Department of Electrical and Computer Engineering
North South University

Senior Design Project

Nearest Neighbor of Different Image Representation

Sajid Wasif

ID# 1912313642

Faculty Advisor:

Dr. Mohammad Ashrafuzzaman Khan

Assistant Professor

ECE Department

Spring, 2023

LETTER OF TRANSMITTAL

June, 2023

To

Dr. Rajesh Palit
Chairman,
Department of Electrical and Computer Engineering
North South University, Dhaka

Subject: **Submission of Capstone Project Report on "Nearest Neighbor of Different Image Representation"**

Dear Sir,

With due respect, I would like to submit my **Capstone Project Report** on **"Nearest Neighbor of Different Image Representation"** as a part of my BSc program. This project was very much valuable to me as it helped me gain experience from practical field and apply in real life. I tried to the maximum competence to meet all the dimensions required from this report.

I will be highly obliged if you kindly receive this report and provide your valuable judgment. It would be my immense pleasure if you find this report useful and informative to have an apparent perspective on the issue.

Sincerely Yours,

....

Sajid Wasif
ECE Department
North South University, Bangladesh

APPROVAL

Sajid Wasif (ID # 1912313642) from Electrical and Computer Engineering Department of North South University, have worked on the Senior Design Project titled "**Nearest Neighbor of Different Image Representation**" under the supervision of Dr. Mohammad Ashrafuzzaman Khan partial fulfillment of the requirement for the degree of Bachelors of Science in Engineering and has been accepted as satisfactory.

Supervisor's Signature

……………………………………….

Dr. Mohammad Ashrafuzzaman Khan

Assistant Professor

Department of Electrical and Computer Engineering

North South University

Dhaka, Bangladesh.

Chairman's Signature

……………………………………….

Dr. Rajesh Palit

Professor

Department of Electrical and Computer Engineering

North South University

Dhaka, Bangladesh.

DECLARATION

This is to declare that this project is my original work. No part of this work has been submitted elsewhere partially or fully for the award of any other degree or diploma. All project related information will remain confidential and shall not be disclosed without the formal consent of the project supervisor. Relevant previous works presented in this report have been properly acknowledged and cited. The plagiarism policy, as stated by the supervisor, has been maintained.

Students' names & Signatures

1. Sajid Wasif

_ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

ACKNOWLEDGEMENTS

I am grateful to our esteemed supervisor, Dr. Mohammad Ashrafuzzaman Khan, Assistant Professor, Department of Electrical and Computer Engineering, North South University, Bangladesh, for his invaluable support, precise guidance and advice pertaining to the experiments, research and theoretical studies carried out during the course of the current project and also in the preparation of the current report. His expert guidance and insightful feedback have played a crucial role in enhancing the quality and scope of my work.

ABSTRACT

Nearest Neighbour of Different Image Representation

This project is a system designed to utilize pre-trained models and feature extraction techniques to identify similarities between images retrieved from various sources. The system calculates distances using distance metrics by collecting and computing feature vectors from a wide range of datasets. The system shows the top 10 images as output based on the highest similarity score as well as reveals the hidden connections between the images and illustrates the potential for cross-domain image similarity. This project can be utilized in fields like content-based image retrieval, cross-domain analysis, and creative exploration.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1 Introduction

The nearest neighbor algorithm is a widely utilized technique in image processing and computer vision for a variety of tasks, such as image retrieval and classification. In the field of image processing, finding the image(s) that are most similar to a given query image based on some predetermined distance metric is usually referred to as finding the nearest neighbor of an image. The method selects the image(s) with the highest rate of visual similarity between the image being searched and a database of reference images. Typically, a distance metric like Euclidean distance or cosine similarity is used to calculate the similarity.

Our project's main objective is to use the extracted features to create a system that allows for similarity search. After taking an input image, our system quickly finds and retrieves the ten closest images based on the similarity of their features and enabling effective content-based image retrieval. Here, we explore a number of innovative CNN models for feature extraction from a given dataset, including VGG16, ResNet101, ZFNet, and MobileNet. Firstly, we proceed with the research project by collecting the required dataset, which consists of a diverse collection of datasets such as persons' activity, different flowers, natural scenes, animals, or several kinds of objects. Then We use multiple CNN architectures such as VGG16, ResNet101, ZFNet, and MobileNet to extract features from images after we obtain the dataset.

We proceed with images through the pre-trained CNN models to extract features, utilizing the learning weights and biases obtained from training on large-scale image classification tasks. We get access to the intermediate feature maps, which serve as rich representations of image features, by eliminating the completely connected layers towards the end of the models. These feature maps are then distorted into 1D arrays to generate a concise representation of the visual content of the images.

Finally, we compared the feature representation of a query image to the entire dataset using the similarity search algorithm. To efficiently compute similarity, we utilize some kinds of distance metrics, including the Euclidean distance and cosine similarity. Then, among images with the highest number of accurate feature representations, the 10 most similar images were displayed in the result.

1.1 Image to Vector

"Image-to-vector" indicates the process of transforming an image into a numerical vector representation. In general, it involves taking the visual data from a picture and turning it into a structured, manageable numerical representation that computers can process quickly. Features extraction, similarity search, and classification are some of these tasks.

In our project the steps we followed to generate vector are:

- Dataset Preparation: we obtained a dataset of images. (Partial ImageNet dataset)
- Pre-trained Models: We used pre-trained versions of ZFNet, MobileNet, ResNet101, and VGG16.
- Image Preprocessing: Images were resized and normalized.
- Feature Extraction: The models processed images to capture various levels of visual features.
- Removing Fully Connected Layers: Fully connected layers, specific to classification tasks, were removed.
- Flattening and Pooling: Intermediate feature maps were flattened or pooled to obtain 1D arrays.

,

1.2 Feature Extraction

To represent and learn essential features from images, we used CNN models that have already been trained, including ResNet-101, VGG16, ZFnet, and Mobilenet. After being trained on large datasets, these algorithms obtained the ability to recognize complex patterns and visual representations. As feature extractors, we used the pre-trained models rather than creating a CNN model from scratch on our relevant dataset. Here are the number of features extracted from an image.

```
1/1 [==============================] - 7s 7s/step
features successfully extracted!
number of image features: 62720
array([[0., 0., 0., ..., 0., 0., 0.]], dtype=float32)
```

Figure 1: Number of Features extracted from an image

We used pre-trained models to pass the input photos through the network, and subsequently, we retrieved the activations from either the final layer or one of the intermediate layers. These activations, which acted as feature vectors, represented the main features and structure of the input images. The retrieved feature vectors were converted into numerical representations in order to ease similarity calculations. We are able to compare and measure the visual qualities of various images using these representations based on how comparable their features are.

The feature vectors act as a simplified version of the images, allowing for quick and efficient retrieval of related images from the collection. Here this figure shows how many features have been successfully extracted.

```
79/79 [==============================] - 4s 37ms/step
features successfully extracted!
(2510, 62720)
```

Figure 1: successfully extracted

1.3 Nearest neighbor of an image

We extracted characteristics from our partial dataset using various CNN models. The ability of these models to identify significant patterns and characteristics in images is well known. The retrieved features were converted into 1D arrays or feature vectors, which are numerical representations of each image's attributes. This conversion enabled fast computation and comparison of feature vectors. The feature vector was extracted from an input image submitted by the user using the same feature extraction procedure. To compare the similarity of the feature vectors, we utilized the cosine similarity metric and the Euclidian distance. They capture the orientation and direction of the feature vectors, resulting in a similarity between the images.

