Retrieval of Images using Reverse Search Engine

Sadhvi Sushravya H S* *Computer Science* *PES University* Bangalore, India

Sathvik K* *Computer Science* *PES University* Bangalore, India

Sanjana Gujjar* *Computer Science* *PES University* Bangalore, India

Sammed D Tapakire* *Computer Science* *PES University* Bangalore, India

Dinesh Singh* *Computer Science* *PES University* Bangalore, India

Abstract—Reverse image search is a rapidly evolving and engaging field, with applications in various industries as well as our everyday lives. Instead of using words as search input, images can be used to return similar images as a result, making it easier and quicker for machine users to find what they are looking for. The capability of web search engines is constantly improving, where users can not only use keywords but also utilize image search features that provide information about the query image, its size, and the websites that use it.

Reverse image search technology is rapidly advancing, and its potential uses continue to grow. By enabling machines to understand images in a more nuanced way, it can enhance various industries and even make our everyday lives more efficient. Overall, the advancement of reverse image search technology has opened up new possibilities in various industries and has made it easier for people to find what they are looking for online. As technology continues to improve, it is likely that we will see even more innovative applications of this technology in the future.

Index Terms—Content Based Image Retrieval (CBIR), Reverse Search Engine (RIS), Convolutional Neural Network (CNN), Feature extraction

I. Introduction

The rapid growth of digital image content has resulted in an increasing need for efficient methods of image retrieval. Reverse image search technology has emerged as a powerful tool for retrieving similar or identical images based on a reference image. Content-based image retrieval (CBIR) is a type of reverse image search that uses the visual features of an image, such as color, texture, and shape, to find other images that are similar in content. Convolutional neural networks (CNNs) have shown great promise in image recognition and feature extraction, making them a natural choice for CBIR applications. In this project, we aim to develop an image retrieval system that utilizes a reverse search engine and a

CNN to accurately and efficiently retrieve images based on their visual features. The system will be evaluated using standard benchmark datasets and compared to other state-ofthe-art image retrieval methods. The results of this project will contribute to the advancement of CBIR technology and have practical applications in areas such as e-commerce, digital forensics, and content creation.

II. Problem Statement

The vast and growing number of images available on the internet has led to an increased reliance on visual search, including the use of reverse image searches.

In the fashion industry, the digitization of the industry has increased consumer options and shortened production cycles, making a system for fast product browsing and inspiration essential. However, text-based search methods are not ideal for fashion design, where visual design is crucial.

To address this issue, a content-based image retrieval system is developed using convolutional neural networks and other techniques to extract features like color, context, and texture from a dataset of fashion industry images.

By converting the query image and database images into feature vectors and using cosine similarity, the system groups similar images and retrieves the most similar ones. The choice of visual features and encoding methods is essential to the effectiveness of image retrieval systems, and machine learning techniques, particularly deep neural networks, can improve the performance of content-based image retrieval.

III. Literature Survey

A. Reverse Image Search for the Fashion Industry Using Convolutional Neural Networks

The researchers were able to define and train their own Convolutional Neural Network (CNN) using a customized dataset and their own parameters. Following successful training and achieving a high level of accuracy, they developed a reverse search engine using Euclidean and Cosine distances. The precision and accuracy of an image search tool are mainly determined by the accuracy and precision of the image features obtained. In the case of CNNs, this implies that the accuracy of the tool is strongly dependent on the correctness of the trained CNN model. Additionally, the researchers observed that in some cases, classes with fewer images performed better than their peers. Therefore, having a large number of training images per class helps prevent the CNN model from overfitting during training.

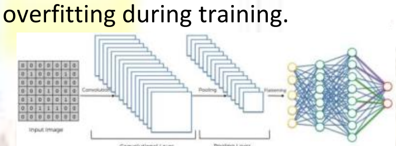

Fig. 1. Core components of Convolutional Neural Network

B. Reverse Image Search Improved by Deep Learning

The proposed image retrieval method presents new possibilities for more precise analysis. To speed up the search process, image data should be organized with relevant information. Real-world examples used in the study were removed from the databases. However, the concept still holds promise for improving reverse image search. Deep learning requires access to large amounts of labeled data, which the current test cases do not include. Nevertheless, the underlying approach can still be applied to extract important insights from big data sets.

