**Content-Based Image Retrieval (CBIR) Digital Image Processing Project**

Muhammad Osama Nusrat 21i-2169

Zeeshan Habib 21i-2193

Abdul Basit Abbasi 22i-2841

******

Department of Artificial Intelligence

FAST National University of Computer and Emerging Sciences Islamabad Campus

**Table of Contents**

[Abstract 4](#_Toc135517305)

[CHAPTER 1 INTRODUCTION 5](#_Toc135517306)

[CHAPTER 2 LITERATURE REVIEW 6](#_Toc135517307)

[CHAPTER 3 METHODOLOGY 8](#_Toc135517308)

[Chapter 5 Discussion 20](#_Toc135517309)

**List of Figures**

[Figure 1 Proposed Block Diagram 10](#_Toc135517320)

[Figure 2 CBIR using deep neural networks 11](#_Toc135517321)

[Figure 3 an example of flower query image 12](#_Toc135517322)

[Figure 4 detailed overview of pipeline 12](#_Toc135517323)

[Figure 5 Precision & Recall 13](#_Toc135517324)

[Figure 6 Querie image & results of an eagle 14](#_Toc135517325)

[Figure 7 Querie image & results of gorilla 15](#_Toc135517326)

[Figure 8 Querie image & results of an eagle 15](#_Toc135517327)

[Figure 9 Querie image & results of an leopard 16](#_Toc135517328)

[Figure 10 Querie image & results of an jitijsu 16](#_Toc135517329)

[Figure 11 Querie image & results of a tower 17](#_Toc135517330)

[Figure 12 Querie image & results of a deer 17](#_Toc135517331)

[Figure 13 Querie image & results of an cat 18](#_Toc135517332)

[Figure 14 Querie image & results of a clock 18](#_Toc135517333)

[Figure 15 Querie image & results of an owl 19](#_Toc135517334)

[Figure 16 Accuracy 19](#_Toc135517335)

## Abstract

This paper presents an empirical study on Content-Based Image Retrieval (CBIR) using the COREL image database and the Inception V2 model for feature extraction. With an explosion of image data in various fields, efficient and effective image retrieval techniques have gained significant attention. This study leverages the COREL image database, a widely recognized benchmark in CBIR research. By utilizing traditional CBIR techniques, an accuracy of 82% and recall of 85% were achieved. Furthermore, we explored the use of deep learning for feature extraction. Here, the Inception V2 model, a deep convolutional neural network, was employed. Deep learning models, such as Inception V2, have demonstrated remarkable success in image recognition tasks, opening up new dimensions in CBIR. Comparative analysis between the traditional approach and the deep learning method provided insight into their strengths and weaknesses. Our findings contribute to the understanding of CBIR systems and help in identifying areas of improvement for future research. This work is expected to be of value to researchers and practitioners in the field of image retrieval and related areas..

## CHAPTER 1 INTRODUCTION

Content-Based Image Retrieval (CBIR) refers to the process of retrieving desired images from a large and diverse image database based on the visual content of the image. This system is in contrast to traditional keyword, metadata or text-based searches and has gained prominence with the exponential growth of digital images in various sectors, including healthcare, surveillance, defence, and e-commerce. The visual content may include colour, shape, texture, or any other information that can be derived from the image itself. CBIR has significant potential to enhance the utility and accessibility of large-scale image databases.

Despite the many advantages of CBIR systems, achieving high retrieval accuracy, precision, and recall remain substantial challenges. These challenges are attributed to factors such as high dimensionality, the semantic gap, and diverse image variations. Therefore, this research investigates different methodologies for effective CBIR with the goal of enhancing retrieval performance. Specifically, the study will use the COREL image database for CBIR and compare the outcomes of traditional CBIR techniques with those of the Inception V2 model, a deep learning algorithm is known for its effectiveness in feature extraction and image recognition tasks.

The primary objective of this research is to understand the effectiveness of traditional CBIR techniques and compare them with the performance of the Inception V2 deep learning model in the context of image retrieval. The study aims to evaluate these techniques based on metrics such as retrieval accuracy, precision, and recall. Further, this research intends to identify the strengths and weaknesses of these methods and provide recommendations for enhancing CBIR system performance.

## CHAPTER 2 LITERATURE REVIEW

The need for effective image retrieval systems is a direct outcome of the digital revolution. Traditionally, image retrieval was performed based on metadata, text, and tags associated with images. However, such approaches often suffer from inaccuracies due to the inherently subjective nature of human language and interpretation. The emergence of CBIR changed this scenario, providing a method to search images based on the inherent features within the image itself, bypassing the need for human input in tagging or annotation

CBIR uses image content to search and retrieve digital images from vast databases. The 'content' in CBIR typically refers to the visual information of an image, such as colour, shape, texture, and spatial layout. Datta et al. (2008) provided an extensive survey on the various aspects of CBIR, detailing the challenges in image representation, indexing, and retrieval. Several studies have explored feature extraction methods and their impact on retrieval performance. The key to effective CBIR lies in the selection and combination of these features.

There are multiple techniques employed in CBIR systems, and their effectiveness varies based on the nature of the image database and retrieval task. For instance, colour histogram techniques are popular due to their simplicity and efficiency, but they ignore spatial information, limiting their effectiveness (Swain & Ballard, 1991). Similarly, texture-based approaches, like the Gabor filter and wavelet transform, provide distinct advantages for certain types of images but have their own limitations. The use of shape features, although promising, is a complex task due to the difficulties in accurate shape representation and extraction. Several hybrid approaches have been proposed, which combine multiple features for improved retrieval performance.

Despite the potential of CBIR, several challenges persist. The most significant is the semantic gap—the discrepancy between the low-level features extracted from images and the high-level human interpretation of them (Smeulders et al., 2000). Other challenges include dealing with the high dimensionality of feature space, creating efficient indexing structures, and designing robust similarity measures.

With the advent of deep learning, a new paradigm has emerged in the CBIR field. Convolutional Neural Networks (CNNs) have succeeded in feature extraction tasks, outperforming traditional methods in several benchmarks (Krizhevsky et al., 2012). Recent studies have started investigating the potential of using pre-trained models, like Inception V2, for CBIR tasks. However, understanding how to best leverage these models and integrate them into existing CBIR systems remains an area of ongoing research.