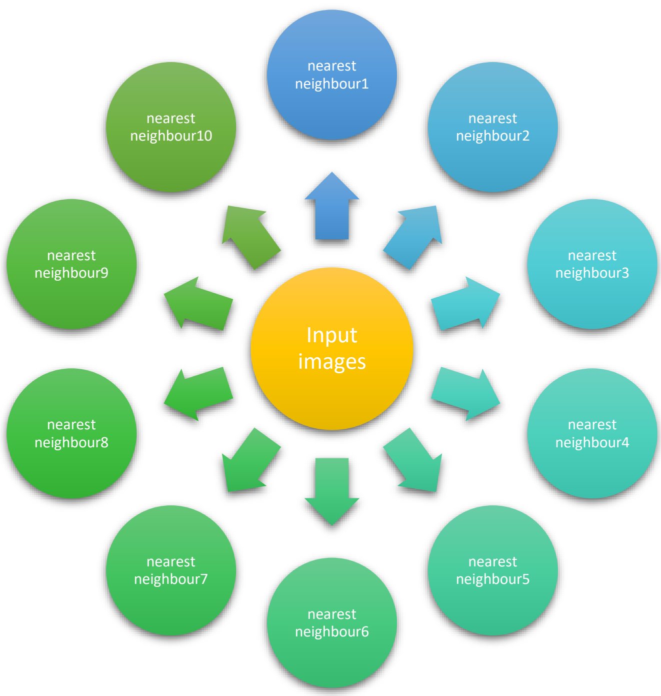


Figure 2: 10 nearest neighbours

Using the cosine similarity scores, we conducted a closest neighbor search to find the photos that were most similar to the query image. The top ten images with the highest cosine similarity scores were selected as the closest neighbors. The system output displayed the 10 nearest images, providing the user with images that were aesthetically comparable to the query image. This allowed users to explore and find related images based on their preferences or interests.

1.4 Euclidean distance

The Euclidean distance is the distance between two points in a multidimensional space measured along a straight line. The Euclidean distance was used as a metric in our research to assess the degree of similarity between feature vectors derived from images. Calculating the Euclidean distance between feature vectors, we were able to quantify the differences in feature representations across images. Images with reduced Euclidean distances were thought to be more comparable because they shared similar feature patterns. Images with greater Euclidean distances, on the other hand, were considered less comparable, because of differences in their feature patterns. Here some of our Euclidean distance result:

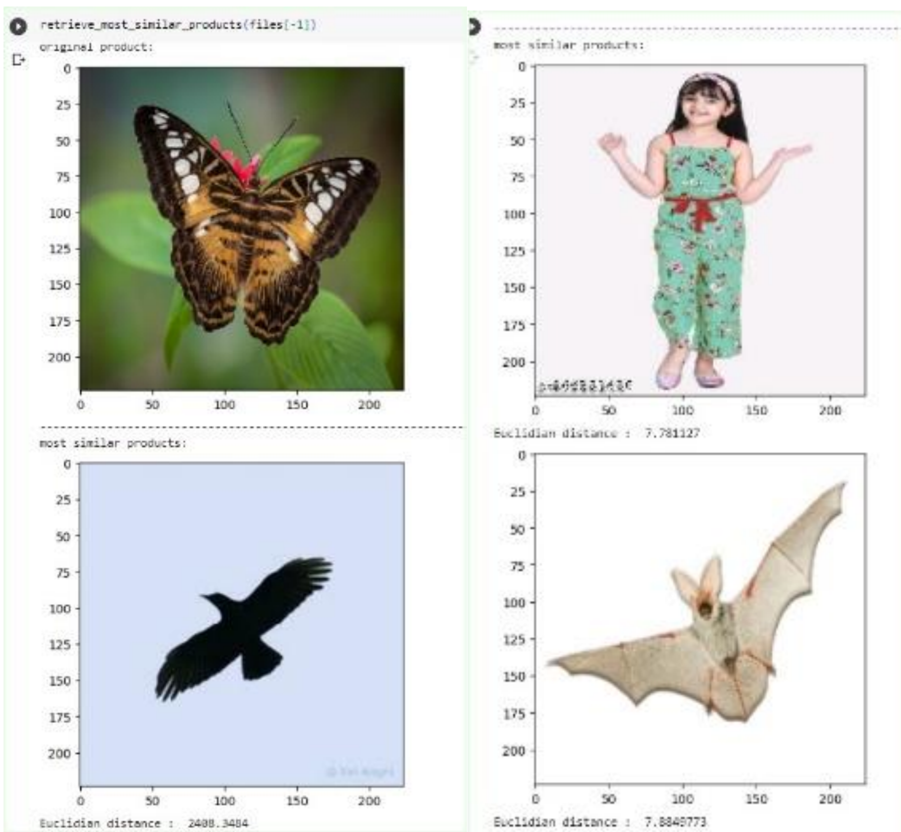


Figure 3: Euclidean distance result

Ultimately, the Euclidean distance was a useful tool in our research for analyzing feature vectors and measuring image similarity.

1.5 Cosine similarity

Cosine similarity was used as a metric in our experiment to assess how similar feature vectors generated from photos were to one another. Cosine similarity computes the cosine of the angle between two vectors to determine how similar their directions are in a multidimensional space. Finding the cosine similarity between feature vectors, we were able to assess the similarity of images based on the position of their feature patterns. Images with higher cosine similarity scores were thought to be more comparable because their feature vectors were closer in direction. Images with lower cosine similarity scores, on the other hand, showed greater variation in their feature patterns. Here some of our Cosine similarity Score:

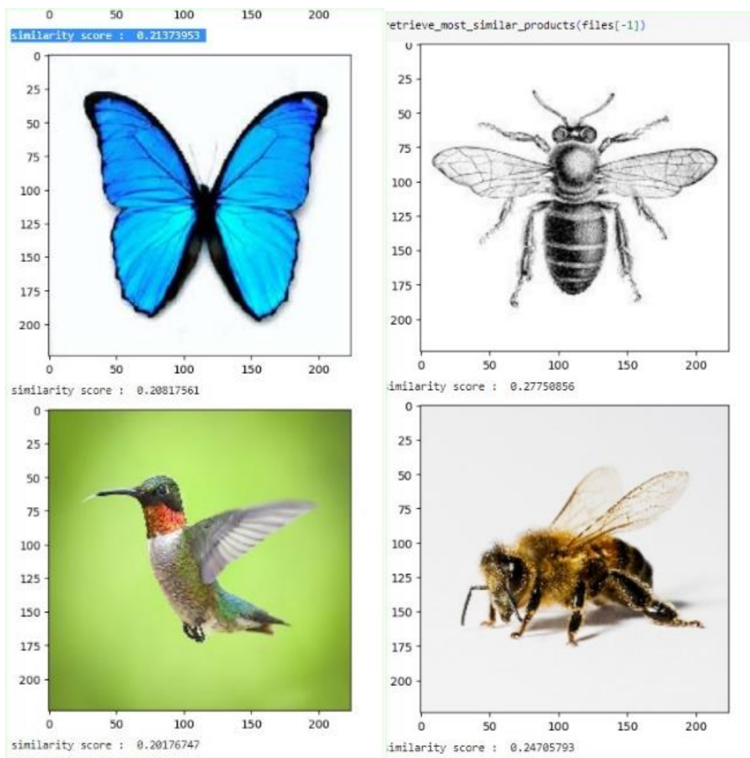


Figure 4: Cosine similarity Score

Using cosine similarity, we were able to capture the semantic similarities between the photographs despite their dimensions or the scale. It is very useful when working with high-dimensional feature spaces because it focuses on the alignment of the feature vectors rather than their exact values.

In our research, Cosine similarity is a useful indicator for identifying visually related photos, improving similarity search systems and facilitating the discovery of visually related images.

1.6 Motivations

Strong interest in computer vision and the useful applications of image analysis and similarity search served as the inspiration for this research. How computers can comprehend and evaluate visuals captivated us. We were intrigued by the developments in deep learning, particularly CNNs, because of how well they performed on image-based tasks. We were interested in learning more about their potential for similarity search and getting practical expertise in this area. We were eager to find out more and add to the body of knowledge in computer vision. The motivation for this project ultimately came from our love of the topic, the practical applications, and our own personal objectives.

1.7 Aims and Objectives

Our project's main objective is to research and put into practice a similarity search engine that locates visually comparable photos within a collection using CNN-based feature extraction.

