario.

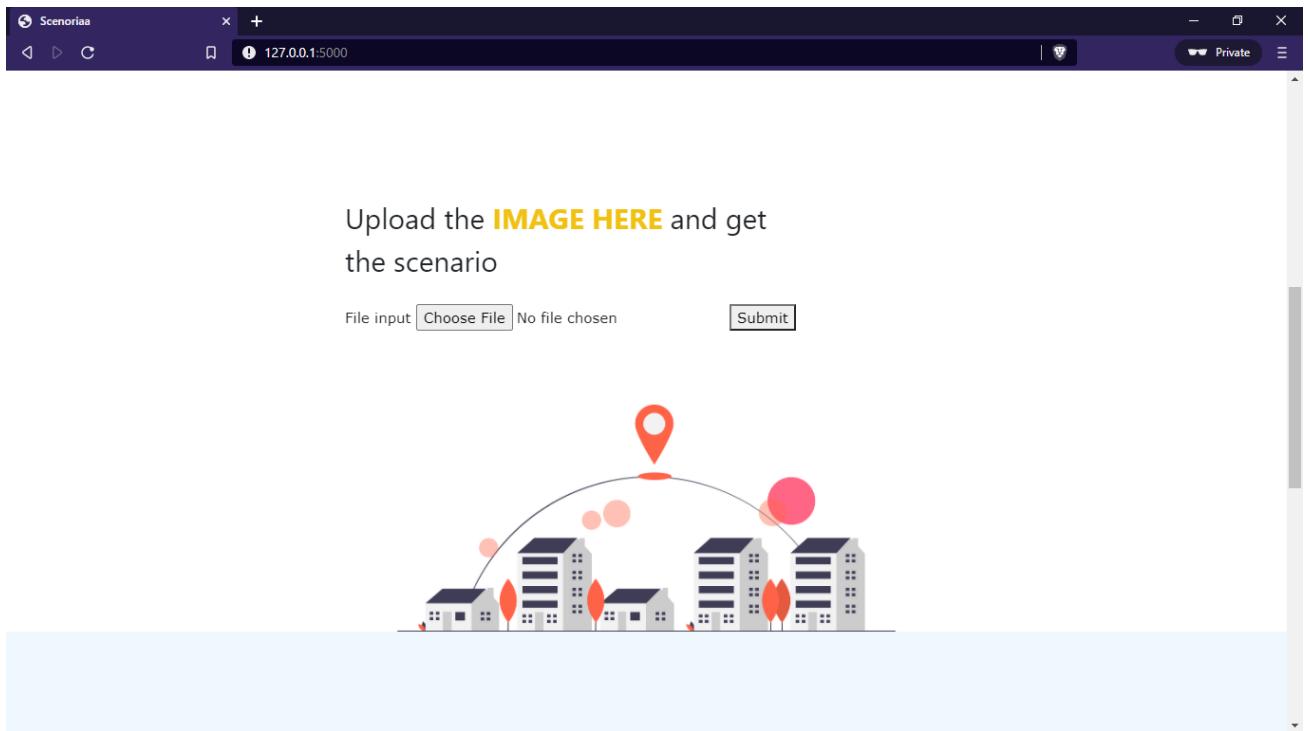


Fig 5.2: uploading image

The Fig 5.2 is the uploading image phase of the project. This page will appear once the user scroll down from the homepage. Here the user gets the option to upload image from the choose file or execute their task as per need. This page was made to be dynamic using bootstrap. In this system the user interface build is very attractive and professional. It is the frontend of the project or it can also be termed as user interface. Here the user gets the multiple options to execute their task as per need. The system has access to the images stored on the system. It is the frontend of the project or it can also be termed as user interface. Here the user has to select the image which he wants to check the scenario from the choose file option. After selected he needs to click on upload image to upload the image. After the image got successfully uploaded user can get the exact scenario of the image, he uploaded. That's how this entire process going to work.

