Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

PROJECT REPORT
on

Smart Interactive Marketing using Machine Learning and Image Processing

submitted by

Name	SAP No.
1) Darshil Modi	60002190025
2) Divya Jain	60002190036
3) Moulik Shah	60002190065
4) Parth Shah	60002190077

under the guidance of

Prof.Sunil Karamchandani

Associate Professor

DEPARTMENT OF

ELECTRONICS AND TELECOMMUNICATION ENGINEERING

Academic Year : 2022-2023

Shri Vile Parle Kelavani Mandal's

Dwarkadas J. Sanghvi College of Engineering

Plot no. U-15, JVPD Scheme, Bhaktivedanta Swami Marg,
Vile Parle (W), Mumbai – 400 056

Shri Vile Parle Kelavani Mandal's
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC Accredited with "A" Grade (CGPA : 3.18)

Department of Electronics and Telecommunication Engineering

This is to certify that the Project Report Stage – II

"_Smart Interactive Marketing using Machine Learning and Image Processing_"

Submitted by:

1. **Darshil Modi**
2. **Divya Jain**
3. **Moulik Shah**
4. **Parth Shah**

Students of **Electronics and Telecommunication Engineering** have successfully completed their **Project Stage – II** required for the fulfillment of **SEM VIII** as per the norms prescribed by the **University of Mumbai** during the First half of the year 2023. The project report has been assessed and found to be satisfactory.

Internal Guide **External Guide**

Head of Department **Principal**

Internal Examiner **External Examiner**

Table of Contents

Chapter No.	Title	Page no.
	Declaration	
	Abstract	
	List of Figures	
	List of Tables	
1	**Introduction**	7
2	**Literature Review**	10
3	**Theory**	13
	Technologies Used	13
	Working Principle	18
	Targeting: Object Detection	18
	Dataset	20
	Block Diagram	20
	Description	24
4	**Results**	32
5	**Application**	33
6	**Project Budget**	34
7	**Future Scope**	35
8	**Conclusion**	31
9	**References**	38

DECLARATION

We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantialextent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Signature

Name

SAP ID:

Date

Abstract

The evolution of consumer-focused marketing strategies is becoming a significant part of the commerce mix. We are in an age of digital transformation that will change the online viewing experience, and further the adoption of technology and data capabilities for marketing solutions. Today, consumers expect a personalized and visually inspiring experience from brands. Shoppable content is the newest trend amongst digital advertisers and continues to rapidly grow in popularity from online ads to social media. While shoppable ads can support sales goals, interactive video takes ads one step further to support all aspects of your marketing campaign.

***Keywords* –** shoppable ads, artificial intelligence, machine learning, e-commerce, object recognition, classification, marketing, computer vision

List of Figures

Fig No	Figure Name	Page No
1	Mean Average Precision	17
2	Cutout Images	19
3	Model Images	19
4	ResNet50 Image Features	20
5	ResNet50 Model Architecture	21
6	VGG-16 Image Features	21
7	VGG-16 Model Architecture	22
8	User Interface Flowchart	23
9	Image Features Graph of ResNet50 and VGG-16	24
10	Fashion Apparel	27
11	Fashion Apparel	27
12	Fashion Apparel	28
13	Web Application	28

List of Tables

Sr. no	Table Name	Page no
1	Project Budget	34

1. Introduction

The epidemic created a large level of disruption to our day-to-day activities, one of which was an increase in the quantity of business performed online and purchases made over the internet. During the height of the pandemic, the number of people making purchases and business transactions online increased significantly. "E-commerce's proportion in global retail trade grew from 14% in 2019 to roughly 17% in 2020," stated the United Nations Conference on Trade and Development (UNCTAD). This indicates a considerable increase when compared to the rate of 14% recorded in the previous year. According to the findings of the study, certain markets have seen an increase in the amount of internet purchasing that is up to fifty percent higher than it was previously.

Zhang, T., Tang, Z., and Han, Z. (2022) also conducted a study, and in their article, they analyse the best online channel structure for multinational corporations, taking into account the impact of live streaming purchasing. The authors present a mathematical model that can assist businesses in determining the most effective channel layout for their online sales operations. This model takes into account the number of channels as well as the types of channels that should be utilised. The model takes into account a wide range of aspects, including customer behaviour, the level of competition, and the cost of operating numerous channels. The writers also share insights into the design and implementation of live streaming shopping, including how to effectively utilise this channel to drive sales and engage with customers. Specifically, the authors discuss how to effectively use this channel to drive sales and engage with customers. In general, the essay offers helpful insights into the ideal online channel structure for international companies and emphasises how important it is for those companies to incorporate live streaming shopping into their sales plans.

A "shoppable ad" is a type of advertisement that consists of a picture that represents products and includes tags (or markups) as well as the capability to complete a purchase. Customers have the ability to view the item's price by moving their mouse cursor over the product tag that is embedded in the image; after doing so, they are given the option of adding the thing to their shopping basket. This takes place the vast majority of the time.

Any company's content integration plan needs to include a combination of shoppable video and interactive intelligence in order to be effective. Any content integration plan must necessarily include this as one of its components. Marketers are coming to the realisation that executing a consumer interaction strategy that is tailored to their target demographic leads to a superior and more relevant user experience, which, in turn, enhances both brand awareness and brand loyalty. This is something that marketers are beginning to realise.

The Interactive Advertising Bureau (IAB) and its Digital Video Centre of Excellence collaborated to carry out a study that was given the title "The Interactive Ad Effect: CTAs in Mobile Video Shoppable Advertising." According to the findings of this study, video advertisements that allow users to interact with products and drive brand lift are successful methods for catching the attention of consumers. When a call to action (CTA) is included in a shoppable advertisement, customers are more likely to pay attention to it and give it more thought. This lays the groundwork for a direct relationship between the advertiser and the customer, which in turn leads to an increase in sales.

According to the findings of a study that was conducted by Koike, E., and Itoh, T. (2015), user attention is drawn more to interactive advertisements that include a call-to-action (CTA). In general, the paper sheds light on the ways in which psychological considerations can be included into the layout of search interfaces in order to improve the efficiency of those interfaces.

The research also shows that viewers pay the most attention to an interactive advertisement in the first five seconds of the ad, particularly if the advertisement includes a call to action (CTA). This highlights how important it is for marketers to present their brand by making a powerful statement right at the beginning of the commercial when they are trying to get people interested in it.

Consumers now have the ability to interact specifically with any item of interest featured inside a movie thanks to object-level recognition, which was previously unavailable to them. In the past, one did not have access to this capability. While this is going on, marketers are gaining access to more complex optimisations, which enable them to delve deeper into each creative object and topic that engages critical consumers.

Anand, M., Kalra, R., and Gupta, A. (2020) conducted a study that provides a complete evaluation of computer vision approaches for smart retailing. The authors describe the ways in which computer vision may be used to analyse visual data such as photographs and videos to improve a variety of areas of commerce. These aspects include inventory management, the customer experience, and marketing. This article presents an overview of the various computer vision techniques that can be utilised in the retail industry. These techniques include object identification, facial recognition, and visual search, among others. The authors also examine the difficulties as well as the potential benefits of working on this subject, and they offer some insights into possible future study avenues. In general, the paper highlights the potential of computer vision techniques in improving different parts of retailing and provides useful insights into the application of these techniques in smart retailing. Additionally, the article illustrates the potential of computer vision techniques to improve numerous areas of retailing.

