A Project Report

On

# Visual Enhancement of E-Commerce Products

Submitted in partial fulfilment of the requirement of

**University of Mumbai**

For the Degree of

**Bachelor of Engineering**

*in*

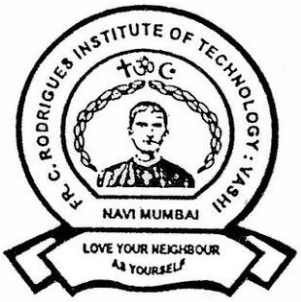
**COMPUTER ENGINEERING**

*Submitted by*

**Akash Kankariya**
**Gaurav Makasare**
**Roselyn Lorson**

*Supervised by*

**Sandhya Pati**




**Department of Computer Engineering**

**Fr. Conceicao Rodrigues Institute of Technology**
**Sector 9A, Vashi, Navi Mumbai - 400703**

**UNIVERSITY OF MUMBAI**

**2019-2020**

# APPROVAL SHEET

This is to certify that the project entitled

"Visual Enhancement of E-Commerce Products"

**Submitted by**


**Akash Kankariya 101623**
**Gaurav Makasare 101629**
**Roselyn Lorson 101648**


**Supervisors :** _____

**Project Coordinator :** _____

**Examiners : 1.** _____

**2.** _____

**Head of Department :** _____

**Date :**

**Place :**

# Declaration

We declare that this written submission for B.E. Declaration entitled "**Visual Enhancement of E-Commerce Products**" represent our ideas in our own words and where others' ideas or words have been included. We have adequately cited and referenced the original sources. We also declared that we have adhere to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any ideas / data / fact / source in our submission. We understand that any violation of the above will cause for disciplinary action by institute and also evoke penal action from the sources which have thus not been properly cited or from whom paper permission have not been taken when needed.

**Project Group Members:**

1. Akash Kankariya, 101623

_____

2. Gaurav Makasare, 101629

_____

3. Roselyn Lorson, 101648

_____

# Abstract

Often visualization of certain items gives us a clear and proper understanding of that item. For every e-commerce store, in order to increase the sales of its products, it is very important that its customers are able to fully understand the features of the product. Furthermore, they should have a clear idea about how the product looks like in their environment, before they actually purchase the product.

Our project - visual enhancement of e-commerce products - aims at providing a solution to display products in a smarter way in any e-commerce store. This solution can be particularly structured into two parts. The first phase would be to use Image processing to clean and make the images consistent across the e-commerce store. To improve the quality of the image, we will analyze the features of the image, like symmetry, clarity, object-camera distance, blur etc. Later, using AI to analyze these images and giving feedback to the clients suggesting them to click a better image of the product. Starting with conversion to a background consistent image, we will try to improvise the rest of the features as well. This will enable us to render good quality images which would be used to showcase products to their end-users. Also, the user will be given the option to allow the improvising changes or not, depending on his/her preferences, the images will be uploaded. This will allow the final decisions to be made by the user only, making our product totally user-oriented.

# Contents

# List of Figures

# Chapter 1

# Introduction

## 1.1 Background

Online shopping has not only increased but it has become a trend. Huge websites like Amazon, Flip-kart, Jabong, eBay are top B2C firms. But whenever any seller wants to list their products on their website, they charge heavily for the same. This point is important because the same thing can be done by the seller itself if he/she has his/her own way of selling their products via the internet.

## 1.2 Motivation

This system empowers each individual who has any products in his arsenalby giving him/her a platform for displaying and describing his/her product in afresh perspective. This makes his products accessible from any corner around theworld. This system aims at providing each individual to have an opportunity tobe an entrepreneur.

## 1.3 Aim and Objective

The aim is to improve the visualisation of e-commerce products using image processing, specifically, foreground extraction to enhance the features of theproducts. The main aim in this first step is to clear out noise and disturbances,and make the favourable qualities of the product even more visible to the endclient. Secondly, we wish to highlight the flaws in any uploaded picture by theuser, by giving the images rating based on artificial intelligence to, in the end, givesuggestions on how the images could've been better.

