

Czech University of Life Sciences Prague

Faculty of Economics and Management

Department of Information Engineering



Bachelor Thesis

Image Filtering using Object Detection

Akshat Vishal Sud

© 2023 CZU Prague

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Economics and Management

BACHELOR THESIS ASSIGNMENT

Akshat Vishal Sud

Informatics

Thesis title

Image filtering using object detection

Objectives of thesis

The main goal of the thesis is to demonstrate the functions of convolutional neural networks and algorithms for detecting and recognizing objects in images using a prototype application.

This will be accomplished using the JavaScript-based ML5 library, and pre-trained network model based on the COCO dataset. The application will take a set of images from the user and return the images that contain particular object.

Methodology

The methodology of the thesis is based on analysis of technical and scientific sources focusing on artificial intelligence, machine learning, and object detection, with particular emphasis on CNN and COCO dataset. Based on the synthesis of the knowledge gained, a pre-trained model will be used, and embedded into a prototype web application. The performance of the model will be assessed in terms of detection performance.

The proposed extent of the thesis

40-50 pages

Keywords

object detection, COCO, filtering, js, ml5

Recommended information sources

CAPELLMAN, Jarred, 2020. Hands-On Machine Learning with ML.NET. Birmingham: Packt Publishing. ISBN 9781789801781.

CYGANEK, Boguslaw, 2013. Object Detection and Recognition in Digital Images: Theory and Practice. Chichester, UK: John Wiley and Sons. ISBN 978-0470976371.

Expected date of thesis defence

2023/24 WS – PEF

The Bachelor Thesis Supervisor

Ing. Petr Hanzlík, Ph.D.

Supervising department

Department of Information Engineering

Electronic approval: 21. 11. 2023

Ing. Martin Pelikán, Ph.D.

Head of department

Electronic approval: 23. 11. 2023

doc. Ing. Tomáš Šubrt, Ph.D.

Dean

Prague on 24. 11. 2023

Declaration

I declare that I have worked on my bachelor thesis titled "Image Filtering Using Object Detection" by myself and I have used only the sources mentioned at the end of the thesis. As the author of the bachelor thesis, I declare that the thesis does not break any copyrights.

In Prague on date of submission

Acknowledgement

I would like to thank my supervisor Ing. Petr Hanzlík, Ph.D. for helping me to gain insight to perform the project. Without the guidance of the professor, I would not have completed the research for which I am grateful.

Image Filtering using Object Detection

Abstract

The primary aspect of the project is associated with the detection of objects within the image by incorporating the COCO dataset. The purpose is to filter a set of user-provided images based on specified target object names. Only the images that contain said objects are a part of the resulting output bucket and can then be further used by the users. The methodology that has been used for the completion of this project includes the use of HTML and JavaScript based ML5 library as the frameworks for the program code that has been written in this study. This project utilizes a pretrained COCO SSD model for object detection. The accuracy level of the program that has been implemented has been tested using a database of new images. Based on this database, it has been identified that the program is quite efficient in identifying the different objects as per the search conducted by the user. The overall accuracy of the pretrained COCO SSD model was 86.2%. Object of interest were not properly detected in challenging scene conditions or when occluded or captured only partially.

Keywords: COCO dataset, Image processing, JavaScript

Filtrování obrázků pomocí detekce objektů

Abstrakt

Primární aspekt projektu je spojen s detekcí objektů v rámci obrazu s využitím datové sady COCO. Účelem je filtrovat sadu obrázků vložených uživatelem na základě zadaných názvů cílových objektů. Pouze obrázky, které obsahují uvedené objekty jsou součástí výsledného výstupu a mohou být uživateli dále používány. Metodologie, která byla použita pro dokončení tohoto projektu, zahrnuje použití HTML a knihovny ML5 založené na a Javascriptu jako rámce pro program, který vznikl v této studii. Tento projekt využívá předtrénovaný model COCO SSD pro detekci objektů. Úroveň přesnosti implementovaného programu byla testována pomocí databáze nových obrázků. Na základě této databáze bylo zjištěno, že program je poměrně účinný při identifikaci různých objektů podle vyhledávání prováděného uživatelem. Celková přesnost předtrénovaného modelu COCO SSD byla 86,2 %. Zájmové objekty nebyly správně detekovány v náročných podmínkách scény, při zakrytí nebo pokud byly zachyceny jen částečně.

Klíčová slova: COCO dataset, Zpracování obrazu, JavaScript

Table of content

Contents

Introduction.....	10
Introduction.....	10
Background of the research	10
Thesis Statement	10
Research rationale.....	10
Significance of the research	11
Justification of the research	11
Conclusion	12
Objectives and Methodology	12
Introduction.....	12
Aims and Objectives	12
Method for this research	13
Research Timeline	16
Summary	17
Literature Review	17
Introduction to the chapter	17
Deep Learning.....	17
Convolutional neural network	18
Artificial Neural Networks	20
JavaScript.....	21
Non-Maximum Suppression (NMS).....	22
COCO dataset	23
Theoretical underpinnings	24
Conceptual framework.....	26
Conclusion to the chapter	27
Practical Part	27
Configuring the application	27
Processes for taking input	28
exploring with the COCO dataset.....	29
Drawing the bounding boxes	31
Filtering the bounding boxes	32
Storing the result.....	33
Methods for providing output.....	34

Implementing the security processes	36
Maintaining the performance	37
Results and Discussion	39
Overall outcome of the project	39
Purpose of JavaScript in the Project	40
Setting conditions with the JavaScript to get input.....	41
Setting conditions with the JavaScript to get output.....	41
Development of the project with limitations	42
Conclusion	43
Recommendations.....	43
Future Scope	44
References.....	45

Introduction

Introduction

The research is focused on a project which is responsible for following the approaches that are maintained by the COCO SSD model. The purpose is to identify certain objects in a set of many images. The project takes a collection of images, and the names of the objects as input through the search bar of the website. The output bucket of the project contains those images which contain the objects that are given by the user. This project is currently limited to the objects available in the COCO dataset.

Background of the research

Object detection is considered to be a technique which is associated with computer vision and is responsible for locating instances of an object within an image. This type of technology uses machine learning or deep learning to provide a result that is meaningful (Yang *et al.* 2019). To be more specific, this type of technology uses the approach of human behavior as humans are able to easily recognize an object within an image due to their instinct which they learned from decades of learning. The project is entirely similar as it is focused on teaching the machines to identify an object within an image (Weerts and Pechenizkiy, 2022). The field of object detection is considered to be complex as it involves identifying features of an object from an image and thus requires complex algorithms which are constantly evolving over the years.

Thesis Statement

The main goal of the thesis is to demonstrate the functions of convolutional neural networks and algorithms for detecting and recognizing objects in images using a prototype application. This will be accomplished using the JavaScript-based ML5 library, and pre-trained network model based on the COCO dataset. The application will take a set of images from the user and return the images that contain particular object.

Research rationale

The COCO SSD holds the utmost importance in the modern era as it was used for developing many current object detection technologies. To be more specific, object detection in an image has become progressively significant in the modern era because of the ascent of PC

vision and picture-handling innovations (Larabi-Marie-Sainte *et al.* 2021). With the assistance of algorithms or a set of rules, machines can perceive and find instances of an item inside a picture or video. This innovation has various functional applications in different fields like security, reconnaissance, self-driving vehicles, advanced mechanics, and medical care. In the security and reconnaissance industry, object location can be utilized to follow and distinguish people or objects of interest progressively. Self-driving vehicles depend on object discovery to distinguish obstacles, people on foot, and different vehicles to guarantee safe driving (Ni *et al.* 2020). In the field of medical care, object discovery is utilized for clinical imaging to identify and find growths, sores, and different irregularities. To summarize, the development of such technology can be made possible by means of frameworks like COCO SSD which can help in the process of detecting objects within an image.

Significance of the research

The research is responsible for identifying objects from a huge number of images. This is utilized for the examination and examination process as both of these stages require object acknowledgement. The library that is utilized by COCO SSD is answerable for distinguishing items from a pool of 80 unique items. These items are now prepared in the library of COCO SSD and in this way are useful for distinguishing the targets in a question of time. This is because of the utilization of profound realizing which is viewed as a critical part of AI (Janiesch *et al.* 2021). The design is to prepare the frameworks in such a way that they can have the option to recognize specific items inside a picture and can handle an immense lump of pictures in a short time.

The improvement of COCO SSD utilizes JavaScript with the coordination of AI which is viewed as a programming language and has been utilized for a long time for its reliance. This innovation is viewed as one of the central advancements of the Internet and uses the methodologies of HTML and CSS (Arnoldovich, 2021). In addition to this, the usage of JavaScript can also be justified as it supports third-party libraries such as COCO SSD.

Justification of the research

The project is considered to be beneficial in the modern era as it is responsible for identifying objects within an image. To be more specific, it can be helpful for various industries all over the world with diverse requirements. One such industry is the automobile industry which

can use the framework of COCO SSD to effectively implement autopilot technology (Hrechuk and Olishkevskiy, 2022). This is due to the specification of autopilot technology which requires to detect incoming objects at a moment's notice in order to avoid any collision. In the medical field, it is responsible for identifying undesirable materials from the body of the human such as tumors or any other external objects. As a result, such technology is responsible for easing the operating procedure for medical experts and helping to save many lives. To summarize, the various benefits that can be brought in by COCO SSD can be helpful in improving the capabilities of various industries all over the world (Rostianingsih *et al.* 2020). This is where the role of the project materializes, which is responsible for processing a huge number of images in the fastest possible manner. However, the library of COCO SSD currently is limited and must be increased in the future to ensure that it can detect more objects to broaden the scope of the project.

Conclusion

The introductory chapter is focused on four critical aspects of a project which include the background of the project, its rationale, significance, and justification of the project. As a result, it is responsible for identifying the needs of the project for which the success factors of the project can be measured. These helped in setting the stage for which the literature review is to be performed, which includes various crucial factors and is required for successful implementation of the project.

Objectives and Methodology

Introduction

This part is centered around distinguishing the points and goals of the venture which guarantees that the undertaking follows a specific example.