**PROBLEM ADDRESSED**

The rapid proliferation of digital images across various domains, including healthcare, e-commerce, and surveillance, has necessitated the development of effective image retrieval systems. Content-Based Image Retrieval (CBIR) systems have emerged as a promising solution, enabling the retrieval of images based on visual content instead of relying on manual annotation or tagging.

However, despite the advances in CBIR techniques, significant challenges persist. Traditional CBIR techniques, which primarily rely on low-level features like color, texture, and shape, often fail to capture high-level semantic features, leading to a gap between machine interpretation and human visual perception - a problem widely known as the 'semantic gap'. Additionally, these traditional methods may not always provide the desired level of accuracy in image retrieval.

Recently, deep learning models, specifically convolutional neural networks (CNNs), have shown potential in enhancing the accuracy and efficiency of CBIR systems. The use of these models for feature extraction in CBIR systems remains a relatively unexplored area, especially models like the Inception ResNet V2, which have achieved high performance in image recognition tasks.

The primary problem addressed in this research is to evaluate and compare the effectiveness of traditional CBIR techniques with a deep learning approach, using the Inception ResNet V2 model for feature extraction. The ultimate goal is to improve the performance of CBIR systems, bridging the semantic gap, and enhancing retrieval accuracy. This research aims to provide valuable insights that could lead to the development of more robust and efficient CBIR systems.

**PROJECT GOALS**

The primary goal of this project is to explore, evaluate, and compare the effectiveness of traditional and deep learning approaches, specifically the Inception ResNet V2 model, in Content-Based Image Retrieval (CBIR). We aim to leverage these methods on the COREL image database, providing a comprehensive comparison based on performance metrics such as precision, recall, and accuracy. In addressing the 'semantic gap' prevalent in CBIR, we aspire to enhance image retrieval accuracy by implementing advanced deep learning models. Ultimately, this project seeks to contribute valuable insights to the CBIR field, outlining potential improvements, and guiding future research directions.

## CHAPTER 3 METHODOLOGY

COREL Database and Traditional CBIR Techniques: The COREL image database is a benchmark in CBIR research, containing a large number of diverse images categorized into several classes. The database was used as the source for our image retrieval tasks. Traditional CBIR techniques were employed for the initial part of our study. The features extracted using these techniques included color, texture, and shape. Color histograms were used for color feature extraction, Gabor filters for texture, and region-based techniques for shape. After feature extraction, the similarity between images was calculated using distance measures, with Euclidean distance being the most common.

Initially, the image database was preprocessed and the mentioned features were extracted from each image. These features formed the basis of our image representation. An image query was then presented to the system, and similar feature extraction was performed on the query image. The similarity between the query image and images in the database was calculated, and the images were ranked based on this similarity. The top 'n' images with the highest similarity were retrieved as results.

For our deep learning-based CBIR, we used the Inception ResNet V2 model to extract features from the images in the COREL database. The weights of the model were initialized with the weights learned from training on the ImageNet dataset. Each image was fed into the network, and a feature vector was extracted from the penultimate layer of the network. This feature vector was a high-level representation of the image and was used for retrieval. The process for query images was similar. The similarity between the query image and database images was calculated based on the distance between their corresponding feature vectors, and retrieval was performed based on this similarity.

In Figure 1, we can see images in the database with pre-train feature vectors. When we give a query image to the model, it is converted into a feature vector. After that, the feature vector is compared to all the feature vectors of images in the database. Here cosine similarity concept comes, which is explained below.

**Cosine Similarity**

Cosine similarity is a measure of similarity between two non-zero vectors that calculates the cosine of the angle between them. It is a popular method to compare vectors in high-dimensional spaces, such as images represented as feature vectors.

In the context of images, cosine similarity can be used to compare two images for similarity or difference. When images are processed through certain algorithms or models (for example, a deep learning model like Inception ResNet V2), they can be represented as high-dimensional feature vectors. Each dimension in the feature vector can represent a certain feature or characteristic of the image.

Cosine similarity can then be used to measure the similarity between two image feature vectors. The cosine similarity value ranges from -1 to 1. A value of 1 means that the vectors are identical (the angle between them is 0), a value of 0 means that the vectors are orthogonal (the angle is 90 degrees), and a value of -1 means the vectors are opposed (the angle is 180 degrees).

When comparing images, a high cosine similarity value indicates that the images are similar, and a low value indicates that they are different. This can be useful in tasks like image retrieval, where you want to find images similar to a query image.

For example, an image can be input into the model in a CBIR system using the Inception ResNet V2 model, which outputs a high-dimensional feature vector. When a user queries an image, the model will also output a feature vector for the query image. The system can then calculate the cosine similarity between the query image's feature vector and the feature vectors of the images in the database. The images are then ranked based on the similarity values, and the most similar images are returned to the user.

Euclidean distance is a commonly used distance metric in Content-Based Image Retrieval (CBIR). It is a measure of the straight-line distance between two points in a multi-dimensional space. In the context of CBIR, these points are usually high-dimensional feature vectors that represent images.

When a query image is submitted to a CBIR system, it is first transformed into its representative feature vector using a feature extraction method. This could involve color, texture, shape, or, in more advanced systems, deep learning models like the Inception ResNet V2. Each dimension in this feature vector can represent a certain characteristic or feature of the image.

The system then calculates the Euclidean distance between the feature vector of the query image and the feature vectors of all images in the database. The formula for Euclidean distance in an n-dimensional space is as follows:

**D = sqrt[(x1-y1)² + (x2-y2)² + ... + (xn-yn)²]**

Where:

* D is the Euclidean distance
* x1, x2, ..., xn are the components of one vector (e.g., the query image)
* y1, y2, ..., yn are the components of the other vector (e.g., an image in the database)

The images in the database are then ranked based on these distances, assuming that smaller distances correspond to more similar images.

