

Table of Contents

[Problem Statement: 3](#_Toc20858)

[Objectives: 3](#_Toc12796)

[Darwin’s Theory: 3](#_Toc2356)

[Natural Phenomena: 3](#_Toc13822)

[Theory: 4](#_Toc572)

[Solution Development: 4](#_Toc10209)

[Code Description: 5](#_Toc20805)

[1. Population Initialization 5](#_Toc6612)

[2. Fitness Evaluation 5](#_Toc30435)

[3. Natural Selection 5](#_Toc27830)

[4. Next Generation 5](#_Toc4268)

[1. Population Initialization: 5](#_Toc20268)

[2. Fitness Evaluation: 6](#_Toc9330)

[3. Natural Selection: 7](#_Toc637)

[Next Generation: 7](#_Toc24118)

[Limitations: 10](#_Toc31536)

# 

# Problem Statement:

The problem being addressed in this assignment is to find an object in actual image by comparing with a smaller image which is actually a part of the actual image and contains the target search object. The assignment requires to search the target object in an optimistic way by developing an evolutionary algorithm based on Darwin’s theory of Natural Selection. The algorithm takes basic ingredients and flow of a natural phenomenon upon which the theory is based, and leads towards a program which can be used to find an object in an image.

# Objectives:

This report is aimed at building an understanding of the Darwin’s theory’s relationship with the algorithm designed to solve the face detection problem. To achieve this objective, the whole process of developing an algorithm, purely based upon the ingredients of the theory, is explained in order to build complete understanding of the thought process behind the program.

# Darwin’s Theory:

In this assignment, we are developing an algorithm based on a theory explaining a natural phenomena. The basic understanding of the theory and its relationship with the algorithm is discussed in details in the sections below.

## Natural Phenomena:

Science is based upon the observation and investigation of natural phenomena happening in the universe. One of the interesting phenomena happening in the universe from hundreds of years is diversity. It refers to variations and similarities present among organisms which lead towards variety in a single specie i.e. organisms belonging to same specie have variant characteristics. This natural reality was closely observed by Darwin who investigated his whole life in collecting evidences on a hypothesis which was developed to explain the existence of diversity among organisms.

## Theory:

According to his hypothesis, a population has diverse characteristics which has a close relationship with the surrounding environment. It means that for an organism to survive in a specific environment, its characteristics help. Those organisms who have favorable characteristics, those characteristics which help its survival in an environment, become the best fit individuals of a population. And then, they are most likely to make a major contribution towards development of next generation. This was called by him, the Natural Selection which simply means that in a population, nature selects best fit organisms according to favorable characteristics required to survive in the environment. In this way, generations after generations, characteristics of organisms change which result in diversity among them. So, in this way, Darwin gave a scientific explanation of a natural phenomena which can be mapped to different situations around us.

# Solution Development:

To solve the face detection problem stated earlier, we take the following ingredients of the Darwin’s Theory.

* Population
* Diversity
* Natural Selection

These ingredients are mapped on the image while developing solution. The first ingredient is taken by taking some random points of the image which show our initial population. This population has diversity as points are representing different areas of the image with different properties. This variation in properties make each point a contributor in diversity of the population. Now, after selecting the population having diversity, the Darwin’s explanation of “Natural Selection” is incorporated to generate the next population of points of the image. This process is repeated, generations are generated based on a criteria of selection which is fitness of the point, until we get the point of interest which is the point where our target image exists in the image.

## Code Description:

The flow of algorithm is divided into four sections and each section applies one element of Darwin’s theory. These sections include the following functions.

1. Population Initialization
2. Fitness Evaluation
3. Natural Selection
4. Next Generation

Each of these functions are discussed in detail in the below section.

### Population Initialization:

Firstly, to map the theory on problem, we must have basic element which is population. As proposed by the theory, there is a population having diversity on the basis of which a new population is generated. Therefore, in coding approach, we initialize the first population through randomization. This python function takes three arguments; rows, columns and population size. To generate a population of points for an image, we have to generate points within dimensions of that image therefore we require the number of rows and columns of the image. Moreover, we specify population size.

