

Content based Image Retrieval using K-Nearest Neighbours and Convolutional Neural Networks

C. Siva kumar¹, O. Bhaskaru², Bhargavi Komatlapalli³, Phani Sai Kumar Areddula⁴, Manikanta Maddali⁵ and Busa Rajasekhar Reddy⁶

¹Associate Professor, Dept. of CSSE, Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College), Tirupati, India

Email: sivakumar.c@vidyanikethan.edu

²Assistant Professor, Dept. of CSE, Gates Institute of Technology, Gooty, Ananthapur, A.P, India

Email: obhaskaru@gmail.com

³⁻⁶UG Scholar, Department of Computer Science and Systems Engineering, Sree Vidyanikethan Engineering College, Tirupati, India

Email: bhargavikomatlapalli99@gmail.com, phanisaiikumarareddula@gmail.com, manikantal62l@gmail.com, rajasekharbusa175@gmail.com

Abstract—Today, there is a lot of digital image material on the Internet. As the need for better image search tools grows, new ways must be found to solve the problem of telling the difference between question pictures and received images. This study presents a new technique to enhance the searching of images in retrieval based on image content. The technique involves using several deep neural network methods, such as ResNet50, InceptionResNetV2, VGG19, InceptionV3, DenseNet121, Xception, and Improved CNN. These models make it easier to find interesting photos. The study tested the technique on the Dataset and a dataset from a Kaggle competition, using metrics like the confusion matrices, F1 scores, and mAP to assess the model performance. The tests show that the proposed system works very well, taking away some of the problems that come with picture searches being unclear and showing how well it can find appropriate images.

Index Terms— ResNet50, InceptionResNetV2, VGG19, DenseNet121, Xception, Improved CNN.

I. INTRODUCTION

The huge amount of digital picture material on the Internet today has created a need for more powerful image search tools than ever before. A lot of tools claim to have fast response times, but it's still hard to tell the difference between question pictures and received images. Combining several deep neural networks in the field of Content-Based Picture Retrieval (CBIR) is a new way to do things that this study offers as an option. The goal is to make picture searches much better. The first step in the method is to create special multi-Convolutional Neural Network (CNN) models using pre-trained designs like ResNet50, InceptionResNetV2, VGG19, InceptionV3, DenseNet121, Xception, and an Improved CNN. These models make it possible to pull out unique picture traits, which makes image searches more accurate. These features are very important for figuring out subtle patterns and traits, which helps with the difficult task of telling the difference between pictures that look

alike. As well as normal datasets like the Dataset, tests are also done on a dataset that the person testing the program got from a Kaggle competition. Several measures, including precision, confusion matrix, and F1 scores, are used to judge how well the suggested system works. The evaluation also includes the quality of the search results, which can be measured with Precision at Different Recall Levels(AP) and Average Precision across Classes(mAP). Experiments per-formed confirm the new system's outstanding performance, demonstrating its ability to effectively address the issues that come with picture search uncertainty. Not only does this study add to the progress made in image search technologies, it also shows how multi deep neural networks can help the field of visual similarity search.

II. RELATED WORKS

In the realm of Image Retrieval, two methods are employed for querying images: concept-based retrieval and visual similarity search.. A concept-based query relies on image-related data like captions, labels, and surrounding text, while the query itself is in textual form. On the other hand, a content-based query involves the image itself, where its content encompasses aspects such as color, shape, texture, local features, and any other pertinent information derived from the image query.The system initiates by extracting features from all images, comprising both query images and those in the image database, utilizing a feature extraction algorithm. Subsequently, it computes the similarity between the query image and all database images. Upon completion, the system presents a collection of images bearing significant similarity to the query image. In similarity search, the process involves identifying the likeness between the query image and those stored in the database, subsequently arranging the results based on similarity, from highest to lowest degree. In Image Classification, each image is assigned to a specific class. However, in Image Retrieval, where the query is an image itself, the returned results may encompass images from various classes. Currently, the predominant technique employed in Image Retrieval involves utilizing similarity search, as depicted in Figure 2. Feature vectors are extracted from the image collection in the database, and these features are subsequently stored in a dedicated feature database. Similarly, the query image undergoes feature vector extraction, following which its feature vector is compared for similarity with those stored in the feature database that enables us to evaluate image similarity by comparing the feature vectors extracted. Various distance measures like Cosine, Euclidean, and Hamming can be employed for this comparison. Consequently, by analyzing these distances, we can identify the top k-images that closely resemble the query image.