Aims and Objectives

The primary aim of the research emphasizes the development of a project which takes images from the users as input and returns the images which contain an object. The objectives of the research are as follows:

- To take an unspecified number of images as an input from the user which can be similar or different in patterns.
- To identify if a certain object or a set of objects are presented in those images.
- To return the images to the output bucket which contains such objects or a set of objects.

Method for this research

The project is to be performed by developing a website which will allow users to upload a set of images. Their end goal is to filter those images which contain a certain object or a set of objects for their own purpose. The development of the project is centered around JavaScript Language which is to be used for developing the pages of the website by making it more dynamic (Wirfs-Brock and Eich, 2020). However, the raw design of the website is to be performed by HTML and CSS, both of which can help in setting the basic structure of the website. To be more specific, HTML is to be used for the project due to the various benefits it provides. This includes faster loading of the pages which can give the output to the users in the best possible time. In addition to this, certain features of the website include an application cache feature which includes uploading huge chunks of data. As a result, it enables the users to upload files regardless of their sizes which is responsible for satisfying the first objective of the project. On the other hand, CSS is helpful for styling the website by arranging its various elements and making sure that it can be operable regardless of the device that is being used by the user.

One of the features of the website is to contain a search box which lets the user input the names of the object/objects. This is the primary input of the website which is to be used to filter the objects from the images. It is where the role of the pre-trained model is useful which is helpful for identifying and exploring the objects from the images with the keywords that are entered from the user. This is to be done by allowing the COCO dataset which contains 80 objects and adding any object which is not part of the dataset eventually leading to showing an error message (Mikael Fernandus Simalango, 2022).

A crucial aspect of the project involves ML5.Net which is considered to be an effective machine-learning library which is being used by the developers of JavaScript. To be more specific, the platform is responsible for the classification of images which are divided into three segments. The first segment is about the classification of the image as to which category it belongs such a vehicle class can contain many objects such as “Bus”, “Train”

and “Boat” along with many others. The second stage is associated with classification and localization which is focused on identifying to which group the objects belong. Only then the final stage can be attained which is focused on particularly detecting the object. Another reason for considering ML5.net is that it supports all types of OS which are widely used all over the world. Moreover, the platform of ML5.net is responsible for collaborating with Tensorflow which is another machine learning platform.

The project also utilizes the platform of Convolutional neural networks which is considered to be a group of architecture and is useful for identifying patterns. In the project, the role of CNN is to identify the patterns within the uploaded images which can prevent any issues regarding overlapping (Li *et al.* 2021). This is performed by involving a three-layered structure in the method for neurons and is liable for recognizing one of the three shades of RGB. It is liable for going about as a channel and is answerable for tracking down the edge of an article. More specifically, it is liable for partitioning the photos into different portions or squares and recognizing the examples of neurons in those separated boxes. To be more exact, the methodology CNN underlines two key ideas which incorporate "convolution" and "pooling". These cycles are rehashed until the objects of the pictures are appropriately recognized. What's more, CNN is additionally useful for supporting both the 2D and 3D networks and consequently, it tends to be useful in recognizing both colorized and non-colorized pictures. This is because of the advantage of 2D lattices which can help in distinguishing non-colorized pictures while 3D frameworks can help in recognizing colorized pictures.

Currently each object is strongly associated with any one label. So, if a football is labeled as a sports ball, it will only detect it if we give the input "sports ball", not "football", but there is a simple solution that can solve this problem with a little bit of hard coding. For example, we are taking input from the user, "ball", "sports ball", "football", or anything.

There will be a mapping function that will map any input related to the ball to its appropriate label, in this case "football" will be mapped to "sports ball". Thus, users can input anything, and a mapper will convert it to a supported keyword. but as of now this mapper function is not implemented. so, the user must have to write down the exact label.

Notwithstanding CNN, the job of ANN likewise assumes a significant part in gathering the goals of the exploration somehow. More specifically, the ANN is liable for guaranteeing that is answerable for fostering an enormous framework in the method for basic components (Abiodun et al. 2019). This is done by dividing it into two segments which include shallow

and deep neural networks. However, the complexity of the project implies that a shallow neural network is sufficient. However, deep neural nets are the recommended choice for the project. One reason behind this is associated with processing the inputs more efficiently than the shallow approach. This is due to the presence of a number of layers which is more than one and is responsible for performing complex tasks which are much more efficient.

Notwithstanding CNN, the job of ANN likewise assumes a significant part in gathering the goals of the exploration somehow. More specifically, the ANN is liable for guaranteeing that is answerable for fostering an enormous framework in the method for basic components (Abiodun *et al.* 2019). This is done by dividing it into two segments which include shallow and deep neural networks. However, the complexity of the project implies that a shallow neural network is sufficient. However, deep neural is the recommended choice for the project. One reason behind this is associated with processing the inputs more efficiently than the shallow approach. This is due to the presence of a number of layers which is more than one and is responsible for performing complex tasks which are much more efficient.

The effectiveness of CNN and ANN can be largely materialized with the BLOB String which works as an intermediary measure. The BLOB is considered to be a binary string which is considered to be more than 200M long (Khan and Shandilya, 2021). The purpose is to establish a link between CNN and ANN by identifying the objects in a proper manner. To be more specific, the application of the CNN method is useful for dividing the image into various layers and converting the image into various layers. This is then converted into a BLOB string which can be used in the input of ANN which is responsible for processing the input which in the current context is the large chunk of images that are being uploaded by the user.

This research has used ml5.js for the COCO SSD model, which is a pre-trained model, that's why as of now this research cannot add new dataset to this model, but COCO SSD itself is a dataset, if we can generate more samples that will fit with the other data inside COCO SSD dataset then we can use that updated dataset to train our model using any other technique, like TensorFlow or sklearn. It would be a challenge for this research to modify the COCO SSD then we can use multimodal or multi network architecture where we can use any other model like YOLO or develop a personalized custom trained model with COCO SSD. This will distribute the load across multiple models, that's why it will give the code more performance, and updating individual models is much easier than a huge monolithic model.

Also, in this architecture we can use multiple systems to distribute the load. so, it's more scalable.

A crucial aspect of the website incorporates the approaches to the search bar which is responsible for allowing the user to put in the objects that are to be identified from the images. The development of such a search box is responsible for using the principles of the loop method along with a nested if (Millan *et al.* 2022). The purpose is to continuously search the images for the objects that are to be presented in the research. If the object is not found in an image it is to be proceeded into the next image until the object is found. In the case, that the object is found within the image it is to be sent in the output bucket so that users can filter the images. The IDE that has been used for the practical part of this research is the VS Code because of the user-friendly interface that is present that helps the developer to reduce the frequency of the errors in the process as well.

To summarize, the entirety of the project is to be done by using the functionalities of JavaScript with the integration of deep learning which is a crucial aspect of machine learning. This is due to the effectiveness of JavaScript which provides the features of public and private classes and thus allows flexibility (Malik *et al.* 2019). Another reason for selecting JavaScript is associated with the four features that are provided by the application which involve Abstraction, Encapsulation, Polymorphism, and Inheritance. All of these are responsible for developing the projects in the most effective manner, which can help in identifying the objects from the images. To conclude, the project can be helpful to various industries all over the world by developing their capabilities, which is the success criteria of the project.

Research Timeline

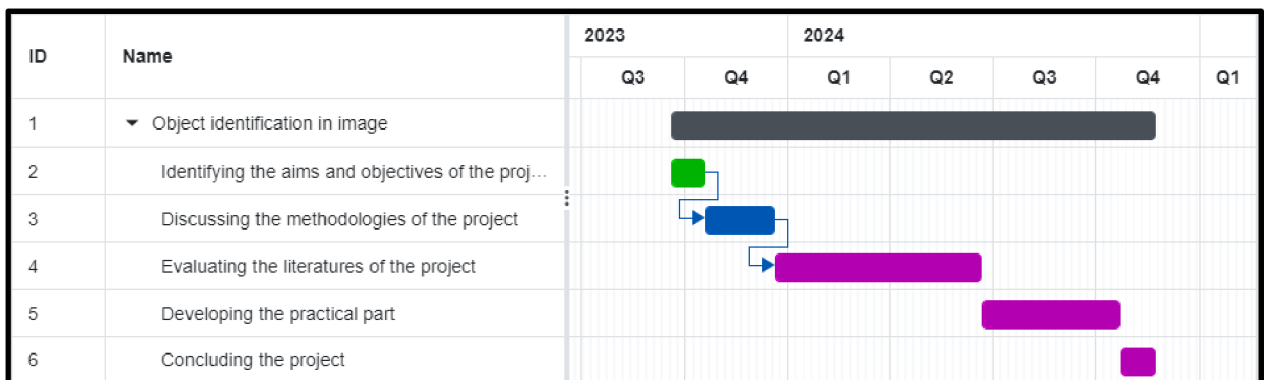


Figure 1: Gantt-Chart

(source: Author)

Summary

This chapter is focused on identifying the aims and objectives of the research, which is the main focus of the project. In addition to this, the chapter is also focused on various methods that are to be used in the project which range programming language that is to be used in the project to the various processes that are to be used in the project. The main goal is to identify the methods that are to be used in the project along with the outcomes all these methods can provide to ensure that the aims and objectives of the project can be met.

Literature Review

Introduction to the chapter

The literature section is centered around the different writings that are related to the ongoing venture. These incorporate the idea of profound learning, CNN, NMS, and JavaScript alongside numerous others. The last segments are centered around significant speculations that are related to the ongoing task and are answerable for supporting the vital writing of the undertaking. The last calculated system is helpful for portraying the exploration through an outline to fulfil the necessities of the task.

Deep Learning

3.2.1 Introduction to Deep Learning

According to the research of Alzubaidi *et al.* (2021), the concept of deep learning is closely associated with machine learning. The purpose is to involve the use of neural networks to solve complex problems. This is due to the architecture of neural networks which emphasizes the structure and function of the human brain by establishing a connection between each of the modules. The end goal is to identify the complex input patterns of the user to solve any particular purpose.

In the opinion of Ahmad *et al.* (2021), deep learning is a subcategory of ML that is generally used in image recognition. This strategy utilizes refined brain networks that are prepared with immense measures of information to distinguish examples and highlights in images. The models created by deep learning can naturally distinguish and order objects in pictures with exceptional precision and speed, making it a powerful device for image identification.

