Image Based Search of Products on Marketplace using Real-World Images

MSc Research Project
MSc in Data Analytics

Dhwani Hingu
Student ID: x19216742

School of Computing
National College of Ireland

Supervisor: Dr. Catherine Mulwa

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Dhwani Hingu
Student ID:	x19216742
Programme:	MSc in Data Analytics
Year:	2021
Module:	MSc Research Project
Supervisor:	Dr. Catherine Mulwa
Submission Due Date:	31/01/2022
Project Title:	Image Based Search of Products on Marketplace using Real-World Images
Word Count:	6942
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Dhwani Hingu
Date:	30th January 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Image Based Search of Products on Marketplace using Real-World Images

Dhwani Hingu
x19216742

Abstract

To enhance the shopping experience of the customer on e-commerce or market place a lot of development has been encountered with respect to technology. One of the enhancements is searching for products through images instead of text, labels, keywords related to that product. Plethora of researches were carried out under image retrieval domain, among which less number of researches were found based on product image retrieval.Also, no research was found where google open image dataset was used for image retrieval. Thus, this research study aims to use google open image dataset for image retrieval, the dataset has real world crowd-sourced images with 600+ classes. Models like CNN, VG16 and ResNet50 were used for image retrieval, also combinations of these models were used for transfer learning and feature extraction. Among all the experiments carried out VGG16 gave remarkable results with accuracy of 71% on google open image dataset.

1 Introduction

The way we purchase online has been transformed by technology. The goal of advancing technology is to enhance the purchasing experience for the buyers. Businesses now have been trying out technology-based solutions to make sure that their customers face less trouble and every step of their shopping experience is simple, smooth, time-saving and intuitive. However, sometimes the most time-consuming task during online shopping is finding a particular product through searching by its text, label or content associated with it. Because most of the time it's difficult to analyze the user's intent behind finding for particular product just by using some textual content. Failure at finding the desired product leads to loss of interest of customers and can change their buying decisions.

With the emerging technologies and solutions developed, it has been discovered that over the decade enormous enhancements have taken place in the field of machine learning and deep learning for retrieving images. Initially, Text based image retrieval(TBIR) was widely experimented, but the results were not beneficial. So, content-based image retrieval (CBIR) came into existence and dominated TBIR. CBIR is an approach where an image as a query is passed in the search bar instead of text to find similar types of images. However, there has been much research found on how CBIR works and its application but still there was a scope to carry out research to enhance shopping experience by retrieving images of products which consist of complex backgrounds. Complex backgrounds are those images which may contain multiple objects in an image or image which do not contain a plain solid color background with one object. Using a deep learning model this research aims to retrieve images by querying real world images. As training images with a complex background is a challenging task compared to images with a plain background. Please refer to Figure 1 which illustrates an object with plain background vs object with complex background.

Figure 1: Plain Background vs Complex Background

Also, according to this report Hutchinson (2017) , there is a demand for image-based search, and Facebook is experimenting with its marketplace. So, Facebook marketplace was one of the motivating examples to conduct this research study, as Facebook marketplace also consists of an enormous number of products and users who buy/sell the products. However, its was not possible to extract real-time data of Facebook marketplace due to ethical concerns, that is why google open image dataset [1] was used in this research study. This google open image dataset was crowd-sourced, consisting of more than 600 classes and 9M images. Rationale for choosing this dataset was, this dataset was unexplored for image retrieval, as most of the researchers used this dataset for object detection or segmentation. So, this research study aims to use google open image dataset for experimenting various deep learning models to retrieve images in order to improve and enhance the shopping experience at marketplace.

1.1 Research Question

RQ: *To what extent can deep learning techniques help to improve product image retrieval?* To enhance experience of people shopping at market place.
As per the literature review conducted it was observed that existing approaches for image retrieval had used images with white plain background and with one object in the image as well as none of the research study was found where google open image dataset was explored for image classification and image retrieval. It was also noticed that to enhance the shopping experience with image retrieval, only clothing or fashion image datasets were considered and other products like toys, electronics, home appliances etc were not

[1] `https://storage.googleapis.com/openimages/web/factsfigures_v4.html`

found to be pondered. Thus, this research study aims to find and solve whether deep learning can help to improve product image retrieval using google open image dataset which consists of real world images which are crowdsourced (9M images) by google with more than 600 classes.

Sub-RQ: *Sub-research question: Can classification of products (i.e. toy television couch Blender, bicycle, laptop, handbag headphones, camera and watch) using pre-trained deep learning models enhance image retrieval?*
Classifying these images into different classes would help to improve image retrieval of the products, as each product holds varied features, which distinguish them from each other. So, by using pre-trained deep learning models can images be classified is the motto of the sub research question. In order to achieve Research Question and Sub-Research Question, research objective are taken into consideration as shown in the Table 1

1.2 Research Objective

ID	Object Description	Evaluation Methods and Metrics
Obj1	Literature Review based on various methods, techniques and approaches used by various researchers	
Obj2	Research Methodology and Design Specification for Image Based Search of Products	
Obj2.1	Modified methodology approach used for Image Based Search of Products	
Obj2.2	Design Specification, Architecture and Process flow	
Obj3	Implementation, evaluation and comparison of deep learning models	CNN, VGG16 and ResNet50
Obj3.1	Implementation, evaluation and comparison of experiment with CNN models	Model Used - CNN. Evaluation Methods - Accuracy, Precision, F1-score and Recall
Obj3.2	Implementation, evaluation and comparison of experiment with transfer learning models	Model Used – VGG16 and ResNet50. Evaluation Methods - Accuracy, Precision, F1-score and Recall
Obj3.3	Implementation and evaluation of experiment with feature extraction and image retrieval models	Model Used – VGG16. Evaluation Methods - Accuracy, Precision, F1-score and Recall
Obj4	Discussion and comparison of the developed model with models discussed in literature review	Evaluation Methods – Accuracy, Precision, F1-score and Recall as well as similarity distance of images

Table 1: Research Objective

2 Related Work

This section of the report covers various studies conducted by various researchers based on their methods, technologies, techniques as well as datasets used for images classification and image retrieval, also their results in terms of accuracy and error rate are studied in order to understand the effectiveness of the models. Overall, these studies by different researchers helped to achieve knowledge about the domain and scope of various techniques and methods to be used for this research.