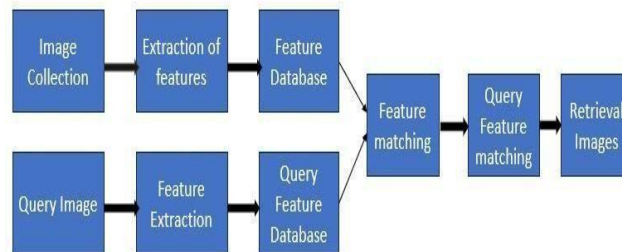


Fig 1: CBIR Interval

In content-based image search systems, the key features relied upon include color, texture, and shape. Hirematch and Jagadeesh Pujari [7] introduced a technique that integrates these features to enhance the efficiency of image search. This approach involves partitioning the image and its elements into equally-sized non-overlapping tiles. Local features pertain-ing to color and texture are derived from histograms that reflect the conditional occurrences between the image tiles and corresponding auxiliary component tiles. A methodology is utilized to compare two images, integrating the principle of identifying the most similar highest priority with a graph that incorporates tiles from both the query and target images, fol-lowing a two-partite structure.. Additionally, the shape feature is extracted by computing the image's edge using gradient vector field. The combination of color and texture attributes within the image and its corresponding elements, along with the shape feature, has yielded robust feature for visual similarity search.

III. METHODOLOGY

Our model comprises two primary components: the training model and the web application.

- Training Model: Initially, we preprocess the data from the database, including normalizing image sizes. Subsequently, we train a pre-existing CNN model on the preprocessed image dataset. Following this, features are

extracted from all database images through the training model. Finally, we utilize K Nearest Neighbors (Annoy) to construct indexes for the image feature vector data. Among the various distance measures between two vectors, we opt for the "angular" metric, which is analogous to the Cosine distance.

- Web Application: The web application enables users to submit query images to initiate image searches within the system. Upon receiving a query image, the system conducts preprocessing and extracts feature vectors from it. These feature vectors are then compared with the image feature vectors stored in the database, yielding the search results. The system outputs a set of k images that are closest to the input image.

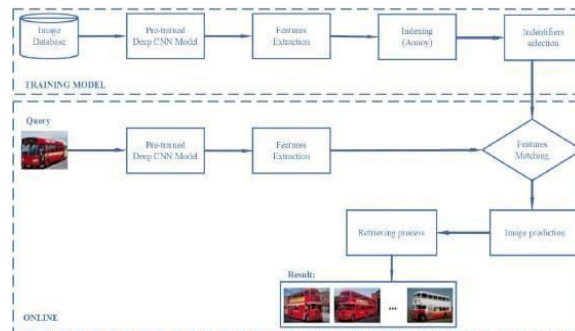


Fig 2: System Architecture

The proposed method enhances Content-Based Image Retrieval (CBIR) by employing a specialized ensemble of Convolutional Neural Network (CNN) models, including ResNet50, InceptionResNetV2, VGG19, InceptionV3, DenseNet121, Xception, and a custom CNN. Leveraging pre-trained structures, these models extract intricate image elements crucial for accurate retrieval. Evaluation on Kaggle and Dataset reveals its efficacy through metrics like accuracy, confusion matrix, F1 scores, average precision (AP), and mean average precision (mAP). By amalgamating advanced neural network techniques, this approach addresses uncertainties in image search, marking a significant stride in CBIR technology. It hinges on crafting representative vectors to encapsulate image features, facilitating robust similarity assessment.