## 1.4 Report Outline

In the world of internet, e-commerce is booming at an exponential speed. From 2000 to 2020 growth is 40X. Due to a recent increase in internet usage, it has benefited the ecommerce platform. Ecommerce provides a platform where sellers upload photos and buyers look at the photo and choose whether to buy or not. So Image plays an important role in the decision. So to benefit the seller we have developed a product which will help them to represent their product better way.

The background is an important feature in the image. No one likes shabby background which has an inconsistent background, unwanted object, different background color, low quality. So to deal with it we are using grabcut which

will remove the foreground form the image and use the constant background which will only highlight the object. Scaling image whether it is good or bad is important, so to get knowledge of it we are using NIMA which rate the image on a scale of 1-10 and give the overall score of attractive it is.

# Chapter 2

# Study Of the System

## 2.1 About the Technique

**Grabcut** uses the rectangular box as input which as foreground in it. Then algorithm segments it iteratively to get the best result. But in some cases, the segmentation won't be fine, like, it may have marked some foreground region as background and vice versa. there are four strokes are available sure foreground, sure background, likely foreground, likely foreground which help in fine-tuning the output image.

**NIMA** is a rating algorithm developed by Google Developers. It uses deep CNN that is trained to predict which images a typical user would rate as looking good (technically) or attractive (aesthetically). It can be used to not only score images reliably and with a high correlation to human perception, but also it is useful for a variety of labor-intensive and subjective tasks such as intelligent photo editing, optimizing visual quality for increased user engagement, or minimizing perceived visual errors in an imaging pipeline.

## 2.2 Various Available Technique

**1.Using frame differencing**

A motion detection algorithm begins with the segmentation part where foreground or moving objects are segmented from the background. The simplest way to implement this is to take an image as background and take the frames obtained at the time t, denoted by I(t) to compare with the background image denoted by B. Here using simple arithmetic calculations, we can segment out the objects simply by using image subtraction technique of computer vision meaning for each pixels in I(t), take the pixel value denoted by P[I(t)] and subtract it with the corresponding pixels at the same position on the background image denoted as P[B].

In mathematical equation, it is written as: P[F(t)]=P[I(t)]-P[B] The background is assumed to be the frame at time t. This difference image would only show some intensity for the pixel locations which have changed in the two frames. Though we have seemingly removed the background, this approach will only work for cases where all foreground pixels are moving and all background pixels are static.[2] A threshold "Threshold" is put on this difference image to improve the subtraction (see Image thresholding).

|P[F(t)]-P[F(t+1)]| > Threshold

This means that the difference image's pixels' intensities are 'thresholded' or filtered on the basis of value of Threshold. [3] The accuracy of this approach is dependent on speed of movement in the scene. Faster movements may require higher thresholds.

For calculating the image containing only the background, a series of preceding images are averaged. For calculating the background image at the instant t,

B(x,y,t)=1/N($\sum_{i=1}^{N} V(x, y, t - i)$)

where N is the number of preceding images taken for averaging. This averaging refers to averaging corresponding pixels in the given images. N would depend on the video speed (number of images per second in the video) and the amount of movement in the video.[4] After calculating the background B(x,y,t) we can then subtract it from the image V(x,y,t) at time t = t and threshold it. Thus the foreground is

$|V(x, y, t) - B(x, y, t)| >$ Th

where Th is threshold. Similarly we can also use median instead of mean in the above calculation of B(x,y,t).

Usage of global and time-independent thresholds (same Th value for all pixels in the image) may limit the accuracy of the above two approaches.

**2. Full reference image QA**

Full-reference image quality assessment (FR-IQA) techniques compare a reference and a distorted/test image and predict the perceptual quality of the test image in terms of a scalar value representing an objective score. The evaluation of FR-IQA techniques is carried out by comparing the objective scores from the techniques with the subjective scores (obtained from human observers) provided in the image databases used for the IQA. Hence, we reasonably assume that the goal of a human observer is to rate the distortion present in the test image. The goal oriented tasks are processed by the human visual system (HVS) through top-down processing which actively searches for local distortions driven by the goal.