#### Assumption 1 (Unique Population):

In this function, we make an assumption which states that in nature, population has unique individuals i.e. each individual has some features which make it distinct in its whole population. To ensure this uniqueness, we randomly generate x and y coordinates and before making them a touple (representing a point on the image), we make sure that generated touple is not already present in the list of initial population. In the whole algorithm, this function is used once, for initializing first population.

### Fitness Evaluation:

After generating a population, the next step is to evaluate it on a criteria which decides its fitness and role in the next generation. As suggested by the theory, an individual’s fitness is evaluated on the basis of its ability to survive a specific environment. In our scenario, the environment according to which we have to make this fitness evaluation is the image which we have to search for in the big image. It means, we are interested in checking to what extent a point (population individual) is matching our point of interest (smaller image containing the object to be searched for in the big image). For this evaluation, we make a python function of fitness evaluation taking three arguments; current generation, large image, small image. This function completes its goal of evaluating fitness of an individual point by performing two tasks. The first task is to take frame from the large image and the second task is to perform correlation to check similarity level.

#### Assumption 2 (Generating Valid Frames):

In first task, the program takes one individual of the population and maps it on the large image. Here it has to create a frame of large image having size similar to the smaller image to compare through correlation.

For these frame, our algorithm makes an assumption which states that the individual point acts as top left corner of the frame and then our program checks, if adding rows and columns (of smaller image) in that point gives a frame within the rows and columns boundaries of the large image or not. If the frame boundaries calculated in the program are within the boundaries, we take that frame. Otherwise, we take the area around that point which has size equal to the size of the small image as in all frames. This validity of frames is ensured in a helper function, frame boundaries.

This approach is used to eliminate the possibility of ignoring the areas which lie on the edges and corners but may contain the point of interest (target image). Moreover, it also makes our overall population a good population in which we instead of ignoring edge point, include them in the population and take the region around them in the first time instead of taking the chance of getting valid region around that point in some other population.

The frames are taken from the large image through a helper function, get frame. This helper function takes all the points of the image within the frame boundaries and return them in the form of a array.

#### Correlation Calculation:

The second task of evaluating correlation between frame of larger image and the small image uses a helper function, correlation coefficient. This function takes two arrays of images and return a single number representing the level of correlation between them. In the end, we store fitness value of each individual of the population for further use in generation of next generation.

### Natural Selection:

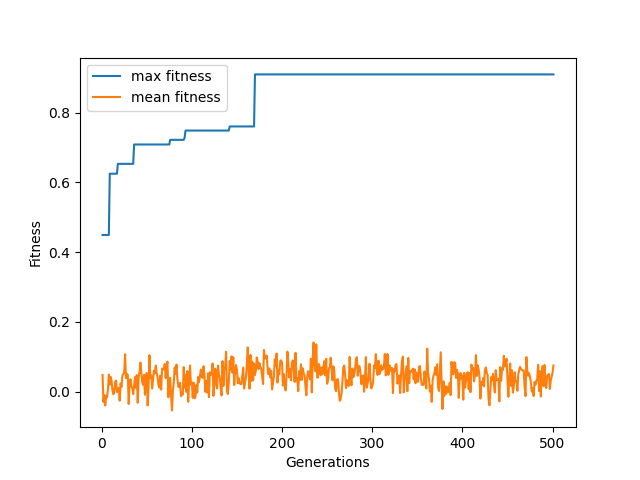
This function takes population fitness evaluated in the previous function and returns a sorted population. This sorting is done by this function on the basis of fitness values against all individuals.