ResNet50:

ResNet-50 comprises a specific type of neural network that is made up of 50-layer architecture with most of them being convolutional layers. It also includes a Max Pool layer and an average pool layer. Residual neural networks can be de-scribed as constructed using the idea of residual blocks.

VGG 19:

VGG19 is a deep learning model with 19 layers that has been trained on a large dataset called Image Net. This dataset consists of over a million images and the pretrained VGG19 network is capable of classifying photos into 1000 different categories such as keyboards, mice, pencils, and various animals.

InceptionResNetV2:

The Inception-ResNet-v2 convolutional neural network replaces the filter concatenation step of the Inception family of architectures with residual connections. It is characterized by a 64-layer deep learning modules, which merges the Incep-tion structure with Residual connections. This fusion aims to expedite data training, simplify optimization, and deliver superior accuracy in the classification of image data.

DenseNet121:

One form of Dense Net is DenseNet-121, which has 121 layers. The code uses PyTorch, a prominent open-source machine-learning package. The code defines Bottleneck, Transition, and DenseNet classes. DenseNet's bottleneck layer is the Bottleneck class. Transition class represents the transition layer, which reduces feature map spatial dimensionality. This class represents the full DenseNet architecture. It shares similarities with ResNet but incorporates several distinctions. DenseNet features a composition of dense blocks and transition layers [20]. In a conventional CNN with L layers, there exist L connections, whereas in DenseNet, the number of connections amounts to $L(L + 1)/2$.

DenseNet offers several advantages:-

-Accuracy: Despite having fewer training parameters than ResNet, DenseNet achieves equivalent accuracy on the ImageNet classification dataset.

- Overfitting: DenseNet demonstrates remarkable efficacy in combating overfitting.

No. of Connections in DenseNet : $L(L+1)/2$ ---(1)

InceptionV3:

The model developed using researchers' ideas has an accuracy of more than 78.1% on the ImageNet dataset. Additionally, InceptionV3 employs more efficient computation methods through factorization technique. The Inception module, integral to InceptionV3, operates as a CNN network conducive to broader training. Unlike the conventional approach of adding more layers (which can lead to overfitting and increased parameters), the Inception module prioritizes width.

Xception:

Initially unveiled in the 2015 ILSVRC competition, Google's Xception, short for "Extreme version of Inception," presents a breakthrough in parameter reduction. Xception has 9 linearly stacked inception modules. It has 22 layers (27 including pooling levels). The final inception module employs global average pooling. A 1x1 convolution with 128 filters is employed for dimension reduction and corrected linear activation.

CNN:

A CNN is a deep learning system ideal for picture detection and processing. It has convolutional, pooling, and fully linked layers.

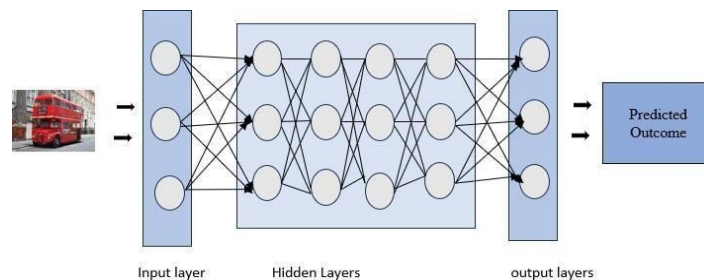


Fig 3: Architecture Of CNN

Implementation:

- 1.Data Collection and Preprocessing: Collect a varied dataset for training and testing, ensuring it's representative. Preprocess by resizing, normalizing pixels, and applying data augmentation for model robustness.
- 2.Feature Extraction with CNN: CNNs, such as VGG16, ResNet, or Inception, are commonly employed in CBIR to extract high- level features from images, enhancing performance.
- 3.Vector Representation: Convert extracted features into a fixed-size feature vector by flattening or employing global average pooling.
- 4.Indexing with KNN: Create an image vector database from the training set, then utilize a pre-trained CNN for feature extraction. Employ KNN for querying similar images in the database.