2.1 Text Based Image Retrieval and Content Based Image Retrieval

Over a decade it has been witnessed in the evolution of image retrieval technology. Text based image retrieval (TBIR) was one of the most common image retrieval systems. It looks for a picture in a database based on its concept, text, tag, label, and other factors. However, it was difficult to convey a picture's whole visual meaning in words, and may yield irrelevant results Nonetheless, TBIR is now obsolete and this was proved by Li et al. (2011). The authors of this paper collected images from the web which had loosely coupled labels for image classification and image retrieval. These images were classified as relevant and irrelevant images starting as clusters, where these clusters are treated as "bags". Author's proposed system was efficient as it was able to effectively exploit loosely labeled images by using robust SVM classifiers.

Content based image retrieval (CBIR) overcame the drawbacks of the TBIR as well as a lot of enhancement can also be seen under CBIR. In CBIR, image as a query is passed in order to get similar images as result. During this search, CBIR analyzes contents of these images instead of tags, text, labels or metadata associated with these images. Images have these sets of features like color, shape and texture and extracting these features for image retrieval was a task. Lots of researchers tried to solve feature extraction problems of image retrieval using various approaches and methods, among them one of the papers was found where Madugunki et al. (2011) proposed comparison of various CBIR techniques with respect to color and texture. The authors used 150 images and used matching techniques, here they calculated distance between two images in order to retrieve images that are similar to the queries image. There are various methods to calculate distance between these images such as Euclidean distance method, City Block Distance, Canberra Distance and Discrete Wavelet Transform(DWT). Among these DWT gave efficient results to them compared to Canberra Distance. This research was helpful to understand different types of methods available to find distance between images for matching similar images.

Another comparative research of ten different distance measures was found where Pasumarthi and Malleswari (2016) used a COREL dataset with 10 different classes such as flowers, animals, beach, scenes, food, people, buses and mountains. They computed a color histogram in HSV space to extract color features of the image and on the other hand wavelet decomposition was adjusted for texture features. Among all distance measures Manhattan distance gave amazing results as well as cosine similarity outperformed. Authors concluded that CBIR is not just based on content of the image but more of it depends on distance between two images and their similarity. Shinde et al. (2015) also carried out an experimental research using color extraction and did not consider shape and texture. They used a machine learning approach – naïve Bayesian classifier to classify images based on color and got 81.24% accuracy.

However, these researchers achieved promising results but there were few limitations such as they did not consider shape and edge feature extraction as well as the datasets they used for training were pretty small. Most of the researchers did their experiments on image retrieval based on color feature extraction, as extracting shape and texture is challenging. On the other hand Varma and Riyaz (2018) did experiments with 1300 images which consisted overall of 17 various types of flowers. They extracted color and edge feature from these images, a hybrid approach was carried out where for color gradient, color histogram was used and gabor wavelet method was used for inner and outer edge, they implemented this research on MATLAB.

2.2 Image Classification for Image Retrieval

As colors cannot be obligated to scale, rotate or translate, therefore they have been one of the basic, traditional and simple methods used for image classifications as well as for image retrieval based on text or by querying images. However, the most difficult part for feature extraction is texture and shape. Shape feature refers to the shape of a specific region that cannot be represented by geometric characteristics and necessitates the use of special attributes and several aspects to specify. In order to identify shape in the image, edge detection and segmentation are the methods used. Lastly, Texture features extraction is one of the vital features and it can be portrayed in terms of scale, contrast, magnitude etc. In 2016, Naveena and Narayanan (2016) explained a study regarding "Image retrieval using combination of color, texture and shape". The author randomly made a custom dataset of 1000 images with 10 different classes and carried image classification using SVM classifier to classify images in one of the classes. However, the author got promising results by comparing results based on individual features as well as combination of features. Overall, this paper was helpful to understand basic features extractions of an image in order to develop reverse image search

On the other hand in 2016 another research study was found with an advancement of technology for image classification and image retrieval, Li et al. (2016) proposed a combination of large scale computing platforms with deep learning models along with feature illustrations in image retrieval as well as classification. The authors used ImageNet dataset, where they first carried out image classification using deep learning technique with batch normalization along with multicrop testing to gain a better performance. This research was helpful to get insight into how deep learning techniques can be used in this research project as the author's 89.45% classification accuracy was promising enough to believe their approach was efficient.

Another research used CNN for image retrieval and classified images based on ranking. Sindu and Kousalya (2019) demonstrates a two phase approach where the first object is detected from the image and based on that multiple objects are extracted, a further relevance ranking algorithm is used to retrieve relevant images. The author used the PASCAL VOC 2007 database which consisted of 2449 images and 20 classes, where SSD networks were implemented with CNN on this dataset. The author got 72.42% mean average precision for all the classes which is pretty good. This research was helpful to understand how SSD works for object detection when there are multiple objects in an image.

Transfer learning (TL) is a study subject that focuses on storing and transferring information obtained while addressing one problem to a different but related problem.

Singh et al. (2018) proposed an investigation study regarding the usefulness of transfer learning on generic images using pre-trained deep learning models. This paper was gave a brief understanding related to traditional transfer learning vs deep learning transfer learning. The author used medical images dataset and implemented various models such as VGG16, VGG19, ResNet50, InceptionV3, Xception, MobileNet and Inception ResNetv2, among which Inception-v3 gave accuracy of 99.45. This investigation study was helpful to understand various deep learning model for transfer learning as well as the author aims to use data augmentation in their future work which motivated this research project to conduct.