Consumers are able to immediately complete a transaction when they use content that can be shopped. This technique accomplishes one of the most significant goals of content marketing, which is to provide answers to questions that customers have at the appropriate time and in the right location. It not only provides a solution to the problem, but it also enables the user to watch a film while concurrently adding items to a shopping cart, which is a very useful combination of features. At the very least, you should be taken to the product page for the item in question and given the opportunity to continue shopping at the same establishment.

The development of shoppable content has made it a great deal simpler for customers to choose and buy products from a wide variety of retailers. The introduction of shoppable content has made this level of ease significantly more accessible. In the not-too-distant future, when it comes to locating, evaluating, and purchasing things from any medium, this strategy will be the one that each and every client favours the most. This is due to the fact that doing everything in this manner will be the simplest and most time-efficient option.

2. Literature Review

E-commerce has seen a rise in the popularity of live-streaming purchasing, and researchers Jie Cai, Donghee Yvette Wohn, Ankit Mittal, and Dhanush Sureshbabu (2018) have investigated the reasons why customers participate in this activity. Live-streaming buying has become a prominent trend in the industry. According to their research, there are primarily two types of reasons people watch live shopping streams: utilitarian and hedonistic. Hedonic motivations are those that are related to the pleasurable experience of buying, such as entertainment and social contact. On the other hand, utilitarian motivations refer to the more pragmatic reasons for purchasing, such as the requirement for knowledge and the desire for ease.

In their 2011 paper, Hamed S. Neshat and Mohamed Hefeeda suggest a system that is based on machine learning and is designed to deliver effective advertising in online videos. The SmartAd system does an analysis of the user's activity and interests, then employs techniques from machine learning to determine which advertisements are most pertinent to display. The authors of the study carried out studies to analyse the efficacy of the system and discovered that, in comparison to conventional approaches to advertising in terms of user engagement and click-through rate, it performed significantly better.

A visual similarity-based interactive product recommendation system for online shopping has been proposed by Jen-Hao Hsiao and Li-Jia Li (2014). This system makes use of image processing techniques. The system uses image processing techniques to extract visual elements from product photos. It then makes recommendations to users based on their browsing history and input regarding products that are visually comparable to those recommendations. The system that is being proposed includes an interactive element that gives consumers the opportunity to submit input on products that have been recommended. This feedback is then used to refine any future recommendations. The authors of the study carried out experiments to evaluate the efficacy of the system and discovered that it surpassed conventional techniques of collaborative filtering in terms of the accuracy of recommendations and the level of user satisfaction. The interactive product recommendation system that is based on visual similarities gives customers a more interesting and personalised experience when they are shopping online.

Due to the fact that interactive video advertising is still in its infant phases of development, this project is of the utmost importance. With the use of artificial intelligence, a piece of software known as Lens AI can quickly transform any photograph or video into a shoppable moment. There are a lot of parallels to be seen between the platform and our concept. A comprehensive and aggregate analysis of audio, visual, and textual information is carried out, in addition to the implementation of marketing technology that positions and shows advertisements that are specifically targeted to the user. These are merely some of the services that can be obtained by using it.

Several studies have been conducted to investigate the application of computer vision and other related technologies within the scope of intelligent interactive marketing. Liu and Li (2021) offered an approach that was based on picture recognition for smart interactive marketing, whereas Anand, Kalra, and Gupta (2020) provided a complete overview of computer vision techniques for smart retailing. Li et al. (2021) conducted a review that focused on the role that smart retailing plays within the context of smart interactive marketing. In addition, Zhang et al. (2020) did a review of smart interactive marketing and its future research directions, and de Pelsmacker, Geuens, and Anckaert (2018) investigated the notion of interactive marketing in the digital age. Both of these studies can be found here. Kietzmann, Bergvall-Kreborn, and Hargitai (2018) gave insights into smart interactive marketing through the use of social media in their last contribution.

Another such site is called MaxTap, and it provides users and companies with interactive shoppable advertisements while they watch videos. The operation of this platform is based on an AI model that recognises frames of interest coupled with object annotations from videos, a computer vision model that recognises the visual pattern of an object, and a complicated search algorithm that performs matching from numerous catalogue photographs.

GOLIVE is an AI-driven platform that brands and retailers can use to enhance engagement and conversions with shoppable video content, receive real-time data regarding the performance of their live shopping shows and the habits of their audience, and consistently deliver an amazing user experience. GOLIVE can do all of these things for brands and retailers.

Zhang, H., Li, X., Chen, Y., and Li, H. (2020) conducted a study that gives a review of smart interactive marketing and suggests future research possibilities in this subject. The writers explore the idea of "smart interactive marketing," which refers to the utilisation of technology such as artificial intelligence and big data in order to increase the efficiency of marketing operations. In this article, several different applications of smart interactive marketing are discussed. These applications include personalised marketing, marketing via social media, and location-based marketing. The authors also share insights on the opportunities and problems present in this sector, in addition to prospective future research directions. In general, the essay offers insightful information regarding the function of technology in contemporary marketing and draws attention to the possible advantages of engaging in smart interactive marketing.

All of the aforementioned platforms are pretty helpful and serve a significant purpose in making the purchasing experience for customers more pleasant and stress-free. However, our project is an advancement in the area since it not only integrates the elements that are considered to be the most useful from the aforementioned platforms into a single project, but it also adds features that are considered to be incredibly vital in order to make the user experience as convenient and joyful as it possibly can be.

3. Theory

The entirety of our software consists of three distinct components: the frontend, the backend, and the machine learning component. ReactJS is a component of the front-end technology, Python (Flask and Rest_API) is a component of the back-end technology, and MongoDB and PostgreSQL are components of the database.

Technologies used for the project:

(1) ReactJS: -

The React.js framework is an open-source JavaScript framework and library developed by Facebook. It's used for building interactive user interfaces and web applications quickly and efficiently with significantly less code than you would with vanilla JavaScript.

In React, you develop your applications by creating reusable components that you can think of as independent Lego blocks. These components are individual pieces of a final interface, which, when assembled, form the application's entire user interface.

Reacts primary role in an application is to handle the view layer of that application just like the V in a model-view-controller (MVC) pattern by providing the best and most efficient rendering execution. Rather than dealing with the whole user interface as a single unit, React.js encourages developers to separate these complex UIs into individual reusable components that form the building blocks of the whole UI. In doing so, the ReactJS framework combines the speed and efficiency of JavaScript with a more efficient method of manipulating the DOM to render web pages faster and create highly dynamic and responsive web applications.

(2) Python: -

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy-to-learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms and can be freely distributed.