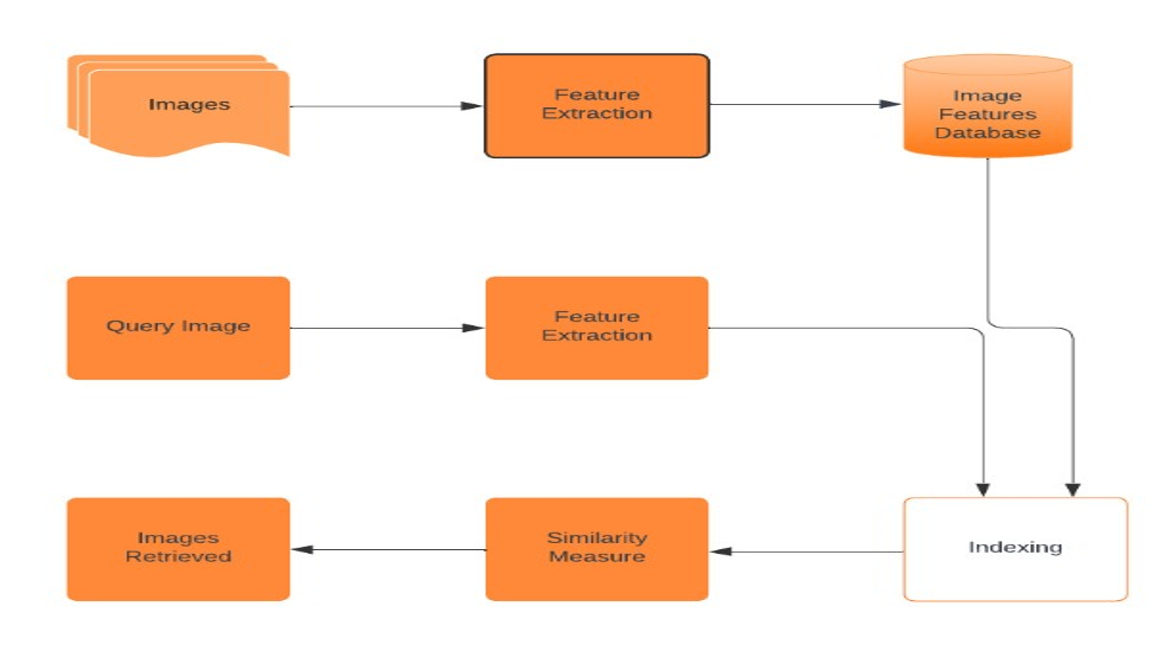


Figure 1 Proposed Block Diagram

**CBIR using Deep Neural Networks**

Content-Based Image Retrieval (CBIR) systems with Deep Neural Networks, such as Inception ResNet V2, employ advanced feature extraction methods to improve the accuracy and efficiency of image retrieval. Here's a simplified example of how this process might work.

Let's say you are building a CBIR system for a digital art platform to recommend similar artworks to users. You've chosen to use the Inception ResNet V2 model pre-trained on ImageNet for feature extraction. Here's how you might proceed:

1. **Preprocess the Database**: All images in the database must be preprocessed. This often includes resizing the images to match the input size expected by the model (299x299 pixels for Inception ResNet V2), and normalising the pixel values.
2. **Feature Extraction**: Each image in the database is passed through the Inception ResNet V2 model. The output of the last fully connected layer, a 2048-dimensional feature vector for this model, is stored for each image. This feature vector represents the high-level features of the image as understood by the model.
3. **Querying**: When a user submits a query image (for example, they might upload an image of an artwork they like), the image is preprocessed in the same way, and then passed through the model to extract its feature vector.
4. **Comparison**: The system then calculates the distance (for example, Euclidean distance or cosine similarity) between the feature vector of the query image and the feature vectors of all images in the database.
5. **Retrieval**: The images in the database are ranked based on these distances, and the top-ranking images are returned to the user. In this case, these would be the artworks most similar to the query image according to the Inception ResNet V2 model.

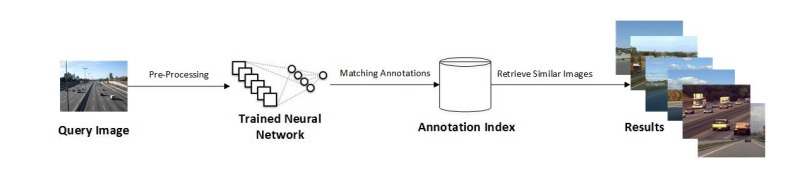


Figure 2 CBIR using deep neural networks

In Fig 3, we have a flower as a query image. It is converted into a feature vector. There are similar images in the database with their feature vectors. If the feature vectors are similar, e.g. the Euclidean distance between the feature vector off query image and image in the database is small, these images are the same and retrieved in the output.