2.3 Image Retrieval Using Deep Learning

Deep learning has infiltrated a wide range of commercial industries and corporate applications. Nonetheless, certain key demands have developed and continue to exist in the absence of a complete answer. “A Survey of Deep Learning: Platforms, Applications and Emerging Research Trends” by Hatcher and Yu (2018) helped enormously to gain domain knowledge and scope for research under deep learning applications. One of the rationale for choosing deep learning technology was it is a rapidly evolving technology that has been successfully used to a wide range of applications and fields, thus it was considered in this research study. Another survey study was found by He (2020) who beautifully articulated “Deep learning in image classification”. One of the steps in this research study was to classify images using some technique instead of getting similar images using pretrained techniques available. This survey was helping to understand various wellknown neuralnetworks and their uses in image classification. Additional, one more survey by gang Zhou et al. (2017) was studied in order to understand what all advancement are taken place under content-based Image Retrieval and it was found even though a lot of research had been done but there was a scope for new research trend in image retrieval. Various feature extraction, distancebased scoring, performance evaluation techniques were studied and were implemented in this research study. A list of various dataset freely available for research were discussed in this survey except google open image dataset. So, this research paper has used google open image dataset for experimenting image retrieval using all the techniques and methods studied through various literature surveys conducted.

Scalable and fast image retrieval by querying images was required for this research study. The activations of Convolutional Neural Networks (CNNs) as features performed exceptionally well in this domain. Tanioka (2019) compared baseline system vs image retrieval system with dot product score, Euclidean distance and Manhattan, cosine score for its accuracy and response time. The authors have used ImageNet and Kaggle’s Dog vs cat dataset, using this dataset the authors have trained a pre-trained VGG-16 model. The limitation of this paper was that the author claims to use ImageNet dataset but it was found that the author had just used 250 images of cats and dogs from Kaggle, however the dataset was pretty small for training images. However, the author’s proposed Manhattan distance and Euclidean shows promising results and was considered as one of the experiments in this research with cosine similarity to check which distance score is better as well as a large dataset from google open image dataset was considered in this research.

Even after using several distance score techniques for image retrieval it was observed that there were various types of issues faced such as rotation invariance, translation Invariance, illumination invariance, scaling invariance etc. Walkoli et al. (2021) demonstrated that by using SIFT with CNN solves the above problems. The author used COCO dataset and used CNN for image classification whereas Euclidean distance for image matching and SIFT for solving mentioned problem. The authors themselves suggest that CNN can be trained with a greater number of classes and use techniques to optimize as well as decrease training time. Also, author suggested to use different dataset as well as huge enough for training, therefore google image dataset with varied classes and large number of images were taken into consideration.

Moreover, Diyasa et al. (2020) used CNN pre-trained model using a small dataset of cats and dogs to classify them and retrieve images of them by extracting features of the cat and dog. The author also rotated cat and dog and passed a query to check whether the images are retrieved or no of the query image is rotated or inverted. However the author got good results and thus proved image augmentation was also needed to experiment on google open image dataset as well as the author considered distance between two images using Euclidean distance.

2.4 Comparison of image retrieval approach used by different researchers

A comparison study was carried out between various authors, dataset used by them and method/approach used by them with respect to image retrieval. Mawoneke et al. (2020) of this paper illustrates image retrieval in fashion dome using Kaggle fashion product image dataset[2] . The author achieved 93% accuracy which looks good but the author did not consider other products like electronics, home appliances, home decor/ furniture etc as well as the images used by the author are simple for training as these images consist of white plain background with a single object in it. Another researcher Boriya et al. (2019) also experimented image retrieval using DeepFashion image dataset [3], again this author did not consider any other products than clothing, also the background of images were not complex. These researchers carried out experiments with six deep learning models among which only CNN gave good results compared to others. Apart from fashion, (Ali and Sharma (2017)) experimented image retrieval using a small dataset[4] and limited classes such as Animal, Butterfly, Facial, Flower. However, these researchers achieved pretty low accuracy and were not prevailing. A detailed comparison of these studies are shown in the Table 2

Research objective Obj1 was achieved by conducting a rigorous Literature review with respect to different approaches used for image retrieval along with different datasets explored for image classification and image retrieval. Comparison of researchers was also conducted which identified gaps for this research study. This marks the conclusion of this chapter as domain knowledge was acquired, scope of research was cleared, gaps were identified as well as novelty was taken care of.

[2]https://www.kaggle.com/paramaggarwal/fashion-product-images-dataset

[3]http://mmlab.ie.cuhk.edu.hk/projects/DeepFashion.html

[4]http://www.vision.caltech.edu/Image_Datasets/Caltech101/

Author	Dataset	Method	Results	Comments
Mawoneke et al. (2020)	Kaggle: Fashion Product Image Dataset	CNN	93%	Only clothing category was considered and background of images were not complex
Boriya et al. (2019)	DeepFashion	Inception, Resnet50, VGG16, CNN, VGG19, MobilNet	CNN(83%)	Only clothing category was considered and background of images were not complex
Ali and Sharma (2017)	Caltech-101 image dataset	SIFT, BFOA and DNN	Classes and accuracy : Animal – 17%, Butterfly - 38%, Facial – 57%, Flower - 95%	Small Data set

Table 2: Comparison of researchers used image retrieval approach

3 Research Methodology and Design Specification

3.1 Introduction

This chapter of the research study emphasizes the methodology used as well as the architecture of the project. The main motive behind this research was to enhance the shopping experience of people through product image retrieval, so an improved and modified version of CRISPDM methodology was designed as shown in the Figure 2 and taken into consideration. Also, the research project fits well into 2-tier architecture and rationale for choosing has been discussed. The flow of the project along the technicalities have also been articulated.

3.2 Image Based Search of Product's Methodology

- **Business Understanding:** Before taking the necessary steps, one must first understand the project. Obtaining information is the first step. So, as seen in the literature review it shows that TBIR is slowly now becoming outdated and CBIR has overcome the limitations/drawbacks of it. Analysis done by Dagan et al. (2021) also proved that an "An Image is Worth a Thousand Terms" and therefore there was a need for this research study to be conducted. When a user is shopping online on an e-commerce platform or marketplace, the user's intent is to get the product what they have visualized in their mind, instead of typing words, tags, labels or content related to image, querying an image might lead to relevant results, which is more time saving than text-based image search.

- **Data Selection and Data Gathering:** During literature review it was witnessed that most of the researchers worked with small datasets for image retrieval, so taking this into consideration, google open image dataset was selected to use in this research study. Rationale for choosing this dataset was, it is an open dataset freely available for the researchers for researching and experimenting purposes. Also the dataset comes with different options to download such as training, testing, validation data, images with or without annotations. It was noticed that google open image dataset was mostly used for object detection and it wasn't used for experimenting image classification or image retrieval. Thus, this research study selected google open image dataset for experimenting image retrieval using real world images readily available. A custom dataset was made with 10 classes(Toy, Television, Couch, Blender, Bicycle, Laptop, Handbag, Headphones, Camera, Watch) of real world products images. In total 4649 images were extracted for training, 1327 for testing and 537 for validations.