Implementation Steps:

- 1.Train or use a pre-trained CNN for feature extraction.
- 2.Extract features from all images in your database using the CNN.
- 3.When a user queries the system with an image, extract features from the query image using the same CNN.
- 4.Apply the KNN algorithm to find the most similar images in the feature space.
6. Evaluation: Evaluate the performance of your CBIR system using metrics like precision, recall, or F1 score, depending on the specific requirements of your project.

$$[\text{Precision} = \{ \text{True positives} \} / \{ \text{True positives} + \text{False positives} \}] \quad (2)$$

$$[\text{Recall} = \{ \text{True positives} \} / \{ \text{True positives} + \text{False negatives} \}] \quad (3)$$

$$[\text{F1 score} = 2 * \{ \text{Precision} * \text{Recall} \} / \{ \text{Precision} + \text{Recall} \}]$$

Algorithm:

1. start
2. Establish a connection
3. Load the necessary libraries.
4. Visualize and resize the images.
5. Convert the images into a NumPy array format and perform normalization to standardize the pixel values.
6. Complete the feature extraction using the CNN model
7. After that retrieve the Image by using the KNN model
8. stop

Expected Outcomes:

Performance metrics commonly used in such studies could include Precision, Recall, F1 Score, Accuracy, and perhaps additional measures specific to image retrieval tasks such as Mean Average Precision (mAP) or Mean

Average Recall (mAR). These metrics would be computed by comparing retrieved images against ground truth labels or annotations. The paper may also present comparative analyses with other existing methods or algorithms to showcase the efficacy and superiority of the proposed approach. Additionally, visualizations, experimental results, and discussions regarding the strengths and limitations of the proposed method would be included to provide comprehensive insights into its performance.

IV. LITERATURE SURVEY

A. Deep Learning for Instance Retrieval: A Survey

Lately, there has been a noteworthy rise in the production and sharing of visual content in fields such as medicine, robotics, and social media. This has created new challenges when trying to search databases for related material. To address this, researchers have been studying Content Based Image Retrieval (CBIR) and have found that AI has made the process of searching for specific instances much easier. There searchers have created new ways of using deep learning techniques such as extracting features, embedding and aggregating them, and fine-tuning networks. Their research focuses on finding relevant works, understanding the relationships between different methods, and identifying challenges and potential future directions in the field of retrieval based on image content.

B. Recent Advance in Visual similarity search: A Literature Survey

There is a lot of study going on in picture search or recovery because of how quickly and easily visual data is being added to the Web. Since text search methods for visual retrieval don't use visual content as a ranking clue, the words used in the text may not match up with the visual content. In the past 20 years, Visual Content Retrieval, which uses visual information to discover valuable photographs, has garnered attention. The purpose gap and the language gap make this kind of problem hard to solve. In the past ten years, many methods for retrieving images based on their information have been created. It is the goal of this study to organize and rate the algorithms that were suggested from 2003 to 2016. In the end, we suggest a few interesting areas for further study.

C. Image retrieval using color, texture, and shape features is facilitated through a content-based approach, employing both the image and its complement

Picture search engines formerly described photos by color, texture, and shape. This research proposes combining these factors to improve photo search. The picture and its corresponding counterpart are separated into tiles, and their complimentary hues are examined to establish color and texture patterns. This local data is maintained for two grid layouts and two image sizes with various characteristics. The method uses the MSHP idea and a bipartite network's adjacency matrix to compare photographs. It also utilizes Gradient Vector Flow fields for shape representation and invariant moments for recording form properties. By combining form, color, and texture data from a picture and its duplicate, effective image retrieval features are achieved. Experiments confirm the effectiveness of this approach. Utilizing image regions as query examples in image retrieval based on content: This research investigates the use of example photos in content based image search. It analyzes the effectiveness of retrieving images with similar low-level attributes based on example photos.