Objectives:

- The feature extraction and similarity search will be conducted using this dataset as the foundation.
- To extract feature vectors from the photos in the dataset, use pre-trained CNN models like VGG16, ResNet101, ZFNet, and MobileNet.
- Employ a closest neighbor search strategy with the feature vectors that were extracted.
- Assess and examine the similarity search system's effectiveness.
- Discuss the advantages and disadvantages of the findings produced by the various CNN models. User-friendly presentation of the similarity search system's findings.
- Create a visual interface or a method to show the query image and the top-k nearest neighbor photos so that people may effectively explore and see them. Produce a thorough report detailing the study, its methodology, experimental design, findings, and analysis.

Chapter 2 Research Literature Review

A literature review is a summary of previous research on a particular topic. Reviewing existing research in a particular field of study is the goal of a literature review. Learning about our field of study which is image recognition is enhanced by conducting a literature review. However, we're looking forward to read three machine learning based approaches that's relevant to our project. This knowledge will be presented in a written report in this chapter.

2.1 Related Paper-1

Trishen Munisami et al. [1] identifying plants by using the image of their leaves. They use a server that performs pre-processing and feature extraction techniques on the image before a pattern matcher matches the information in the images to the information in the database. They extracted the length and width of the leaf, its perimeter, its area, and a distance map along the vertical and horizontal axes, among other properties. They used the nearest neighbor classifier to implement and test it on 640 leaves belonging to 32 different species of plants. Their rate of accuracy was 83.5%. They improved their recognition accuracy by 87.3% by employing the color histogram.

2.2 Related Paper-2

In the second paper named "Local Aggregation for Unsupervised Learning of Visual Embeddings", Chengxu Zhuang et al. [2] used a neural network or Local Aggregation (LA) method which nonlinearly embed inputs in a smaller space and they identified close neighbors and background neighbors. They identified two sets of neighbors for an xi and its embedding vi. They used Bi for Nearest-neighbor based identification and Ci for Robustified clustering-based identification, to identify close neighbors, they applied an unsupervised clustering algorithm. They Followed the methods of AlexNet and VGG16 architectures, to add batch normalization (BN) layers in their experiment. They used K-nearest neighbor (KNN) classification results using the embedding features. With all methods, Local Aggregation (LA) performs much better than alternative methods. LA trained ResNet-50 achieves 60.2% top-1 accuracy on ImageNet classification. For the LA method, they consistently see performance gains from both overall deeper structures and from early layers to deeper layers within an architecture. Rather than using AlexNet we use resnet101v2 for our project.

2.3 Related Paper-3

In the third paper named "NEAREST NEIGHBOUR STRATEGIES FOR IMAGE UNDERSTANDING", This paper was published by Sameer Singh students of University of Exeter EX4 4PT, UK. They use Nearest neighbor methods provide an important data classification tool for recognizing object classes in pattern recognition domains. Their main objective of this paper is to develop two versions of the nearest neighbour method. First model will resolve conflicts in the k-nearest neighbour rule, second is closest average distance of samples of classes involved. They described traditional nearest neighbour rule for their recognition. The results shown in their paper for our two nearest neighbour models are extremely encouraging.

2.4 Related Paper-4

Firstly, a paper named "Image reconstruction from ResNet101 semantic feature vector" was read. This paper was published by V´ıt LIST´IK1, Dept. of Cybernetics, Faulty of Electrical Engineering, Czech Technical University in Prague. V´ıt LIST´IK et al [1] want to prove that although it is possible to reconstruct the image from the semantic feature vector. The task is to generate an image from the semantic feature vector which will be very similar to the original image. The task is the same as for autoencoder with a difference of using pre-trained CNN. They used only the images, not the labels. The original dataset consists of 14M labeled images. They are using random subsets of the dataset. They used pre-trained ResNet101 for the extraction. Based on their results they concluded that is not possible to reconstruct the private information. Using this method, they were able to accurately identify 95.2 percent. Rather than using reconstruct the image from the semantic feature vector cosine similarity was our choice for this project.

2.5 Related Paper-5

Maize leaf disease is classified by Mohammad Syarief et al. [4] using the seven CNN models AlexNet, VGG16, VGG19, GoogleNet, Inception-V3, residual network 50 (ResNet50), and ResNet101. They used 200 photos divided into four classes in their experiment. They divide their work into two stages: first, they extract features, and then they classify images. They also use machine learning techniques including decision trees, support vector machines, and k-Nearest Neighbors. According to their outcomes, AlexNet and support vector machines had the best classification, with accuracy, sensitivity, and specificity of 93.5%, 95.08%, and 93%, respectively.

2.6 Related Paper-6

Honey Janoria et al. [5] used deep learning-based approaches to extract features from skin cancer images and use them to detect the type of skin disease using machine learning classifiers. A transfer learning model was developed that used a VGG-16-layer CNN architecture to extract 1000 features and a support vector machine, decision tree, linear discriminate analysis, and K-Nearest Neighbor algorithm for linear classification. In their experimental results, the VGG-16 CNN model with the K-Nearest Neighbor algorithm had the highest accuracy of 99%.

2.7 Related Paper-7

Xiao Ke et al [6] proposes an end-to-end pedestrian detection and re-identification model in real-world scenes that can effectively solve issues such as inadequate pedestrian expression ability, occlusion, diverse pedestrian attitudes, and the difficulty of small-scale pedestrian detection. The original images are processed with a non-overlapped image blocking data augmentation method and then input into the YOLOv3 detector to obtain the object position information. They used a LCNN-based pedestrian re-identification model to extract the features of the object, and the eigenvectors of the object and the detected pedestrians are calculated to determine whether they can be marked as target pedestrians. The method they used is lightweight and end-to-end, which can be applied to real scenes.

Chapter 3 Methodology

This chapter gives an overview of the different parts of the work chronologically. It mainly discusses the work's theories, techniques, and step-by-step workflow. To complete this part of our project, we have used Resnet101v2, vgg16, ZFNet, and MobileNet models which is an open-source machine learning model.

3.1 Workflow

The figure depicts the suggested method's complete process diagram.

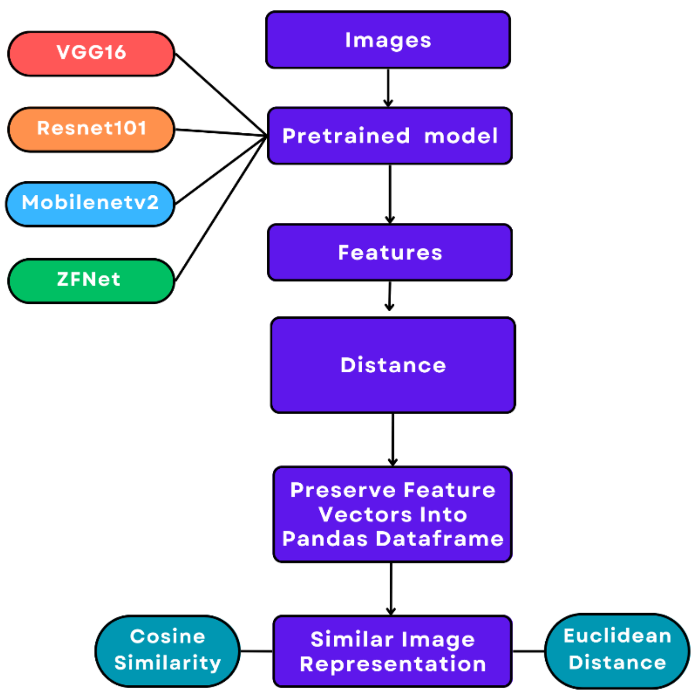


Figure 5: Nearest Neighbour of Different Image Representation.