Often, programmers fall in love with Python because of the increased productivity it provides. Since there is no compilation step, the edit-test-debug cycle is incredibly fast. Debugging Python programs is easy: a bug or bad input will never cause a segmentation fault. Instead, when the interpreter discovers an error, it raises an exception. When the program doesn't catch the exception, the interpreter prints a stack trace. A source-level debugger allows inspection of local and global variables, evaluation of arbitrary expressions, setting breakpoints, stepping through the code a line at a time, and so on. The debugger is written in Python itself, testifying to Python's introspective power. On the other hand, often the quickest way to debug a program is to add a few print statements to the source: the fast edit-test-debug cycle makes this simple approach very effective.

(3) MongoDB

MongoDB is a document database with the scalability and flexibility that you want with the querying and indexing that you need.

- MongoDB stores data in flexible, JSON-like documents, meaning fields can vary from document to document and data structure can be changed over time
- The document model maps to the objects in your application code, making data easy to work with
- Ad hoc queries, indexing, and real-time aggregation provide powerful ways to access and analyse your data
- MongoDB is a distributed database at its core, so high availability, horizontal scaling, and geographic distribution are built-in and easy to us

- MongoDB is free to use. Versions released prior to October 16, 2018, are published under the AGPL. All versions released after October 16, 2018, including patch fixes for prior versions, are published under the Server-Side Public License (SSPL) v1.

(4) SQL

Structured Query Language (SQL) is a standardized programming language that is used to manage relational databases and perform various operations on the data in them. Initially created in the 1970s, SQL is regularly used not only by database administrators, but also by developers writing data integration scripts and data analysts looking to set up and run analytical queries. The term SQL is pronounced sequel. SQL is used for the following: modifying database table and index structures; adding, updating, and deleting rows of data; and retrieving subsets of information from within relational database management systems (RDBMS) -- this information can be used for transaction processing, analytics applications and other applications that require communicating with a relational database. SQL queries and other operations take the form of commands written as statements and are aggregated into programs that enable users to add, modify or retrieve data from database tables. A table is the most basic unit of a database and consists of rows and columns of data. A single table holds records, and each record is stored in a row of the table. Tables are the most used type of database objects, or structures that hold or reference data in a relational database. Other types of database objects include the following: Views are logical representations of data assembled from one or more database tables. Indexes are lookup tables that help speed up database lookup functions. Reports consist of data retrieved from one or more tables, usually a subset of that data that is selected based on search criteria. Each column in a table corresponds to a category of data -- for example, customer name or address -- while each row contains a data value for the intersecting column. Relational databases are relational because they are composed of tables that relate to each other. For example, a SQL database used for customer service can have one table for customer names and addresses and other tables that hold information about specific purchases, product codes and customer contacts. A table used to track customer contacts usually uses a unique customer identifier called a key or primary key to reference the customer's record in a separate table used to store customer data, such as name and contact information.

(5) JavaScript

JavaScript is a programming language that is commonly used to add interactive elements to web pages. It is a client-side language, meaning that it runs directly in the user's web browser and does not require any additional software or server-side processing. JavaScript is often used to create dynamic content and user interface elements, such as drop-down menus, image galleries, and interactive forms. It is also commonly used in conjunction with other technologies, such as HTML and CSS, to create modern web applications.

JavaScript is a dynamic computer programming language. It is lightweight and most commonly used as a part of web pages, whose implementations allow client-side script to interact with the user and make dynamic pages. It is an interpreted programming language with object-oriented capabilities.

JavaScript was first known as **LiveScript,** but Netscape changed its name to JavaScript, possibly because of the excitement being generated by Java. JavaScript made its first appearance in Netscape 2.0 in 1995 with the name **LiveScript**. The general-purpose core of the language has been embedded in Netscape, Internet Explorer, and other web browsers.

The ECMA-262 Specification defined a standard version of the core JavaScript language.

- JavaScript is a lightweight, interpreted programming language.
- Designed for creating network-centric applications.
- Complementary to and integrated with Java.
- Complementary to and integrated with HTML.

Open and cross-platform

The front end operates independently, while our backend API collaborates with our Machine Learning Model and engages in conversation with other APIs through the use of GET/POST/PATCH requests. The following components make up our user interface: Login, Registration, Home, Movies, and TV Shows, and Navigation Menu. Within the Multiplayer option is a Shopping Cart Section, which enables us to compile a wish list of things based on the contextual advertisements that we view.

In the Machine Learning Section, we developed an Image Retrieval System. The fundamental idea behind this system is to assess how similar a query is to the data and features that are already stored in the database, and then to return the items that are the most relevant to the query. When we use an image retrieval system, we extract every aspect of the photos from the database and preserve it. After that, when an image query is received, its features are extracted, and a comparison is made between those features of the query image and the features of the photos in our database to establish the degree of similarity. Finally, the user is shown with the top n photos that are connected to their search.

Images are analysed in order to extract features from them. There are three different CNN architectures that we are able to evaluate and contrast: ResNet 50, VGG 16, and Xception. The performance metric known as Mean Average Precision @ K is the one that is utilised to evaluate and contrast the three models. It is a representation of the overall average Precision at K for all searches. The definition of precision at K is the number of relevant items in the top K items divided by the number of things in the top K.

```
map_resnet = np.mean(top_10_simiarity_scores_resnet)
map_resnet

0.4127090911938036

[86] map_vgg = np.mean(top_10_simiarity_scores_vgg)
map_vgg

0.4938162150096625
```

Fig-1: Mean Average Precision

From the results above we can decode that the ResNet model had 41% similarity when tested while the VGG model had 49% similarity when tested so we can clearly say that in this case the VGG model performed better in comparison to the ResNet model.

Working Principle

Any picture or video may be turned into shoppable moments by artificial intelligence in a matter of seconds. Our Platform enables the monetization of web traffic by connecting advertisers and publishers for the purpose of achieving mutually beneficial outcomes. This is accomplished through the facilitation of interactive in-image or in-video buying experiences that are spread across the internet. Any graphic content can be converted into an internet store as part of our mission here. Our platform is able to add value to each ad impression by using intelligent product identification and advertisements that are contextually appropriate. In its most basic form, our venture is the provision of a service. An application for the web will be developed utilising React, JavaScript, and various other APIs. MongoDB, which will also serve as the backend for the project, will be used to manage the data collecting and keep it updated. When a user logs into our Web App, they are given the option to stream any video that is stored on our servers. While the movie is playing, the machine learning-based recommender will hunt for the same type of goods available on multiple websites such as Amazon or Flipkart using the backend data, which will be issued a unique identifier. Twilio and WebRTC will be utilised in order to achieve the goal of merging videos sourced from a variety of platforms. Our one-of-a-kind AI model examines videos in order to identify potentially intriguing frames and label things. A computer vision model has the ability to recognise the visual pattern of an object. Using a sophisticated search algorithm, millions of catalogue photos are compared and matched. Taking into consideration the ever-increasing popularity of OTT platforms, social media, and online shopping, our solution makes use of computer vision models to provide customers with an interactive and straightforward shopping experience. Thanks to our artificial intelligence, users may discover what their favourite celebrities are wearing and using as they are watching the show. Our one-of-a-kind AI algorithm searches through videos for interesting frames and annotates the items that are relevant. Product Tagging: Our product recognition system enables quick product tagging of things that are either an exact match, visually similar, or complement detected products. These items can be tagged as having one of three attributes: perfect match, visually similar, or complement. Contextually Relevant Ads: Our one-of-a-kind associative advertising algorithm taps into the thoughts of people by forming the required connection between the brand and discovered things that are relevant to it. This allows us to serve ads that are contextually relevant.