With deep learning, PC frameworks are equipped for identifying and sorting objects in images with practically identical accuracy to that of people, delivering an important innovation for different applications, going from independent vehicles to clinical imaging.

3.2.2 Features of Deep Learning

According to Mathur *et al.* (2020), deep learning is a kind of ML that imitates the human brain to display and resolve complex issues. Innovation has acquired immense prevalence as of late because of its capacity to learn and enhance its own overwhelming information. These are made out of different layers of interconnected hubs that interact and decipher information. Besides, deep learning calculations are equipped for understanding meaning they can gain from unstructured information without express programming.

From the perspective of Moen *et al.* (2019), deep learning has the capability to analyze images the means of technological advancements, empowering machines to perceive and characterize articles and sounds with more prominent exactness than at any time in recent memory. It has likewise taken huge steps in natural language handling, permitting machines to decipher and answer human language. Additionally, deep learning calculations can deal with a lot of information and scale to satisfy the needs of quality applications. They can likewise consequently separate applicable elements from raw information, decreasing the requirement for manual component designing. In general, deep learning is a strong and adaptable innovation that can possibly upset numerous enterprises, like medical care, money, and transportation.

Convolutional neural network

3.3.1 Understanding Convolutional neural network.

According to O'Shea and Nash, (2015), CNN is considered to be an architecture of ANN. CNN is responsible for solving image-driven pattern recognition tasks using simple architecture.

On the other hand, sGu *et al.* (2018), highlights the importance of deep learning the approaches of which are followed by CNN. These include visual and speech recognition along with natural language processing all of which are the result of extensive development. As for CNN, it is responsible for performing layered design, activation functions and optimization all of which are crucial for fast computation that can be beneficial for the project.

3.3.2 Architecture of Convolutional neural network

CNN is also known as the Convolutional Neural Network, which is considered to be a class of neural networks. This specific architecture specializes in analysis of images, through the application of a grid-like topology, which is used to digitize the image that is trying to be analyzed within this specific network. This network follows a specific architecture, where the concerned image is being converted into binary information, which is then passed through different layers to analyze the image after processing it.

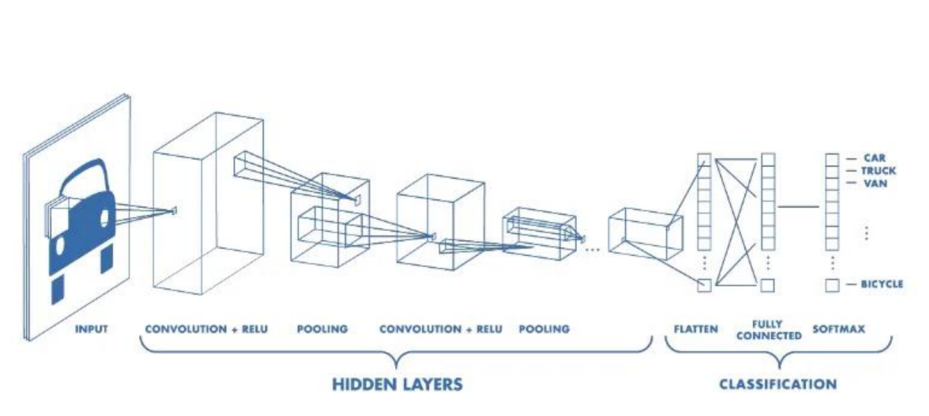


Figure 2: Architecture of CNN

(source: For evaluation from the internet)

This architecture represents the entire process that is implemented within the functions and operations of a CNN and how this specific topology analyses the image that is being fed into the network.

According to Agarap, (2017), the effectiveness of the operational capabilities of CNN is largely dependent on the architecture it follows. To be more specific, the architecture of CNN is mostly the integration of linear SVM (Support Vector Machine) which is responsible for image classification. As a result, this is responsible for identifying the images that belong to different classes of the COCO dataset. To summarize, the architecture of CNN is 2 layered which uses the max pooling model which is helpful for supporting the stages of the project. Lu *et al.* (2017), emphasizes the performance of CNN as it utilizes the flexible dataflow architecture. This is also known as FlexFlow which is responsible for identifying any mismatch by means of feature map and synapse parallelism. In addition to this, it significantly reduces the workload of the design which is 2-10 times faster and 2.5-10 times more efficient. This is largely due to the scalability of the model which is responsible for identifying the objects within the image faster and providing faster output to the bucket.

3.3.3 Application of Convolutional neural network

Sun *et al.* (2022), highlighted some measures in which CNN can be applied in the practical field. This is largely due to the development of computer technology and the ever-increasing usage of machine learning. One such application is associated with image processing by improving the accuracy of image classification. This CNN model can also be used to classify images which are similar in nature and are considered to be complex. The reason behind this is largely due to the usage of the VGG model which provides the usage of batch normalization. This can be used to standardize data and the associated process is considered to be faster for which it plays a viable role in the project. To be more specific, the main purpose of the project is to process large chunks of images as fast as possible which can be helped by the batch normalization process.

Artificial Neural Networks

3.4.1 Overview of Artificial Neural Networks

According to Rosa *et al.* (2020), the ANN is considered to be a model which has been used extensively in the modern era. The foremost reason behind this is associated with optimizing the techniques to speed up the learning algorithm. As a result, it is helpful for regularization of the techniques that are responsible for generating quality data. For this reason, the integration of the model is largely viable for the project as it can help to generate good quality output in an optimized manner. Adding to this, the processes of ANN are also helpful for processing the images as fast as possible which is responsible for supporting the objectives of the project.

In the opinion of Schmidhuber (2015), the ANN is considered to be the application of machine learning that is responsible for recognizing patterns. This can be achievable by two processes which include shallow and deep learning, both of which are considered to be effective for the project.

3.4.2 Application of Artificial Neural Networks

According to Negassi *et al.* (2020), the application of ANN can be reflected in its usage in the medical sector. To be more specific, it can be used to analyze cystoscopic images which are responsible for identifying cancer. Accordingly, it is valuable for opening the degree that permits medical services experts to perform satisfactory therapy measures. This is to a great extent because of the use of AI which is liable for distinguishing malignant growth cells from the human body. Moreover, it is additionally significant for further developing the drug cycle of malignant growth which can be advantageous for people from here on out. The

research performed by Jha *et al.* (2020), generally upheld the use of ANN in the clinical area. One such measure can be recognized as the utilization of Atrous Spatial Pyramid Pooling which is utilized to catch data inside a specific organization. Subsequently, it permits clinical associations to recognize any hint of undesired materials that are available in the human body. It permits the medical services area to give satisfactory therapy measures to further develop the general ailment of the patients.

JavaScript

3.5.1 Understanding JavaScript

According to the research of De Leeuw, (2015), the usage of web technology in the modern era is considered to be extensive. This is where the role of JavaScript materializes and is responsible for the development of such web technology. To be more specific, JavaScript is responsible for providing a lightweight and cross-platform programming platform. As a result, it is considered to be the most effective programming language and is being used worldwide.

JavaScript began as a scripting language to improve web interactivity and has since evolved into a versatile and powerful programming language. This thesis delves into the evolution of JavaScript, its key features, and its role in shaping the modern web development landscape. The research looks at how JavaScript affects user experience, server-side development, and the emergence of new frameworks and libraries. It also investigates the challenges and opportunities that JavaScript presents in terms of security, performance, and cross-browser compatibility.

Client-side frameworks and libraries such as React, Angular, and Vue.js have transformed web development. The role of JavaScript in facilitating the development of single-page applications (SPAs) and the rise of component-based architectures is examined in this section.

JavaScript on the Server

With the introduction of Node.js, JavaScript's influence has expanded to server-side development. This chapter examines the advantages and disadvantages of server-side JavaScript, including its asynchronous programming model. From the perspective of Selakovic and Pradel (2016), the programming platform that is provided by JavaScript has gained popularity due to certain reasons. These include performance issues where the role of optimization plays a major role. In addition to this, JavaScript can also be used regardless

of the operating system along with server-side development. Moreover, it also supports control structures and statements, both of which can be crucial aspects of the project for which it is used for developing the website.

3.5.2 Classes in JavaScript

According to Wagner (2016), JavaScript is considered to be an object-oriented programming language which is used for developing web pages and applications. This is largely due to the usage of classes that are responsible for developing such websites. One such class is associated with the constructor class which is useful for creating and initializing objects within the class. This class is considered to be new and thus does not recognize the browser of Internet Explorer 11 or before. Adding to this, the constructor class is considered to be a crucial feature of ECMAScript 6 for which it is incorporated in the project. Moreover, such classes are also responsible for implementing a data model which is considered to be an essential part of MVC architecture. To summarize, the implementation of such a class is considered to be complex but in exchange is considered to be effective for which it largely suits the objectives of the project.

In the opinion of Wei and Ryder, (2015), another class of JavaScript is known as a first-class function. To be more specific, the first-class functions are considered to be variables in the programming language. This is helpful for assigning the functions to any other variable or can be passed as an argument or can be returned by another function. To be more specific, such a class is responsible for providing flexibility that is responsible for performing the most complex of tasks. Adding to this, the requirement of the project is considered to be complex as it is responsible for processing huge chunks of images where the role of first-class function materializes. In addition to this, it is also responsible for precise analysis of the program which can be beneficial for the project by providing adequate output to the users.

Non-Maximum Suppression (NMS)

3.6.1 Understanding NMS

According to the research of Hosang *et al.* (2017), the NMS is considered to be an algorithm which is responsible for detecting objects. To be more specific, the NMS algorithm emphasizes post-processing which is responsible for merging detections of the same object within an image. The algorithm is considered to be simple but precise. This is due to the importance of post-processing which is responsible for capturing the image in such a manner

that the objects within that can be effectively identified. This is where the role of the bound box materializes which is then responsible for highlighting the objects in the image. In addition to this, NMS can also integrate with the COCO dataset by exploring it with its objects.