### Next Generation:

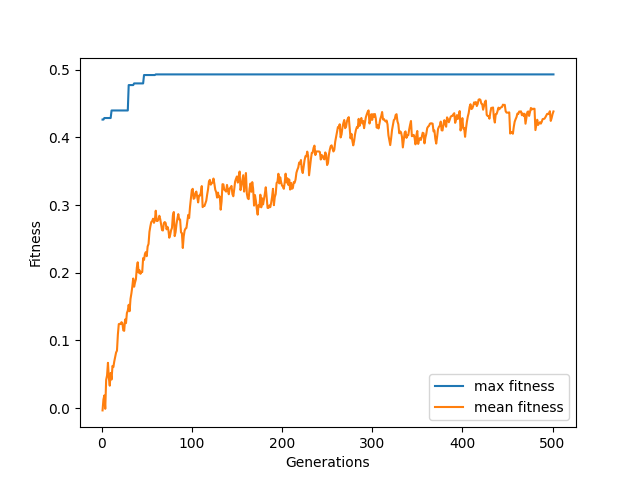
This is the most important function of the program. It takes sorted population, rows and columns of the large image as argument and performs two main tasks to generate the next generation on the basis of current generation which is sorted on each individual for its fitness values. These tasks include crossover and mutation. The crossover is performed by a helper function in which a chunk of bits of the parents is interchanged. And the mutation is also performed in a helper function which alternates the value of bit at specific position of the child. Here we make new generations by combining two parents and generating two children. For example, in sorted order, point 1 and 2 act as parent 1 and parent 2 to create child 1 and 2 and then point 3 and 4 are used to create child 3 and 4.

#### Assumption 3 (Crossover):

Crossover is used in two ways. In the first way we take one random number as tooka and partition the combined binary of parents and then merge accordingly. It results in increasing the fitness of single individual as shown in the figure below.

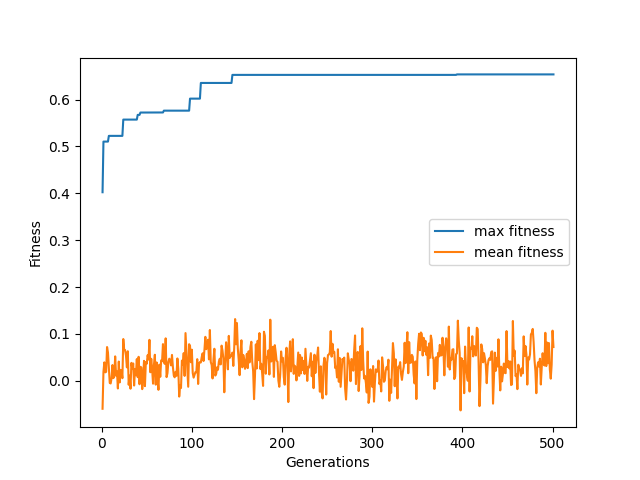


In other case, we take two random numbers as tooka\_1 and tooka\_2 and partition the parent bits in three portions and then merge. According to our assumption, in the case of two tookas, the result is improvement in the mean fitness of population which can be shown in the figure attached below.

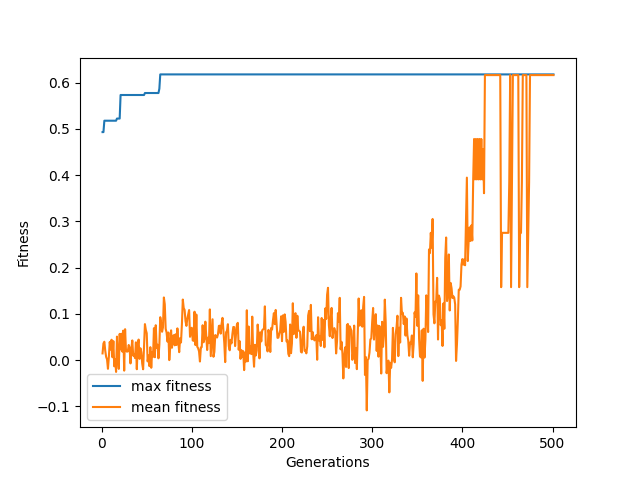


#### Assumption 4 (Mutation):

There are two cases in mutation also. In the first case, we take mutation rate up to two percent and mutation occurs in the middle. It results in improving the individual fitness instead of population fitness as shown in the figure below.



In the second case, mutation occurs randomly. In this case the average fitness of the population increases as shown in the figure below.



# Limitations:

There are following limitations in the coded algorithm.

1. We are unable to get a combined result of gradually improving mean fitness and maximum fitness of generations. Although, we reached to getting an improving curve in both cases separately, we could not integrate them in a single code.
2. In the case of maximum fitness of generations, we gradually reach to some value of fitness but after that the curve becomes horizontally straight showing that for long number of generations there is no improvement in the maximum fitness of any individual generation. And then suddenly, we get the match for best fit in one generation and curve goes straight up vertically.