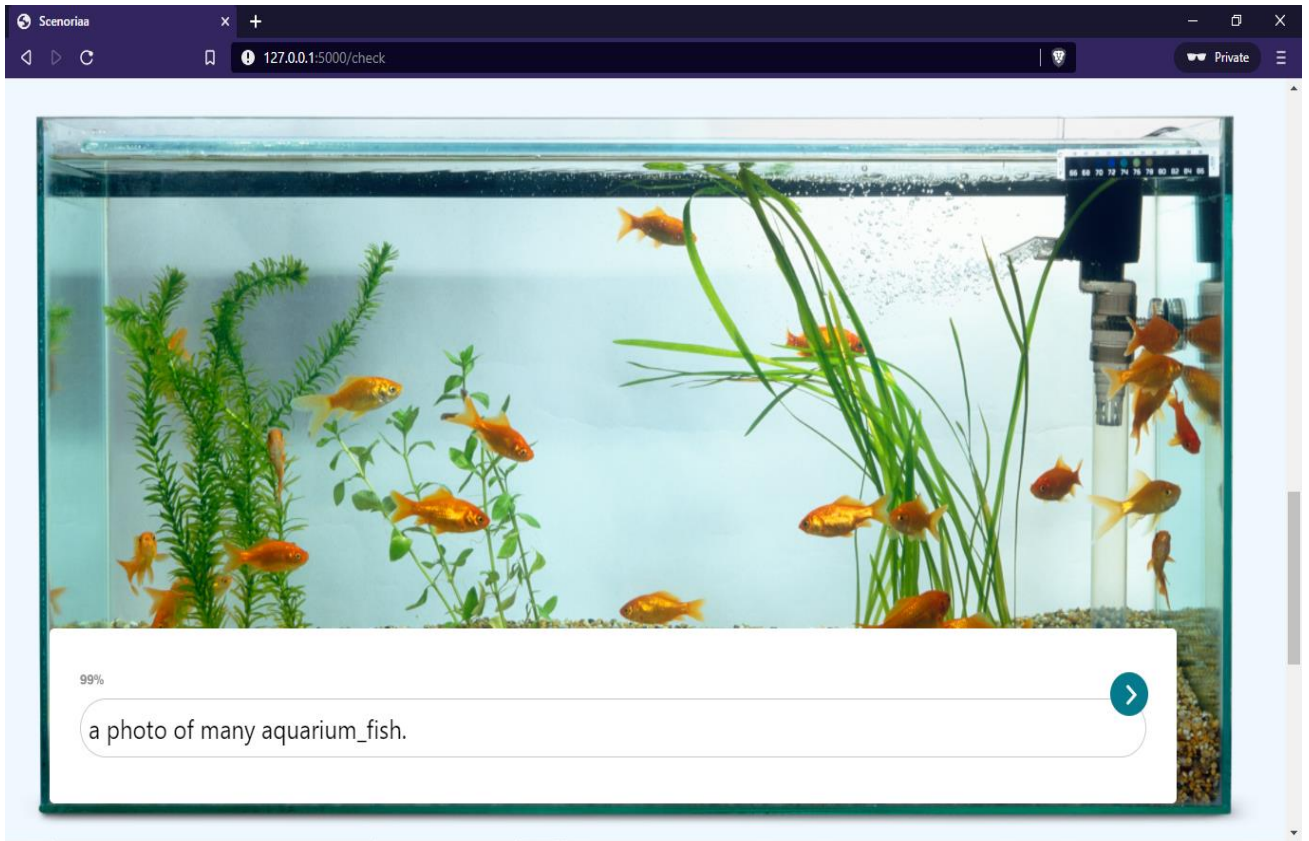


Fig 5.3: Result 1

The Fig 5.3 is the result page which generally is the output of the website. The user uploaded a image of an aquarium. It gives the exact scenario as a photo of many aquarium fish, and also it gives how much exactly both the images and scenario matched.