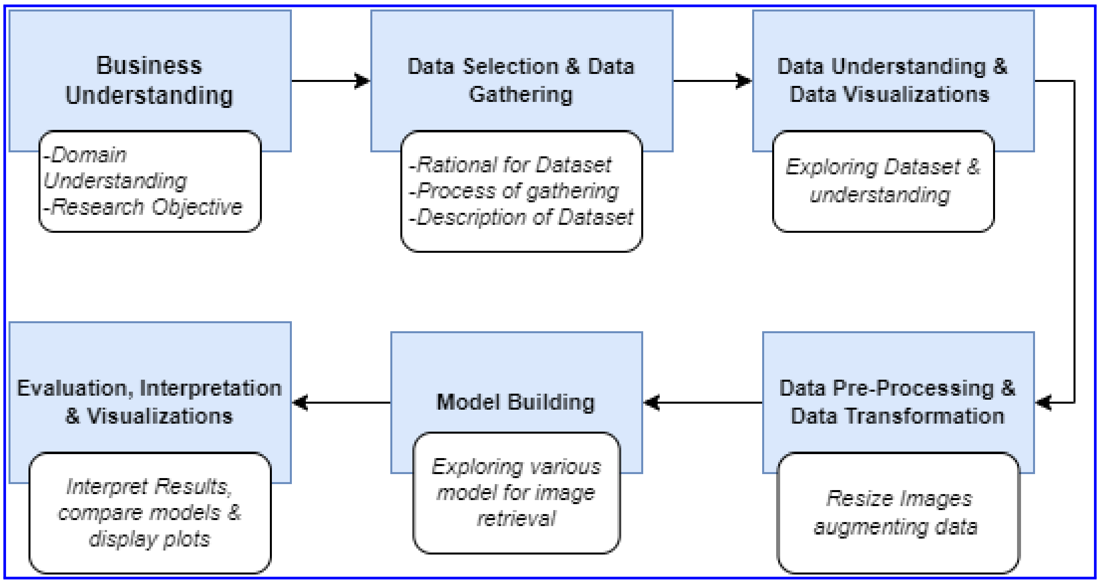


Figure 2: Methodology Used for Image Based Search of Products

- **Data Understanding and Data Visualizations:** Once the dataset was selected for any project, it's important to explore the whole dataset and to understand it. After downloading the data, it was noticed that there was a different folder for all the labels of the images. Image name was the same as Label text file name and the label file consisted of the class of the image, coordinates of the box(XMin, XMax, YMin, YMax). Also, the size of all the images were different.

- **Data Pre-Processing and Data Transformation:** After exploring the custom dataset, it was understood that there was the need to do some data pre-processing as well as some transformation was required. Images were resized to 224*224, so that all images can get trained in the same size. ImageDataGenerator class was used to create copies of images in the training dataset. These copies of images are shifted, flipped, zoomed etc. This was done in order to increase the model's performance as the performance of deep learning models usually increases with a large number of data provided.

- **Model Building:** It's a process where various experiments are carried out with various types of model using the data which is ready after pre-processing. Based on the literature review study it was analyzed that few models like CNN and VGG gave promising results on different datasets, although these dataset were pretty small in size. So, this research study has experimented CNN, VGG16 and Resnet50 on Google open image dataset for image retrieval. During model building, parameters for building the models were adjusted in order to experiment which set of parameters give good results. Base model CNN was applied with adjusting some parameters, further VGG16 and ResNet50 were used as feature extractor models.

- **Evaluations, Interpretations and Visualizations:** After experimenting with various models, results will be evaluated based on different evaluation methods such as accuracy, precision, recall, loss etc. Also, some visualizations in the form of graphs and plots are displayed in order to get a better visual picture of the model's performance.

3.3 Project Design Specification

This section seeks to visually depict the many stages that must be completed in order. In general, there are two types(Two-tiered or three-tiered) of architecture design for any project. In this research a 2-tier architecture has been considered, where the client layer is the customer using a marketplace to shop products using image-based search. On the other hand, the Business logic layer, also called a data layer, consists of all the tools, methods, techniques and models required to build the system. Rationale for choosing 2-tier architecture instead of 3-tier architecture was the amount of data extracted, the system did not have multiple dataset nor merging of two different dataset was done.

