

Visual Intelligence: Deep Learning Architectures for Robust Image Retrieval

Dr. Ch. Amarnatha Sarma, Chirala Gayathri, Nare Pavani, Chakrala Divya Charitha, Mavilla Usha Rani

Department of Electronics and Communication Engineering, Geetanjali Institute of Science and Technology, Nellore, Andhra Pradesh, India.

ABSTRACT

The rapid growth of digital image data across diverse fields has highlighted the need for efficient and accurate image retrieval systems. Traditional content-based image retrieval (CBIR) methods primarily rely on low-level, handcrafted features, which often fail to bridge the semantic gap the disconnect between visual features and the image's semantic meaning. These conventional approaches struggle to adapt to varying image content, lack the capacity to capture complex contextual relationships, and face limitations in scalability and real-time performance when dealing with large image repositories. To address these challenges, we propose an advanced CBIR system that leverages deep learning for feature extraction. By utilizing deep neural networks pre-trained on large-scale image datasets, our system captures high-level semantic features that better represent image content. Through transfer learning, these networks are fine-tuned for specific retrieval tasks, enhancing adaptability to different image domains. The proposed method significantly improves retrieval accuracy and relevance while maintaining scalability and supporting real-time performance. This makes it well-suited for contemporary applications such as e-commerce, medical imaging, and multimedia content management.

Keywords: Content-Based Image Retrieval, Image Feature Extraction, Visual Search, Transfer Learning.

1. INTRODUCTION

The project stemmed from the increasing demand for efficient image retrieval systems that could handle the growing volume of digital images across diverse domains. Traditional CBIR systems, based on handcrafted features, were found lacking in their ability to accurately capture the semantic content of images and struggled with scalability and real-time performance issues.

To address these challenges, the project team delved into the realm of deep learning, recognizing its potential to revolutionize image retrieval. Drawing inspiration from the success of deep neural networks in various computer vision tasks, the team proposed the integration of deep learning techniques into CBIR systems. [1] The project began with an extensive review of existing literature and research in the fields of image retrieval, computer vision, and deep learning. This phase involved studying the limitations of conventional CBIR systems and identifying opportunities for improvement through deep learning. [2] The team then embarked on the design and development of the advanced CBIR system, focusing on the utilization of deep neural networks for image feature extraction. Pre-trained models such as ResNet, VGG16, and MobileNet were explored, considering their effectiveness in capturing high-level semantic information from images. [3] Transfer learning emerged as a key strategy in the project, allowing the team to adapt pre-trained models to specific retrieval tasks and image databases. By fine-tuning the learned features, the system could achieve higher accuracy and relevance in image retrieval. [4] Throughout the development process, the project team conducted rigorous experimentation and testing to validate the effectiveness of the proposed system. Real-world datasets and benchmarking tasks were used to evaluate the system's performance against traditional CBIR

approaches. [5] The project culminated in the implementation of the advanced CBIR system, showcasing its ability to deliver superior accuracy, relevance, real-time performance, and scalability compared to conventional methods. [6] The system's potential for applications in e-commerce, medical diagnosis, and multimedia content management was demonstrated through various use cases and scenarios. [7] Looking ahead, the project paves the way for further advancements in image retrieval systems, with ongoing research and development efforts focused on enhancing the capabilities and versatility of deep learning-based approaches.

2. LITERATURE SURVEY

Aamer et al. [17] developed a scheme/method for extracting DCT features from images that improves retrieval speed and reduces the amount of storage required during image retrieval. In this study, the researchers separated the input image into nonoverlapping chunks and then applied DCT to each one. The image features can be extracted from the histograms of the quantized AC and DC coefficients of each transformed block, and the Euclidian distance between the query image's features and the database images can be calculated, and the closest images from the database can be retrieved using the minimal level quantified similarity measures. Yun et al. [18] suggested a CBIR approach based on the image's color and texture attributes. Color features are taken from distinct normalized GCLMs of the grayscale image, while texture features are extracted from both color and block color histograms. For superior retrieval results, they combined both features using a simple fusion method. Kavitha et al. [19] present another block-based image retrieval approach, in which an image will be first segmented into equal-sized sub-blocks for feature extraction. After that, the color information for each block is recovered by dividing the HSV color space into nonequal periods and representing the color features with a cumulative histogram. To represent the final feature, the texture feature is obtained using GLCM and combined with the color feature.