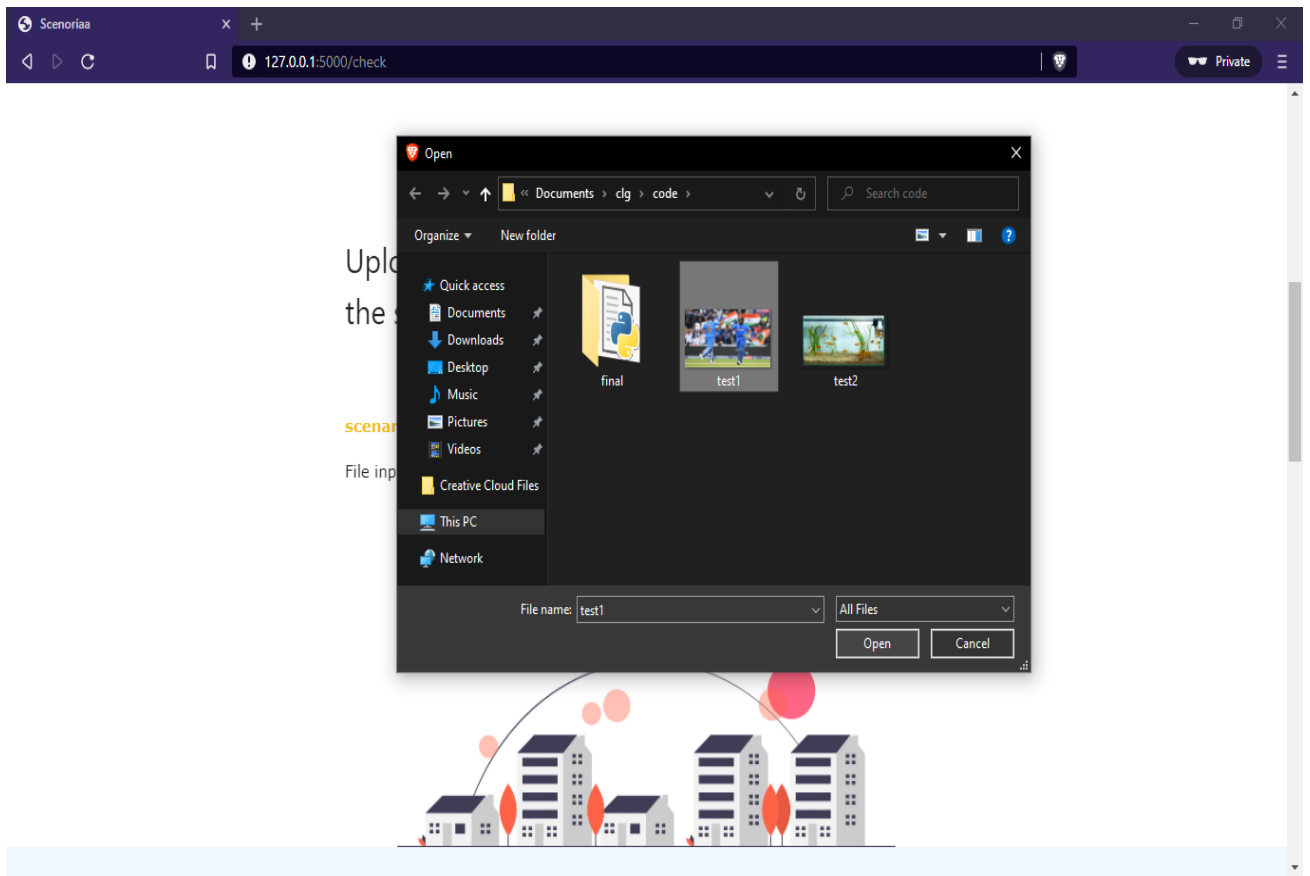


Fig 5.4: selecting image

The Fig 5.4 shows that the user has to select the image from his system. The user can select only one image at a time. After selecting the image, user needs to upload the image to get the scenario. The fig 5.4 shows how to upload the image on the website.