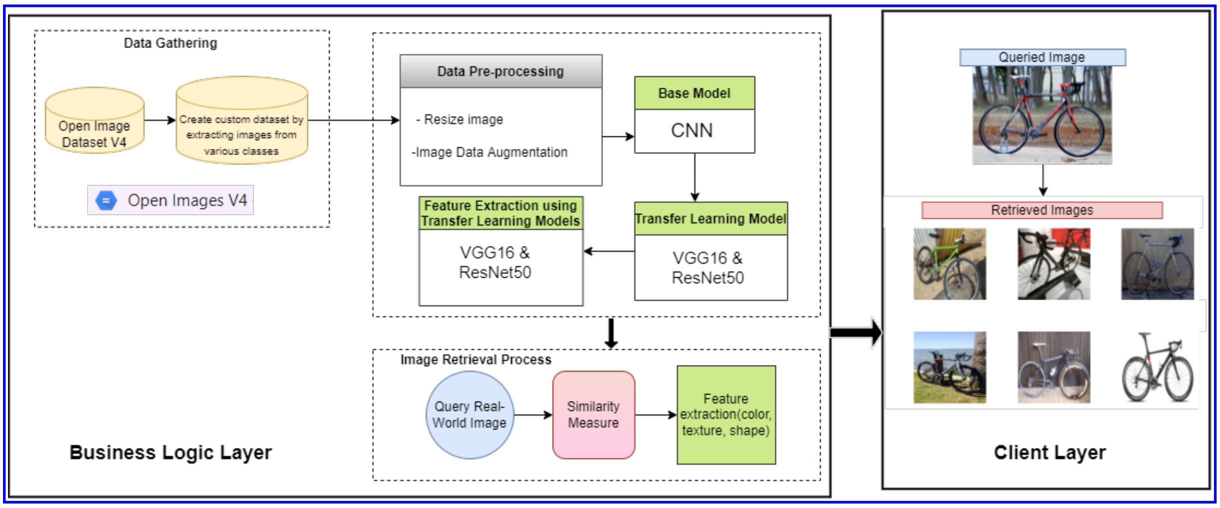


Figure 3: Process Flow

So, as seen the in the Figure 3 during data gathering phase, the data was extracted from google open image dataset. 10 different categories/classes were considered and based on those 4649 pictures for training, 1327 for testing, and 537 for validations were retrieved. After downloading and creating custom dataset, these images were pre-processed by rescaling the size of the image as well as data augmentation was carried out using ImageDataGenerator class. Once these images were pre-processed a base model was applied to experiment how the system works without labels, transfer learning and feature extraction. Results of base model CNN were evaluated on the basis on accuracy, precision and recall. Another experiment was taken into consideration using VGG16 and ResNet50 for transfer learning and their results were compared and interpreted. One More experiment was performed using VGG16 and ResNet50 for feature extraction and then again results were compared and interpreted. Based on the above training and successful experiments, image retrieval process was conducted, where an unseen/untrained image was passed and their features were extracted as well as similarity measure was checked where distance between two images were calculated and images with less distance as well as with similar features were retrieved and displayed at the client layer.

3.4 Conclusion

By designing own modified methodology as per the requirement of the research along with project design and explaining them the research objective(Obj2, Obj2.1 and Obj2.2) were achieved as discussed in the introduction chapter earlier. This brings to conclude this chapter of the research by meeting all objective set.

4 Implementation, Evaluation and Results

4.1 Introduction

This chapter of the research study emphasis on the various experiments carried out for enhancing the product image retrieval using various deep learning model. A detailed explanation of all experiments with respect to all necessary steps carried out for implementing the model, further interpreting and evaluating the model performance and comparison of the model's results was articulated. Rational for choosing particular deep learning model for all experiments was also stated. Even for evaluation accuracy, precision, recall and f1-score were used for assessing the model's performance.

4.2 Experiment with CNN model

4.2.1 Implementation of CNN model

Building a convolution neural network is an excellent approach to utilize deep learning to categorize photos (CNN). Also based on literature review conducted, it was observed that CNN gave notable results, thus CNN as base model is considered as the very first experiment in this research. However, with the help of Keras Python package, creating a CNN model was very straightforward and simple. Pixels are used by computers to see pictures and a matrix of pixels are created. So, with the help of the kernel these pixels of the matrix are multiplied and later on their values are added up, however this process is repeated until all pixels of the image have been traversed. In this research, a CNN model with following set of parameters have been created:

- Num of layers: This means how many layers of Conv2d are to be considered for building the model. This research study has used 2 or 3 Conv2d layers.
- Num of filters: The work of filters is to detect features in an image. This parameter will build a number of filters in each Conv2d layer. Here, 32 layers of filters have been considered.

- Filter size: Based on the filter a matrix is created. Here, filter size is adjusted as (3,3) or (5,4) with other combinations of parameters in order to acquire best performing model

- Initializer: After setting the size of the kernel, the kernel needs to be initialized. Here glorot_uniform was used to initialize the kernel

- Activation function: Here, Relu activation function is used because pictures are non-linear, the relu activation function is applied after the convolutional procedure to achieve non-linearity. if the input is positive, the Relu function will output it immediately; otherwise, it will produce 0

- Dropout: In order to make sure that the model does not fall under over-fitting, a dropout parameter is used. Here, 0.2 and 0.7 were adjusted to prevent the model from over-fitting

- Optimizers: Optimizers are techniques or approaches that adjust the characteristics of your neural network, such as weights and learning rate, to decrease losses. Here, one of the best optimizers – Adam is used to reduce the loss.

4.2.2 Evaluation, Results and Comparison of CNN Model

As explained above, the parameters of the CNN model were adjusted in a way that 6 different combinations were created in order to get the best performing model. Throughout, for all the CNN models, 10 epochs were considered with a batch size of 64. Also, for all the combinations, the performance was examined based on accuracy, precision and recall as well as a confusion matrix was plotted to cross check whether the model was able to predict the true values or not. As seen in the Figure 4 these combinations were compared in the end and it was comprehended that by adjusting kernel size (3,3) and setting dropout rate as 0.7 gave good results around 29.76% which is pretty low but highest compared to different set combinations adjusted. As this model did not perform that good, it was decided to go further with another experiment by using VGG16 and ResNet50 for transfer learning.