The workflow describes the whole process of Nearest Neighbour of Different Image Representation. After giving the input image, the output will be shown 10 Nearest Neighbour for this image with an estimated accuracy

3.2 Dataset

For this project we are going to use the ImageNet dataset. But now we used our own dataset. The dataset into 10 classes and Every class contains 100 images. Total 1000 image we use for our project in this course. The dataset contains car, bird, dog,cat and other varieties picture , which was not efficient enough so we want to gradually shift the dataset to the ImageNet dataset as soon as possible. Here is the link of our dataset. https://www.kaggle.com/competitions/imagenet-object-localization-challenge/data?select=ILSVRC

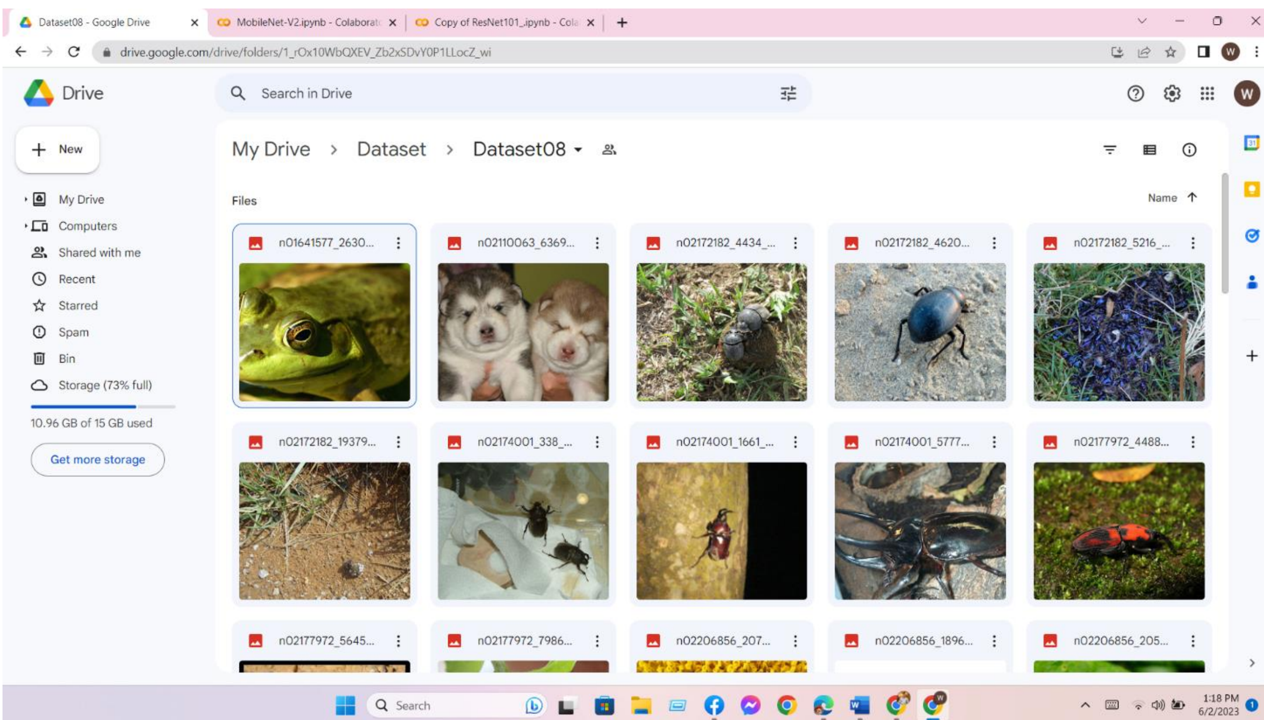
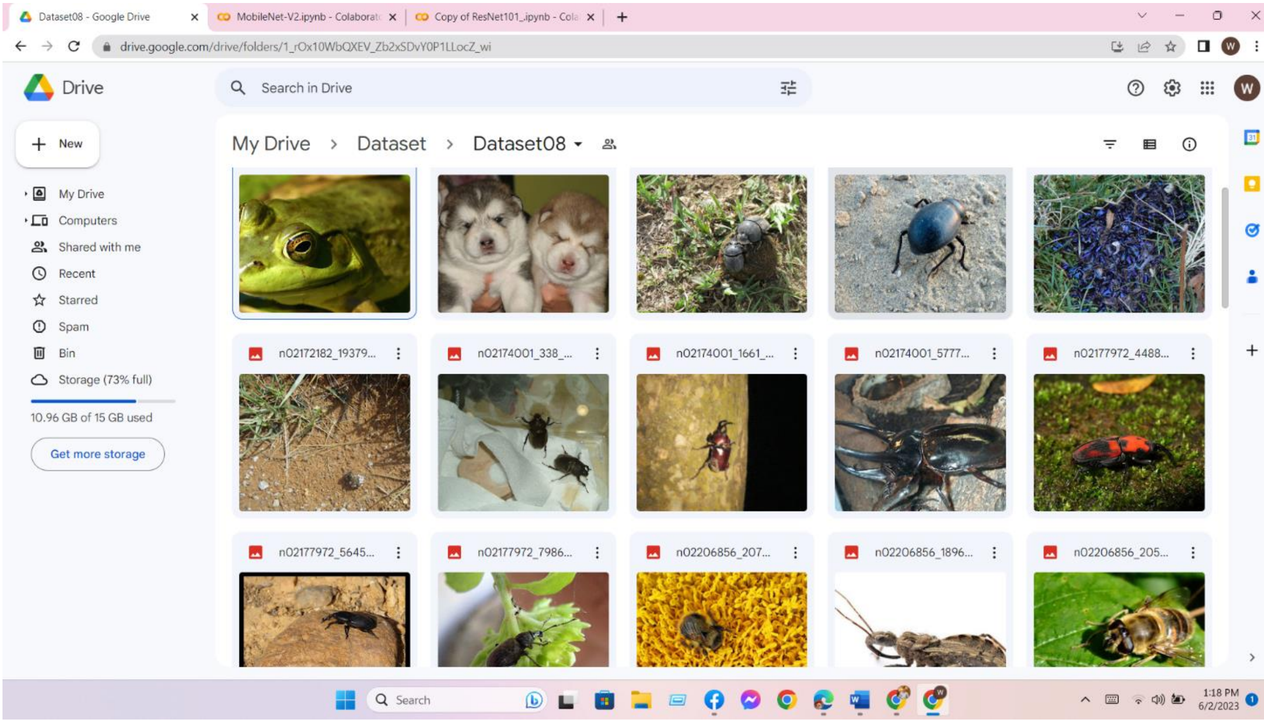