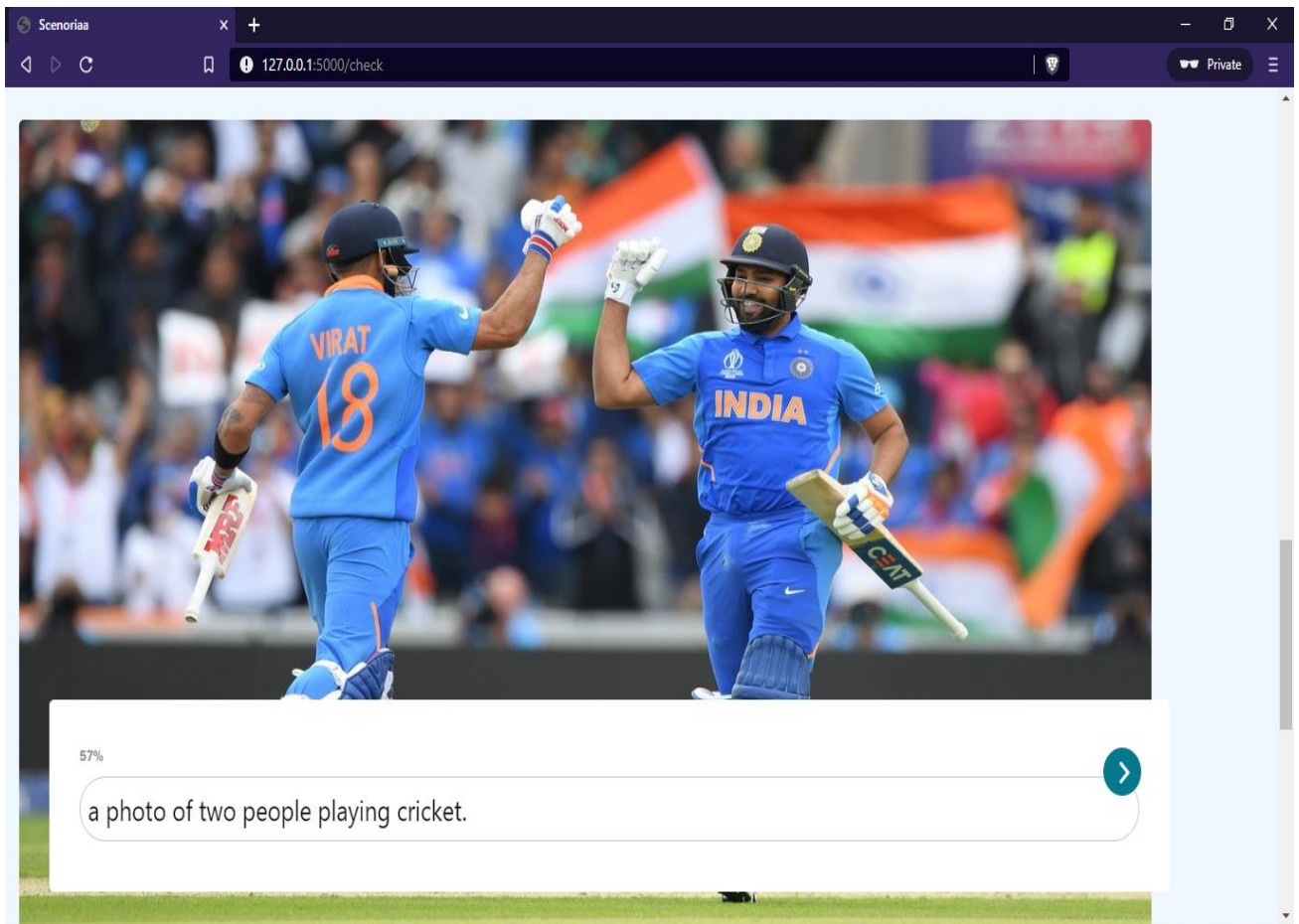


Fig 5.5: Result 2

The fig 5.5 is another example of how this website works. Here the user searching a photo of two people playing cricket, the website shows the result. But it did not give exact result like the player's name and everything that's why it shows matching percentage as 53.

CHAPTER 6

CONCLUSION AND FUTURE WORK

CONCLUSION

The model has its own dictionary consists of classes and scenarios. As it has its own dictionary, whenever user gives a image the model will compare that photo with the photos which is stored on the class on the dictionary and then gives the desired result. For instance, if the user wants to find how much amount does the picture and scenario matches, he can get that also. This will lead to the improvement in the quality of images and that's how this going to work.

FUTURE WORK

There are many reverse image search engines are there but this model will definiely help in finding the absolute scenario of the image. It will also come handy for visually impaired people. Even though they have many ways to learn, this will really help them in building their confidence. As this also gives the correctness it will be useful to verify whether this is a correct scenario or not. This will also help to improve the quality of the image and will be definitely helpful for the people who loves to surf images .

REFERENCES

- [1] AlecRadford, JongWookKim, ChrisHallacyAdityaRamesh,GabrielGoh,SandhiniAgarwal, Girish Sastry, Amanda Aspell, Pamela Mishkin,Jack Clark, Gretchen Krueger,IlyaSutskeverLearning Transferable Visual Models From Natural Language Supervision] – (2021)
- [2] Socher, R., Karpathy, A., Le, Q. V., Manning, C. D., and Ng,A. Y. Grounded compositional semantics for finding anddescribing images with sentences. Transactions of theAssociation for Computational Linguistics, 2:207–218, (2014).
- [3] Scheuerman, M. K., Paul, J. M., and Brubaker, J. R. Howcomputers see gender: An evaluation of gender classification in commercial facial analysis services. Proceedingsof the ACM on Human-Computer Interaction, 3(CSCW):(1–33, 2019.)
- [4] Socher, R. and Fei-Fei, L. Connecting modalities: Semisupervised segmentation and annotation of images usingunaligned text corpora. In 2010 IEEE Computer SocietyConference on Computer Vision and Pattern Recognition,pp. 966–973. (IEEE, 2010.)
- [5] D. Ping Tian, “A review on image feature extraction and representation techniques,” International Journal of Multimedia and Ubiquitous Engineering, vol. 8, no. 4, pp. 385–396, 2013. [15] D. Zhang and G. Lu, “Review of shape representation and description techniques,” Pattern Recognition, vol. 37, no. 1, pp. 1–19, 2004.
- [6] R. Datta, J. Li, and J. Z. Wang, “Content-based image retrieval: approaches and trends of the new age,” in Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval, pp. 253–262, ACM, Singapore, November 2005.
- [7] Z. Yu and W. Wang, “Learning DALTS for cross-modal retrieval,” CAAI Transactions on Intelligence Technology, vol. 4, no. 1, pp. 9–16, 2019.
- [8] N. Ali, D. A. Mazhar, Z. Iqbal, R. Ashraf, J. Ahmed, and F. Zeeshan, “Content-based image retrieval based on late fusion of binary and local descriptors,” International Journal of

Computer Science and Information Security (IJCSIS), vol. 14, no. 11, 2016.

[9] N. Ali, *Image Retrieval Using Visual Image Features and Automatic Image Annotation*, University of Engineering and Technology, Taxila, Pakistan, 2016.

[10] B. Zafar, R. Ashraf, N. Ali et al., “Intelligent image classification-based on spatial weighted histograms of concentric circles,” *Computer Science and Information Systems*, vol. 15, no. 3, pp. 615–633, 2018.