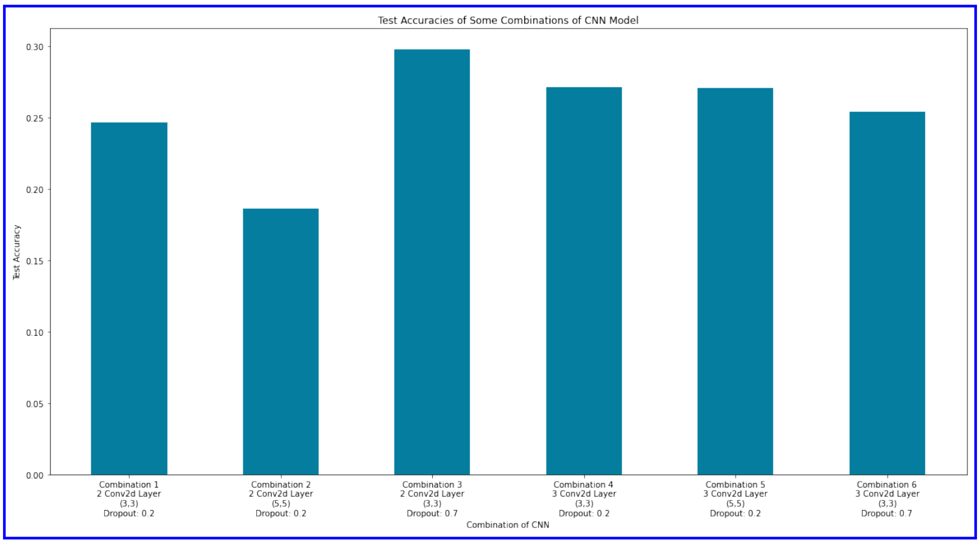


Figure 4: Comparison of CNN Models

4.3 Experiment with Transfer Learning Models

4.3.1 Implementation of Transfer Learning Models

Rational for performing transfer learning was because it's a process in which a model learned on one problem is applied in some way on a second problem that is linked to it as well it gives improved performance and saves lots of time. Reason for choosing VGG16 was, it has 13 convolution layers and 3 fully connected layers, which all together makes 16 layers, thus it's an amazing architecture, freely available and lastly it's been giving benchmarking results when used for transfer learning. On the other hand, ResNet50 was also chosen for transfer learning, because ResNet exemplifies depth of the deep learning architecture. ResNet consists of multiple basic blocks , these blocks are like residuals they can be in hundreds or even more and a network can be formed furthermore can be trained. A function was created to train models for transfer learning; the last fully connected layers were adjusted as to whether the last fully connected layer will be trainable or not. If it's said to be true, it would be trainable and their respective results are recorded and evaluated based on accuracy, precision and recall. Similarly, if in function it is said to be false then the last fully connected layer won't be trainable and likewise the results are accessed and examined. In the end all the models with tweaking the parameters are compared as seen in the Figure 5

4.3.2 Evaluation, Results and Comparison of Transfer Learning Models

Again 10 epochs were considered with a batch size of 64. It was observed that keeping the last fully connected layer as true for VGG16 gave remarkable results. When the first epoch was running it gave accuracy around 15.77%, loss around 3.7%. But as the epochs were running slowly the accuracy also kept improving and after the last 10th epoch accuracy was 96.88%. However, the test accuracy was also pretty good, around 71.5% . As seen in the figure 5 evaluation methods were used to check the model's performance, for all the classes, the model was able to predict from image and classify whether from which category the image belongs to. Also, a confusion matrix as seen in the Figure 6b was plotted to get a clear idea of how many classes the model was able to predict correctly. Thus, VGG16 with last layer trainable gave promising results compared to VGG16 model where last layer was not trainable as well as RestNet50 did turn out to be performing good, as when the ResNe50 model was trained by keep last layer as not trainable gave accuracy around 26% and when the last layer was trainable gave accuracy around 33%. Both of these ResNet50 models gave low accuracy compared to what VGG16 gave, thus for further experiment of image retrieval VGG16 by training the last layer was considered. As shown in the Classification Report Figure 6a, the VGG16 model was able to interpret and classify images when the last four layers were trainable. The model was able to predict and classify bicycles and the least it was able to predict was watch and coach class. However, overall the model was able to classify and predict the class of the object in-spite of having multiple objects in one image