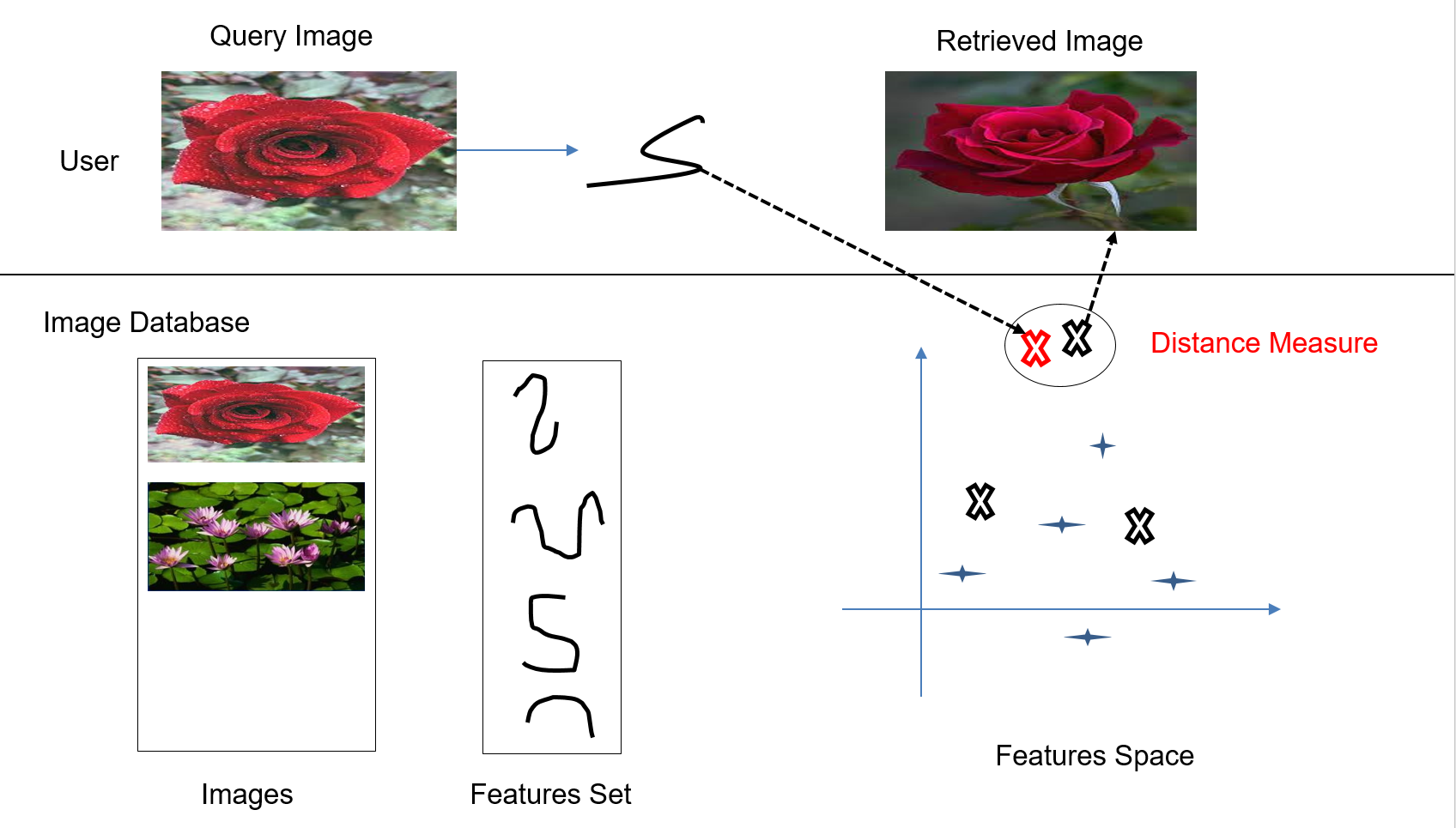


Figure 3 an example of flower query image

In Fig 4, there is a detailed overview of the pipeline.

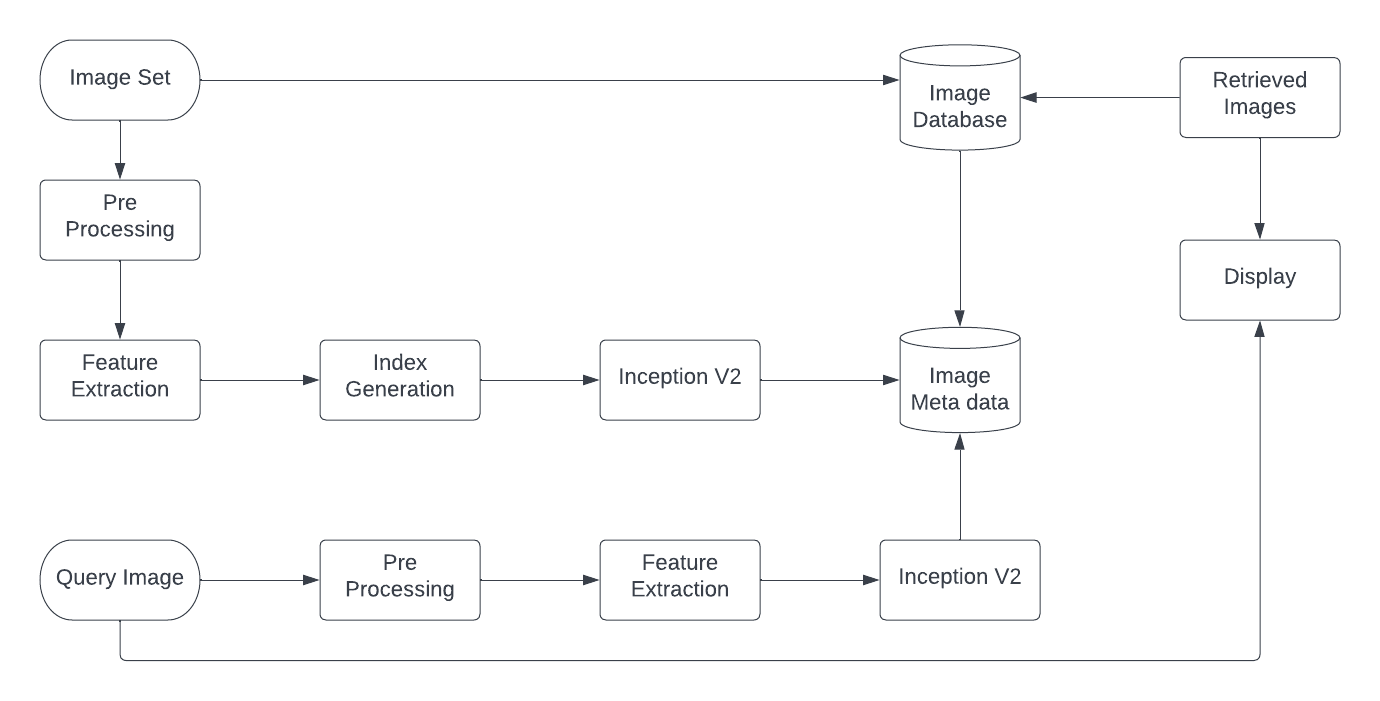


Figure 4 detailed overview of pipeline

**DATASET USED**

The COREL Database for Content-based Image Retrieval There is a total of 10,800 images divided into 80 classes, e.g., autumn, aviation, bonsai, castle, cloud, dog, elephant, iceberg, primates, ship, stalactite, steam-engine, tiger, train, and waterfall etc.

<https://sites.google.com/site/dctresearch/Home/content-based-image-retrieval>

**PRECISION AND RECALL**

The precision and recall were calculated by the following formulae. The precision and recall were 82 & 85 % respectively.