A. Targeting: Object Detection

Target audience via pictures and videos they consume and serve them relevant ads on their desktop and mobile websites.

At the very exact moment when the content catches the audience's eye, our platform automatically tags detected objects with products from the brand's assortment that are visually like the detected ones.

Fig-2: Cutout Images

Fig-3: Model Images

We here also referred to a study carried out by P. F. Felzenszwalb, R. B. Girshick, D. McAllester and D. Ramanan (2010). The authors propose a framework for representing objects as a collection of parts and learning to classify each part individually. The method is evaluated on several standard datasets, and the results demonstrate that it achieves state-of-the-art performance in object detection. The article also provides insights into the design of part-based models and the learning algorithms used to train them. Overall, the article contributes to the development of more accurate and efficient object detection methods for computer vision applications.

B. Dataset

Till now we have taken 2 datasets into consideration, FarfetchListings which is a dataset of 180k luxury fashion products, this dataset will help in Creating some Fashion GANs, Price prediction based on the image, and Logo/brand detection. The other Dataset we worked with is Deep Fashion Dataset is a large-scale clothes database for Clothing Category and Attribute Prediction. This dataset contains 289.222 diverse clothes images from 46 different categories.

C. Block Diagram

1) Machine Learning

1. ResNet 50

```
image_features_resnet
 175592: array([[0.0053292 , 0.00871991, 0.0031104 , ..., 0.00054069, 0.03176047,
         0.0066896 ]], dtype=float32),
 2843: array([[0.00957112, 0.03670508, 0.00109281, ..., 0.0328091 , 0.00145665,
         0.01942545]], dtype=float32),
 70219: array([[0.011113  , 0.00489337, 0.01447277, ..., 0.00393461, 0.0107916 ,
         0.01450461]], dtype=float32),
 31995: array([[0.0045037 , 0.00621989, 0.00019922, ..., 0.01236122, 0.00380718,
         0.08717059]], dtype=float32),
 72462: array([[0.00775348, 0.00414437, 0.        , ..., 0.01634505, 0.01586486,
         0.04887246]], dtype=float32),
 34324: array([[0.01192532, 0.02448948, 0.01585401, ..., 0.00385921, 0.01255545,
         0.00319375]], dtype=float32),
 174560: array([[0.00801292, 0.00294437, 0.00053329, ..., 0.0142288 , 0.00825194,
         0.08920507]], dtype=float32),
 48684: array([[0.00296431, 0.00283637, 0.        , ..., 0.0260499 , 0.0073208 ,
         0.06547643]], dtype=float32),
 128392: array([[0.00954026, 0.01775294, 0.0003269 , ..., 0.0122135 , 0.00338945,
         0.04803375]], dtype=float32),
 51238: array([[0.01144404, 0.01073372, 0.00125966, ..., 0.00471686, 0.00018374,
         0.00830228]], dtype=float32),
 169088: array([[0.00957264, 0.01463745, 0.        , ..., 0.023917  , 0.00135024,
         0.02210595]], dtype=float32),
 85630: array([[0.        , 0.00054493, 0.06815649, ..., 0.00572339, 0.00771748,
         0.00276129]], dtype=float32),
 65086: array([[0.00641859, 0.0092345 , 0.00309291, ..., 0.02417375, 0.00542802,
         0.0659291 ]], dtype=float32),
```

Fig-4: - ResNet50 Image Features

Residual Network is referred to as ResNet. A deep network with 50 layers is called ResNet 50. It contained a unique component, or unique connection, known as the skip connections, which made it possible for it to train these densely layered models quickly. The inputs are immediately connected to the output layer in a skip connection. The network is then compelled to simulate h(x)-x, where x is the input and h(x) is the hypothesis function. Since we are modelling the residuals, this process is known as residual learning (errors). ResNet is also available in deep networks with 152 layers and 34 layers.

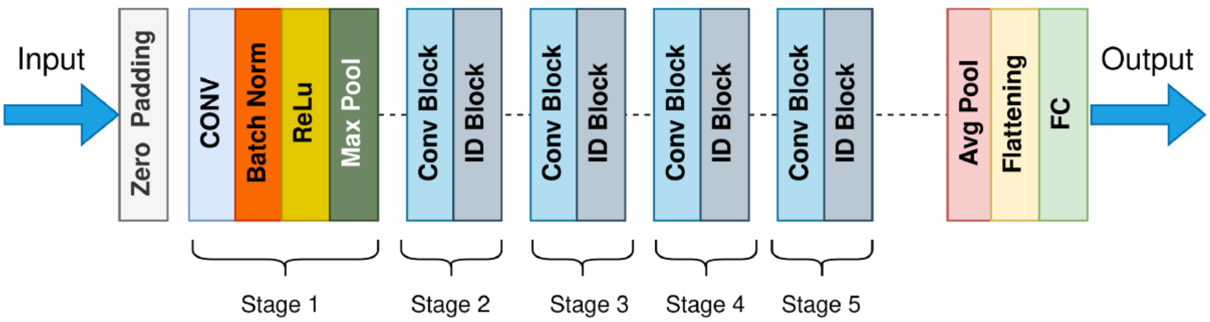

Fig-5: ResNet50 Model Architecture

2. VGG 16

```
image_features_vgg

{532: array([[0.        , 0.012774, 0.        , ..., 0.        , 0.        , 0.        ]],
      dtype=float32),
 130463: array([[0.         , 0.         , 0.00516033, ..., 0.         , 0.         ,
        0.04651381]], dtype=float32),
 135152: array([[0.         , 0.         , 0.00602107, ..., 0.00173344, 0.         ,
        0.01610303]], dtype=float32),
 34231: array([[0., 0., 0., ..., 0., 0., 0.]], dtype=float32),
 175592: array([[0.00165546, 0.         , 0.02455219, ..., 0.         , 0.         ,
        0.00875581]], dtype=float32),
 2843: array([[0.         , 0.00881973, 0.         , ..., 0.         , 0.         ,
        0.         ]], dtype=float32),
 70219: array([[0.02763467, 0.02661736, 0.01692335, ..., 0.02713561, 0.         ,
        0.04910623]], dtype=float32),
 31995: array([[0.         , 0.00169285, 0.         , ..., 0.         , 0.         ,
        0.01719502]], dtype=float32),
 72462: array([[0.         , 0.         , 0.         , ..., 0.         , 0.         ,
        0.02062305]], dtype=float32),
 34324: array([[0.         , 0.02363858, 0.00909978, ..., 0.         , 0.         ,
        0.         ]], dtype=float32),
 174560: array([[0.         , 0.         , 0.         , ..., 0.         , 0.         ,
        0.03290765]], dtype=float32),
 48684: array([[0.         , 0.         , 0.         , ..., 0.         , 0.         ,
        0.03846119]], dtype=float32),
 128392: array([[0.         , 0.02765273, 0.         , ..., 0.         , 0.         ,
        0.01016376]], dtype=float32),
 51238: array([[0.         , 0.03626903, 0.         , ..., 0.         , 0.         ,
```

Fig-6: VGG-16 Image Features.