D. DOLG: Deep Orthogonal Fusion of Local and Global Features for Single-Stage Image Retrieval

Image retrieval is one of the most important steps in getting photos from a library that look like the question image. When retrieving pictures, it is common to first find candidate images through a similarity search that uses global image features. Then, the candidates are re-ranked by using their local features. Traditionally, memory problem solving has used either global or local image representation learning. However, in this study, we propose a one-stage solution that combines both local and global information within pictures. We introduce a framework called Deep Orthogonal Local and Global (DOLG) for capturing complete picture information. The local and global representations are then combined and summed to create a final image representation. This approach can be taught using visual labels and our model performs well on the Revisited Oxford and Paris datasets.

V. RESULTS

A. Performance Metrics

We can decide the algorithm performance by based on the Performance Metrics measures. We use CNN algorithm because it has better accuracy value.

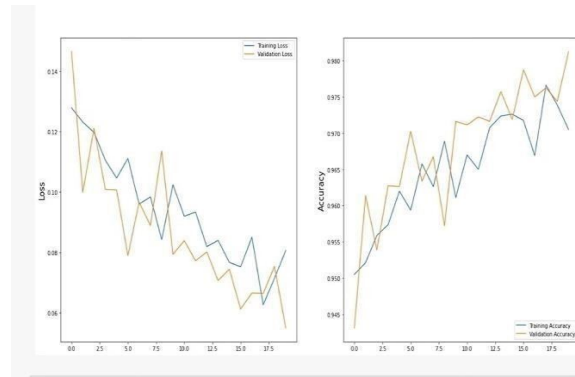


Fig 4: Comparision of Training and Testing Datasets

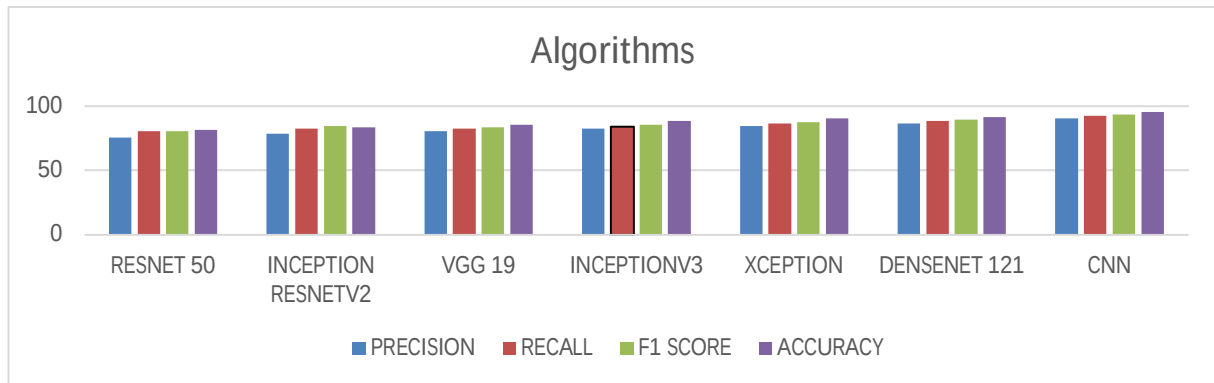


Fig 5: Performance Metrics Of Algorithms

VI. CONCLUSION

According to the results, this study is a big step toward solving the problem of picture search uncertainty in this era of rich digital material. Improving the quality of image searches has been shown to be very effective when custom mixed Convolutional Neural Network (CNN) in retrieval based on image content. Using information from CNN models that have already been trained, the new models are very good at pulling unique features that are needed to pick out minor patterns and traits. This helps solve the problem of telling the difference between images that look alike. The system's impressive performance is backed by a rigorous evaluation using well-known Dataset and a dataset collected by the system itself. Various metrics like accuracy, confusion matrices, F1 scores, Precision at Different Recall Levels (AP), and Average Precision across Classes (mAP) demonstrate the effectiveness of the method. Besides scientific achievements, this research reveals how multi deep neural networks can push Content-Based Image Retrieval. This study has effects that go beyond its local setting; it helps picture search tool ways to find similarly similar images on the internet in the future.