As constructed by the research of Rothe *et al.* (2015), the NMS is considered to be a key post-processing stage which is used in many computer applications. To be more specific, such measures are considered to be crucial as it is responsible for using bounding boxes which are crucial for highlighting the objects within an image. However, it is not perfect as in some instances it detects nearby objects for which it can give inaccurate results and thus can lead to faulty output.

3.6.2 Importance of NMS in image processing

According to Hussain and Agarwal (2015), the role of NMS is considered to be effective as it is responsible for processing images. The purpose is to analyze the image and give appropriate output which is the current objective of the project. One such measure is known as detecting the edges of the image even if it does not possess a background. To be more specific, edge detection is considered to be an essential aspect which is crucial for identifying the boundaries of an image. As a result, it can prevent the detection of any objects which are outside the boundary of the image. It helps to maintain the structural property of an image which can act as a base by which the objects are to be detected within the images.

Zhu *et al.* (2020), highlighted another crucial aspect of NMS is associated with identifying the summary of an image which includes identifying the objects within. To be more specific, the NMS provides the benefits of post-processing which is responsible for removing redundancy and low-quality sections of the images. This is due to following the principles of machine learning that are responsible for constant improvement which is responsible for learning the patterns of the objects from the previous images. In addition to this, NMS also uses the principles of various training models to predict the boundaries of an image which helps in maintaining the confidence score of the objects which is crucial in properly identifying the objects.

COCO dataset

3.7.1 Understanding the COCO Dataset

COCO is the acronym for “Common Objects in Context” which is a dataset that contains large-scale image recognition that is essential to increase the accuracy of the comparisons

that are made through the program (COCO, 2023). This dataset consists of more than 330,000 images all of which belong to the 80 object categories and 5 captions that are responsible for describing the scene. The importance of this dataset is reflected in its widespread usage in the research field of computer science. It is used for object detection by dividing it into two main categories while one category is responsible for organizing the images into a directory the other is focused on the annotations of the images which are generally of JSON format. This is the dataset that is to be used in the project from which the images are to be set side by side with that of the uploaded images of the users.

Sharma *et al.* (2021), highlighted another crucial aspect of the COCO dataset which is responsible for classifying the images into various segmentations. In addition to this, the COCO dataset is responsible for identifying the objects where the role of image classification plays a vital role. This is done by following the principles of deep learning which is considered to be an application of machine learning.

3.7.2 Importance of the COCO dataset in the project

As per Caesar *et al.* (2018), the COCO dataset is viewed as powerful on account of example acknowledgement which is a critical achievement component of a venture. More specifically, it permits contrasting a picture from its dataset and the transferred picture of the clients. This is for the most part finished by distinguishing the examples of the items inside the picture. This is finished by distinguishing the items and limiting the classes which is essential for recognizing the sizes and states of the articles that are introduced in the picture. Furthermore, it overlooks non-urgent components inside the pictures, for example, grass and sky which have no sort of shape.

According to the viewpoint of Zhou *et al.* (2017), the viability of the COCO dataset is viewed as critical as it very well may be utilized with the end goal of article recognition. This is because of the compelling utilization of profound learning calculations which are ordinarily utilized for fostering the abilities of the framework. Furthermore, the dataset is additionally liable for utilizing the methodologies of CNN which is answerable for supporting the undertaking by perceiving the examples of a picture.

Theoretical underpinnings

The primary focus of the project is on detecting objects within an image as fast as possible. The dataset is responsible for storing the shapes of certain objects that are to be contrasted which allows the user to identify the presence of an object in that image. This process is

considered to be complex and for this reason, only a single theory cannot be helpful for justifying the research.

Feature-based object detection theory

The first theory of object detection is associated with feature-based which is responsible for detecting an object by recognizing its features. This theory supports the principles of convolutional neural networks for which this theory can be ideal for the project. To be more specific, the theory highlights that the objects which are inside the bounding box can be detected by recognizing their patterns (Chu and Cai, 2018). The reason behind this is the shape of the object that is different from others which enables the algorithms to detect the object with precision. To summarize, such a model has proven its worth in the earlier systems and for this reason this theory is considered to be ideal for the project.

Deep learning-based object detection

The detection of an object within an image can be justified by means of deep learning-based object detection theory. This is due to the usage of deep learning responsible for identifying the objects within an image (Zaidi *et al.*, 2022). Adding to this, the concept of deep learning is considered to be a branch of machine learning for which the theory can be effective for the project. As for the theory, it is responsible for teaching the machines, which is the most complex part of the project. It is largely due to the monotonous nature of the machines as these are developed for performing such jobs. For this reason, the capabilities of object detection models that have been developed over the years are somewhat limited as these are subjected to follow certain patterns and thus the provided output is erroneous. This is where the concept of deep learning plays an effective role which can be used to teach machines about detecting objects more accurately.

Multi-scale object detection

The detection of an object is considered to be a complex endeavor and thus erroneous output can be considered a common factor. This is where the role of multi-scale objection plays a crucial role as the approach utilizes various layers which allow for accurate image detection (Kim *et al.*, 2018). For this reason, the approaches of the theory are considered to be robust and can provide accurate output. Adding to this, the theory is responsible for maintaining the principles of convolutional neural network which is one of the requirements of the project and also justifies the necessity of the theory. Moreover, this theory is responsible for supporting the principles of the COCO dataset and for which it is most suitable for the

project. Despite this, the association of various layers implies that it can increase the processing time of the images which is negligible and cannot impact the project greatly.

Conceptual framework

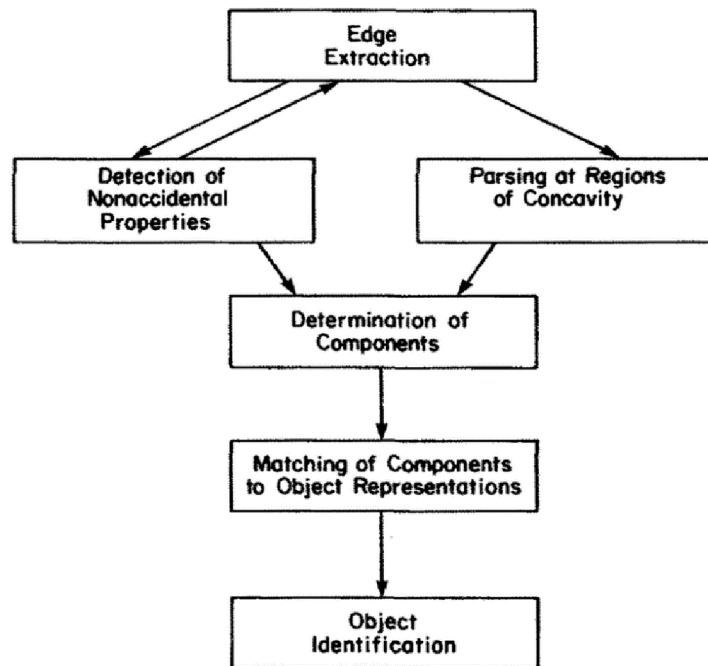


Figure 3: Conceptual framework of object detection

(Source: Morris, 2017)

Object location is viewed as a bunch of cycles that are liable for guaranteeing that an ideal result can be given to the client. This should be possible by separating the edges of the jumping box that is answerable for holding the articles. Regardless of this, typically the jumping box might contain other superfluous components which are not pieces of the article. More specifically, the COCO dataset contains no data in regard to those components for which quitting those components is important. The following stage is essential and is liable for deciding the parts inside a picture which can make the method involved with distinguishing the picture conceivable. This stage is liable for deciding the bounding boxes that contain the item which is available in the COCO dataset. This jumping box can be laid out assuming the item introduced in the picture matches the portrayal of the article which is introduced in the COCO dataset.

Conclusion to the chapter

This chapter is centered around different written works that have been done already about the parts that are to be utilized in the undertaking. It incorporates the programming language that will be utilized in the advancement of the site which can permit the clients to transfer the pictures. This incorporates HTML and JavaScript which can be useful for keeping up with the security of the site and can make the site more advantageous for its clients. Furthermore, the writing additionally incorporates the different systems that are to be utilized simultaneously. These incorporate the systems of CNN and ANN, the two of which support the standards of AI 5 and hence are great for the task. Likewise, writing about the COCO dataset is additionally examined in the part that is answerable for looking at the articles. The last segment of the hypothesis and applied structure features the appropriate speculations and general idea of the venture individually.

Practical Part

The main purpose of the project is to detect objects within the image by exploring them with the COCO dataset. However, the project has certain limitations. This is because of the limitation of available resources because processing many images could significantly slow down the performance of the website. In the current setup, the website allows the users to upload about 100 images at a time or 500 Mb of total uploaded images. This is considering the fact that the capacity of the server is limited to the way in which the website is to be stored. Adding to this, the filtering of images is also responsible for taking a significant amount of time for which the output time can be increased by a significant margin. These limitations must therefore be present for the feasibility of the project.

Configuring the application

The website used for object detection is programmed with JavaScript. On the other hand, HTML is used for creating and structuring the website along with CSS or Cascading Style Sheet that is used for designing the website for this reason, the considered application for the project must be the latest or any other version of Visual Studio that is available on the official website of Microsoft. It is due to the fact that Visual Studio is responsible for supporting all these types of programming languages which can make the project possible. However, integrating the COCO dataset is mostly not supported by Visual Studio. For this

reason, it is necessary to install Microsoft Visual C++ as it allows the installation of COCO API. Moreover, Visual Studio for Windows initially did not support a Machine Learning model for which the training of the model can be done by following certain steps or by using ML.NET. It can be done by using the CNTK model for which Node.JS must be installed, which can be useful for evaluating the CNTK model. To summarize, this configuration of Visual Studio can allow the project to be completed with minimal barriers and this maintains the effectiveness of the project.

Processes for taking input.