A relatively simple architecture. Consists of 2 or 3 convolutional layers followed by a pooling layer, then again 2 or 3 convolutional layers and a pooling layer. The VGG architecture reaches up to either 16 or 19 layered networks depending on the variant. (In this case, we have used VGG 16, i.e., the 16-layered variant.

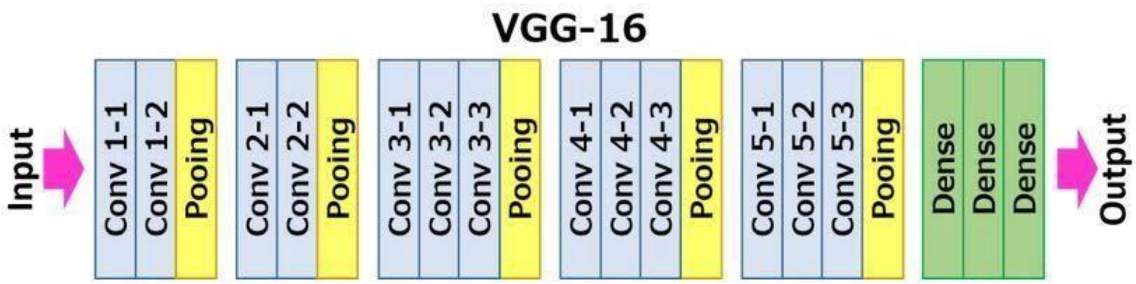

Fig-7: VGG-16 Model Architecture

2) Development

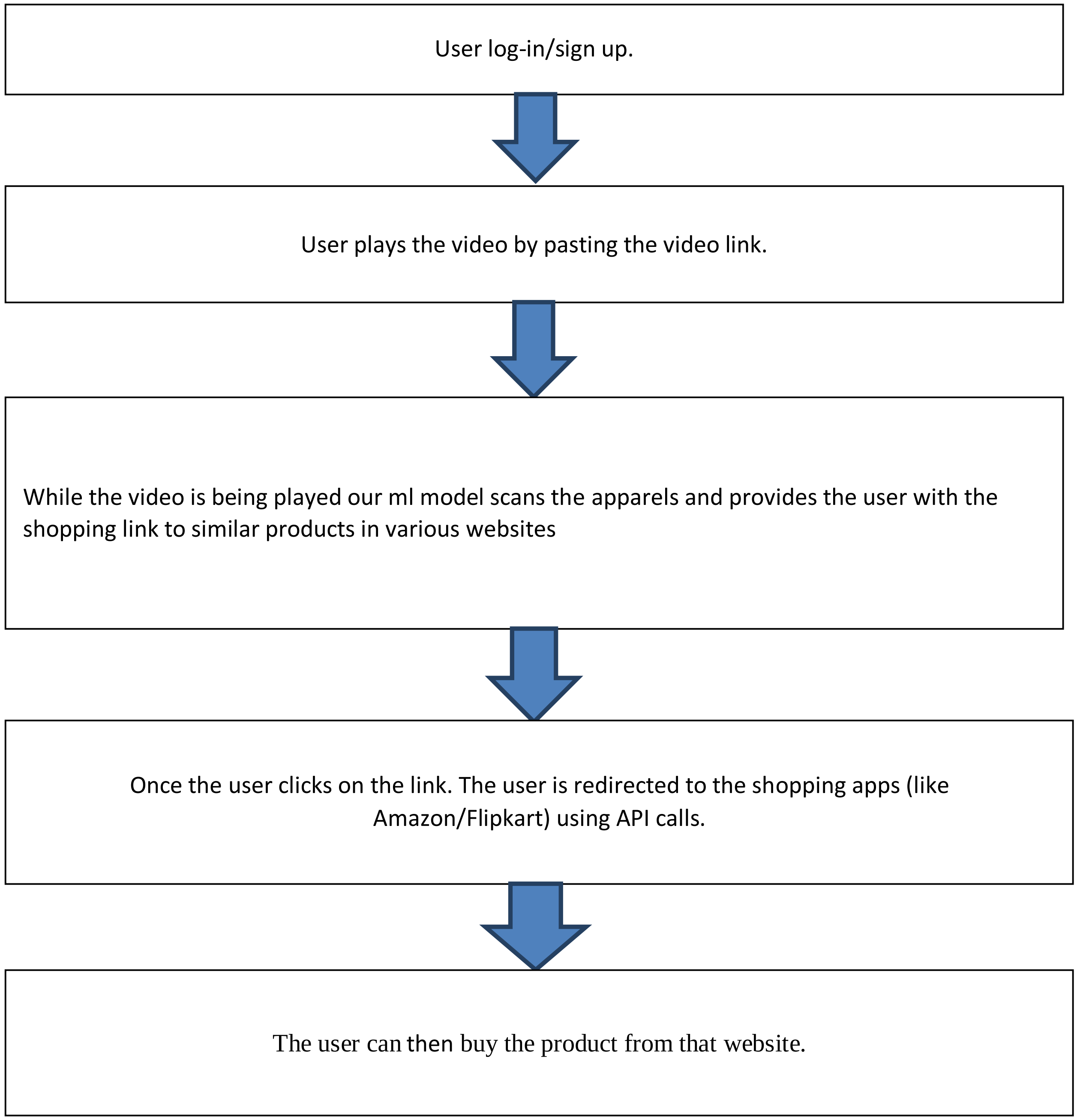

Fig-8: - User Interface Flow Chart

Description

(A) Machine Learning

(i) <u>Model Training</u>

Convolutional Neural Networks, Transfer Learning, ResNet50, and VGG 16 Architectures are some of the ones we deal with in this section to determine image similarity. First, we will perform database maintenance and investigate our dataset. After that, we develop a category for the extraction of features and the identification of images that are most like to one another. The architecture that we use is VGG-16 with ImageNet weights, and the architecture that we use for ResNet 50 is ImageNet weights. Images of 224 and 244 are used as input for the VGG 16 and ResNet 50 models, and we then pre-process these images in the appropriate manner. The next step is to put our feature extractor through its paces using 10,000 images.

Calculate the similarity between the query features and all of the other images in image_features_vgg and image_features_resnet by using Euclidean Distance. After that, we plot the results from all three models and prepare a comparison.