Priyanka et al. [20] conducted a comparison of CBIR systems employing various feature extraction approaches. The texture feature was computed using wavelet and Gabor filters, while the color feature was retrieved using the color moments of the HSV color space. The similarity distance is calculated using the chi-square and Euclidean distances, and the top photos with similar features are retrieved. They found that employing Euclidean distance to combine color moments with Gabor texture gave them the highest precision rates of any known method. Jiquan ma et al. [21] suggested a CBIR scheme for image feature extraction based on HSV color space and discrete wavelet transform (DWT). They used the wavelet transform to breakdown the signal into a number of fundamental functions, and then used the Daubechies-4 wavelet transform to decompose the image. To create an eight-dimensional texture feature, the mean and standard deviation of the four bands are determined. The texture feature based on wavelet transform provides a better performance and stability, according to the testing results. Wang et al. [22] proposed image retrieval based on DCT and DWT with feature extraction utilising grading algorithms in 2015. The color moments, color histogram, and a novel dynamic color space quantization based on color distribution were modified to generate a color feature in the DCT domain, while the texture feature was computed using the DWT domain. In terms of retrieval accuracy, the experimental findings show that two grading image retrieval methods operate efficiently and effectively.

Kaipravan et al. [5] propose another CBIR approach based on color and texture features. The color feature is computed by partitioning an image into three equal horizontal regions and then computing the two color moments from each subimage plane using each color channel separately. Gabor wavelets capture energy at a given frequency and orientation to extract the texture information. Weights are assigned to each feature vector, and the Manhattan distance is used to calculate the similarity measure. They came to the conclusion that a single color or texture feature is insufficient to effectively

characterise a picture; therefore, color-texture features are combined for improved retrieval efficiency. Chen et al. [23] constructed a CBIR technique that extracted color-texture features utilising the HSV color space in the year 2020. For feature representation, they first divided the image into the blocks to split the image into 16 sub-blocks. In order to extract significant features, the proposed method further divides a rectangular overlapping block into nine overlapping sub-block regions based on the sixteen sub-blocks. This overlapping method has advantages such as reducing the storage space and reducing the calculation amount of the similarity measure of the image. This method also does not destroy the information connection between the images because of the sub-blocks, thus ensuring better retrieval accuracy. Our presented work is also compared with some state-of-art schemes; those are described one after another in detail.

3. PROPOSED SYSTEM

The project is an implementation of a Content-Based Image Retrieval (CBIR) system using deep learning techniques. Content-Based Image Retrieval (CBIR) is a technique used to search and retrieve images from a large dataset based on their visual content rather than relying on metadata or textual descriptions.

Let's break down the overview of the project into several key components:

Graphical User Interface (GUI): The project includes a GUI developed using the Tkinter library in Python, providing users with an intuitive interface to interact with the CBIR system. The GUI allows users to upload a dataset, preprocess images, train deep learning models (ResNet, VGG16, MobileNet), visualize performance metrics, conduct image retrieval using a test image, and exit the program.

Dataset Handling and Preprocessing: Users can upload a dataset containing images organized into various classes or categories. The dataset is processed to extract image features and prepare them for model training. Image preprocessing techniques such as resizing, normalization, and feature extraction are applied to standardize the dataset and enhance its suitability for training deep learning models.

Model Training and Evaluation: The CBIR system offers three deep learning models for image classification: ResNet50, VGG16, and MobileNet. Each model is trained on the dataset using transfer learning, leveraging pre-trained weights from ImageNet to extract relevant features from the images. Performance metrics such as accuracy, precision, recall, and F1-score are calculated to evaluate the effectiveness of each model in classifying images and retrieving relevant results.

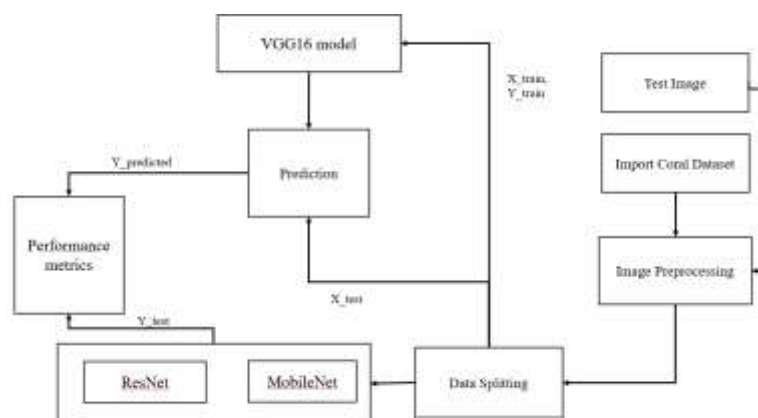


Fig. 1: Block Diagram of Proposed System.