FUTURE SCOPE

In the future, custom mixed Convolutional Neural Network (CNN) models will be used in more areas and datasets, which will make them more flexible. More study could look into making the models better so they can be used on a larger scale and in real time, which would mean that they could be used for more than just retrieving images and videos. Leveraging state-of-the-art technologies such as artificial intelligence and natural language processing, could potentially enhance the intelligence and utility of systems through collaborative efforts. This study paves the way for more new ideas and suggests ways to improve and expand the use of material- Based Image Retrieval in the constantly changing world of digital material.

REFERENCES

- [1] R M Haralick, K Shangmugam, etc. Texture feature for image classification [J].IEEE Transaction on Systems,1973, SMC-3 (6): 768- 780

- [2] Anne H , Solberg S , Jain A K.Texture Fusion and Feature Selection Applied to SAR Imagery[J].IEEE Transaction on Geoscience and Remote Sensing, 1997, 35(2) : 475-478I.
- [3] A. W. Smeulders, M. M. Worring, A. Gupta, and R. Jain, ContentBased Image Retrieval at the End of the Early Years, IEEE Trans. Pattern Anal. Machine Intell., vol.22 no.12, pp1349-1380, 2000.
- [4] Deng, Y., Manjunath, B.S. Unsupervised segmentation of colortexture regions in images and video[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2001,Vol:23(8),800~ 810
- [5] Duan, Guoyong & Yang, Jing & Yang, Yilong. (2011). ContentBased Image Retrieval Research. Physics Procedia. 22. 471–477. 10.1016/j.phpro.2011.11.073.
- [6] Kannan, S., V., Udayakumar, E., & Dhinakar, P. (2021). Analysis of convolutional recurrent neural network classifier for COVID-19 symptoms over computerised tomography images. International Journal of Computer Applications in Technology, 66(3-4), 427-432.
- [7] P.S. Hirematch, Jagadeesh Pujari (2007). Content base image retrieval base on color, texture and shape feature using Image and its complement. IJCSS, International journal of computer science and security, vol 1, issue 4, Dec 2007,pp .25-35..
- [8] D. N. F. Awang Iskandar James A. Thom S. M. M. Tahaghoghi (2008). Content-based Image Retrieval Using Image Regions as Query Examples. CRPIT Volume 75- Database technologies
- [9] Shabaz Basheer Patel, Anand Sampat, “Semantic image search using queries”, Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on. IEEE, 2017
- [10] M. Yang, D. He, M. Fan, B. Shi, X. Xue, F. Li, E. Ding, and J. Huang, “Dolg: Single-stage image retrieval with deep orthogonal fusion of local and global features,” in ICCV, 2021, pp. 11 772–11 781.
- [11] N. Khosla, and V. Venkataraman, “Building Image-Based Shoe Search Using Convolutional Neural Networks”, CS231N Course Project Reports, (2015)
- [12] Agnes Borras, Francesc Tous, Josep Lladós, Maria Vanrell, High-Level Clothes Description Based on ColorTexture and Structural Features, In: Lecture Notes in Computer Science, Iberian Conference, Pattern Recognition and Image Analysis (2013)
- [13] Lukas Bossard, Matthias Dantone, Christian Leistner, Christian Wengert, Till Quack, Luc Van Gool, Apparel Classification with Style”, In: Computer Vision–ACCV 2012, Springer (2013)
- [14] Bui Thanh Hung, Prasun Chakrabarti. (2022) Parking Lot Occupancy Detection Using Hybrid Deep Learning CNN-LSTM Approach. Algorithms for Intelligent Systems book series (AIS)
- [15] Bui Thanh Hung, Vijay Bhaskar Semwal, Neha Gaud, Vishwanth Bijalwan. (2021) Violent Video Detection by Pre-trained Model and CNN-LSTM Approach. Proceedings of Integrated Intelligence Enable Networks and Computing, Springer Series in Algorithms for Intelligent Systems
- [16] C. Siva Kumar. (2023) Obtrusion unmasking of Machine Learning- Based Analysis of Imbalanced Network Traffic 2023 International Conference on Computer Communication and Informatics (ICCCI - 2023), 979-8-3503-4821-7/23/\$31.00 ©2023 IEEE.