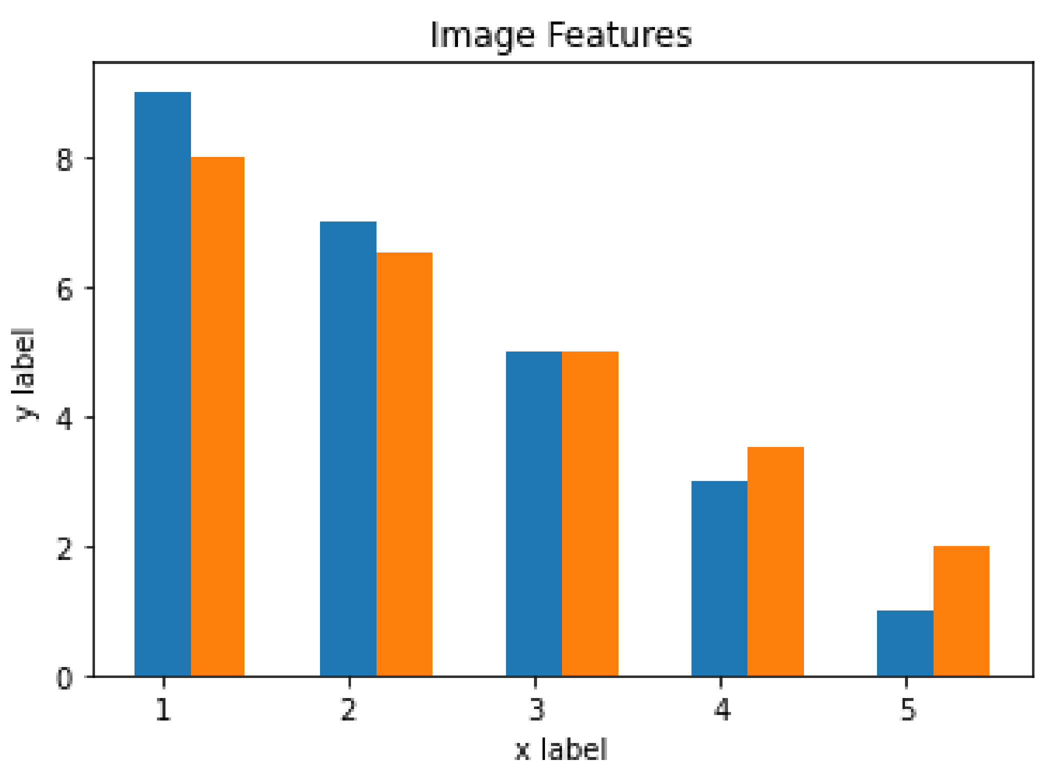

Fig-9: Image Features graph of ResNet50 and VGG-16

We have displayed the Histogram of the Image Features of ResNet 50 and VGG-16 in the graph that is located above. The graph that is coloured blue represents ResNet 50, and the graph that is coloured orange represents VGG-16. The graph illustrates the dictionary values of the characteristics that relate to their respective arrays.

In this alternate approach, we will attempt to differentiate between a number of different types of apparel. It is clear that the initial classes in the ImageNet dataset that were used to train the ResNet34 model are distinct from the classification categories that we have employed. In order for transfer learning to take place, it is necessary to replace the final layer of the network with a new linear layer that contains the same number of activations regardless of the classes that are present in our dataset. Because there are 46 different types of apparel in our scenario, the new layer has the same number of activations. The newly added layer uses randomization to do the initialization of the weights and does not include any pre-trained weights. As a consequence of this, prior to being trained, our model will produce random data; nevertheless, this does not imply that the data produced is entirely arbitrary. The weights of the original model will be maintained for all of the other layers, and all of the layers will have a strong ability to recognise common visual concepts such as gradients and basic geometric forms. For this reason, when we are fine-tuning our model so that it can distinguish between various kinds of clothing, we only freeze the top layer of the network. If we do this, we will be able to optimise the weights of the lowermost layer without having any impact on the weights of the levels that are deeper.

The research that was conducted by Yan, Y., Wang, S., Zhang, M., and Guo, L. (2018), which focuses on the application of visual recognition and recommendation approaches in smart interactive marketing, served as a reference for our work. The authors describe the ways in which these strategies might be utilised to improve the efficacy of marketing efforts by analysing visual data such as photographs and videos. This article presents an overview of the various visual identification and recommendation approaches that can be utilised in marketing, such as the classification of images and videos, the detection of objects, and content-based recommendation systems. The authors also examine the difficulties as well as the potential benefits of working in this subject, and they offer some insights into possible future study avenues. In general, the paper focuses on the potential of visual recognition and suggestion techniques to improve the efficacy of marketing efforts and provides helpful insights into the application of these techniques in smart interactive marketing. In addition, the article highlights the potential of these approaches to improve the effectiveness of marketing activities.

(ii) Learning Rate

a. Learning Rate Finder:

Now let's go through the data in our Data Loader and gradually increasethe learning rate with each mini batch to observe how the value of loss changes with the change in learning rate. Our goal is to find the most efficient learning rate that will allow the network to converge faster. This point is the steepest slope of the loss curve. The points of extremums (min and max) and flat parts of the curve correspond to the learning rates that do not allow the network to learn, as the loss at these points does not improve.

b. Discriminative Learning Rates:

After training all layers of the network we need to review the learning rates again, as after a few batches of training with a relatively high learning rate the old learning rate is not appropriate anymore and likely needs to be decreased as the learning slows down.
The Top-3 accuracy of our model is 88.6%, which is 6% higher than the benchmark accuracy.The Top-5 accuracy of our model is 94.1%, which is 4% higher than the benchmark accuracy.

c. Evaluation of a User-Specified Dataset:

Finally, we'll examine the model's performance with user-supplied photos. I used the phone's camera to take 98 photos of my personal attire. Let's load the photographs and see if the model can accurately classify them. First on the user-specified data, the model's accuracy is 62%, which is lower than on the Deep Fashion Dataset. It is still effective for a 46-class classification model, though. The user dataset contains photographs that are very dissimilar to the ones used to train the model. For instance, while photographs from the Deep Fashion dataset show a human wearing an item, user images only show a clothing item, making it more difficult to scale the clothing. Since it is challenging for a machine to determine trousers' lengths in relation to the human body, almost all of the pants in the user dataset were categorized as shorts. However, the model taught the fundamental ideas and could be applied in a range of settings related to fashion.

In the Pictures below we are taking different images as an input like Shoe, Heel and Dress and with our trained model we recommend similar products of different brands, and this helps us in classifying the different types of similar products available and works as a recommendation system for us. These modelscan then be taken into consideration for classification with Deep Fashion Model and then be used for optimal match between the two products and categories and then integrate it with our custom API for calling it in Development part.

Fig-10: Fashion Apparel

Fig-11: Fashion Apparel

Fig-12: Fashion Apparel

(B) Development

We created a web app in the react video player that allows users to play any video by providing the URL. The user can go to our web app and enter the URL of the video they want to watch. Our webapp then loadsthe video and displays it on our user interface.

There are several things that could be done to improve and expand upon this web app. Some potential ideas include Adding support for different video formats and protocols, such as MP4, WebM, and RTSP,to ensure that users can play a wider range of videos. Implementing search and discovery features to helpusers easily find and access specific videos, such as by keyword or category. Adding social features, suchas the ability to share videos with friends and post comments, to create a more engaging and interactive user experience.