Figure 6: sample picture of our ImageNet dataset



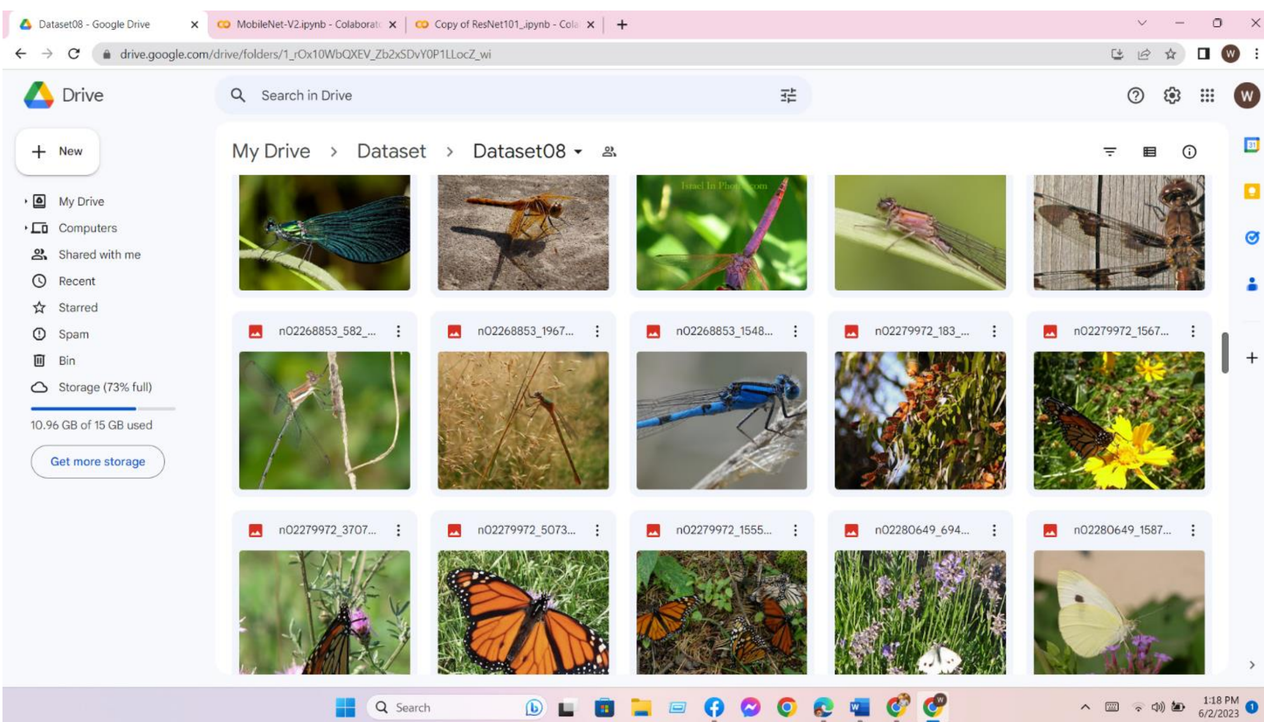


Figure 6: sample picture of our ImageNet dataset

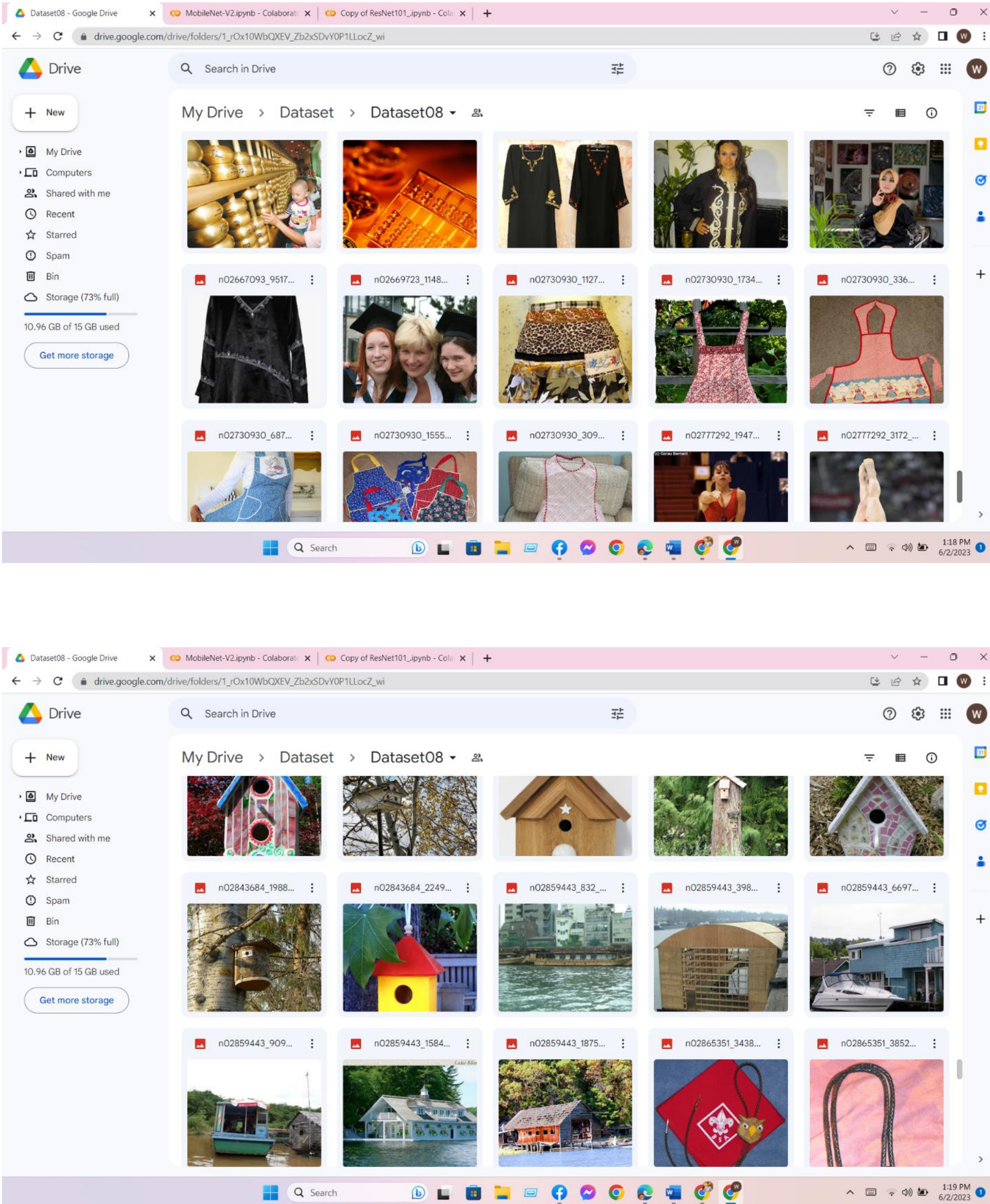


Figure 6: sample picture of our ImageNet dataset

3.3 Data Pre- Processing

Pre-processing images is a critical stage in enhancing the effect of picture classification. Because the CNN learning method coordinates the execution of our machine learning activity, we classified and resized the images for training and testing during the image pre-processing step.

3.4 Model Installation

To begin, we use google Collab. Following that, we downloaded the Resnet101v2 and vgg16 Additionally, we employed the transfer learning method, which retains the parameters from the previous layer, to eliminate the final layer of the Resnet101v2 model and retrain a new layer.

3.5 Convolutional Neural Network (CNN)

CNNs are multi-layer artificial neural networks capable of both unsupervised and supervised feature extraction and classification. A CNN is composed of a series of convolutional and pooling layers used to extract features, followed by one or more fully connected layers used for classification. Sparse connectivity and weight sharing are characteristics of convolutional layers. A convolutional layer's units receive their inputs from a small rectangular subset of the previous layer's units. Additionally, the nodes of a convolutional layer are grouped in weighted feature maps. Each feature map's inputs are tiled to correspond to overlapping regions of the previous layer, equating the aforementioned procedure to convolution, while the shared weights within each map correspond to the kernels. Convolution produces nonlinearities at the element level when the output is passed through an activation function. Following this is a pooling layer that subsamples the previous layer by aggregating small rectangular subsets of values. Max or mean pooling is used to substitute the maximum or mean value for the input values. Following that are a series of fully connected layers, the final one having a unit count equal to the class count. This section of the network performs supervised classification and receives as input the values from the previous pooling layer that comprise the feature set. The CNN is trained using back propagation and the gradient descent method.

3.6 Machine Learning Models

Given that this is a classification experiment, we selected well-known classifiers that are appropriate for our project.

3.6.1 Resnet101v2

ResNet-101 is a convolutional neural network that is 101 layers deep. we can load a pretrained version of the network trained on more than a million images from the ImageNet dataset. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of 224-by-224. Using the dataset as a training set, the machine can detect 10 nearest neighbours for this image.

3.6.2 VGG16

VGG16 is a convolutional neural network (CNN) architecture that is widely regarded as one of the best available vision model architectures. Rather than having a large number of hyperparameters, VGG16 focused on 3x3 convolution layers with stride 1 and always used the same padding and maxpool layer of a 2x2 filter with stride 2. Throughout the architecture, it maintains this arrangement of convolution and max pool layers. Finally, the output is handled by two FC and a SoftMax. The 16 in VGG16 refers to the sixteen weighted layers contained within. This is a truly massive network with over 138 million parameters.