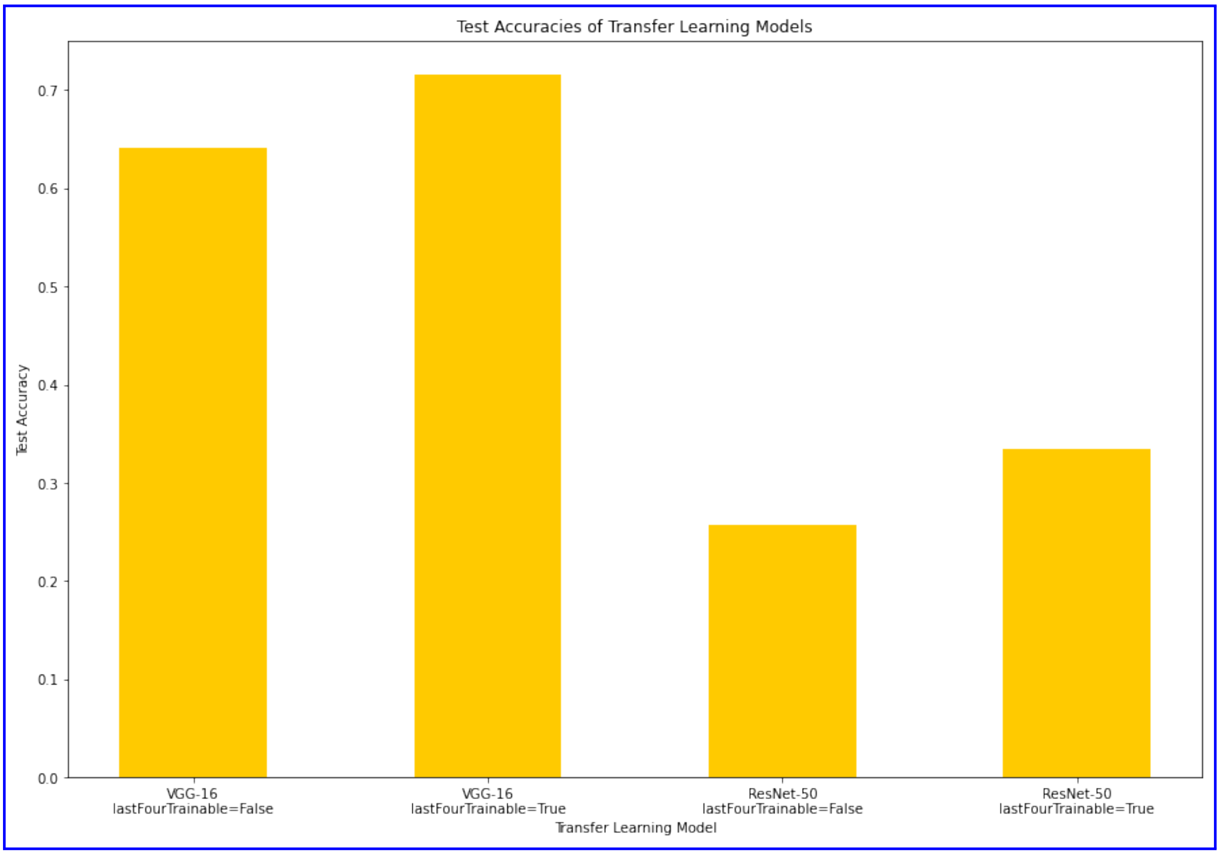


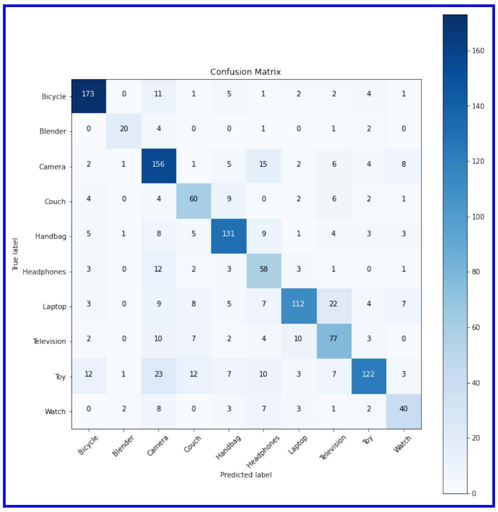
Figure 5: Comparison of VGG16 and ResNet50 Models

```
Test Loss:  1.7210322618484497
Test Accuracy:  0.7151469588279724
Classification Report
              precision    recall  f1-score   support

     Bicycle       0.85      0.86      0.86       200
     Blender       0.80      0.71      0.75        28
      Camera       0.64      0.78      0.70       200
       Couch       0.62      0.68      0.65        88
     Handbag       0.77      0.77      0.77       170
  Headphones       0.52      0.70      0.59        83
      Laptop       0.81      0.63      0.71       177
  Television       0.61      0.67      0.64       115
         Toy       0.84      0.61      0.71       200
       Watch       0.62      0.61      0.62        66

    accuracy                           0.72      1327
   macro avg       0.71      0.70      0.70      1327
weighted avg       0.73      0.72      0.72      1327
```

(a) VGG16 Classification Report



(b) VGG16 Confusion Matrix

Figure 6: Evaluation of VGG16 Model

4.4 Experiment of Image Retrieval using Feature Extraction

4.4.1 Implementation of Image Retrieval using Feature Extraction

Image retrieval can benefit from feature extraction in deep learning models. In this experiment, VGG16 and ResNet50 that were trained for transfer learning were used to extract features. As seen, in the previous section VGG16 gave better results compared to other models, so feature extraction for image retrieval was carried out on VGG16. The output layer was used before classification in order to extract features. Further, the output weights were then placed in front of the model's classification layer. Later, a vector of features was created for training dataset and validation as a DataFrame. Once the model is finished, this DataFrame was saved as pickle. Finally, the model was ready to retrieve images, here an image was given and was compared based on its feature vectors and the images with similar vector features as well as images who had less distance between them were retrieved. In order to get vectors and distance between images functions were used, Cosine similarity was used for getting similarity between two images with respect to distance.

4.4.2 Evaluation and Results of Image Retrieval using Feature Extraction

In the above sections of evaluation, the model's performance was done on the basis of accuracy, precision, recall and f1score. But, here in the image retrieval evaluation section, the model's performance was interpreted based on the similar image retrieved when an unseen/ untrained image was queried. The similarity was checked with the help of feature vector and distance measure which was explained in the previous section. As per the Figure 7 , it's been interpreted and observed that the model performed excellent, as approx. 95% to 99% similarity was witnessed between the queried image and the retrieved image. Also, the model was able to detect main class of the object, if the image was blurred(as shown in the Figure 7e) or had multiple objects in the image (laptop and person as shown in the Figure 7f). Thus, overall VGG16 outperformed while building a transfer learning model as well as showed prominent results while extracting features from an image for image retrieval.

(a) Bicycle class

(b) Blender class

(c) Television class

(d) Camera class

(e) Headphones class

(f) Laptop class

Figure 7: Products Images Retrieved

4.5 Conclusion

By Implementing all the experiments and evaluating them as well as comparing them, the best fit model was achieved. Thus, research objective (Obj3, Obj3.1, Obj3.2 and Obj3.3) as discussed in introduction chapter – sub-section : 1.2 were attained and this abide to conclude this chapter

5 Discussion and Comparison

In this research study, with the help of literature review, a handful of comparison studies by different researchers were studied based on the methods and dataset they used. From which scope of research for this research was cleared as well as reason for choosing deep learning models like CNN, VGG16 and ResNet was also acquainted. Further, these models were experimented with google open image dataset and their results were interpreted, evaluated and compared. Based on the evaluation, the best model was achieved to enhance the shopping experience of the product image retrieval with high similarity between the queried image and retrieved images. This developed model was then compared with the models studied in the literature review section 2.4. As seen in the Table 3, it depicts that the research study model gave remarkable results in comparison of complexity of the images used, novelty with respect to unexplored dataset. Lastly, this marks to the conclusion of this chapter by meeting research objective discussed in the introduction chapter.

Author	Dataset	Method	Results	Comments
Mawoneke et al. (2020)	Kaggle: Fashion Product Image Dataset	CNN	93%	Only clothing category was considered and background of images were not complex
Boriya et al. (2019)	DeepFashion	Inception, Resnet50, VGG16, CNN, VGG19, MobilNet	CNN(83%)	Only clothing category was considered and background of images were not complex
Ali and Sharma (2017)	Caltech-101 image dataset	SIFT, BFOA and DNN	Classes and accuracy : Animal – 17%, Butterfly - 38%, Facial – 57%, Flower - 95%	Small Data set
Current Research Study	Google Open Image Dataset	CNN, VGG16 and ResNet50	VGG16(71%)	All classes gave similarity between images from 95% to 99% with varied set of categories and with complex background

Table 3: Comparison of developed model with model discussed in the literature review