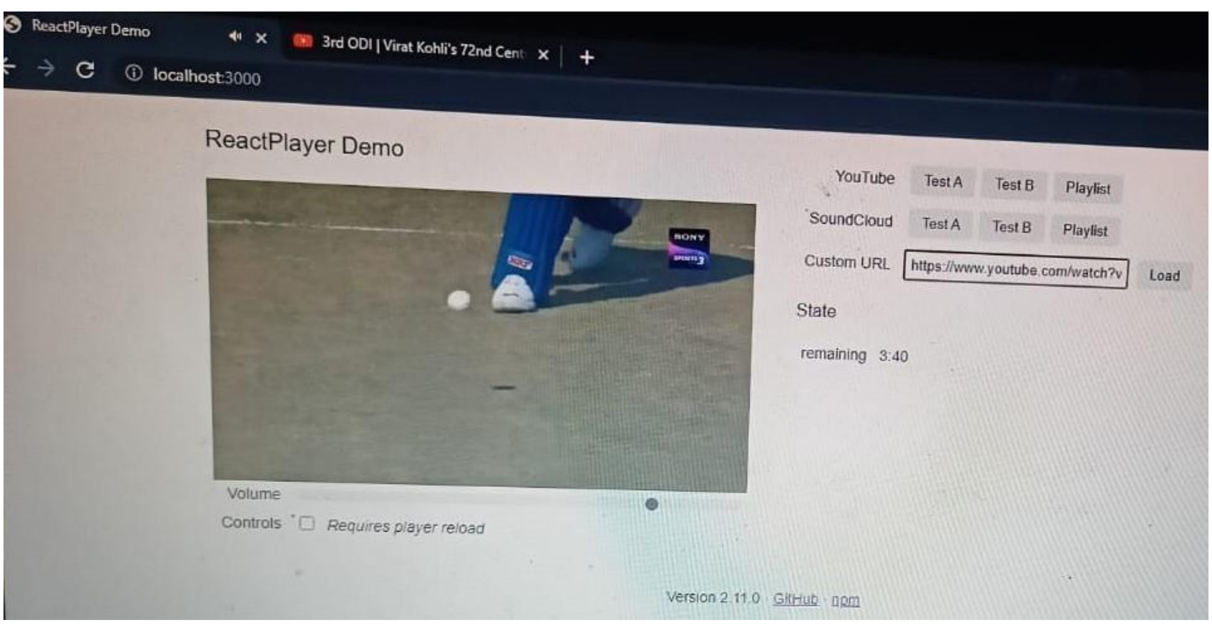

Fig-13: Web Application

We also referred to a study that was carried out by A. S. Rao, S. Bhat and N. D. Chandavarkar(2017) where this article discusses the development of a computer vision application that aims to improve online shopping. The authors describe the design and implementation of the application, which uses computer vision techniques to analyze product images and extract information such as color, size, and texture. The application then uses this information to provide users with more accurate and relevant search results. The article also presents the results of user testing, which showed that the application improved the overall shopping experience for users. Overall, the article demonstrates the potential of computer vision techniques in improving online shopping and provides insights into the design and implementation of such applications.

Integrating with other video platforms, such as YouTube and Vimeo, to allow users to access and play videos from these sources within your app. Developing mobile apps for iOS and Android to allow users to access and play videos on their mobile devices. Implementing support for live streaming, to allow users towatch and interact with live video content. Adding support for different languages and regional settings, tomake the app accessible to users around the world.

We developed a video playing web app made on React. It is a highly interactive and user-friendly application designed for seamless video playback. The app is built using React.js which allows for efficient updates and rendering of components. The user interface is intuitive, making it easy for users to navigate and find the videos they want to watch. The
app is also optimized for performance, with fast load times and smooth playback. The user can login to our site and play any video of his/her choice. Users can play any video by simply providing the video link in our video section. While the video is being played our ml model provides the user with the product link. Our machine learning model uses computer vision and image recognition technology to identify the products in the video and display them to the user in real-time.

The user can view information about the product, such as its price, product description on various shopping applications. They can also add the product to their shopping cart for later purchase. Moreover, the user can also filter the product results based on various criteria, such as brand, category, price range, and more. This allows them to quickly find the products they are interested in and make informed purchasing decisions. The product information displayed to the user is regularly updated to ensure accuracy and relevance. This enhances the user experience by providing them with relevant information about the products they see in the video.

We have integrated our ml model with our front-end using REST API. The REST API enables seamless integration of the machine learning model and the website's front-end. The front-end can send API requests to retrieve information about the products in the video, and the API can respond with the necessary information. As a result, the front-end can quickly display the relevant information to the user, resulting in a smooth and efficient user experience. Using REST APIs also makes scaling the system simple. The machine learning model can be updated or improved, and the front-end can continue to retrieve data from the API without requiring any changes. Furthermore, REST APIs provide a safe and dependable means of transferring data between the machine learning model and the front end. To ensure that only authorized users have access to product information, the API can implement authentication and authorization.

4. Results

Learning by machine, software development, and feedback from users are all components of this project. Training convolutional neural networks for image similarity using Transfer Learning, ResNet50, and VGG 16 Architectures is part of the project. The model is trained to recognise features and locate photos that are most similar to one another by utilising the Euclidean Distance. It is reported that the accuracy of the model is 62% when it is tested using a dataset that is given by the user. This accuracy is lower than the accuracy that can be attained using the Deep Fashion dataset. Despite this, the model can still be useful when applied to a 46-class categorization system.

The project entails the building of a web app in React, in addition to the machine learning component, which enables users to play any video by supplying the URL of the video. To produce a more interesting and interactive experience for users, the web app might be enhanced by the addition of support for a wider variety of video formats and protocols, the incorporation of search and discovery capabilities, and the provision of social features. The models that were built in the machine learning component might be combined with the web app to recommend comparable products from different companies. This could aid in identifying the various sorts of comparable products that are already available and function as a recommendation system. The finished product may support the incorporation of a bespoke API for use in making calls to the development phase.

It is possible to draw the conclusion, given the findings, that the VGG model performed better than the ResNet model in this case, with a similarity score of 49%, as compared to 41% for the ResNet model. In addition, the accuracy of our model's Top-3 predictions was found to be 6% higher than the accuracy of the benchmark, and the accuracy of our model's Top-5 predictions was found to be 4% higher than the accuracy of the benchmark. On the other hand, it was discovered that the accuracy of the model on data provided by the user was 62%, which is a smaller percentage than that on the Deep Fashion Dataset. In spite of this, one could argue that it is still useful for a classification model with 46 categories.

A visual product search option is made available to users of the platform. This means that consumers can now shop by clicking on an item inside a video that they are interested in purchasing. It is also possible to use the platform to compare the costs and qualities of the products that are accessible on the internet and select the one that is most suitable for the consumer because the platform includes all of the similar products that are currently available. The project might potentially be used in the manner of an online shopping website, in which case users would not be required to enter in their search query but rather would be able to play any video and purchase the item of their choice simply by clicking on it. In addition, the initiative has the potential to serve as a platform for marketing.