3.6.3 ZFNet

ZFNet is a type of neural network that is used for image recognition tasks. It was created to improve upon an earlier neural network called AlexNet, which was the first neural network to win a large-scale computer vision competition called the ImageNet Challenge.in our project we used ZFNet for object detection.To train ZFNet, a large database of labeled images is needed. This database is used to train the network in a process called supervised learning. The network is then tested on a separate set of images to evaluate its accuracy. In addition to training the network, there are several techniques that can be used to improve its performance

3.6.4 MobileNetV2

MobileNetV2 is a convolutional neural network architecture that seeks to perform well on mobile devices. It is based on an inverted residual structure where the residual connections are between the bottleneck layers. The intermediate expansion layer uses lightweight depth wise convolutions to filter features as a source of non-linearity. As a whole, the architecture of MobileNetV2 contains the initial fully convolution layer with 32 filters, followed by 19 residual bottleneck layers.

Chapter 4 Result, Analysis and Discussion

Our system extracted features from the dataset with impressive results using multiple CNN models such as VGG16, ResNet101, ZFNet, and MobileNet. We were able to generate rich and meaningful representations of the images in the dataset by utilizing the potential of these four pre-trained models. These representations were then transformed into 1D arrays or feature vectors, preserving the essential characteristics of each image. For accurate and relevant results, the system demonstrated effective feature extraction capabilities and advanced similarity metrics. It analyzed image similarities using cosine similarity and Euclidean distance, allowing for more accurate retrieval of similar images. It identified the ten images with the highest similarity scores or shortest distances to the input image, indicating similarities in feature representations. Users can use this retrieval process to find visually similar images and gain a better understanding of their underlying patterns and content. The results include ten closest images, cosine similarity scores, and additional information like class/category or data.

Therefore, our project demonstrated the ability to use feature extraction and nearest neighbor search to identify visually similar images from a dataset, allowing users to discover additional information based on a given input image.

Here are some examples of our results & analysis:

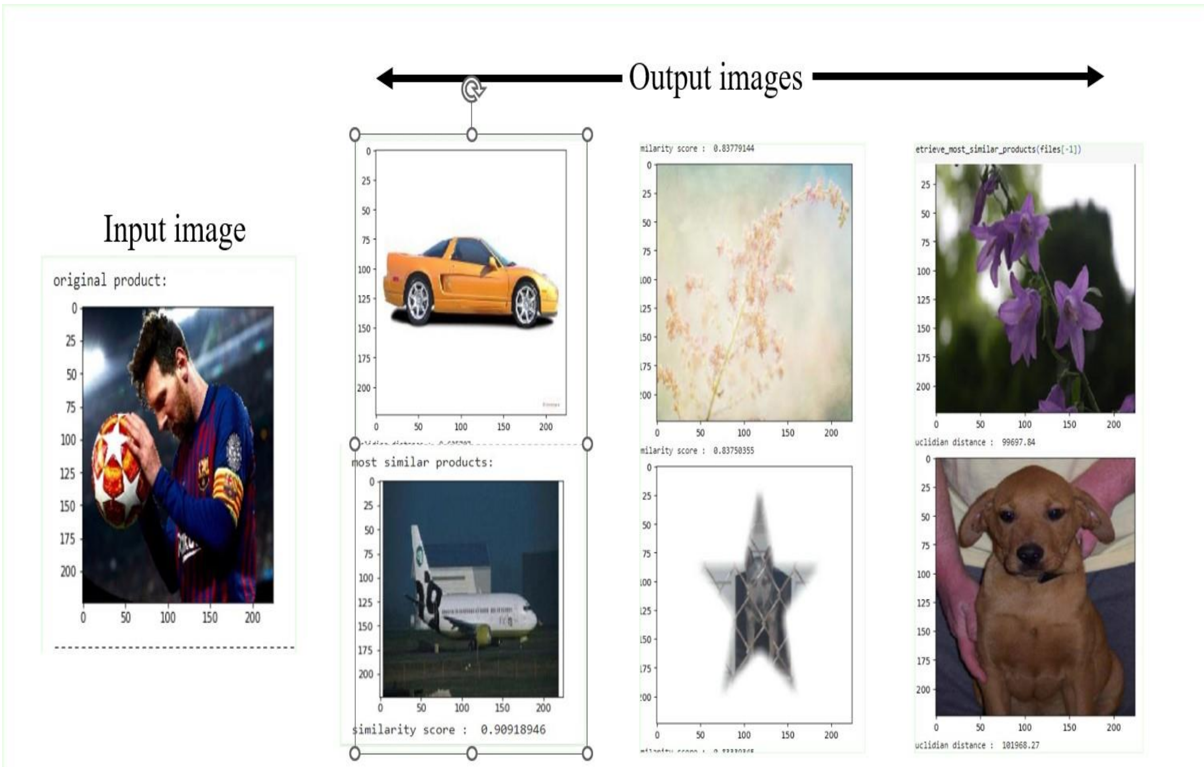


Figure 7: results

Here the sequence of outcomes shows how pixel values affect object detection. In contrast to the given image, the collected features differ in their properties and nature. In the Parts of the input image that are extracted as feature vectors, and interesting outputs from various natures objects, and so on with similar features to the given input can be generated. This shows that similarity can be defined by many things that humans cannot simply tell or explain, but they are similar by their features as demonstrated by this system. The outcomes of this system are a clear example of similar things in nature that are entirely different but similar to the object that not only humans but also the system it generates can understand.

Examples of our results & analysis:

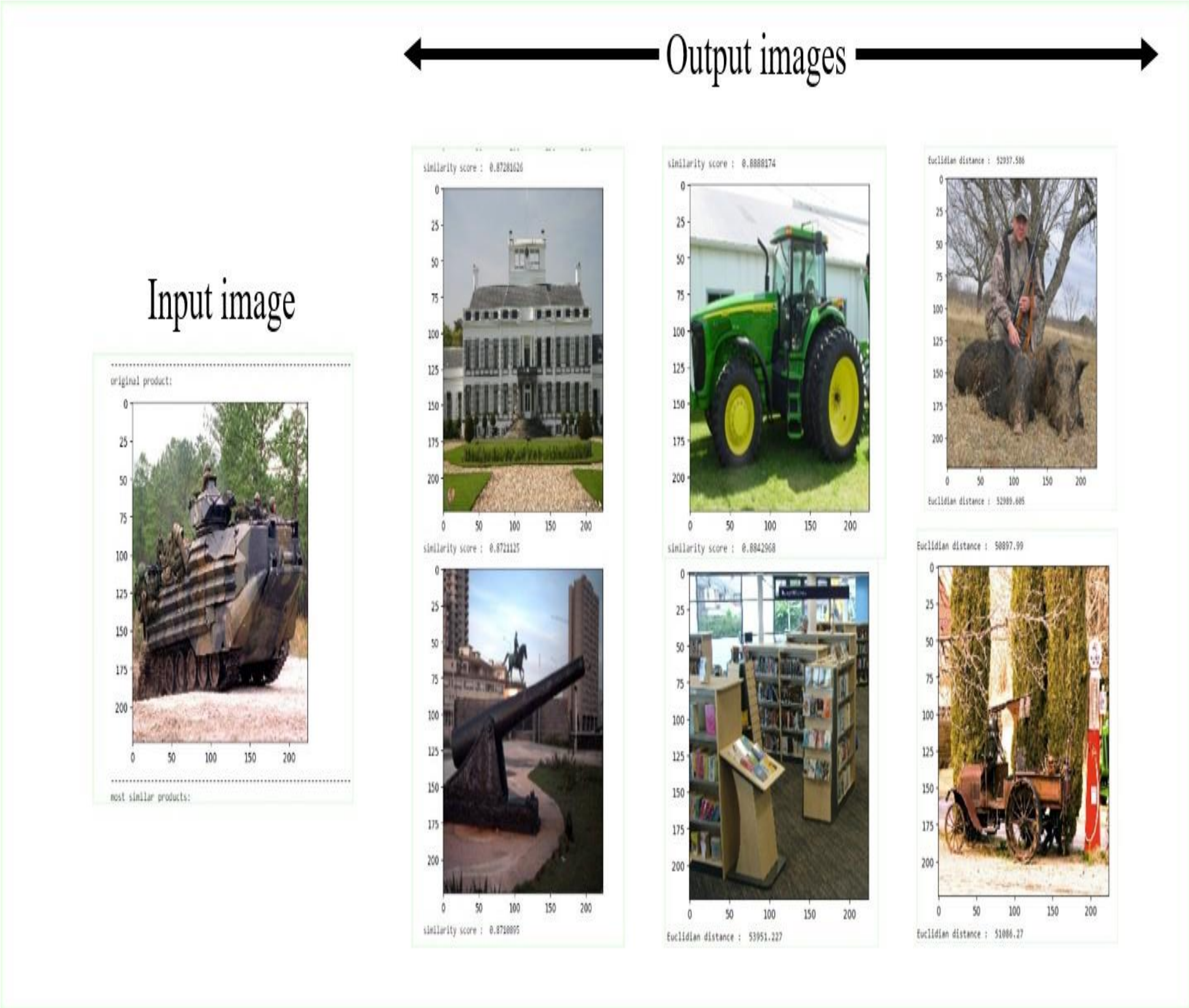


Figure 7: results

For this result, the extraction of feature vectors from images and the resulting similar output demonstrates how object detection can be varied with intriguing similarities from various fields. It also provides a chance to explore and observe similarities on a large scale that may appear to be completely unrelated but are actually very similar which with the human eye cannot be defined and can't be explained by any sort of logic.

Examples of our results & analysis:

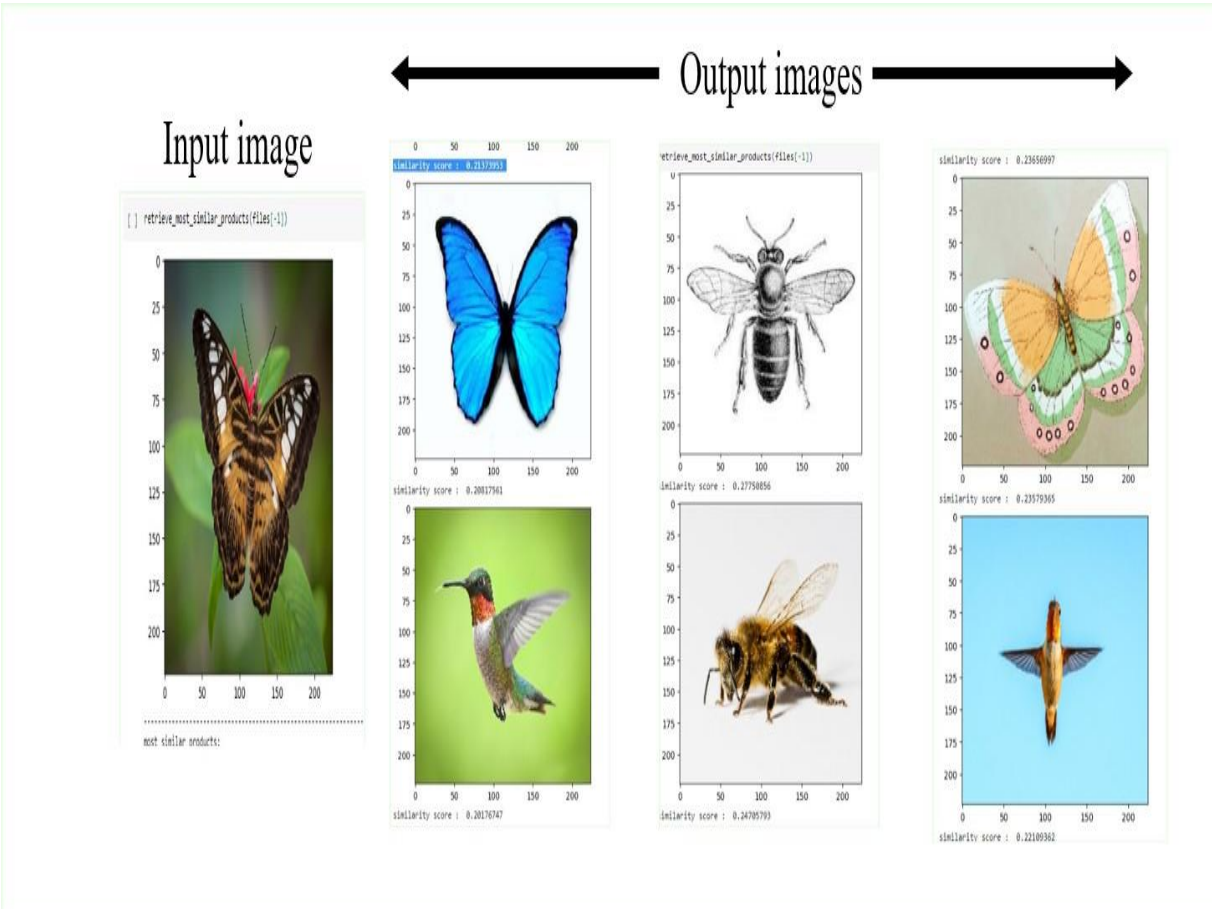


Figure 7: results

Compared to the input image, where contents can appear in their true form, object detection outputs might differ in size, shape, and nature, even among objects of the same species. This reveals feature vectors' remarkable manipulation capabilities in object detection algorithms, allowing them to distinguish and organize objects based on similar features.

Examples of our results & analysis:

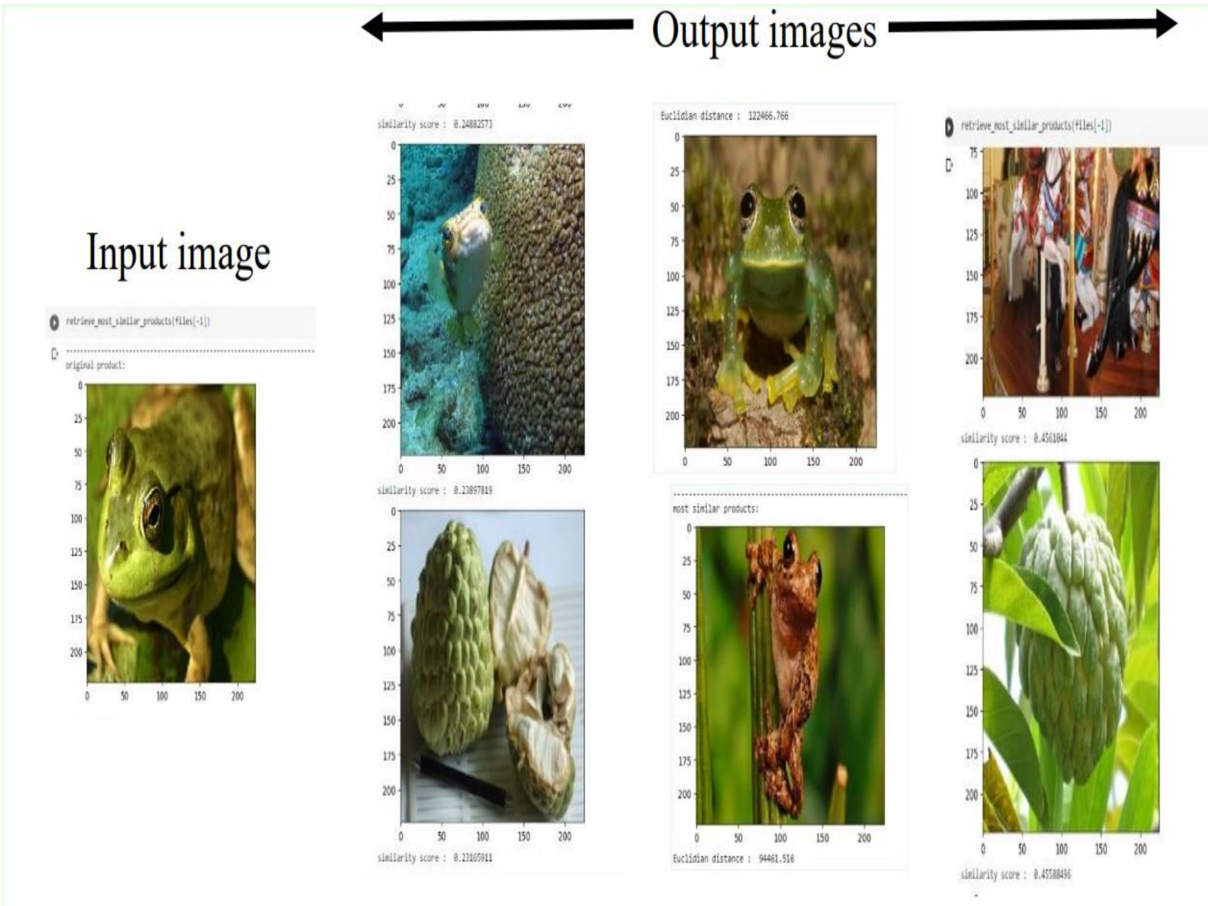


Figure 7: results

Here the image outputs demonstrate how a single object's size, shape, and color can represent an entirely new object of the same size, shape, and color. The Custard apple and sea fish represent frogs' closest objects, allowing them to locate objects with similar features in size, color, or shape, enabling them to locate similar objects. The findings indicate that these features are crucial for our visual understanding and object recognition.