[11] G. Qi, H. Wang, M. Haner, C. Weng, S. Chen, and Z. Zhu, “Convolutional neural network based detection and judgement of environmental obstacle in vehicle operation,” *CAAI Transactions on Intelligence Technology*, vol. 4, no. 2, pp. 80–91, 2019.

[12] U. Markowska-Kaczmar and H. Kwaśnicka, “Deep learning—a new era in bridging the semantic gap,” in *Bridging the Semantic Gap in Image and Video Analysis*, pp. 123–159, Springer, Basel, Switzerland, 2018.

[13] F. Riaz, S. Jabbar, M. Sajid, M. Ahmad, K. Naseer, and N. Ali, “A collision avoidance scheme for autonomous vehicles inspired by human social norms,” *Computers & Electrical Engineering*, vol. 69, pp. 690–704, 2018.

[14] H. Shao, Y. Wu, W. Cui, and J. Zhang, “Image retrieval based on MPEG-7 dominant color descriptor,” in *Proceedings of the 9th International Conference for Young Computer Scientists ICYCS 2008*, pp. 753–757, IEEE, Hunan, China, November 2008.

APPENDIX

A. Source Code

MAIN.PY

```
import os, clip, torch, json
from PIL import Image
from torchvision.datasets import CIFAR100
from flask import Flask, render_template, request

app = Flask(__name__)
device = "cuda"# if torch.cuda.is_available() else "cpu"
# Load modelcls
model, preprocess = clip.load('ViT-B/32', device)
print("{} mode loaded done".format('ViT-B/32'))
path = ''

@app.route('/')
def home():
    return render_template("index.html")

@app.route('/blog')
def blog():
    return render_template("blog.html")

@app.route('/check', methods=['POST'])
def check():
    file = request.files['inputFile']
    global path
    path = "static/temp/"+file.filename
    image_data=file.save(path)
    # Setup
    torch.cuda.empty_cache()
    print("{} is setup done".format(device))

    # Load dictionary
    text = list()
    with open('model\dictionary.txt', 'r') as f:
        text = json.loads(f.read())

    print("dictionary of {} loaded done".format(len(text)))

    #text to weight
    text_inputs = torch.cat([clip.tokenize(c) for c in text]).to(device)
    text_features = model.encode_text(text_inputs)
```

```

print("dictionary to torch done")
image_input = preprocess(Image.open(path)).unsqueeze(0).to(device)
# Calculate features
with torch.no_grad():
    image_features = model.encode_image(image_input)

# Pick the top 5 most similar labels for the image
image_features /= image_features.norm(dim=-1, keepdim=True)
text_features /= text_features.norm(dim=-1, keepdim=True)
similarity = (100.0 * image_features @ text_features.T).softmax(dim=-1)
values, indices = similarity[0].topk(1)
sen = text[indices[0]]
return render_template('index.html', image=path, sen=sen, score=int(values[0].tolist()*100))

@app.route('/predict', methods=["GET", "POST"])
def predict():
    global path
    torch.cuda.empty_cache()
    print(request.form['sen'])
    text_inputs = torch.cat([clip.tokenize(request.form.get('sen'))]).to(device)

    text_features = model.encode_text(text_inputs)
    print("dictionary to torch done")
    print(path)
    image_input = preprocess(Image.open(path)).unsqueeze(0).to(device)
    # Calculate features
    with torch.no_grad():
        image_features = model.encode_image(image_input)

    # Pick the top 5 most similar labels for the image
    image_features /= image_features.norm(dim=-1, keepdim=True)
    text_features /= text_features.norm(dim=-1, keepdim=True)
    similarity = (100.0 * image_features @ text_features.T).softmax(dim=-1)
    values, indices = similarity[0].topk(1)
    return render_template('index.html', image=path, sen=request.form['sen'], score=int(values[0].tolist()*100))

app.run(debug=True)

```

Index.html

```

<!DOCTYPE html>
<html lang="en">
<head>

    <title>Scenoriaa</title>
<!--

DIGITAL TREND

https://templatemo.com/tm-538-digital-trend