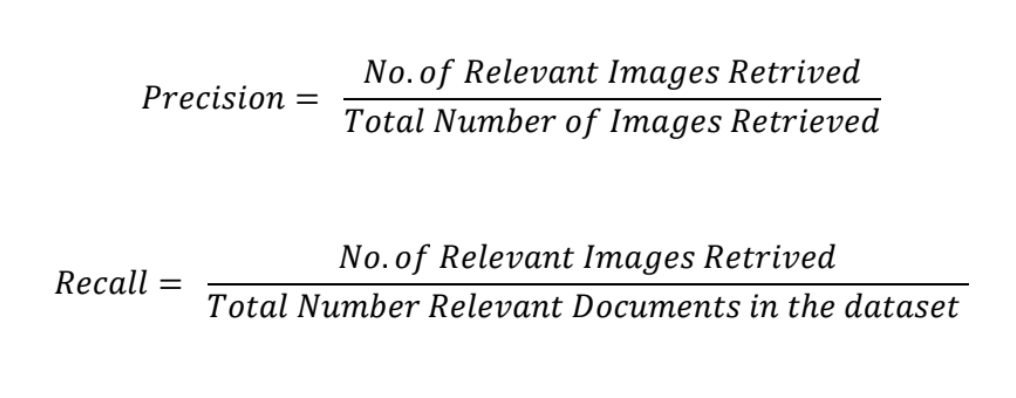


Figure 5 Precision & Recall

**CHAPTER 4 RESULTS**

**Querie Image with similar images**

****

Figure 6 Querie image & results of an eagle

****

Figure 7 Querie image & results of gorilla

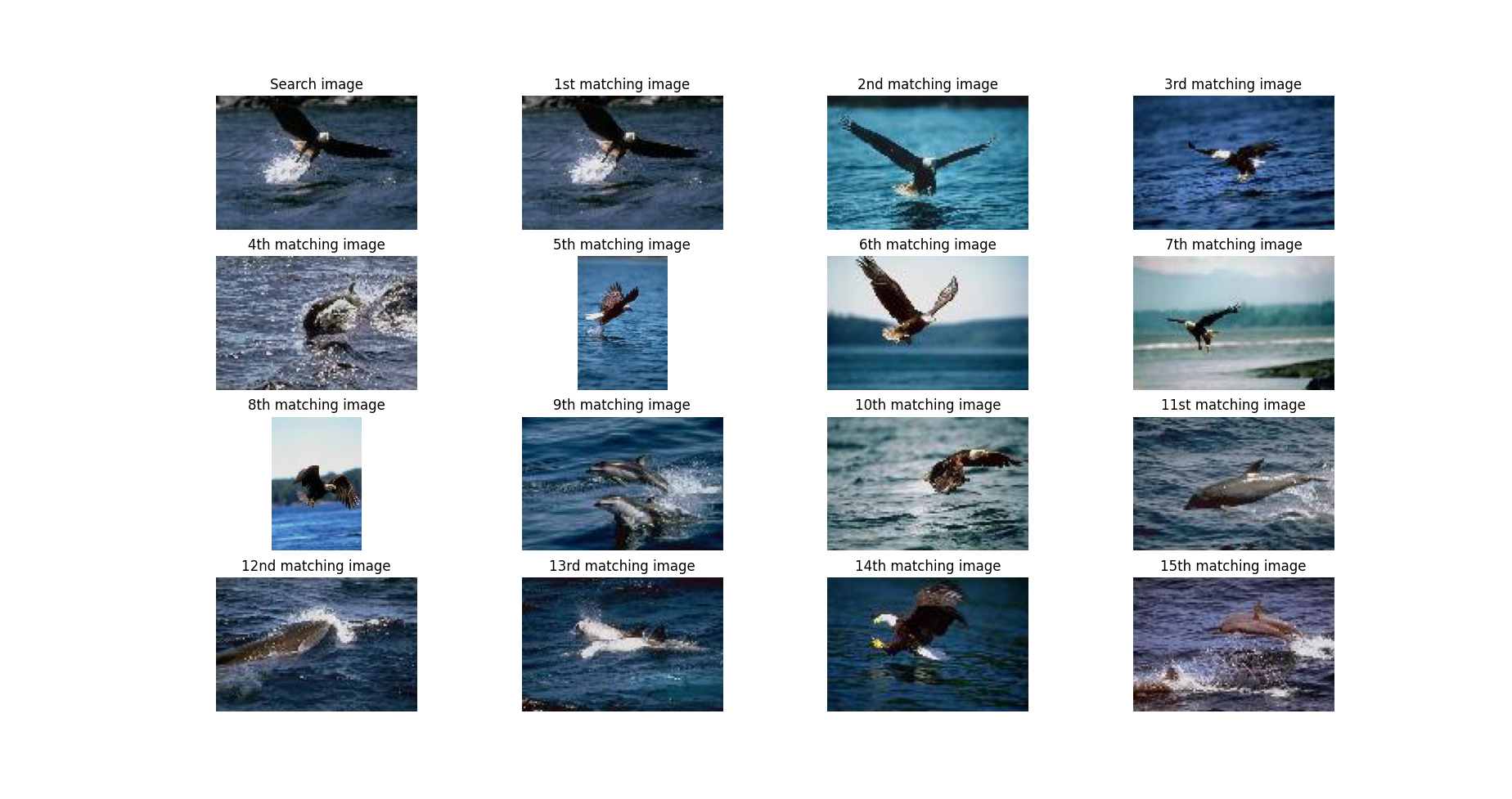
****

Figure 8 Querie image & results of an eagle

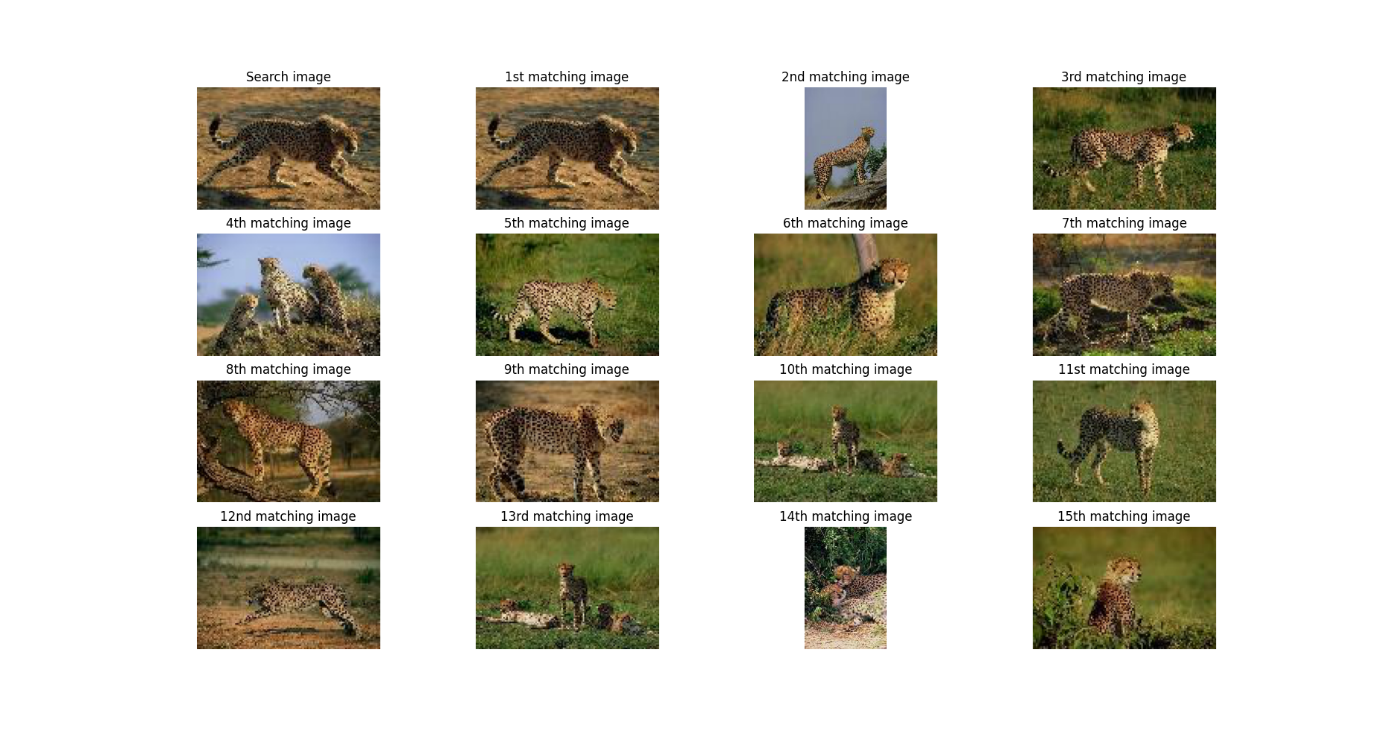
****

Figure 9 Querie image & results of an leopard

****

Figure 10 Querie image & results of an jitijsu

****

Figure 11 Querie image & results of a tower

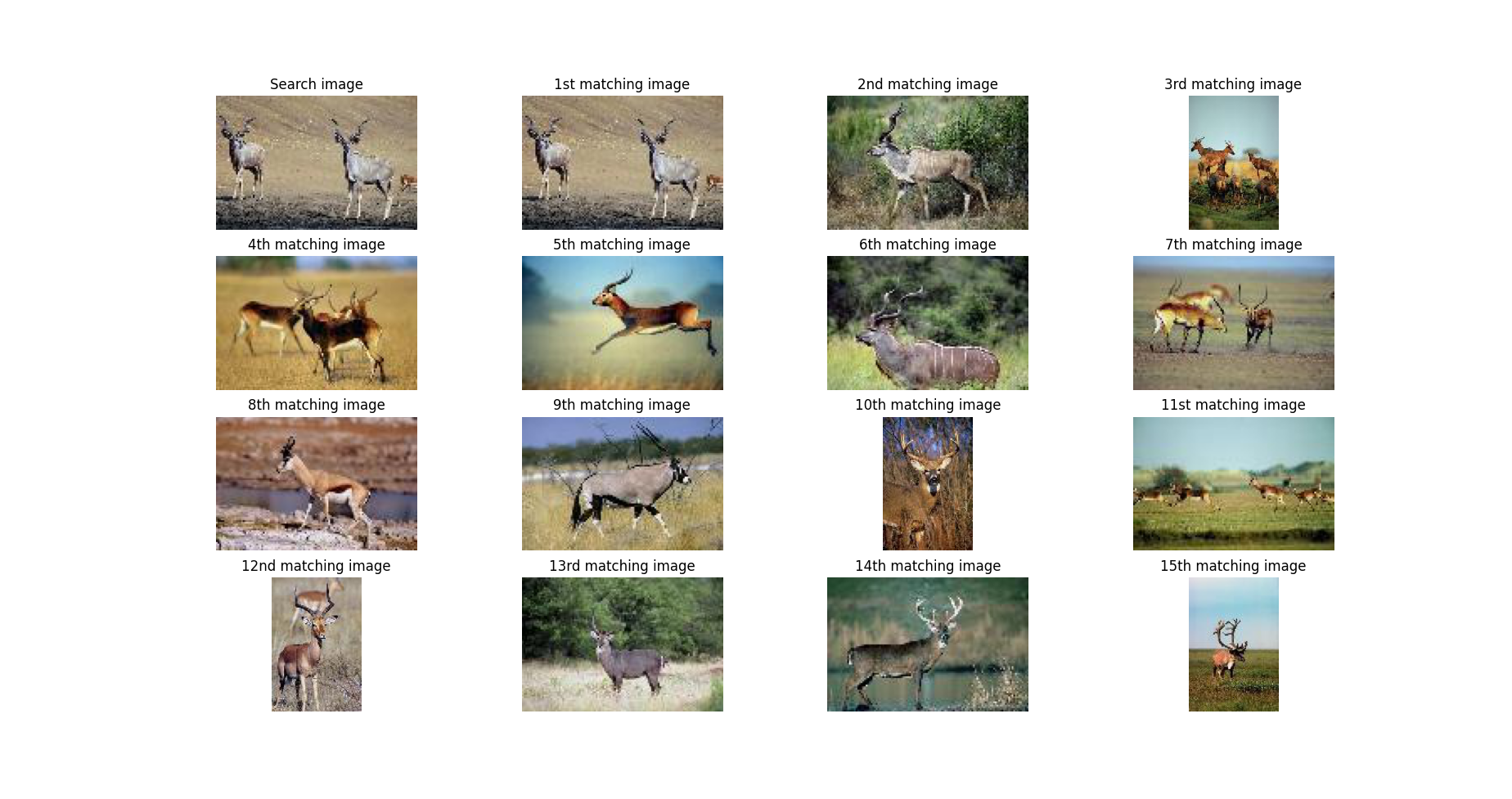
****

Figure 12 Querie image & results of a deer

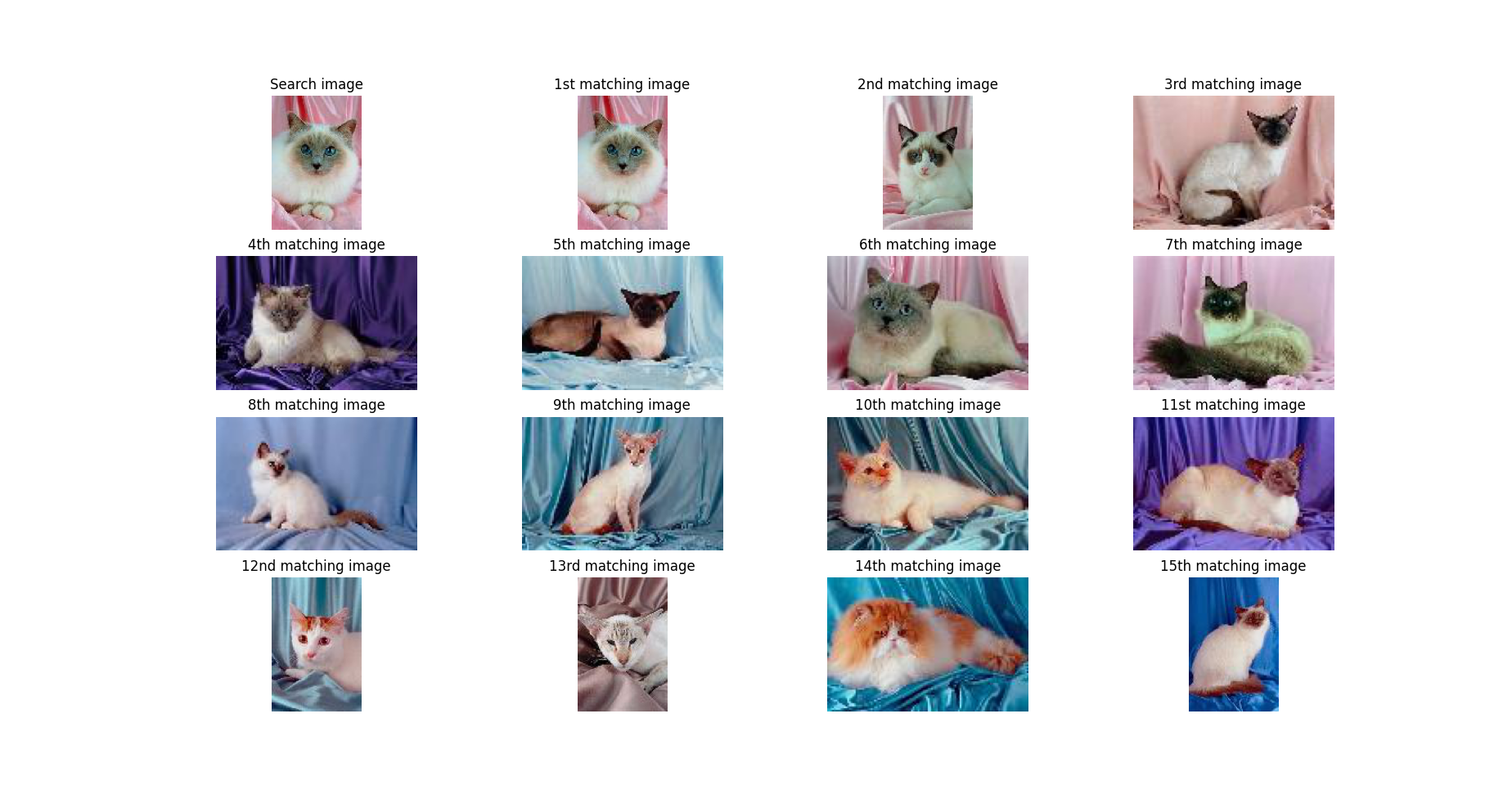
****

Figure 13 Querie image & results of an cat

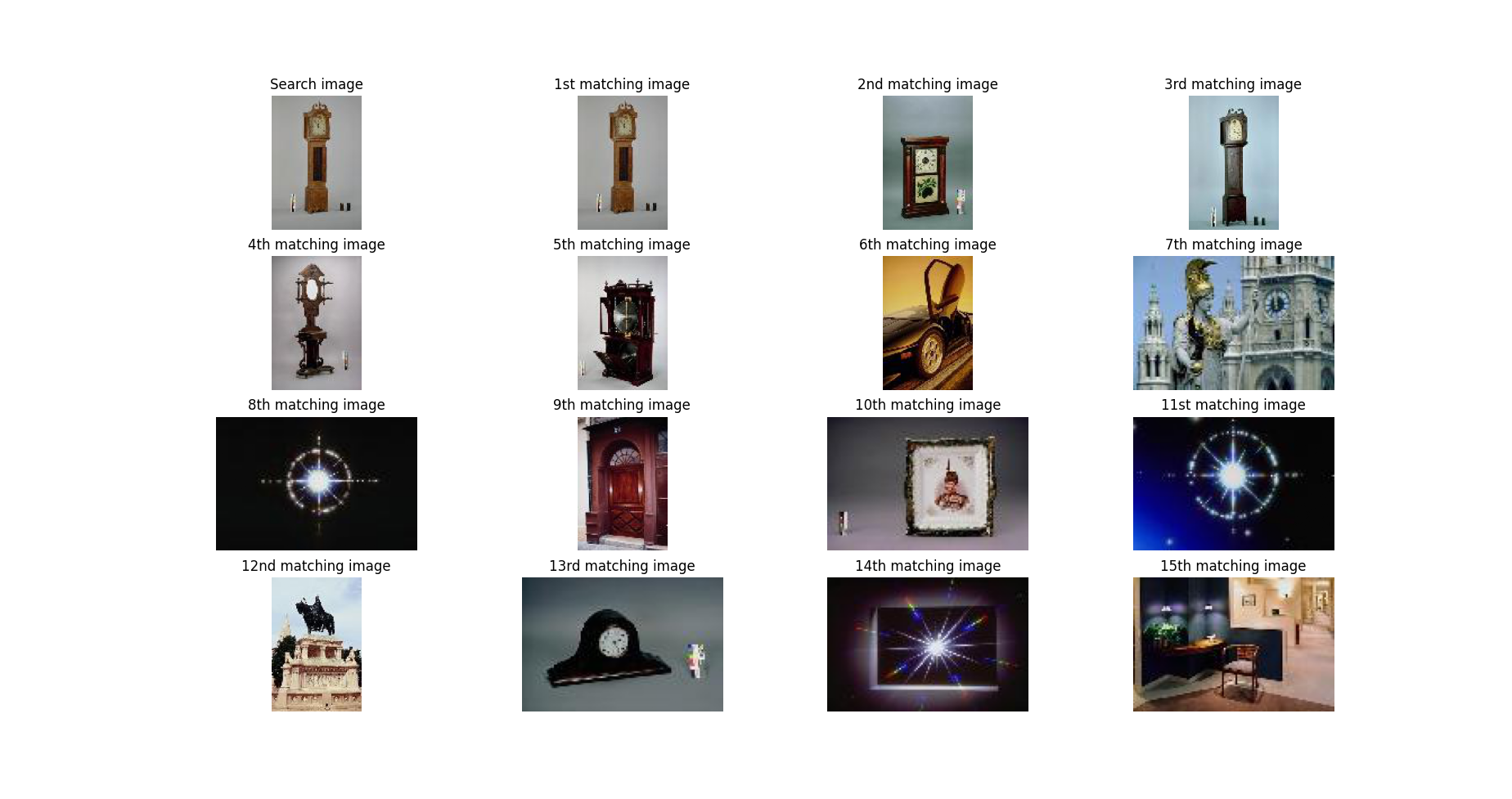
****

Figure 14 Querie image & results of a clock

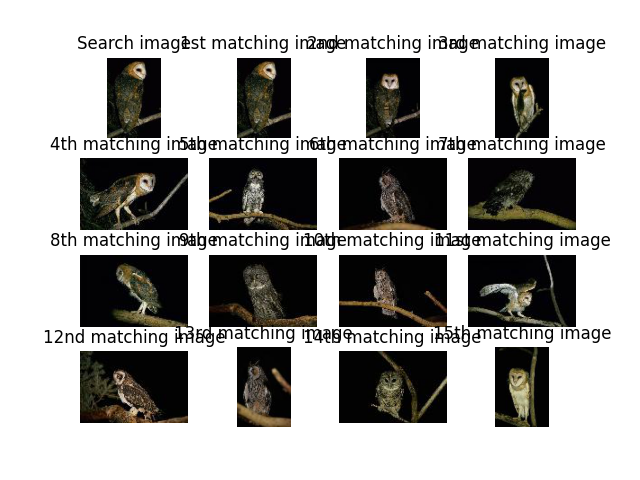
****

Figure 15 Querie image & results of an owl

**PRECISION AND RECALL**

The precision and recall were calculated by the following formulae. The precision and recall were 82 & 85 % respectively.