**3. No reference image QA**

The aim of no-reference image quality assessment (NR-IQA) techniques is to measure the perceptual quality of an image without access to the reference image. In this letter, a novel NR-IQA measure is introduced in which quality-aware statistics are used as perceptual features for the quality prediction. In the method, the distorted image is converted to grayscale and filtered using gradient operators. Then, the speeded-up robust feature (SURF) technique is employed to detect and describe keypoints in obtained images. The SURF interest point detection method is affected by distortions in the filtered image. Therefore, it can be used to reflect the decreased attention of the human visual system caused by image distortions.

## 2.3 Related Works

Related work (expected at least 5 research papers brief description with comparison)

1. **Automatic Foreground Extraction For Images And Videos locate the foreground object coarsely by a salience detection algorithm [1]**
   This system refines the object by Weighted Kernel Density Estimation(WKDE) and graph cut algorithm. Salient object detection is a technique of detecting the object in the image that catches our visual attention. After salience detection, they compute the initial probability of each pixel belonging to the foreground using Bayes estimation and difference of pixel colours. Then we use WKDE to refine the probability map and at last, use graph cut to cut out the foreground. The algorithm used is WKDE, graph cut is used to extract the foreground object.

2. **Foreground Target Extraction in Bounding Box Based on Sub-block Region Growing and Grab Cut [2]**
   The region of interest for the network detection based on deep learning algorithms exists a problem of background interference, which goes against the feature extraction of the region of interest. In order to solve the problem of background interference faced by the region of interest, this paper proposes a foreground extraction algorithm for the target detection based on deep learning algorithms in combination of the sub-block region growing algorithm and the Grab Cut algorithm. Firstly, this algorithm pre-processes the colour image through the Retinex algorithm to improve the image contrast; secondly, it extracts the features of the image sub-blocks (including dominant hue, secondary hue, colour and luminance, etc.) and constructs the similarity measurement function between different sub-blocks; thirdly, it defines the seed region, sub-block region growing and termination criterion of the region growing and achieves the coarse segmentation of the foreground; finally, it re-extracts the coarse segmentation results that have been processed through the open operation by means of the Grab Cut algorithm to obtain the foreground segmentation results of the input image.

3. **Efficient Foreground Segmentation Using an Image Matting Technology [3]**
   In this paper, an algorithm is proposed where the Gaussian Mixture Model is first used to model the static background regions. The boundary box of the foreground regions is determined via inter-frame change

detection and SIFT feature analysis, and Grabcut algorithm(an image matting technology) is then used to obtain the optimal segmentation of foreground moving objects.In order to solve the problems for these existing methods, they proposed a novel algorithm for foreground segmentation by employing SIFT sampling and GrabCut technique, where the Grabcut algorithm can combine intensity information with local information to obtain more accurate segmentation results. An important feature of the GrabCut lies in its comprehensive consideration of the distributions of foreground objects and to reduce computation cost, sift trimap is used.

4. **4.Planar Background Elimination In Range Images[4]**
   In this paper, they have captured the objects placed over a supporting plane (e.g. a table or the floor). The only restriction applied is that the object must be entirely above the plane, and the scanner should not capture data below the plane. This condition is easily attainable, both in a laboratory or in the field. Based on this precondition, we developed a method to automatically detect this supporting plane.In these planar background detector uses several techniques, like a variation of the Hough transform, robust methods and pre-condition of no information below the supporting plane.

5. **Background Subtraction Algorithm Based Human Motion Detection [5]**
   This paper presents a new algorithm for detecting moving objects from a static background scene to detect a moving object based on background subtraction. It sets up a reliable background updating model based on statistical. After that, morphological filtering is initiated to remove the noise and solve the background interruption difficulty. At last, the shadow effect is removed by combining projection analysis with shape analysis. Finally, moving human bodies are accurately and reliably detected. This method also has a very good effect on the elimination of noise and shadow, and be able to extract a complete and accurate picture of moving the human body.