```

```

-->
<meta charset="UTF-8">
<meta http-equiv="X-UA-Compatible" content="IE=Edge">
<meta name="description" content="">
<meta name="keywords" content="">
<meta name="author" content="">
<meta name="viewport" content="width=device-width, initial-
scale=1, maximum-scale=1">

<link rel="stylesheet" href="static/css/bootstrap.min.css">
<link rel="stylesheet" href="static/css/font-awesome.min.css">
<link rel="stylesheet" href="static/css/aos.css">
<link rel="stylesheet" href="static/css/owl.carousel.min.css">
<link rel="stylesheet" href="static/css/owl.theme.default.min.css">

<!-- MAIN CSS -->
<link rel="stylesheet" href="static/css/templatemo-digital-trend.css">
<link rel="stylesheet" href="https://www.w3schools.com/w3css/4/w3.css">

</head>
<body>

<!-- MENU BAR -->
<nav class="navbar navbar-expand-lg">
  <div class="container">
    <a class="navbar-brand" href="index.html">
      <i class="fa fa-line-chart"></i>
      Scenoriaa
    </a>

    <button class="navbar-toggler" type="button" data-
toggle="collapse" data-target="#navbarNav" aria-controls="navbarNav" aria-
expanded="false"
      aria-label="Toggle navigation">
      <span class="navbar-toggler-icon"></span>
    </button>

    <div class="collapse navbar-collapse" id="navbarNav">
      <ul class="navbar-nav ml-auto">
        <li class="nav-item">
          <a href="#about" class="nav-
link smoothScroll">Scenario</a>
        </li>
        <li class="nav-item">
          <a href="#project" class="nav-
link smoothScroll">Matching score</a>
        </li>
        <!-- <li class="nav-item"> -->
          <!-- <a href="/blog" class="nav-link">Blog</a> -->
        <!-- </li> -->
        <li class="nav-item">

```

```

        <a href="login.html" class="nav-link contact">Login</a>
    </li>
</ul>
</div>
</div>
</nav>

<!-- HERO -->
<section class="hero hero-bg d-flex justify-content-center align-items-center">
    <div class="container">
        <div class="row">

            <div class="col-lg-6 col-md-10 col-12 d-flex flex-column justify-content-center align-items-center">
                <div class="hero-text">

                    <h1 class="text-white" data-aos="fade-up">We are here to visualize this world for you</h1>

                    <a href="contact.html" class="custom-btn btn-bg btn mt-3" data-aos="fade-up" data-aos-delay="100">Let us visualize together!</a>

                    <!-- <strong class="d-block py-3 pl-5 text-white" data-aos="fade-up" data-aos-delay="200"><i class="fa fa-phone mr-2"></i> +99 080 070 4224</strong> -->
                </div>
            </div>

            <div class="col-lg-6 col-12">
                <div class="hero-image" data-aos="fade-up" data-aos-delay="300">

                    
                </div>
            </div>
        </div>
    </div>
</section>

<!-- ABOUT -->
<section class="about section-padding pb-0" id="about">
    <div class="container">
        <div class="row">

            <div class="col-lg-7 mx-auto col-md-10 col-12">
                <div class="about-info">

```

```

        <h2 class="mb-4" data-aos="fade-
up">Upload the <strong>IMAGE HERE</strong> and get the scenario</h2>
        {% if sen %}
            <br><br><strong>scenario </strong>{{sen}}</p
>

        {% endif %}
    </div>

    <div class="about-image" data-aos="fade-up" data-aos-
delay="200">

        <form method="POST" action="/check" enctype="multipart
/form-data">

            <label for="inputFile">File input</label>
            <input type="file" name="inputFile">
            <button type="submit">Submit</button>
        </form>
        
    </div>
</div>

</div>
</section>

<!-- PROJECT -->
<section class="project section-padding" id="project">
    <div class="container-fluid">
        <div class="row">

            <div class="col-lg-12 col-12">

                <h2 class="mb-5 text-center" data-aos="fade-up">
                    Please take a look through our
                    <strong>featured Digital Trends</strong>
                </h2>

                <!-- <div class="owl-carousel owl-theme" id="project-
slide"> -->

                <div class="item project-wrapper" data-aos="fade-
up" data-aos-delay="100">

                    {% if image %}
                        

                        <div class="project-info">
                            <small>{{score}}%</small>

                            <h3>

```

```

<form action="/predict" method="POST">
    <span><input class="w3-
input w3-border w3-round-
xxlarge" type="text" id="myfile" name='sen' value="{{sen}}"></span>
    <i class="fa fa-angle-
right project-icon"></i>
    </form>
    </h3>
</div>
{% endif %}
</div>

<!-- </div> -->
</div>

</div>
</div>
</section>

<footer class="site-footer">
    <div class="container">
        <div class="row">

            <div class="col-lg-5 mx-lg-auto col-md-8 col-10">
                <h1 class="text-white" data-aos="fade-up" data-aos-
delay="100">We make creative <strong>brands</strong> only.</h1>
            </div>

            <div class="col-lg-3 col-md-6 col-12" data-aos="fade-up" data-aos-
delay="200">
                <h4 class="my-4">Contact Info</h4>

                <p class="mb-1">
                    <i class="fa fa-phone mr-2 footer-icon"></i>
                    +99 080 070 4224
                </p>

                <p>
                    <a href="#">
                        <i class="fa fa-envelope mr-2 footer-icon"></i>
                        hello@company.com
                    </a>
                </p>
            </div>

            <div class="col-lg-3 col-md-6 col-12" data-aos="fade-up" data-aos-
delay="300">
                <h4 class="my-4">Our Studio</h4>