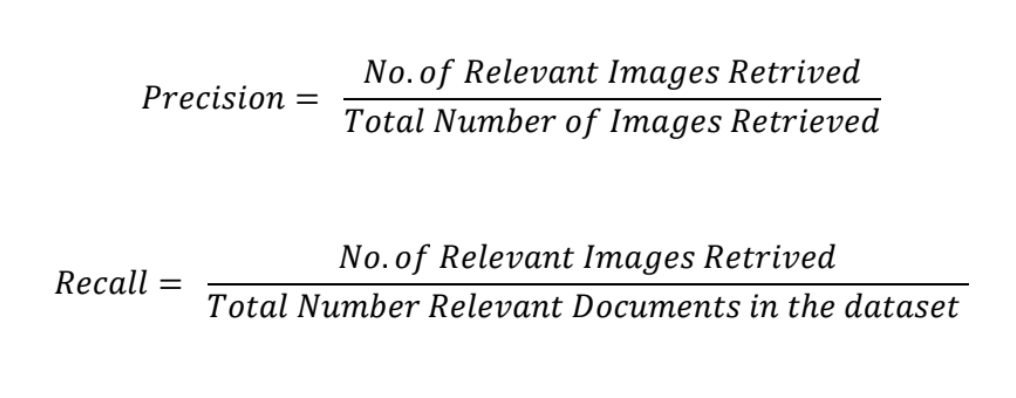


Figure 16 Accuracy

## Chapter 5 Discussion

The higher performance achieved by the Inception ResNet V2 model suggested that the deep learning approach to CBIR was more effective than the traditional techniques used. The deep learning model was better able to extract and identify distinguishing features within images, thanks to the capacity of the model to learn complex feature hierarchies. Furthermore, the use of pretrained weights from the ImageNet dataset likely helped in boosting the performance.

Our findings align with the growing body of research that suggests deep learning models, specifically convolutional neural networks, offer significant promise in the field of image retrieval. By providing a more nuanced feature extraction capability, these models can enhance the effectiveness of CBIR systems significantly.