5. Application

1) The platform provides the user with a visual product search option i.e., shoppers can now click on a product in the video that they wish to buy.
2) Since the platform enlists all the similar products available, it can also be used to compare the prices and qualities of the product available on the internet and select the one that best suits the customer.
3) The project can also be used like an e-commerce website where you don’t have to type the query but play any video and buy the product you like by just clicking on it
4) The project can also be used as a marketing platform.
5) Through this project, we can also provide sellers with insights into popularity of different products via analyzing click through rate (CTR) ratio.
6) By identifying the products clicked on, the project can provide user with a ‘virtual try-on’ option and help customers easily decide on the products they want to buy.

6. Project Budget

If the services of free servers discontinue, then we will have to purchase cloud services.

HEROKU CLOUD SERVICES:

Cloud Service	Cost	Time Duration
Eco Dynos	$5	1000 hours/month
Mini postgres	$5	month

7. Future Scope

There are several areas where this approach could be further developed and expanded:

- Expansion of shoppable content: As shoppable content is becoming increasingly popular among digital advertisers and consumers alike, there is a lot of scope for the project to expand into new areas and platforms. For example, the project could explore creating shoppable content for new social media platforms or for specific types of products.
- Personalization: With customers expecting a personalized experience from the brands they consume, the project could focus on further improving the level of personalization offered by shoppable content. This could involve using data analytics and artificial intelligence to tailor shoppable content to individual consumers based on their preferences and behavior.
- Integration with emerging technologies: The project could also explore integrating shoppable content with emerging technologies such as augmented reality, virtual reality, or voice assistants. This would not only enhance the interactive and personalized nature of shoppable content but also create new opportunities for brands to engage with consumers.
- Optimization of marketing strategies: Interactive video ads have been identified as a way to help all areas of a marketing strategy. The project could focus on optimizing marketing strategies using interactive video ads, which could include conducting A/B testing to determine the most effective ad formats, analyzing consumer behaviour data to inform targeting strategies, and using machine learning algorithms to optimize ad delivery.
- The use of data and analytics to personalize the shopping experience for customers is becoming increasingly important. The project could explore ways to incorporate customer data to create more personalized and targeted shoppable content and interactive video ads.
- Shoppable content and interactive video ads can be integrated into various digital platforms and media channels. The project could explore ways to optimize this integration to increase reach and engagement across different platforms and channels.
- Measuring the effectiveness of shoppable content and interactive video ads is critical to understanding their impact on sales and overall marketing performance. The project could focus on developing new metrics and measurement frameworks to assess the effectiveness of these tactics.

- As new technologies emerge, there may be opportunities to incorporate them into shoppable content and interactive video ads. For example, virtual and augmented reality could be used to create more immersive shopping experiences for customers.
- With the increasing use of customer data in marketing, there are important ethical considerations to be addressed. The project could explore ways to ensure that shoppable content and interactive video ads are developed and implemented in an ethical and transparent manner.

Overall, the future scope of the project involves continued innovation and exploration in the use of shoppable content and interactive video ads to create more personalized, engaging, and effective marketing experiences for customers.

8. Conclusion

This paper presents a comprehensive approach to building a deep learning model for fashion product classification and a corresponding video playing web application that seamlessly integrates with the model to provide real-time product recommendations. The model is trained on the widely used Deep Fashion Dataset and achieves a remarkable Top-3 accuracy of 88.6% and Top-5 accuracy of 94.1%. Moreover, the model's efficacy is further demonstrated through its evaluation on a user-specified dataset, where it attains an accuracy of 62%, thus highlighting its potential for a wide range of fashion-related applications. The video playing web app, built on React, efficiently leverages REST APIs to provide users with relevant product links, descriptions, and other essential information while the video is playing, facilitating well-informed purchasing decisions.

In the end we work on extending our study on paper by Liu, M., & Li, D. (2021) which proposes an image recognition-based approach for smart interactive marketing. The authors suggest using computer vision technology to recognize product images and analyze customer preferences. They also discuss how this approach can be used to improve customer experience and increase sales.

With scope for further expansion, the app can be augmented with features such as support for different video formats, improved search and discovery capabilities, social features, and more, thereby enhancing user experience and improving the app's overall utility. In essence, this paper showcases the immense potential of deep learning models and their integration with web applications to provide real-time recommendations, thus augmenting user experience. The research presented here lays a strong foundation for further exploration and development in the domain of fashion e-commerce.

9. References

[1] Wang, C.L. (2021), "New frontiers and future directions in interactive marketing: Inaugural Editorial", Journal of Research in Interactive Marketing, Vol. 15 No. 1, pp. 1-9.

[2] Andrej Novák, George Lăzăroiu, Tomáš Kliestik (2021), “Live Shopping in the Metaverse: Visual and Spatial Analytics, Cognitive Artificial Intelligence Techniques and Algorithms, and Immersive Digital Simulations”, Linguistic and Philosophical Investigations, Vol. 21, 187-202.

[3] Wen-Huang Cheng, Sijie Song, Chieh-Yun Chen, Shintami Chusnul Hidayati, Jiaying Liu (2021), “Fashion Meets Computer Vision: A Survey”, ACM Computing Surveys, Volume 54 Issue 4, Article No.: 72, pp 1–41.

doi: https://doi.org/10.1145/3447239

[4] Eriko Koike, Takayuki Itoh (2015), "An Exploratory Apparel Product Search Interface Adopting Shopping Psychology", The Journal of Society for Art and Science, Vol. 14, No. 5, pp.170-179.

[5] Jen-Hao Hsiao, Li-Jia Li (2014), “On visual similarity based interactive product recommendation for online shopping”, 2014 IEEE International Conference on Image Processing (ICIP).

doi: 10.1109/ICIP.2014.7025614

[6] Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester, Deva Ramanan (2010), "Object Detection with Discriminatively Trained Part Based Models", IEEE TPAMI, Vol. 32, Issue 9, 1627 - 1645.

doi: 10.1109/TPAMI.2009.167

[7] Jie Cai, Donghee Yvette Wohn, Ankit Mittal, Dhanush Sureshbabu (2018), “Utilitarian and Hedonic Motivations for Live Streaming Shopping”, TVX '18: Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video, Pages 81–88. doi: https://doi.org/10.1145/3210825.3210837

[8] Abhishek S. Rao, Shatanand Bhat, Naveen D Chandavarkar (2017), “Development of an application for better online shopping using computer vision”, International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS).
doi: 10.1109/ICECDS.2017.8389649

[9] Tao Zhang, Zhongjun Tang, Zhongya Han (2022), “Optimal online channel structure for multinational firms considering live streaming shopping”, Electronic Commerce Research and Applications, Volume 56, Issue C.
doi: https://doi.org/10.1016/j.elerap.2022.101198

[10] Hamed S. Neshat, Mohamed Hefeeda (2011), “SmartAd: A smart system for effective advertising in online videos”, IEEE International Conference on Multimedia and Expo.
doi: 10.1109/ICME.2011.6011961