```

```

        <p class="mb-1">
            <i class="fa fa-home mr-2 footer-icon"></i>
            Av. Lúcio Costa - Barra da Tijuca, Rio de Janeiro - RJ, Brazil
        </p>
    </div>

    <div class="col-lg-4 mx-lg-auto text-center col-md-8 col-12" data-
aos="fade-up" data-aos-delay="400">
        <p class="copyright-text">Copyright &copy; 2020 Your Company
        <br>
        <a rel="nofollow noopener" href="https://templatemo.com">Design: Te
mplateMo</a></p>
    </div>

    <div class="col-lg-4 mx-lg-auto col-md-6 col-12" data-aos="fade-
up" data-aos-delay="500">

        <ul class="footer-link">
            <li><a href="#">Stories</a></li>
            <li><a href="#">Work with us</a></li>
            <li><a href="#">Privacy</a></li>
        </ul>
    </div>

    <div class="col-lg-3 mx-lg-auto col-md-6 col-12" data-aos="fade-
up" data-aos-delay="600">
        <ul class="social-icon">
            <li><a href="#" class="fa fa-instagram"></a></li>
            <li><a href="#" class="fa fa-twitter"></a></li>
            <li><a href="#" class="fa fa-dribbble"></a></li>
            <li><a href="#" class="fa fa-behance"></a></li>
        </ul>
    </div>

</div>
</div>
</footer>

<!-- SCRIPTS -->
<script src="static/js/jquery.min.js"></script>
<script src="static/js/bootstrap.min.js"></script>
<script src="static/js/aos.js"></script>
<script src="static/js/owl.carousel.min.js"></script>
<script src="static/js/smoothscroll.js"></script>
<script src="static/js/custom.js"></script>

</body>
</html>

```

```

<!DOCTYPE html>
<html lang="en">
<head>

    <title>Digital Trend - Blog</title>
<!--

DIGITAL TREND

https://templatemo.com/tm-538-digital-trend

-->
    <meta charset="UTF-8">
    <meta http-equiv="X-UA-Compatible" content="IE=Edge">
    <meta name="description" content="">
    <meta name="keywords" content="">
    <meta name="author" content="">
    <meta name="viewport" content="width=device-width, initial-
scale=1, maximum-scale=1">

    <link rel="stylesheet" href="static/css/bootstrap.min.css">
    <link rel="stylesheet" href="static/css/font-awesome.min.css">
    <link rel="stylesheet" href="static/css/aos.css">
    <link rel="stylesheet" href="static/css/owl.carousel.min.css">
    <link rel="stylesheet" href="static/css/owl.theme.default.min.css">

    <!-- MAIN CSS -->
    <link rel="stylesheet" href="static/css/templatemo-digital-trend.css">

</head>
<body>

    <!-- MENU BAR -->
    <nav class="navbar navbar-expand-lg position-absolute">
        <div class="container">
            <a class="navbar-brand" href="index.html">
                <i class="fa fa-line-chart"></i>
                Digital Trend
            </a>

            <button class="navbar-toggler" type="button" data-
toggle="collapse" data-target="#navbarNav" aria-controls="navbarNav" aria-
expanded="false"
                aria-label="Toggle navigation">
                <span class="navbar-toggler-icon"></span>
            </button>

            <div class="collapse navbar-collapse" id="navbarNav">
                <ul class="navbar-nav ml-auto">
                    <li class="nav-item">
                        <a href="index.html#about" class="nav-link">Studio</a>

```

```

        </li>
        <li class="nav-item">
            <a href="index.html#project" class="nav-
link">Our Works</a>
        </li>
        <li class="nav-item">
            <a href="blog.html" class="nav-link active">Blog</a>
        </li>
        <li class="nav-item">
            <a href="contact.html" class="nav-
link contact">Contact</a>
        </li>
    </ul>
</div>
</div>
</nav>

<!-- BLOG -->
<section class="blog section-padding">
    <div class="container">
        <div class="row">

            <div class="col-lg-12 col-12 py-5 mt-5 mb-3 text-center">

                <h1 class="mb-4" data-aos="fade-
up">Digital Trend Blog</h1>
            </div>

            <div class="col-lg-7 col-md-7 col-12 mb-4">
                <div class="blog-header" data-aos="fade-up" data-aos-
delay="100">
                    

                    <div class="blog-header-info">
                        <h4 class="blog-category text-info">Creative</h4>

                        <h3><a href="blog-
detail.html">The Key to Creative Work is Knowing When to Walk Away</a></h3>
                    </div>
                </div>

                <div class="col-lg-5 col-md-5 col-12 mb-4">
                    <div class="blog-sidebar d-flex justify-content-
center align-items-center" data-aos="fade-up" data-aos-delay="200">
                        

                        <div class="blog-info">
                            <h4 class="blog-category text-danger">Design</h4>