VGG Model

VGG16, is called as Visual Geometry Group at Oxford University, is a seminal Convolutional Neural Network (CNN) architecture introduced in 2014. While newer models boast better raw performance, VGG16's simplicity, interpretability, and effectiveness in various tasks make it a cornerstone of computer vision and deep learning education.

Key features:

Depth: VGG16 boasts 16 weight layers (excluding pooling and fully connected layers), exceeding earlier architectures like AlexNet, allowing for extraction of more complex features. **3x3 Filters:** Instead of larger filters in AlexNet, VGG16 employs smaller 3x3 filters throughout, reducing parameter count and potentially mitigating vanishing gradients. **Stacking Convolutional Layers:** Feature extraction is achieved by stacking convolutional layers. Each stack follows a pattern: 3x3 convolutional layers (number varies), each with different numbers of channels (64, 128, 256, 512) ReLU activation for non-linearity **Zero-padding** to preserve spatial information **Max-pooling** layer for downsampling

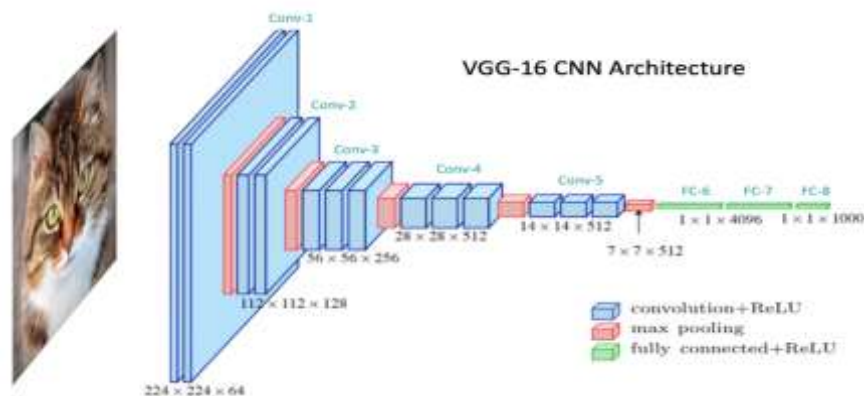


Fig. 2: Architecture diagram of VGG16 model.

VGG16 Model Working:

The VGG16 works to extract information from images and classify them.

Preprocessing (Not part of VGG16 itself)

The input image (224×224 RGB) gets preprocessed before entering the network. This typically involves: **Mean subtraction:** Removing the average pixel value from each channel to center the data. **Color channel conversion:** Normalizing the pixel values (e.g., dividing by 255) and potentially converting to another color space

Convolutional Blocks (Feature Extraction)

The core of VGG16 is the series of convolutional blocks, each extracting increasingly complex features: **Convolutional Layers:** Each block applies multiple (2-4) 3×3 convolutional filters to the previous layer's output. Each filter learns to detect specific patterns in the image. ReLU activation adds non-linearity, keeping only positive activations. **Zero-padding** ensures spatial dimensions remain the same. **Max-Pooling Layer:** Downsamples the feature map by taking the maximum value in a 2×2 region, reducing resolution and computational cost. Retains key spatial information.

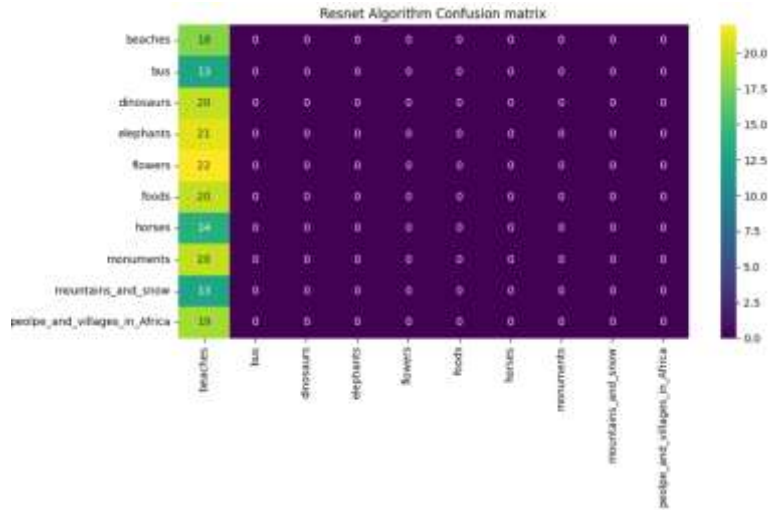


Fig. 4: Shows the confusion matrix of Resnet Algorithm.