The main objective of the input preprocessing is to prepare the input images in a way that is suitable for the pretrained network. It includes various substages, which include resizing the image, cropping the image and improving the quality of the image. The purpose of this is to make the image compatible as all the images may not be of the same size and dimensions (Usmonov, 2021). The second stage is about detecting the objects within the image by using the principles of convolutional neural networks. In the next stage, the detected images are to be classified into various categories. The final stage includes drawing a bounding box around the object or objects which allows for easy identification of the model. However, there are some constraints as some images may be larger than others and thus can contain many objects. Other than this, the complexity and the quality of the image due to distortion may complicate the process which can increase the processing time of the image. In the next stage, the detected images are to be classified into various categories. The final stage includes drawing a bounding box around the object or objects which allows for easy identification of the model. However, there are some constraints as some images may be larger than others and thus can contain many objects. Other than this, the complexity and the quality of the image due to distortion may complicate the process which can increase the processing time of the image.

```
// user input
let images = null;
let objects = [];
```

Figure 4: Taking user input.

(source: Author)

In the case of the current project, the input processing can be divided into various sections all of which play their role. The first input measure is associated with taking the input from

the users which is to be done by two lines of code. The first line is responsible for setting the value of the images as null while the second line is responsible for setting the value of the objects as empty array. The objects variable is responsible for storing the value of the objects that are to be inputted by the user. However, the objects must be presented in the COCO dataset otherwise finding the object is not possible unless the original pretrained ML5 model is extended using the transfer learning. However, the objects must be presented in the COCO dataset otherwise finding the object is not possible unless the usage of transfer learning process of training the network.

```
// load the coco ssd model using ml5 object detector
function preload() {
  coco_model = ml5.objectDetector("cocossd");
  console.log("COCO SSD loaded successfully!");
}
```

Figure 5: Initializing COCO SSD model.

(source: Author)

The second step is loading the ML5 object detector which is accomplished by the preload function. In the case of successful initialization, the console will prompt an output message of successful loading.

exploring with the COCO dataset

The info cycle is liable for taking pictures from the client to guarantee that they can measure up to the COCO dataset. This examination should be possible in different habits, one of which incorporates the consideration of an item locator that is pre-prepared. The distinguishing proof of the item can then be measured up to the articles that are accessible in the COCO dataset (Padilla et al., 2020). It is to be finished by Crossing point over-Association which is answerable for deciding the comparative score.

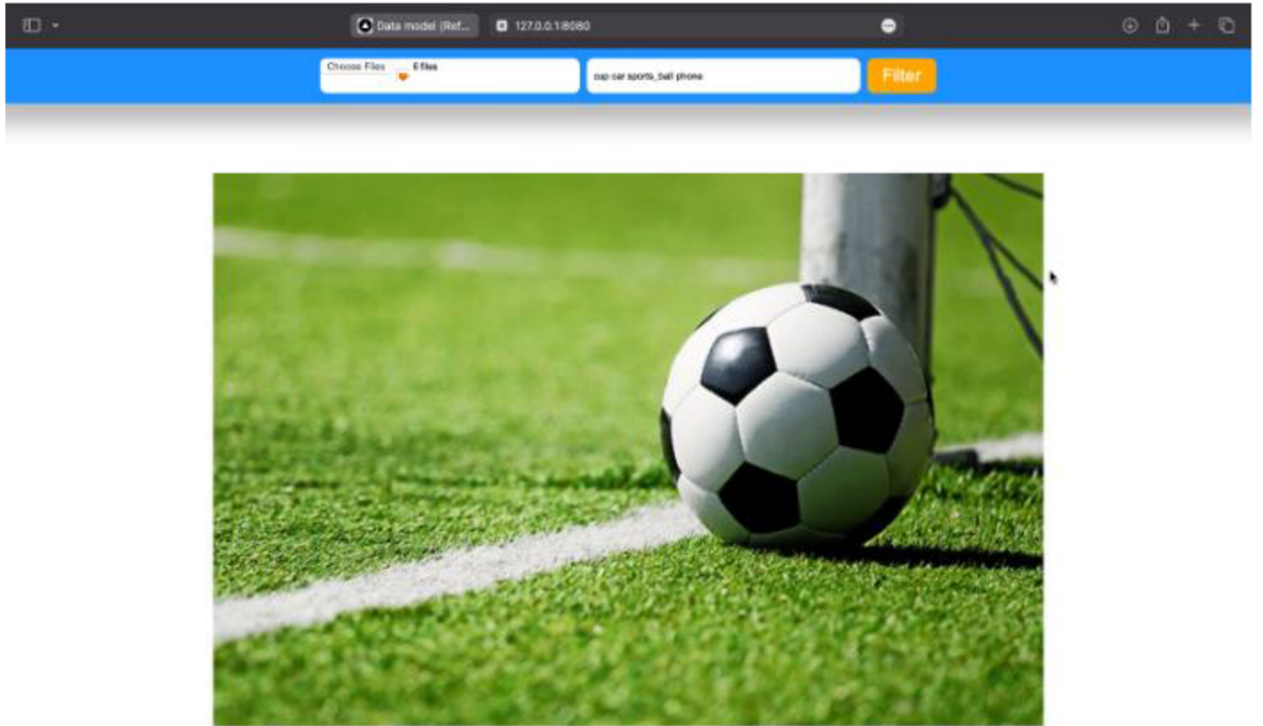


Figure 6: Website Homepage

(source: Author)

Moreover, the COCO dataset is liable for utilizing a semantic division model which is useful for marking every pixel in the item. In addition, the examination should likewise be possible physically which isn't considered for the venture as it is answerable for human connection. On account of the ongoing venture, the use of the COCO dataset suggests that the pre-prepared model is to be utilized for correlation that is liable for performing compelling examinations. In any case, the use of a pre-prepared model might comprise certain difficulties as it can affect the exactness of the item identification process. Notwithstanding this, this approach can be viewed as successful as it is liable for giving the results sooner rather than later.

```

// process the result if any object got detected
function onDetected(error, results, img) {
  if (error) {
    console.error(error);
  }

  // process through all the objects detected in a single image
  for (let i = 0; i < results.length; i++) {

    // if the image contain atleast any one object the image will considered as output
    if (objects.includes(results[i].label)) {

      // create a new img element to show it on the result list
      let x = document.createElement("img");
      x.src = img;

      document.getElementById("all_img").append(x);
      return;
    }
  }
}

```

Figure 7: Loop conditions for object detection

(source: Author)

The comparison with the COCO dataset is the next stage of the project which will be the output of the project. The first condition of the project is about providing an error message in the case the object is not detected in the uploaded image of the user which can be used in the html document. This will lead to the prompt that will show an “error message”. Another condition is detecting all the objects in a single image which will prompt the console to show that image in the result of the image which is to be done by the “document. getElementBYId” code.

Drawing the bounding boxes

The purpose of the bounding box is to ensure that each of the detected objects can be stored in the outlined by a particular box with the highest probability. This helps us understand where the AI sees the object of interest with the highest probability. This is where the role of JavaScript excels as it can help with drawing the bounding box the AI sees objects if matched with the user preference can be displayed. It includes various stages that include creating a canvas element on the page and getting the context of the canvas element (Adhikari and Huttunen, 2021). However, drawing such canvas elements must be done with boxes of various sizes and aspect ratios to ensure that the objects can be stored within the box.

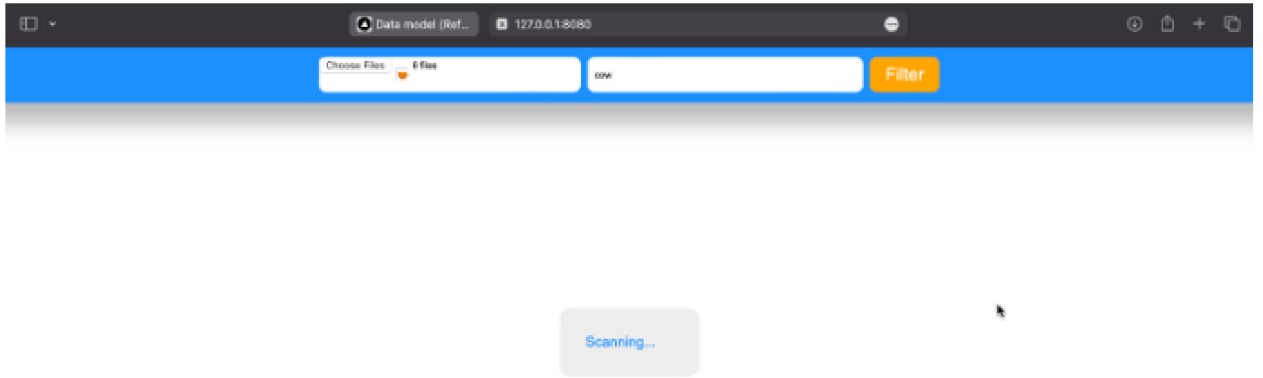


Figure 8: Image scanning on the website

(source: Author)

The determination of the line width can also be done by JavaScript. These processes can be repeated as many times as possible to ensure that several bounding boxes can be drawn. Despite this, it is likely that an image may not consist of more than 5 objects that are to be input by the user.

Filtering the bounding boxes

The creation of the bounding boxes is considered to be one side of the coin while filtering those is the other side. This can be done by the filter function of JavaScript which is responsible for making a callback function with the help of arguments (El Jurdi *et al.*, 2020). The output of this function can be of two types as it can return true or false which can help to filter the bounding boxes more efficiently. To be more specific, the project contains bounding boxes in a variety of sizes. For this reason, the filtering function is necessary which can provide a bounding box for a particular size. The program that has been developed for this project as also addressed a method that could be used to detect the objects as per the size of the images that are present in the dataset and provide output as per the search of the user accordingly.


```
// process the input image using the coco ssd model
function detect(inp, img) {
  if (inp) {
    coco_model.detect(inp, (err, res) => onDetected(err, res, img));
  } else {
    alert("Invalid image source!");
  }
}
```

Figure 9: Processing the uploaded images.

(source: Author)

The filtering of the bounding boxes can also be augmented by processing the image using the COCO SSD model. In this case, the input image contains the specified object it will provide an output with the resolution of the image with the image itself that contains the object. On the other hand, the console will provide an invalid message about “invalid image source”.

Storing the result

The primary output of the project is responsible for providing images that contain certain objects which are to be entered by the user. To store such output the usage of TensorFlow’s object-detection library can be used. One reason for choosing the library is associated with providing a simplistic approach that is useful for storing the data without much hindrance (Rukanskaitė, 2021). In addition to this, the library also provides an easy-to-use API that is responsible for detecting objects within the image.