# Chapter 3

# Proposed System

## 3.1 Problem Statement

We require a solution which will serve as an add-on to the out-house company.It would need to have a couple of features which would improve the visualisationof products. The first feature would be to use Artificial Intelligence and ImageProcessing to analyze the images uploaded by the user. This feature will rate thequality of images of the user and give them suggestions to click it better. Theseimages will be used in the e-commerce store we'll create for the user and a goodquality image can increase the chances of conversion to sales. The second featurewould be the generate the AR view of the product. The user can choose to optout of this feature. This would allow to give a more realistic view of the productalong with giving him the freedom to use the traditional image viewing system.This feature would increase the exposure and likeability of the product, in turn,increasing chances of conversion

## 3.2 Scope

This system has a wide scope as the number of individuals starting theirbusiness will increase in the future and so will the need for those businesses toexpand as soon as possible. Having an application of their own, without havingany third party involved for providing such service, gives the business opportunityto grow and flourish

## 3.3 Proposed System

There are multiple ways through which we can extract the foreground. We are using the Grabcut algorithm. In a grabcut algorithm, we first provide a rectangle that defines the boundary of the foreground. Then through the Gaussian Mixture Model (GMM) grabcut predict whether the pixel is from the background or foreground. Grabcut algorithm also provides an option for fine-tuning. Fine-tuning options are

1. Sure Foreground
2. Sure Background
3. Probable Foreground
4. Probable Foreground

Through fine-tuning, anomalies can be removed from the image. Rating image is important to know how good the image is. We are using NIMA which rates the image on the scale of 1-10, where 1 is for worst and 10 is for best. In general, image quality assessment can be categorized into full-reference

and no-reference approaches. If a reference "ideal" image is available, image quality metrics such as PSNR, SSIM, etc. have been developed. When a reference image is not available, "blind" (or no-reference) approaches rely on statistical models to predict image quality. The main goal of both approaches is to predict a quality score that correlates well with human perception. In a deep CNN approach to image quality assessment, weights are initialized by training on object classification related datasets (e.g. ImageNet) and then fine-tuned on annotated data for perceptual quality assessment tasks.d data for perceptual quality assessment tasks.

# Chapter 4

# Design Of the System

## 4.1 Requirement Engineering

### 4.1.1 Requirement Elicitation

1. Customers providing decent quality pictures of products.
2. Good enough device for clicking pictures.
3. A good processing system application for computing and providing results.

### 4.1.2 Software lifecycle model

The software lifecycle model that has been used is the **Agile Model**. Each build is incremental in terms of features; the final build holds all the features required.

The initial basic requirement of the system was to clean the image. This process involved cleaning the image using image processing and enhancing it. The next process involved removing the background of the image. For efficiency and better performance in less amount of time, different algorithms were tried and tested and the best suited one was the Grab Cut algorithm. The entire processing and funcioning was run on a command prompt, but with icremental developments, a graphic user interface had been developed for a smoother user experience.

Also feauters such as selecting images of same category from different folders had been added rather than selecting them from a set fo mixed images. For better understanding of the quality of the image a rating algorithm has been implemented. Using this feature it is possible that if the image is not of good quality that is if the rating is low, the respected image can be replaced for a better quality image.The advantages of the Agile Model are that it is a very realistic approach to software development, it delivers early partial working solution, little or no planning required, etc.