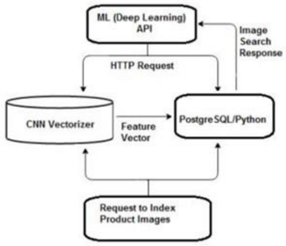

Fig. 2. Deep Learning Model Architecture

C. Image Retrieval System Based on Colour Layout Descriptor and Gabor Filters

The authors employed a CBIR approach that relied on CLD and Gabor texture descriptors, with color, shape, and texture being among the visual factors exhibited in the images. CLD texture was effective in capturing the distribution of texture with similar colors in an image, while Gabor texture was useful in representing the color characteristics of pixels. The performance of the proposed system exceeded that of the two aforementioned methods in terms of retrieval precision, regardless of whether the images were grayscale textures, color textures, or natural color images.

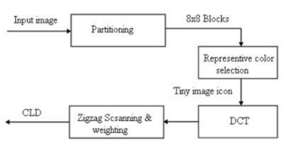

Fig. 3. A block diagram of the CLD Architecture

D. Content-based Image Retrieval using color features of salient regions.

The proposed technique uses a color evaluation strategy to extract salient portions and identify dominant colors for each significant location. A binary map is then created to represent the spatial distribution of the dominant hues within the salient area, which is divided into smaller sub-blocks. Each sub-block is assigned a cost of either one or zero depending on the number of pixels that contain the dominant color. The binary map is useful in identifying the geographic distribution of the dominant color within the relevant area for objects with the same color. The technique also includes a binary map similarity comparison method, which outperforms traditional color-based methods in retrieving color images from various databases. Furthermore, the retrieval results are closely related to the image contents because the approach focuses on the exceptional regions. The proposed method, which employs basic and binary operations, is well-suited for item-based color image retrieval in online and mobile applications.

Fig. 4. Results of experiments on retrieving colour images using (a) LBA, (b) the saliency-based method, and (c) the suggested method.

IV. DATA

We have a large dataset accessible through e-commerce platforms, containing manually labeled goods, high-quality images, and product descriptions. Each item has a unique ID, which can be used to retrieve metadata from styles/ID.json and images from images/ID.jpg. Styles.csv also includes fundamental product categories and display names. Kaggle suggests several potential applications, such as training an image classifier or an NLP classifier using the product descriptions to obtain the master category, predicting elegant labels, generating images using GANs, and improving visual similarity. The dataset includes four files: images, styles, images.csv, and styles.csv.

Fig. 5. Images in the dataset

V. METHODOLOGY

A project management methodology comprises principles, tools, and techniques that aid in the planning, execution, and management of projects. The layout technique of contentbased image retrieval (CBIR) focuses on the high-level image capabilities derived from convolutional neural networks (CNN) in image retrieval, also known as instance retrieval. This method extracts features like color, texture, and shape from images to retrieve them from a collection. CNN extracts relevant photo features in multiple layers and converts visual content into abstract semantic concepts using deep learning frameworks. With the increasing number of online production and distribution networks, along with a growing number of available images, CBIR has become popular for retrieving images from vast, unlabelled collections. However, the efficient and rapid access of these large image databases and the extraction of similar images from a given photo pose significant challenges, which can be addressed through the development of effective deep learning models focused on CNNs. Several studies have shown promising results for various CBIR tasks using observational research and yield key lessons for enhancing CBIR performance. Google's image search is one of the most well-known CBIR methods based on image characteristics.