```

```

        <h3><a href="blog-
detail.html">Why Truly Accessible Design Benefits Everyone</a></h3>
    </div>
</div>

    <div class="blog-sidebar py-4 d-flex justify-content-
center align-items-center" data-aos="fade-up" data-aos-delay="300">
        

        <div class="blog-info">
            <h4 class="blog-category text-success">lifestyle</h4>

            <h3><a href="blog-
detail.html">Be Humble About What You Know</a></h3>
        </div>
    </div>

    <div class="blog-sidebar d-flex justify-content-
center align-items-center" data-aos="fade-up" data-aos-delay="200">
        

        <div class="blog-info">
            <h4 class="blog-category text-primary">Coding</h4>

            <h3><a href="blog-
detail.html">The Mistakes I Made As a Coding Beginner</a></h3>
        </div>
    </div>

</div>

<div class="col-lg-5 ml-auto mt-5 pt-5 col-md-6 col-12">

    
</div>

<div class="col-lg-5 mr-auto mt-5 pt-5 col-md-6 col-
12 newsletter-form">
    <h4 data-aos="fade-up" data-aos-
delay="200">Email Newsletter</h4>

    <h2 data-aos="fade-up" data-aos-
delay="300">Let's stay up-to-date. We'll share you all good stuffs.</h2>
    <form action="#" method="get" enctype="multipart/form-
data">

        <div class="form-group mt-4" data-aos="fade-up" data-aos-
delay="400">

```

```

        <input name="email" type="email" class="form-control"
            id="email" aria-
describedby="emailHelp" placeholder="Please enter your email" required>

        <small id="emailHelp" class="form-text text-
muted">We'll NOT share your email address to anyone else.</small>

    </div>

    <div class="form-group form-check" data-aos="fade-
up" data-aos-delay="500">
        <input name="monthly" type="checkbox" class="form-
check-input" id="monthly">

        <label class="form-check-
label" for="monthly">Please send me a monthly newsletter.</label>
    </div>

    <button type="submit" data-aos="fade-up" data-aos-
delay="500" class="btn w-100 mt-3">Sign up</button>
    </form>
</div>

</div>
</div>
</section>

<footer class="site-footer">
    <div class="container">
        <div class="row">

            <div class="col-lg-5 mx-lg-auto col-md-8 col-10">
                <h1 class="text-white" data-aos="fade-up" data-aos-
delay="100">We make creative <strong>brands</strong> only.</h1>
            </div>

            <div class="col-lg-3 col-md-6 col-12" data-aos="fade-up" data-aos-
delay="200">
                <h4 class="my-4">Contact Info</h4>

                <p class="mb-1">
                    <i class="fa fa-phone mr-2 footer-icon"></i>
                    +99 080 070 4224
                </p>

                <p>
                    <a href="#">
                        <i class="fa fa-envelope mr-2 footer-icon"></i>
                        hello@company.com
                    </a>
                </p>

```

```

    </div>

    <div class="col-lg-3 col-md-6 col-12" data-aos="fade-up" data-aos-
delay="300">
        <h4 class="my-4">Our Studio</h4>

        <p class="mb-1">
            <i class="fa fa-home mr-2 footer-icon"></i>
            Av. Lúcio Costa - Barra da Tijuca, Rio de Janeiro - RJ, Brazil
        </p>
    </div>

    <div class="col-lg-4 mx-lg-auto text-center col-md-8 col-12" data-
aos="fade-up" data-aos-delay="400">
        <p class="copyright-text">Copyright &copy; 2020 Your Company
        <br>
        <a rel="nofollow noopener" href="https://templatemo.com">Design: Te
mplateMo</a></p>
    </div>

    <div class="col-lg-4 mx-lg-auto col-md-6 col-12" data-aos="fade-
up" data-aos-delay="500">

        <ul class="footer-link">
            <li><a href="#">Stories</a></li>
            <li><a href="#">Work with us</a></li>
            <li><a href="#">Privacy</a></li>
        </ul>
    </div>

    <div class="col-lg-3 mx-lg-auto col-md-6 col-12" data-aos="fade-
up" data-aos-delay="600">
        <ul class="social-icon">
            <li><a href="#" class="fa fa-instagram"></a></li>
            <li><a href="#" class="fa fa-twitter"></a></li>
            <li><a href="#" class="fa fa-dribbble"></a></li>
            <li><a href="#" class="fa fa-behance"></a></li>
        </ul>
    </div>

</div>
</div>
</footer>

<!-- SCRIPTS -->
<script src="static/js/jquery.min.js"></script>
<script src="static/js/bootstrap.min.js"></script>
<script src="static/js/aos.js"></script>
<script src="static/js/owl.carousel.min.js"></script>
<script src="static/js/custom.js"></script>

```

```
</body>  
</html>
```