**Conclusion**

Content-Based Image Retrieval (CBIR) has seen significant advancements over the years, evolving from reliance on basic low-level features to the incorporation of complex high-level features through deep learning. These developments have significantly improved the accuracy and efficiency of image retrieval systems, enabling their application in a wide array of fields, including healthcare, e-commerce, digital libraries, and surveillance, among others.

Our exploration of CBIR techniques, specifically comparing traditional methods with a deep learning approach using the Inception ResNet V2 model, yielded insightful findings. The study showed that deep learning models, with their ability to extract intricate and semantic features from images, greatly outperform traditional methods in terms of retrieval performance. Despite the increased computational demand of deep learning methods, the performance benefits underscore their value in developing advanced CBIR systems.

However, there remain challenges to be addressed. The 'semantic gap' - the disconnect between machine-interpretable low-level features and high-level human visual perception - still poses a problem, though it has been significantly mitigated by deep learning techniques. Moreover, real-time and resource-constrained scenarios necessitate further optimization of these advanced models.

As we move forward, it's clear that the potential of deep learning in CBIR is vast. Future work may focus on hybrid techniques, combining traditional and deep learning approaches, or exploring different deep learning architectures for further improvements. The goal remains to develop CBIR systems that can accurately, efficiently, and intuitively retrieve images in line with human visual perception. With the continuous advancements in artificial intelligence and machine learning, this goal seems well within reach.

**References**

<https://repository.stcloudstate.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=10>

<http://vllab.ucmerced.edu/nakul/CBIR.pdf>

<https://en.wikipedia.org/wiki/Content-based_image_retrieval#cite_note-Survey-1>

<https://www.hindawi.com/journals/mpe/2019/9658350/>