Chapter 5 Impacts of the Project

5.1 Environmental Effect

The environmental impact of my project is that it finds the nearest neighbor of an image using CNN can be evaluated from multiple perspectives. Firstly, the training of the CNN model requires a significant number of computational resources, which can result in increased energy consumption and associated carbon emissions. To mitigate this impact, it is recommended to use energy-efficient hardware and optimize the CNN model architecture to reduce the training time and resource utilization. In addition, the deployment of the model to serve image similarity queries can also have environmental implications. Running the model and processing image data requires computational resources and energy consumption. However, this impact can be mitigated by optimizing the deployment infrastructure, such as using efficient servers and reducing network latency. Another important aspect to consider is the sustainability of the underlying image data used for inference. If the image data is obtained from unsustainable sources, such as unethically harvested or illegally obtained images, the project's sustainability is compromised. Therefore, it is important to ensure that the image data used for the project is legally and ethically sourced, and the project is compliant with relevant privacy and data protection regulations. Finally, it is important to consider the project's social sustainability, as well. The technology used in the project may have potential social impacts, such as job displacement, privacy concerns, or perpetuating bias and discrimination. Thus, it is essential to proactively address these issues and ensure that the project is inclusive, transparent, and accountable to all stakeholders

5.2 Sustainability

The usefulness, viability, maintainability, and scalability of a project determine its sustainability. A project is more likely to be maintained over time if it meets a genuine need and offers value to its users. Code quality, documentation, and ease of adding new features or making changes are all factors in maintainability. As the number of images and queries increases without performance degrading, scalability is essential. The Nearest Neighbor of Different Image Representation project has a solid foundation for sustainability because it uses convolutional neural networks to identify the ability to accurately match images with similar features has implications for tasks like product recommendations and medical imaging. Machine learning algorithms for image recognition and classification can reduce human error and improve efficiency in various industries. Automating the image analysis process saves time and resources, contributing to a more sustainable future. The knowledge and insights gained from the project can be used to develop advanced models and techniques, enabling further advancements in the field. By sharing findings and contributing to the collective knowledge base, the project contributes to a community of researchers and practitioners dedicated to advancing machine learning and computer vision. Overall, the project has great potential to contribute to a more sustainable future by using machine learning to improve image recognition and classification image's closest neighbors.

Chapter 6 Project Planning

Figure 8: Gantt chart.

Chapter 7 Usability, Manufacturability of the system

7.1 Usability

The usability of a project is crucial for its intended audience, such as developers and data scientists with a basic understanding of machine learning and deep learning techniques. The project offers a user-friendly interface for entering data and choosing parameters, making it accessible through web or mobile apps. It can be used in various applications, such as image search engines, recommendation systems, and content-based image retrieval systems. The project can also be integrated with other deep learning algorithms and technologies, such as convolutional neural networks (CNNs) and image recognition software, to provide more accurate and reliable results. Overall, the usability of Nearest Neighbor of Different Image Representation using deep learning is high, as it offers a user-friendly interface, can be used in various applications, and can be enhanced through integration with other technologies.

7.2 Manufacturability

Manufacturability is the ability of a project to be produced and scaled for mass production. The Nearest Neighbor of Different Image Representation project may not be suitable for mass production due to its data analysis nature, but it can be implemented in industries like healthcare, automotive, and e-commerce. It involves scalability, automation, reliability, quality control, and resource availability. The design must consider manufacturing process limitations and ensure resources are available for efficient production. Ensuring manufacturability is crucial for producing a project at a reasonable cost and meeting quality standards for widespread commercial use.

In conclusion, the usability and manufacturability of the Nearest Neighbor of Different Image Representation project using deep learning are essential factors to consider during the development process. By ensuring the project is user-friendly and efficiently scaled, it can become a valuable tool for data scientists and developers.

Chapter 8 Conclusions

8.1 Summary

The system successfully demonstrated the use of pretrained models and feature extraction techniques to identify similarities between images from different types and nature. Computing feature vectors and measuring distances, the system accurately identified related images. The findings showcase the potential of cross-domain image similarity in applications such as content-based image retrieval and creative exploration. Further advancements and integration of deep learning techniques could enhance the system's performance and broaden its applications. Overall, this project may contribute to the field of cross-domain image analysis and provides valuable insights for future research and practical use.

8.2 Limitations

Due to Google Collab limitations, I was able to add a maximum of 4000 images to my data set.

8.3 Future Improvement

In terms of future work, I am interested in exploring how the features vector influences object detection and using more diverse CNN models to improve similarity score and Euclidian distance.

References

[1] http://poseidon2.feld.cvut.cz/conf/poster/poster2018/proceedings/Poster_2018/Section_IC/IC_057_Listik.pdf

[2] Local Aggregation for Unsupervised Learning of Visual... - Google Scholar

[3] NEAREST NEIGHBOUR STRATEGIES FOR IMAGE UNDERSTANDING (psu.edu)N. Barla, "v7labs," v7, 15 March 2022. [Online]. Available: https://www.v7labs.com/blog/semantic-segmentationguide#:~:text=Semantic%20Segmentation%20follows%20three%20steps, by%20creating% 20a%20segmentation%20mask. [Accessed 24 April 2022].

[4] KHLOE, "Image Recognition Applications," 16 January 2022. [Online]. Available: https://www.datasciencesociety.net/image-recognition-applications-7-essential-future-uses/. [Accessed 25 April 2022].

[5] R. I. S. B. S. Nishat Tasnim, "A Convolution Neural Network Based Classification Approach

[6] for Recognizing Traditional Foods of Bangladesh from Food Images," *A Convolution Neural Network Based Classification Approach for Recognizing Traditional Foods of Bangladesh from Food Images,* p. 4, 2020.

[7] Pawangfg, "Residual-networks-resnet-deep-learning," geeksforgeeks, 27 January 2022.

[8] [Online]. Available: https://www.geeksforgeeks.org/residual-networks-resnet-deep-learning/. [Accessed 22 April 2022].

[9] A. Thite, "Introduction to VGG16," Great Learning, 1 October 2021. [Online]. Available:

[10] https://www.mygreatlearning.com/blog/introduction-to-vgg16/. [Accessed 23 April 2022].

[11] On Vectorization of Convolution Layer in Convolution Neural Networks (CNNs) | by Sanghvirajit | Analytics Vidhya | Medium

[12] Scaling nearest neighbors search with approximate methods. (jeremyjordan.me)

[13] https://link.springer.com/article/10.1007/s11042-012-1289-4

[14]http://poseidon2.feld.cvut.cz/conf/poster/poster2018/proceedings/Poster_2018/Section_IC/IC_057_Listik.pdf

[15]J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li,... - Google Scholar

[16]Local Aggregation for Unsupervised Learning of Visual... - Google Scholar

[17]Unsupervised Learning | Neural Computation | MIT Press