### 4.1.3 Requirement Analysis

#### 4.1.3.1 UML diagrams/ DFDs based on the project

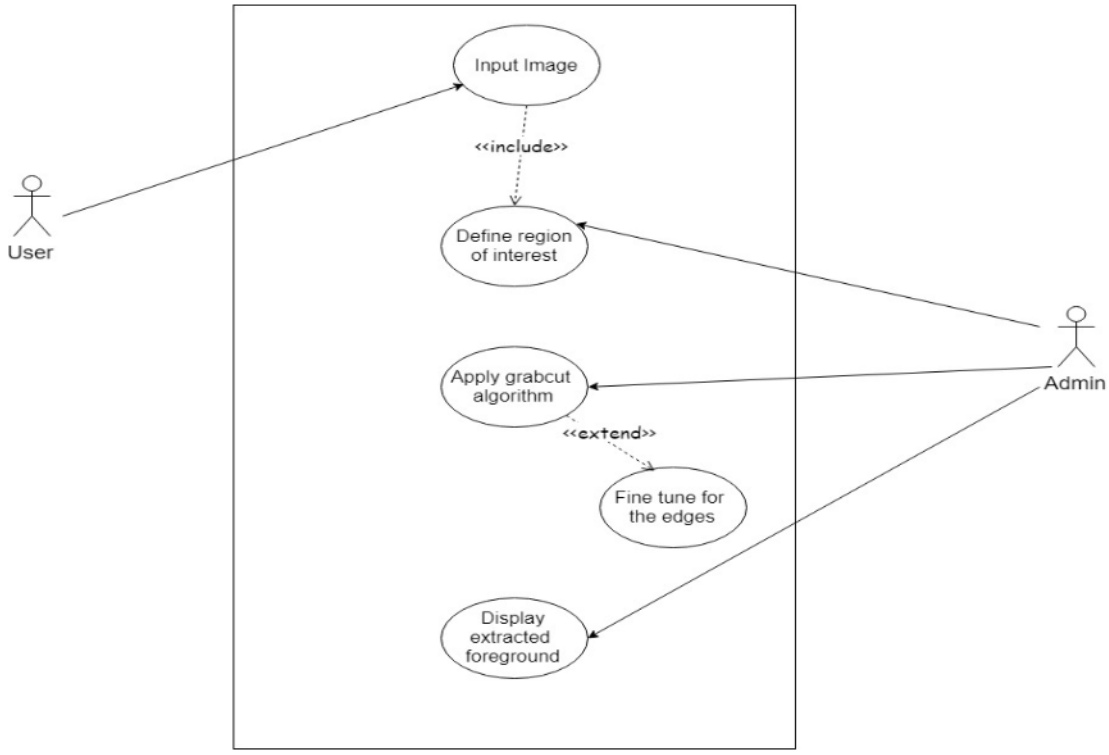



Figure 4.1: Use Case Diagram

#### 4.1.3.2 Hardware and software requirement

- Window 7 and above
- Python 3.7
- OpenCV 4.1.2
- Tensorflow 2.0

## 4.2 System architecture

### 4.2.1 UI/UX diagram

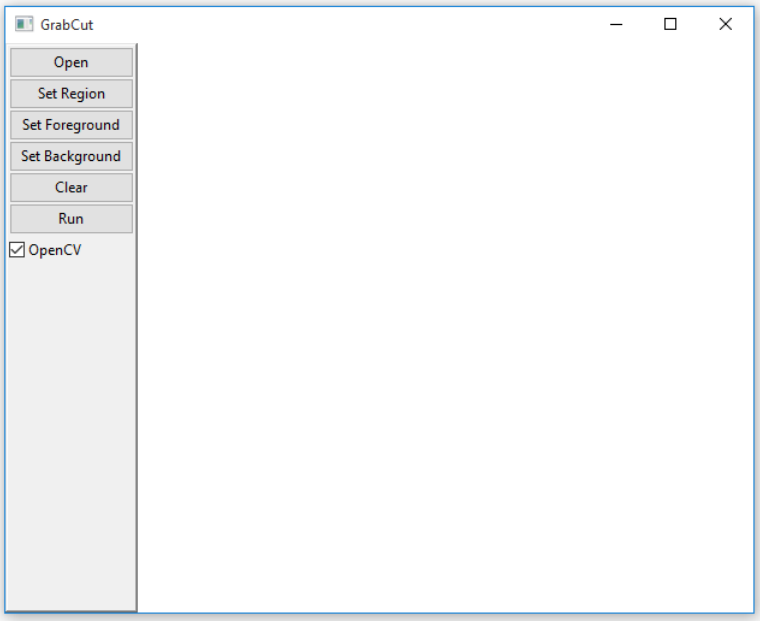



Figure 4.2: User interface

### 4.2.2 Block Diagram

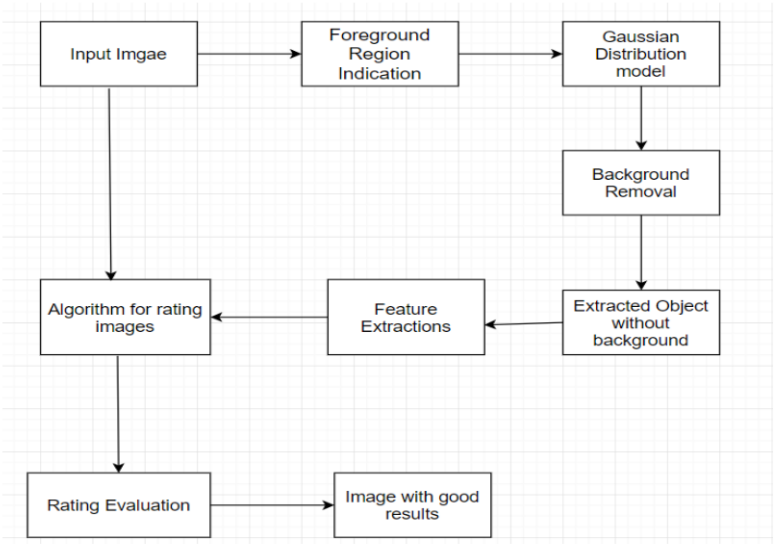



Figure 4.3: System Architecture Block Diagram

# Chapter 5

# Result and Discussion

## 5.1 Screenshots of the System

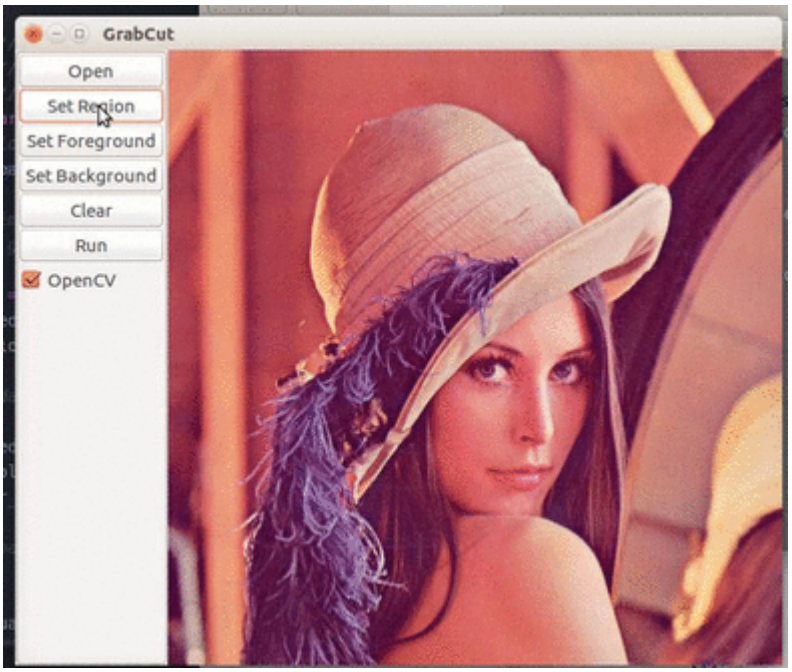



Figure 5.1: GUI

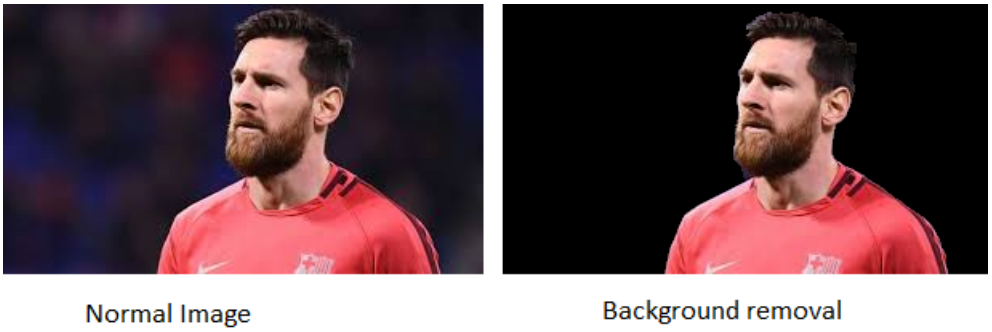



Figure 5.2: Output of GrabCut

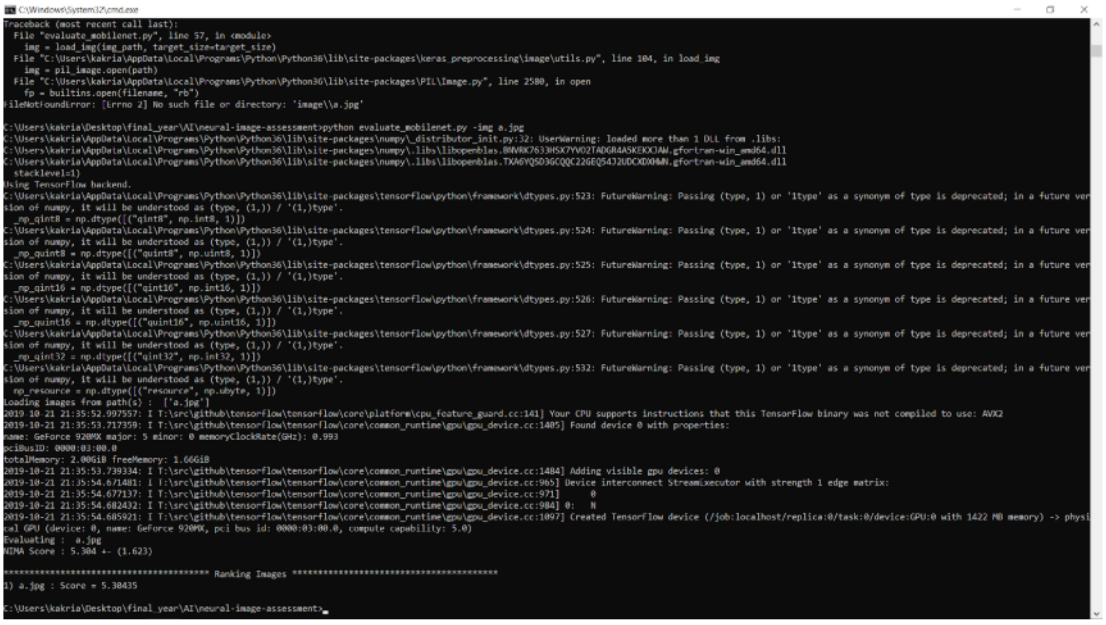



Figure 5.3: Rating Output

## 5.2 Sample Code

### Background Removal

```
if (self.rect_or_mask == 0):                # grabcut with rect
    bgdmodel = np.zeros((1, 65), np.float64)
    fgdmodel = np.zeros((1, 65), np.float64)