6 Conclusion and Future Work

The aim of this research was to improve and enhance the shopping experience of the people shopping at the marketplace. When a user is shopping at the marketplace, often they have already decided what they want to buy, they have visualized the product in their mind or sometimes they might even have images of the products they are looking for. So, retrieving that particular product with the help of text, labels or content associated with it will be time consuming and in most cases might lead to failure for not getting the desired product. However, this might lead to loss of customers which might affect business directly. Literature review was carried, gaps were discussed and studied as well as scope of research was cleared. Based on the literature review, a dataset was found and it was witnessed that the dataset chosen for this research study was novel, as none of the research was found where researchers had used google open image dataset for image retrieval. Once the dataset was found, the dataset was explored to understand it and a further unique and modified methodology approach was designed inspired from CRISP-DM methodology. Moreover, to understand the research project in depth, project design was designed in order to get a clear execution plan before implementation of the model. Once these methodology approach and design specification objectives were met, implementation of the model was initiated. For model building the dataset was pre-processed and data augmentation techniques were used. Once the data was ready, various experiments were conducted, starting with applying CNN as base model. The CNN model's parameters were then adjusted and 6 combinations of the models were executed, further this combination of CNN models were compared and evaluated. However, the CNN model did not perform well and another experiment was conducted with transfer learning techniques where VGG16 and ResNet50 models were used. It was u found out that by applying the transfer learning model the accuracy got better and the model was able to predict better compared to CNN model. Among VGG16 and ResNet50, VGG16 gave accuracy of 71% which was exceptional. Moving forward, the last experiment was conducted which was an essential experiment of image retrieval with feature extraction and distance measure between images. Here, the already trained VGG16 model was used for further extracting features of the image and retrieving similar product images based on queried product images. This experiment was also successful as similarity between the images were found to be between 95% to 96%, which means the model was able to predict what particular image is of and was able to classify and retrieve similar types of product images from the dataset. This marks the conclusion that the given research was able to solve the problem statement as the models were able to retrieve images based on similarity and classify them into respective categories.

Future Scope: This research has just extracted 10 classes (i.e. toy television couch, blender, bicycle, laptop, handbag, headphones, camera and watch) from google open image dataset for image retrieval but in future large sets of classes/categories can be considered. Also, different deep learning models can be used further as well as different approaches and methods can be experimented. To enhance the shopping experience of the customers on the marketplace, clicking images of the product in real-time and retrieving similar products can be implemented in future.

Acknowledgement

I would like to give my gratitude to my supervisor Dr. Catherine Mulwa for her constant support throughout the semester as well as assisting me with all my queries. I would also like to thank my parents for financially supporting my Masters. Lastly, I would like to thank my brother for patiently listening out to my queries.

References

Ali, A. and Sharma, S. (2017). Content based image retrieval using feature extraction with machine learning, *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1048–1053.

Boriya, A., Malla, S. S., Manjunath, R., Velicheti, V. and Eirinaki, M. (2019). Viser: A visual search engine for e-retail, *2019 First International Conference on Transdisciplinary AI (TransAI)*, pp. 76–83.

Dagan, A., Guy, I. and Novgorodov, S. (2021). An image is worth a thousand terms? analysis of visual e-commerce search, *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '21, Association for Computing Machinery, New York, NY, USA, p. 102–112.
URL: *https://doi.org/10.1145/3404835.3462950*

Diyasa, I. G. S. M., Alhajir, A. D., Hakim, A. M. and Rohman, M. F. (2020). Reverse image search analysis based on pre-trained convolutional neural network model, *2020 6th Information Technology International Seminar (ITIS)*, pp. 1–6.

gang Zhou, W., Li, H. and Tian, Q. (2017). Recent advance in content-based image retrieval: A literature survey, *ArXiv* **abs/1706.06064**.

Hatcher, W. G. and Yu, W. (2018). A survey of deep learning: Platforms, applications and emerging research trends, *IEEE Access* **6**: 24411–24432.

He, Z. (2020). Deep learning in image classification: A survey report, *2020 2nd International Conference on Information Technology and Computer Application (ITCA)*, pp. 174–177.

Hutchinson, A. (2017). Facebook is testing out image-based search for related products.
URL: *https://www.socialmediatoday.com/news/facebooks-testing-out-image-based-search-for-related-products/506841/*

Li, H., Su, P., Chi, Z. and Wang, J. (2016). Image retrieval and classification on deep convolutional sparknet, *2016 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pp. 1–6.

Li, W., Duan, L., Xu, D. and Tsang, I. W.-H. (2011). Text-based image retrieval using progressive multi-instance learning, *2011 International Conference on Computer Vision*, pp. 2049–2055.

Madugunki, M., Bormane, D. S., Bhadoria, S. and Dethe, C. G. (2011). Comparison of different cbir techniques, *2011 3rd International Conference on Electronics Computer Technology*, Vol. 4, pp. 372–375.

Mawoneke, K. F., Luo, X., Shi, Y. and Kita, K. (2020). Reverse image search for the fashion industry using convolutional neural networks, *2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP)*, pp. 483–489.

Naveena, A. K. and Narayanan, N. (2016). Image retrieval using combination of color, texture and shape descriptor, *2016 International Conference on Next Generation Intelligent Systems (ICNGIS)*, pp. 1–5.

Pasumarthi, N. and Malleswari, L. (2016). An empirical study and comparative analysis of content based image retrieval (cbir) techniques with various similarity measures, *3rd International Conference on Electrical, Electronics, Engineering Trends, Communication, Optimization and Sciences (EEECOS 2016)*, pp. 1–6.

Shinde, S. R., Sabale, S., Kulkarni, S. and Bhatia, D. (2015). Experiments on content based image classification using color feature extraction, *2015 International Conference on Communication, Information Computing Technology (ICCICT)*, pp. 1–6.

Sindu, S. and Kousalya, R. (2019). Two phase content based image retrieval with relevance ranking on classified images, *2019 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1–6.

Singh, S., Ho-Shon, K., Karimi, S. and Hamey, L. (2018). Modality classification and concept detection in medical images using deep transfer learning, *2018 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, pp. 1–9.

Tanioka, H. (2019). A fast content-based image retrieval method using deep visual features, *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, Vol. 5, pp. 20–23.

Varma, N. M. and Riyaz, A. (2018). Content retrieval using hybrid feature extraction from query image, *2018 International Conference on Information , Communication, Engineering and Technology (ICICET)*, pp. 1–4.

Walkoli, D., Sali, V., Patil, S., Sonawane, R. and Mahalakshmi, B. (2021). Content-based image retrieval using sift and cnn, *2021 Asian Conference on Innovation in Technology (ASIANCON)*, pp. 1–5.