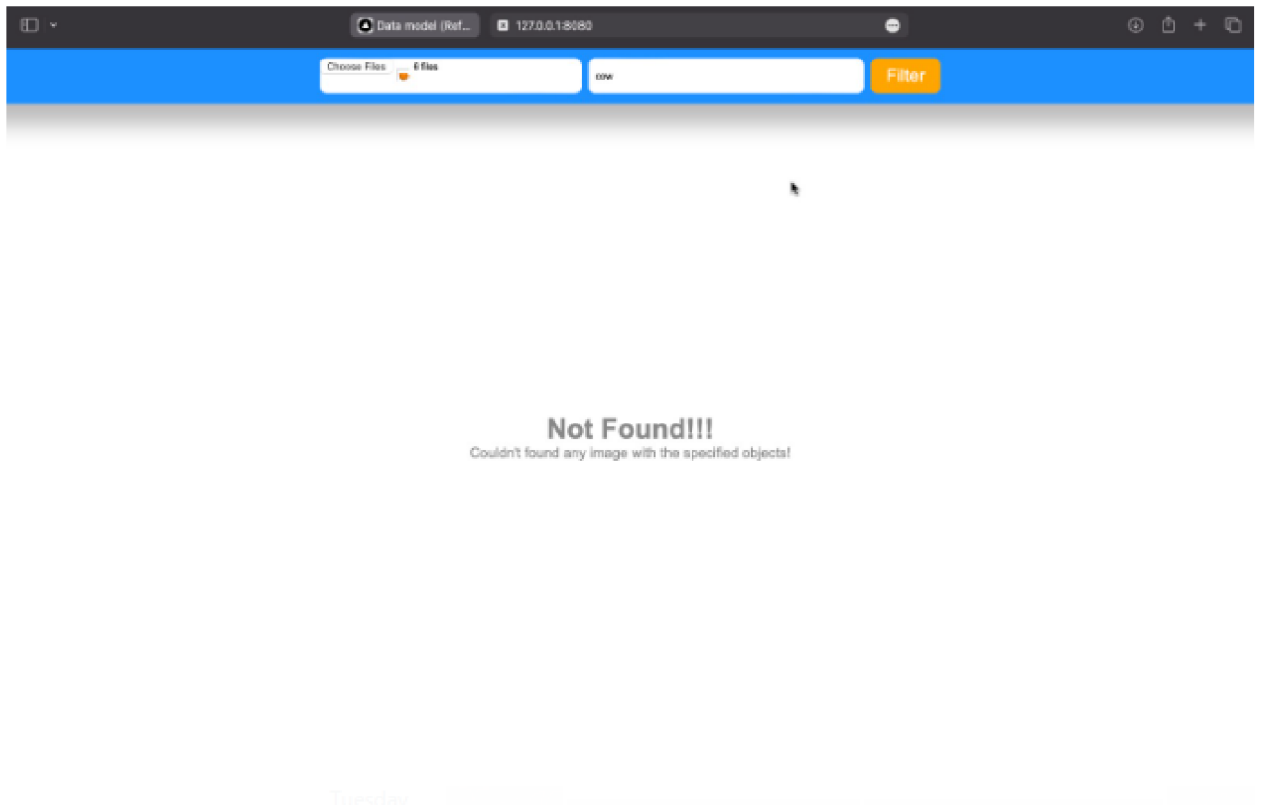


Figure 10: Output shown on the website.

(source: Author)

However, the benefits of the library can be made possible through its proper installation where the role of npm plays a major role. The installation of the library enables the usage of the object detector class which provides a number of methods. These strategies are valuable for recognizing objects inside a picture. The strategy that is helpful for distinguishing the item is known as identifying for which the class is picked for the task to give appropriate results. Moreover, the class likewise gives a save technique that is essential for saving the result in a specific document design. It incorporates JSON and CSV designs which are open through different applications that are accessible on the web.

Methods for providing output

The result of the task incorporates the pictures that contain specific objects. These should be given in such a way that they can be effectively deciphered by the client. The primary phase of the undertaking incorporates bringing in the essential libraries that are significant for giving the result. For example, it incorporates TensorFlow js-object-identification which assists with identifying objects inside the picture in the best way (Javadi et al., 2021). The subsequent stage incorporates making another article locator which should be possible by

the distinguish technique. The following stage is tied in with stacking the model so that articles can be distinguished as quickly as could really be expected. The last stage is tied in with giving pictures that contain explicit items which are attributed by the client. This large number of stages is critical for giving specific pictures that contain the items. These are to be given in JSON or CSV design both of the record configurations can be gotten to by different applications. On account of the ongoing task, TensorFlow article discovery can be helpful as it is answerable for identifying the articles with a Programming interface that likewise upholds these standards of convolutional brain organization. Adding to this, such strategies are liable for supporting JPEG and GIF designs which are two of the most well-known pictures organizes that can be transferred by the client.

```
// filter button click event handler
filterBtn.onclick = () => {

  // atleast one object name is required to further proceede
  if (keywords.value === "") {
    return alert("please enter objects to search");
  }

  // split the input to get different objects from single input
  objects = keywords.value.split(" ");
  document.getElementById("all_img").innerHTML = "";

  // loop through all the inputted images to check one by one
  for (let i = 0; i < imgs.files.length; i++) {
    let file = imgs.files[i];

    if (file) {
      // create a browser file reader that is required by the coco ssd mode
      // to process the image on the client browser environment
      const reader = new FileReader();

      // use the file reader to create a image object
      reader.onload = function (e) {
        const img = new Image();
        img.src = e.target.result;

        // after preparing the image it will start the detection process
        img.onload = function () {
          detect(img, e.target.result);
        };
      };
    };
  };
};
```

Figure 11: Providing the output

(source: Author)

The output is to be provided by taking the objects and images from the user in the means of one text box. The other box is about taking the input of the images from the user by enabling them to upload them. The first condition is about taking at least one object as input from the user otherwise it will prompt the user to enter the objects. This is where the primary

functionality of the loop function materializes and holds various conditions. These include processing the image on the client browser, using the file render and preparing the image to start the detection process.

Implementing the security processes

The project is developed in such a manner that it can provide output to the users in the most timely manner. The first security measure is associated with using a secure connection by using encryption so that images cannot be accessed by any third party. It can be achieved by using a secure connection that is generally provided by an SSL certificate in the form of HTTPS (Oakes *et al.*, 2019). In addition to this, strong authentication can also maintain the security of the website for which the users must be able to register themselves in the website.

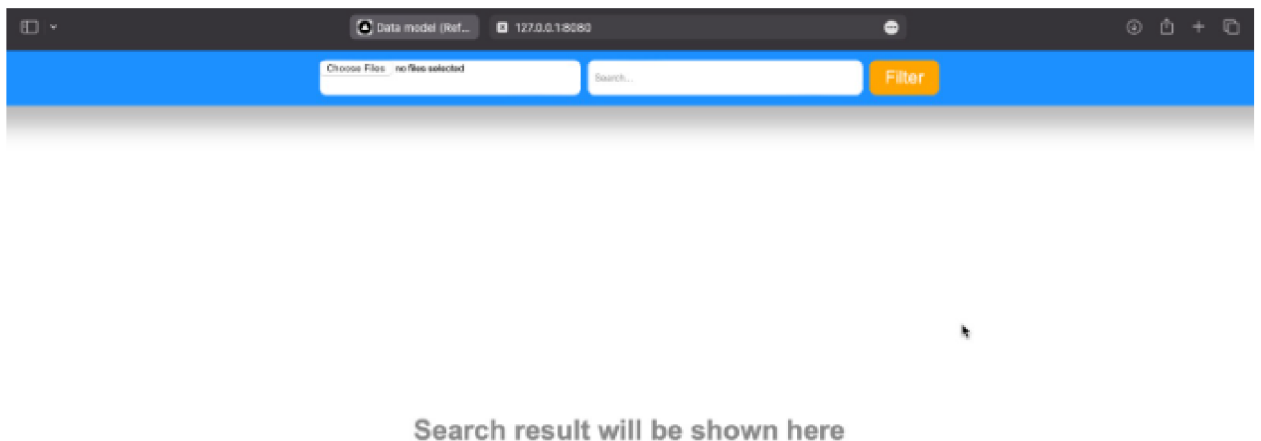


Figure 12: Output tray

(source: Author)

However, the primary concern in this regard is that the users must be provided with the resources that they require for uploading the images and downloading the output to prevent unauthorized data access. Another security measure involves the validation of inputs that are crucial. It can prevent entering malicious codes on the website which can lead to loss of data. Despite all these security measures, the integration of a web application firewall is a necessity as it can prevent any unauthorized incursion. In addition to this, it is also useful for

preventing SQL injection and cross-site scripting which could compromise the security of the website. Moreover, the security features of the firewall must be updated on a constant basis so that any third party cannot benefit by extracting the key features of the website.

Maintaining the performance

The performance of the website is considered to be a key factor that is responsible for ensuring the success of the project. In the current project, the output time of the project is considered to be a key aspect as it can influence the decision of the users about using the project (Rahi *et al.*, 2019). It can be achieved through various factors, one of which includes the use of a lightweight model. Adding to this, the usage of JavaScript allows various object detection models which include Tensorflow.js and provides good accuracy to maintain the performance of the website.

```
nav {
  background: ■ dodgerblue;
  width: 100%;
  height: 64px;
  display: flex;
  align-items: center;
  justify-content: center;
  box-shadow: 0 0 24px 24px □ #0004;
}
```

```
button {
  width: 80px;
  height: 40px;
  background: ■ orange;
  color: ■ white;
  border-radius: 8px;
  border: none;
  font-size: 20px;
  font-weight: light;
}
```

```
.obj-inp {  
  width: 300px;  
  height: 32px;  
  background: #white;  
  border-radius: 8px;  
  margin-right: 8px;  
  border: none;  
  padding: 4px 8px;  
}
```

Figure 13: CSS for website designing

(source: Author)

The maintenance of the website can be managed by designing with the CSS which is also known as Cascading Style Sheet to optimize the website. These can be helpful for designing the properties of the navigation bar, buttons and the input box that will contain the objects. For instance, the background of the navigation bar has a 64-pixel height with a width of 100 per cent which covers all the length of the webpage. This is similar to the case of buttons and input boxes which will store its various properties such as color, margin and padding along with others.