VI. MODEL ARCHITECTURE

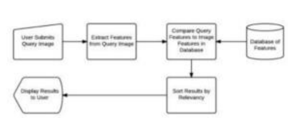

Fig. 6. Core components of an image search system

When a user submits a query image, reverse search engine extracts the features of the query image and also that of the images in the database and compares the features. Sorting of images from the database happens at the next step based on cosine similarity and the most similar images are retrieved.

VII. FEATURE EXTRACTION MODELS

A. Variational Autoencoders

Variational autoencoders (VAEs) are a type of neural network used for unsupervised learning and generative modeling. They consist of an encoder network that maps input data to a latent space, and a decoder network that maps the latent space back to the original input data. The VAE model aims to learn a probability distribution of the input data in the latent space and generate new samples by sampling from that distribution. This allows for the generation of new data that is similar to the original input data but not identical. VAEs are useful in applications such as image and speech generation, data compression, and anomaly detection.

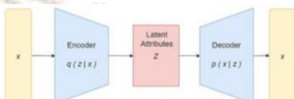

Fig. 7. Variational Autoencoders

B. VGG 16

VGG 16 is a convolutional neural network (CNN) architecture commonly used for image classification tasks. It consists of 16 layers, including convolutional layers with 3x3 filters, max pooling layers, and fully connected layers. VGG 16 is known for its simplicity and uniformity in design, with all convolutional layers having the same number of filters and max pooling layers of the same size. This architecture has been trained on the ImageNet dataset, resulting in high accuracy in image classification tasks. Its popularity and effectiveness have led to its use in various computer vision applications, such as object detection and segmentation.

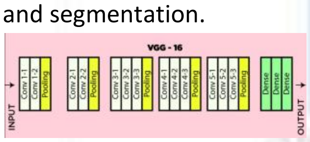

Fig. 8. VGG-16 Architecture

VIII. CLASSIFICATION MODELS

A. VGG 19

VGG 19 is a convolutional neural network architecture that consists of 19 layers, including convolutional, pooling, and fully connected layers. The VGG 19 model has been widely used in computer vision tasks such as object recognition and image classification due to its outstanding performance on large-scale image recognition challenges, such as the ImageNet dataset. The VGG 19 model is designed with a focus on simplicity and increasing depth, which enables it to learn hierarchical features from input images effectively. The architecture has a uniform structure, where each block of layers consists of a set of convolutional layers followed by a max-pooling layer. Overall, the VGG 19 model is a powerful tool in the field of computer vision and has paved the way for further advancements in deep learning.

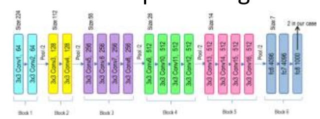

Fig. 9. VGG-19 Architecture

B. RGB Classifier

RGB (Red, Green, Blue) Classifier is a machine learning model used for image recognition tasks. It takes an image as input and classifies it into a specific category or label. The model extracts features from the image in the form of pixel values and uses them to predict the image's class. The RGB color model is used because it is the most common color model for digital images. The classifier is typically trained on a dataset of labeled images to learn the patterns and features associated with each class. Once trained, it can be used to classify new, unseen images. The RGB classifier is widely used in various applications, including object detection, face recognition, and autonomous vehicles.

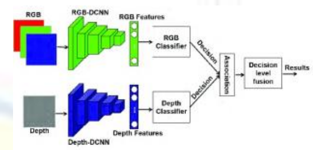

Fig. 10. RGB Classifier Architecture

C. Cosine Similarity

Cosine similarity is a measure used to determine the similarity between two vectors in a high-dimensional space. It is commonly used in classification models to compare documents or data points based on their similarity to each other. The cosine similarity calculates the cosine of the angle between two vectors and generates a value between -1 and 1. A value of 1 indicates that the two vectors are identical, while a value of 0 indicates no similarity, and a value of -1 indicates complete dissimilarity. In classification models, cosine similarity is used to cluster similar data points together and separate dissimilar data points. By using cosine similarity, the classification model can identify patterns and similarities between data points, leading to more accurate predictions and classifications.