    cv.grabCut(self.img2, self.mask, self.rect, bgdmodel, fgdm
                self.rect_or_mask = 1
elif self.rect_or_mask == 1:                # grabcut with mask
    bgdmodel = np.zeros((1, 65), np.float64)
    fgdmodel = np.zeros((1, 65), np.float64)
    cv.grabCut(self.img2, self.mask, self.rect, bgdmodel, fgdm
```

### Image Rating

```
img = load_img(img_path, target_size=target_size)
        x = img_to_array(img)
        x = np.expand_dims(x, axis=0)

        x = preprocess_input(x)

        scores = model.predict(x, batch_size=1, verbose=0)[0]
```

```
mean = mean_score(scores)
std = std_score(scores)

file_name = Path(img_path).name.lower()
score_list.append((file_name, mean))

print("Evaluating : ", img_path)
print("NIMA Score : %0.3f +- (%0.3f)" % (mean, std))
print() ‘
```

## 5.3 Testing

### 5.3.1 Unit Testing

1. **Capturing the Frame**
   This unit involves capturing the frame or a rectangular boundary across the image inorder to focus on the most probable part of the image where the foreground is present. The expected output is the region with the rectangular boundary along with all the pixels inside it.It is obeserved that the pixels are getting captured inside the box and a clear region with a frame is ready to work with.

2. **Grab Cut algorithm Testin**
   The expected outcome of this unit is to focus and capture only the foreground within the captured frame. It is observed that the pixels in the foreground are eliminated and only the foreground remains intact.

3. **Fine Tuning**
   This unit involves clearing out the minute unwanted pixels or the incorporate the missing pixels that is deciding whether the pixels are the most suitable background or the most suitable foreground. The expected output can be observered on reiterating the fine tuning.

4. **Nima algorithm Testing**
   This unit involves providing an image which is expected to get evaluated and a rating is to be generated for it. It is observed that the expected function is being performed and a rating is generated for the supplied image.