Results and Discussion

Overall outcome of the project

The undertaking is liable for giving specific results which can be useful for the clients as it permits them to handle gigantic lumps of pictures. More specifically, the reason for the task is to guarantee that pictures can be handled in the quickest measure of time. The ongoing task is hoping to accomplish this through a site which is created in such a way that it keeps up with security norms. The gathered pictures are processed following the standards of CNN and ANN design. Additionally, the COCO dataset has been utilized for contrasting the pictures which likewise lessens the heap of the site as the COCO library is viewed as an outer substance.

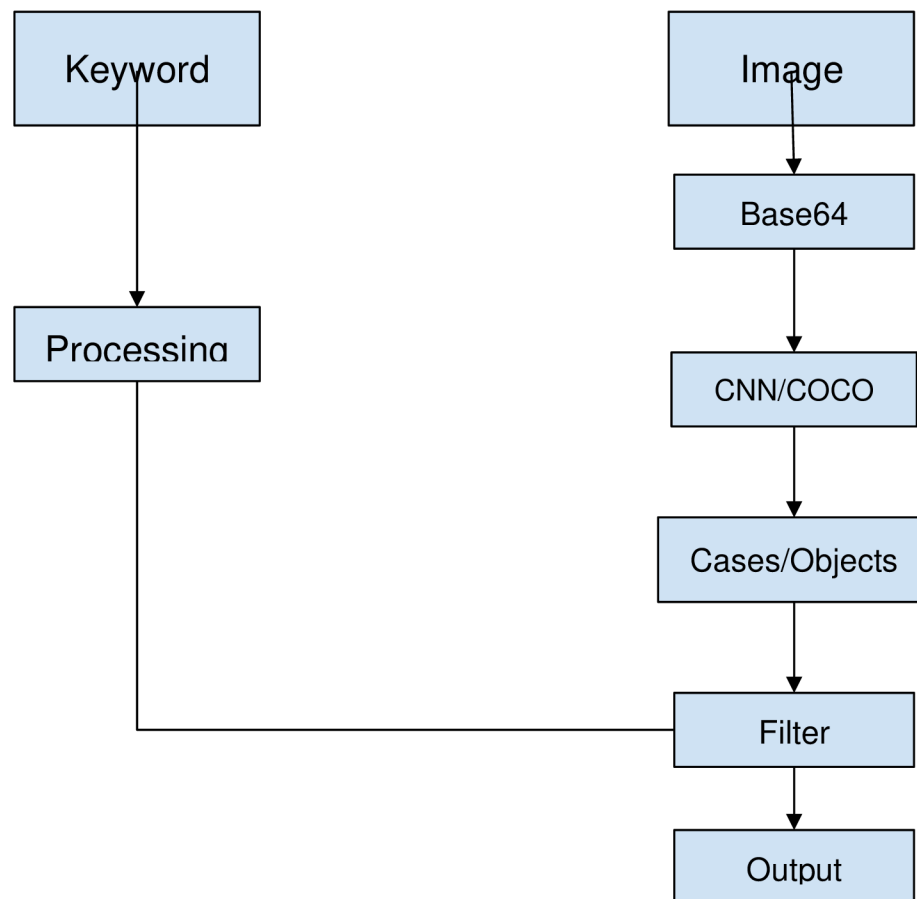


Figure 14: Flowchart

(Source: Author)

The images that are to be uploaded by the user are to be converted by means of base 64 conversion method. It is to be done via COCO SSD/CNN model which uses various classes/objects to do so. Only then the images can be filtered out in which the role of external preprocessing plays a crucial role which is in text format and thus lead to the formation of output.

Purpose of JavaScript in the Project

The entirety of the project is to be done by JavaScript which is used for the purpose of coding and is to be stored in a file. The file is responsible for storing all the coding materials that range from the initialization of the coco model and the HTML element that is to be used for the project to the ultimate provision of the output in a convenient manner (Williams and Moser, 2019). The other aspects of the coding include the initialization of the COCO model which is the main purpose of the project. This will allow the comparison between the objects of the COCO dataset with the entered objects of the users. Another perspective is about the usage of the loop function that is responsible for continuously using the search function until the object is to be found in the image. Only then this can be transferred to the output bucket of the project which can be shown to the user. Otherwise, it will provide an erroneous output to the user.

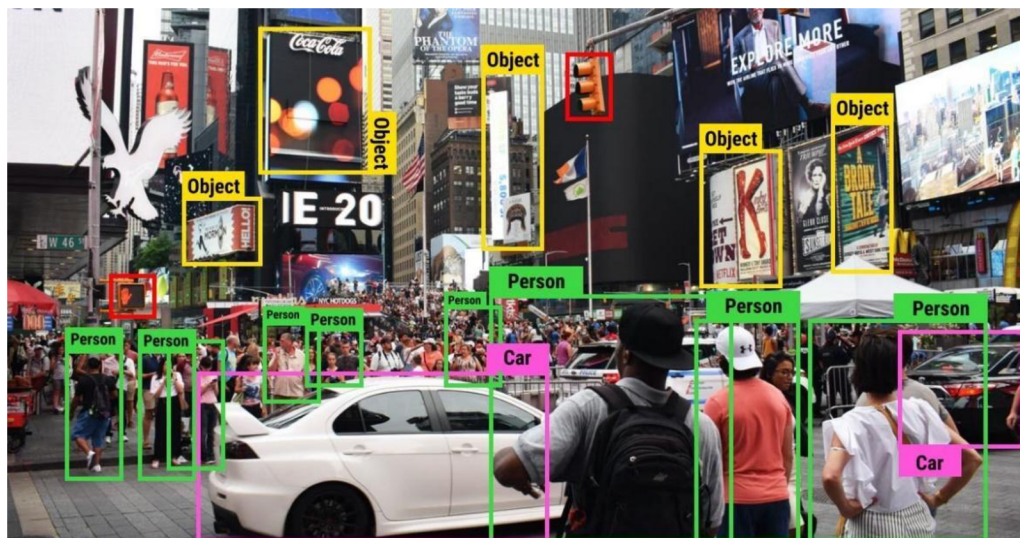


Figure 15: Code Testing

(source: For evaluation from the internet)

The program is developed in such a manner that it can be able to detect objects from an image. The objects must be presented in the COCO data set which is the primary condition of the code. Behind the scenes, the code will identify the objects within the image and

highlight them with certain colors which is based on their properties. For instance, from the above picture, the code will identify the person object based on their appearance which is same for the Car object. The code will also highlight other objects which are not presented in COCO dataset and thus name them as objects which are highlighted in yellow. There are a total of 58 car images. I upload them all and enter the keyword "car". It filters out 50 images. So, the accuracy is 86.2%. Among the images that we used for testing, some of the car images have fancy graphics working on it. That is making some noise on the main object. There are a few weird looking cars that we usually don't see in our regular life. During training time, the lack of this kind of image is currently confusing the model. Some car images are taken from a very unusual angle that is normally used in promotional content. That's not clearly exposing all the car features. Few images contain half portions of the car, not the full car. That's why the model couldn't detect those cars. The solution to this problem is with all those common and perfectly taken photos we should also use some unusual photos for training. Some noise should also be added in a few of the training datasets, that will make the model more flexible. Noise means the program could also add some fancy graphics on a few training images, so that they don't look like a common car. Thus, our model will also learn that behind those fancy graphics it's also a car. Not an unknown object.

Setting conditions with the JavaScript to get input

A crucial aspect of the usage of JavaScript is that it allows comparison among the various entities of the project that are responsible for providing accurate output. This is where the role of the loop function materializes for which it is considered to be an effective element of the project. On the other hand, various conditions are also needed to ensure that the users provide valid input (Bartlett, 2022). This can be helpful for ensuring that users can't be able to enter junk data which could compromise the security of the website. This is due to the development of technology that is responsible for the formation of scripts which can interfere with the security of the website. In addition to this, it is also responsible for validating that the users can only input the name of the objects. In the case a blank output or garbage data is entered it can simply prompt the user to enter the name of the objects properly.

Setting conditions with the JavaScript to get output

The implications of JavaScript are not only limited to getting input from the users but are also responsible for providing accurate output to the users. This is done by using the loop

function that is responsible for continuously exploring the input of the user with the COCO dataset so that the objects can be highlighted for the user. To be more precise, the function is responsible for providing various types of output from the user (Svekis *et al.*, 2021). For instance, it may be the case that an image may contain only a single object from many objects that are entered by the user. This is to be identified and shown by one of the conditions of the loop function. Similarly, another image may contain two/three objects both of which must be identified and shown by the two other conditions. In another instance, it might be the case that an image may not contain any of the entered objects of the image. For this, those images are not part of the output bucket that is to be provided to the users which can be satisfied by another condition of JavaScript.

Development of the project with limitations

The project is developed considering the fact that objects can be detected from huge chunks of images which can be helpful for real-life instances. During the development of the project major issues regarding the importing of the COCO dataset never materialized as it is done by some simple coding process (Linneberg and Korsgaard, 2019). However, the project may have certain limitations which are likely to be improved in the future. One such limitation is associated with the limited number of images that can be currently uploaded. It is due to the fact that the images are certain sizes and due to the server constraint, not more than 100 images or about 500 Mb worth of images could not be uploaded at the moment. Another crucial aspect of the project involves the presence of similar objects in the image which can provide inaccurate output. This improvement can be helpful for increasing the overall accuracy of the project, which can help in real-life examples. In addition to this, the design of the website can also be developed in such a manner that it is responsible for affecting the overall capabilities of the project.