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Fig. 11. Cosine Similarity

VIII. RESULTS AND CONCLUSION

A. Features of the images were extracted

Features were extracted using VGG16, but, accuracy of prediction with VGG16 is very low so that is 76.2%, so we worked with different models, we used variational autoencoders where the accuracy is 85% and extracted a feature vector of every image and stored it in the database for comparision.

```
Accuracy and Loss of the VGG16 Model are:
Loss : 0.8629643
Accuracy : 0.7622254
```

Fig. 12. Accuracy and Loss of VGG16

```
Test loss: 0.6111644506454468
Test accuracy: 0.8560606
```

Fig. 13. Accuracy and Loss of Autoencoders

B. Classes were predicted

Using the autoencoders, we classified all the images into various different classes namely t-shirts, shirts, jeans, skirts and kurtas.

```
[[3.83370977e-16 1.01592554e-11 1.05140084e-10 4.71214861e-07
  9.99999523e-01]]
The Model predicts that it is tshirts
```

Fig. 14. Class Prediction

C. Cosine Similarity

Cosine similarity is used to compare the feature vectors of the input image with every other image in the dataset by grouping them by same class, as well as any photos with a cosine similarity value below a threshold.

References

[1] Gede Susrama Mas Diyasa, Alfath Daryl Alhajir, Amir Muhammad Hakim and Moh. Fathur Rohman. “Reverse Image Search Analysis Based on Pre-Trained Convolutional Neural Network Model.” 2020 Information Technology International Seminar (ITIS) Surabaya, Indonesia, October 14-16, 2020.

[2] Paras Nath Singh and Tara P. Gowdar. “Reverse Image Search Improved by Deep Learning.” 2021 IEEE Mysore Sub Section International Conference (Mysore Con).

[3] Kudzai Felix Mawoneke, Xin Luo, Youqin Shi and Kenji Kita. “Reverse Image Search for the Fashion Industry Using Convolutional Neural Networks.” 2020 IEEE 5th International Conference on Signal and Image Processing.

[4] “Large scale reverse image search A method comparison for almost identical image retrieval” Mathieu Gaillard 1 , Elod Egyed-Zsigmond˝ 1,2 1 Universite de Lyon INSA Lyon, mathieu.gaillard@insa-lyon.fr´ 2.Universite de Lyon LIRIS, INSA Lyon elod.egyed-zsigmond@insa-´ lyon.fr References.

[5] D. Garcia-Gasulla, F. Pares, A. Vilalta, J. Moreno, E. Ayguade, J. Labarta, U. Cortes and T. Suzumura, “On the Behavior of Convolutional Nets for Feature Extraction”, Journal of Artificial Intelligence Research archive Volume 61, Issue 1, 2018, pp. 563-592.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks” Communications of the AcM Volume 60, Issue 6, 2017, pp.8490.

[7] Desai, Padmashri and Pujari, Jagadeesh and Kinnikar, Anita, “An image retrieval using combined approach wavelets and local binary pattern,” B.V.B. College of Eng. and Tech. Hubli, Karnataka, Aug. 2016.

[8] Dhvani Kansara et Al. “Optimizing Reverse Image Search by Generating and Assigning Suitable Captions to Images”, Springer 2020, Advanced Computing Technology and Applications, PP 621-631.

[9] I. G. S. M. Diyasa, A. D. Al Hajir, A. M. Hakim and M. F. Rohman, ”Reverse Image Search Analysis Based on Pre-Trained Convolutional Neural Network Model,” 2020 6th Information Technology International Seminar (ITIS), 2020.

[10] L. Tsochatzidis, K. Zagoris, N. Andrikidis, A. Karahaliou, L. Costar idou, I. Pratikakis. Computer-aided diagnosis of mammographic masses based on a supervised content-based image retrieval approach Pattern Recognition, 71 (2017).

[11] W. Yu, K. Yang, H. Yao, X. Sun, P. Xu. Exploiting the complementary strengths of multi-layer cnn features for image retrieval, Neurocomputing, 237 (2017).