### 5.3.2 Integration Testing

All the units are combined and a system a generated where an image is taken and the region of choice is indicated. Next the Grab Cut algorithm is implemented on it which gives the extracted foreground which is further enhanced using fine tuning. In the end the NIMA algorithm is run over the extracted image a rating is generated. The obeserved functionality is clearly as expected.

### 5.3.3 Blackbox Testing

In this type of testing it is obeserved that for some images that contain inconsistent background, transparent objects or too many objects, the observed output contains pixel discrpencies. This is a performance error which rarely occurs for only certain type of images.

# Chapter 6

# Conclusion & Future Scope

Thus, We have done image processing and made an artificial intelligence ratingsystem for this term. The motive is to improve the visualisation of e-commerceproducts and result in better conversion to sales. We have focused on imageprocessing, specifically, foreground extraction of the products and further ratingthe images to give suggestions on how the images could have obtained even higherratings. Our further goal is to make an Augmented Reality system that would allowthe end client to have a three dimensional (3D) view of the products providing anall new way to observe what they are paying for. This section of the project mayhave its challenges as to modelling it seamlessly but if done right, this new featurecan change the way we observe, visualise and buy anything online.

# References


[1] Z. Tang, Z. Miao, Y. Wan, and J. Li, "Automatic foreground extraction for images and videos," in *2010 IEEE International Conference on Image Processing*, pp. 2993–2996, 2010.

[2] Y. Yu, J. Zhu, and J. Pei, "Foreground target extraction in bounding box based on sub-block region growing and grab cut," in *2018 14th IEEE International Conference on Signal Processing (ICSP)*, pp. 344–349, 2018.

[3] X. Gong and Z. Li, "Efficient foreground segmentation using an image matting technology," in *2013 International Conference on Computational and Information Sciences*, pp. 750–753, 2013.

[4] A. Vrubel, O. R. P. Bellon, and L. Silva, "Planar background elimination in range images: A practical approach," in *2009 16th IEEE International Conference on Image Processing (ICIP)*, pp. 3197–3200, 2009.

[5] R. Jain and S. Kaur, "Impact of work environment on job satisfaction," *International Journal of Scientific and Research Publications*, vol. 4, no. 1, pp. 1–8, 2014.

# Acknowledgement

Success of a project like this involving high technical expertise, patience and massive support of guides, is possible when team members work together. We take this opportunity to express our gratitude to those who have been instrumental in the successful completion of this project. We would like to show our appreciation to **Ms. Sandhya Pati** for their tremendous support and help, without them this project would have reached nowhere. We would also like to thank our project coordinator **Mrs. Rakhi Kalantri** for providing us with regular inputs about documentation and project timeline. A big thanks to our HOD **Dr. Lata Ragha** for all the encouragement given to our team. We would also like to thank our principal, **Dr. S. M. Khot**, and our college, **Fr. C. Rodrigues Institute of Technology, Vashi**, for giving us the opportunity and the environment to learn and grow.

**Project Group Members:**

1. Student 1 Name, Roll No

_____

2. Student 2 Name, Roll No

_____

3. Student 3 Name, Roll No

_____

# Appendix A : Timeline Chart

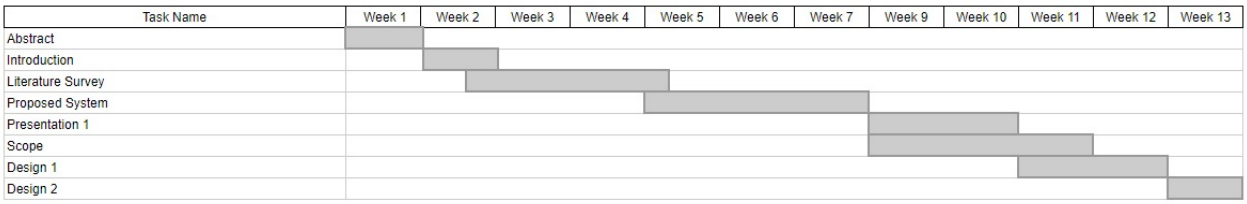



**Figure 6.1: Timeline**

# Appendix B : Publication Details

**Paper Published in the following journal**

1. International Research Journal of Engineering and Technology (IRJET) Date: 03 March 2020

**Authors**

1. Akash Kankariya
2. Gaurav Makasare
3. Roselyn Lorson