Conclusion

The primary purpose or the goal of the project is to detect a certain number of objects within the image that have numerous applications. To be more precise the project is responsible for exploring the COCO dataset that contains an extensive number of images which are to be considered as the object. The main goal of the thesis is to demonstrate the functions of convolutional neural networks and algorithms for detecting and recognizing objects in images using a prototype application. This was accomplished using the JavaScript-based ML5 library, and pre-trained network model based on the COCO dataset. The application will take a set of images from the user and return the images that contain particular objects. This is to be collated with the entered object of the user which may be one or more than one. In the current project, the number of images is about 100 and the size is worth 500 Mb due to the capacity of the server. It is despite the fact the project is currently running in the local host but is considered to be entered into the live server. For this reason, the project is to be developed in such a manner that it can be too much of a burden to the server. To be more precise, the goal of the project is focused on uploading huge chunks of data along with the processing of those that consume an enormous amount of bandwidth. For this reason, it is necessary that the project can consume a limited number of resources for the user which is crucial for maintaining the performance of the project. This program has received the existing output in this study through analyzing scores of images that were present within the dataset of COCO SSD which were useful in assessing the accuracy of the program that has been written and developed for this project. This has therefore created a blueprint for this study to get the desired output and results through the program developed.

Recommendations

The project covers certain aspects that are responsible for providing the outputs in the fastest time possible. The purpose is to improve certain aspects of the project to ensure the improvement of its overall performance. Nevertheless, the project uses certain measures which can be improved in the future to improve the detection success rate. One such measure is associated with the usage of the COCO dataset which contains about 330 thousand images that can be used for comparison with the user input. However, there are certain datasets that support many more images than the COCO dataset such as ImageNet which supports about

14 million images. The integration of such a model is responsible for improving the scope of the project as it can process many more images that are entered by the user.

Future Scope

Another crucial aspect of the project is the processing of images, which is not sufficient in the context of real-life projects. To be more specific, the project can be improved by incorporating the processing of videos which can be helpful for real-life projects. It can be done by the BDD100k library which has certain advantages when correlated with the COCO dataset. This advantage is associated with processing a certain number of videos, which is around 100 thousand hence the name suggests. Alternatively, the size of the library is about 1.8Tb which is much larger than that of the COCO dataset and thus can be integrated into the project. On the other hand, the project can also be improved by integrating the ImageNet with the BDD100k which can allow the processing of images and videos alike and hence can be applied in real-life projects.

References

- Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Umar, A.M., Linus, O.U., Arshad, H., Kazaure, A.A., Gana, U. and Kiru, M.U., 2019. Comprehensive review of artificial neural network applications to pattern recognition. *IEEE access*, 7, pp.158820-158846.
- Agarap, A.F., 2017. An architecture combining convolutional neural network (CNN) and support vector machine (SVM) for image classification. *arXiv preprint arXiv:1712.03541*.
- Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J. and Ahmad, F., 2021. Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Transactions on Emerging Telecommunications Technologies*, 32(1), p.e4150.
- Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M. and Farhan, L., 2021. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, 8, pp.1-74.
- Arnoldovich, T.A., 2021, September. OVERVIEW SYNTACTIC AND SEMANTIC TOOLS OF HTML AND CSS TO CREATE AN ACCESSIBLE WEB RESOURCE. In *The 2 nd International scientific and practical conference—Topical issues of modern science, society and education*|(September 5-7, 2021) SPC—Sci-conf. com. ual, Kharkiv, Ukraine. 2021. 624 p. (p. 132).
- Bartlett, J., 2022. Basic JavaScript Syntax. In *Programming for Absolute Beginners: Using the JavaScript Programming Language* (pp. 119-132). Berkeley, CA: Apress.
- Caesar, H., Uijlings, J. and Ferrari, V., 2018. Coco-stuff: Thing and stuff classes in context. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1209-1218).
- COCO, 2023. Common Objects in Context. [online]. COCO. Available at: <https://cocodataset.org/#home> [Accessed 25 September 2023].
- De Leeuw, J.R., 2015. jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior research methods*, 47, pp.1-12.
- Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J. and Chen, T., 2018. Recent advances in convolutional neural networks. *Pattern recognition*, 77, pp.354-377.

Hosang, J., Benenson, R. and Schiele, B., 2017. Learning non-maximum suppression. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4507-4515).

Hrechuk, D. and Olishevskiy, I.H., 2022. Autopilot technology in vehicles.

Hussain, Z. and Agarwal, D., 2015. A comparative analysis of edge detection techniques used in flame image processing. International Journal of Advance Research In Science And Engineering IJARSE, (4), pp.3703-3711.

Janiesch, C., Zschech, P. and Heinrich, K., 2021. Machine learning and deep learning. Electronic Markets, 31(3), pp.685-695.

Jha, D., Riegler, M.A., Johansen, D., Halvorsen, P. and Johansen, H.D., 2020, July. Doubleu-net: A deep convolutional neural network for medical image segmentation. In 2020 IEEE 33rd International symposium on computer-based medical systems (CBMS) (pp. 558-564). IEEE.

Khan, Z.I. and Shandilya, V.K., 2021. Effect Of Blob And String Data Type On Time And Space Consumption In Recognition Based Image Authentication Scheme. Webology (ISSN: 1735-188X), 18(6).

Larabi-Marie-Sainte, S., Alskireen, R. and Alhalawani, S., 2021. Emerging applications of bio-inspired algorithms in image segmentation. Electronics, 10(24), p.3116.

Li, Z., Liu, F., Yang, W., Peng, S. and Zhou, J., 2021. A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems.

Linneberg, M.S. and Korsgaard, S., 2019. Coding qualitative data: A synthesis guiding the novice. Qualitative research journal, 19(3), pp.259-270.

Lu, W., Yan, G., Li, J., Gong, S., Han, Y. and Li, X., 2017, February. Flexflow: A flexible dataflow accelerator architecture for convolutional neural networks. In 2017 IEEE International Symposium on High Performance Computer Architecture (HPCA) (pp. 553-564). IEEE.

Malik, R.S., Patra, J. and Pradel, M., 2019, May. NL2Type: inferring JavaScript function types from natural language information. In 2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE) (pp. 304-315). IEEE.

Mathur, P., Srivastava, S., Xu, X. and Mehta, J.L., 2020. Artificial intelligence, machine learning, and cardiovascular disease. Clinical Medicine Insights: Cardiology, 14, p.1179546820927404.

Mikael Fernandus Simalango Post author July 30, 2022. What Object Categories / Labels Are in Coco Dataset? [online]. Amikelive. Available at: <https://tech.amikelive.com/node-718/what-object-categories-labels-are-in-coco-dataset/> [Accessed 22 September 2023].

Millan, J., Phakamad, B., Rhodes, S. and Tipon, C., 2022. Instacart-Accessibility usability evaluation of grocery delivery website for the visually impaired.

Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M. and Van Valen, D., 2019. Deep learning for cellular image analysis. *Nature methods*, 16(12), pp.1233-1246.

Negassi, M., Suarez-Ibarrola, R., Hein, S., Miernik, A. and Reiterer, A., 2020. Application of artificial neural networks for automated analysis of cystoscopic images: a review of the current status and future prospects. *World journal of urology*, 38, pp.2349-2358.

Ni, J., Chen, Y., Chen, Y., Zhu, J., Ali, D. and Cao, W., 2020. A survey on theories and applications for self-driving cars based on deep learning methods. *Applied Sciences*, 10(8), p.2749.

O'Shea, K. and Nash, R., 2015. An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458.

Rosa, J.P., Guerra, D.J., Horta, N.C., Martins, R.M., Lourenço, N.C., Rosa, J.P., Guerra, D.J., Horta, N.C., Martins, R.M. and Lourenço, N.C., 2020. Overview of artificial neural networks. *Using artificial neural networks for analog integrated circuit design automation*, pp.21-44.

Rostianingsih, S., Setiawan, A. and Halim, C.I., 2020. COCO (creating common object in context) dataset for chemistry apparatus. *Procedia Computer Science*, 171, pp.2445-2452.

Rothe, R., Guillaumin, M. and Van Gool, L., 2015. Non-maximum suppression for object detection by passing messages between windows. In *Computer Vision—ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part I 12* (pp. 290-306). Springer International Publishing.

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. *Neural networks*, 61, pp.85-117.

Selakovic, M. and Pradel, M., 2016, May. Performance issues and optimizations in javascript: an empirical study. In *Proceedings of the 38th International Conference on Software Engineering* (pp. 61-72).

Sharma, D.K., 2021, March. Information Measure Computation and its Impact in MI COCO Dataset. In *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)* (Vol. 1, pp. 1964-1969). IEEE.

Sun, Y., Zheng, W. and Ren, Z., 2022, April. Application of convolutional neural network in image processing. In International Conference on Multi-modal Information Analytics (pp. 375-383). Cham: Springer International Publishing.

Svekis, L., van Putten, M. and Percival, R., 2021. JavaScript from Beginner to Professional (Vol. 544). Packt Publishing.

Wagner, G., 2016, December. Introduction to simulation using JavaScript. In 2016 Winter Simulation Conference (WSC) (pp. 148-162). IEEE.

Waldman, A.E., 2019. Power, process, and automated decision-making. Fordham L. Rev., 88, p.613.

Weerts, H.J.P. and Pechenizkiy, M., 2022, March. Teaching responsible machine learning to engineers. In Proceedings of the Second Teaching Machine Learning and Artificial Intelligence Workshop (pp. 40-45). PMLR.

Wei, S. and Ryder, B.G., 2015. Adaptive context-sensitive analysis for JavaScript. In 29th European Conference on Object-Oriented Programming (ECOOP 2015). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Williams, M. and Moser, T., 2019. The art of coding and thematic exploration in qualitative research. International Management Review, 15(1), pp.45-55.

Wirfs-Brock, A. and Eich, B., 2020. JavaScript: the first 20 years. Proceedings of the ACM on Programming Languages, 4(HOPL), pp.1-189.

Yang, Z., Liu, S., Hu, H., Wang, L. and Lin, S., 2019. Reppoints: Point set representation for object detection. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 9657-9666).

Zhou, X., Gong, W., Fu, W. and Du, F., 2017, May. Application of deep learning in object detection. In 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS) (pp. 631-634). IEEE.

Zhu, W., Lu, J., Li, J. and Zhou, J., 2020. Dsnet: A flexible detect-to-summarize network for video summarization. IEEE Transactions on Image Processing, 30, pp.948-962.