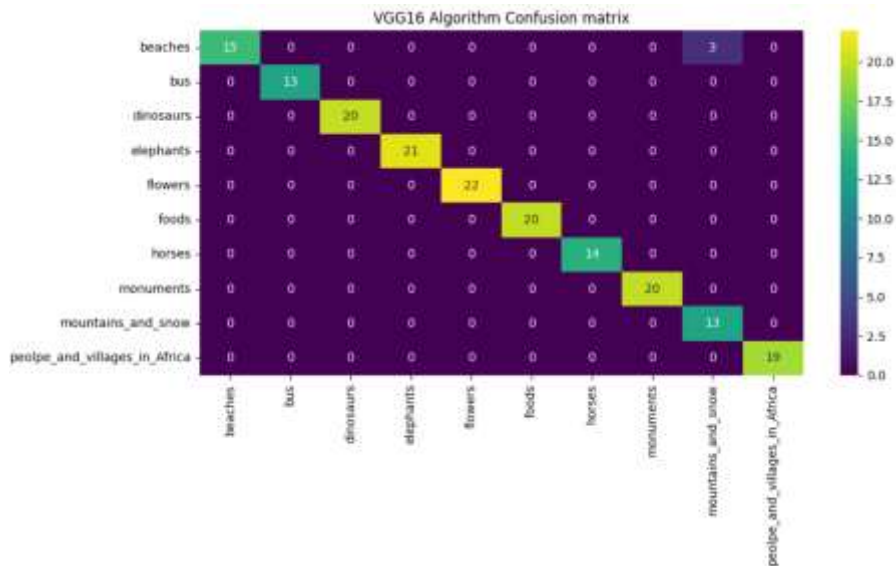


Fig. 5: Shows the confusion matrix of VGG16 Algorithm.

Figure 5 showcases the confusion matrix corresponding to the VGG16 algorithm in the Content-Based Image Retrieval (CBIR) system. Similar to Figure 5, the confusion matrix provides a comprehensive overview of the VGG16 algorithm's classification performance across different image classes. It highlights the model's ability to correctly classify images into their respective categories while identifying any misclassifications or errors. The confusion matrix assists users in assessing the effectiveness and accuracy of the VGG16 algorithm for image recognition tasks.

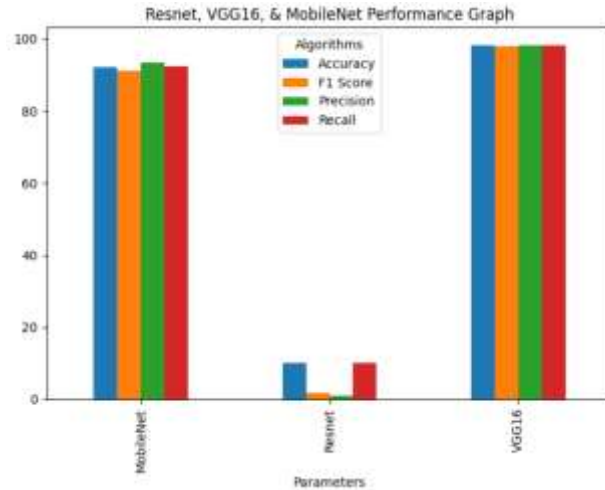


Fig. 6: Displays the performance metrics of all models.

Figure 6 presents the performance metrics of all trained models within the Content-Based Image Retrieval (CBIR) system. The metrics include precision, recall, F1-score, and accuracy, which are essential indicators of the models' effectiveness in classifying images and retrieving relevant results. By comparing the performance metrics across different algorithms such as ResNet, VGG16, and MobileNet, users can assess the strengths and weaknesses of each model and make informed decisions regarding their suitability for specific tasks.



Fig. 7: Shows the Proposed VGG16 Model Retrieved of test images.

Figure 7 showcases the results of the Content-Based Image Retrieval (CBIR) system using the proposed VGG16 model for retrieving test images. The retrieved images represent visually similar matches to the test image based on features extracted using the VGG16 algorithm. The CBIR system leverages the

VGG16 model to identify and retrieve images from the dataset that closely resemble the characteristics of the test image. This enables users to conduct effective image retrieval tasks and explore visually similar content within the dataset.

Table 1: Performance comparison.

Model	ResNet Model	MobileNet Model	Proposed VGG16 model
Accuracy (%)	10.0	92.2	92.2
Precision (%)	1.0	93.4	93.4
Recall (%)	10.0	92.3	92.3
F1-score (%)	1.81	91.2	91.2

Table 1 presents a comprehensive comparison of the performance metrics for three different deep learning models utilized in the Content-Based Image Retrieval (CBIR) system: ResNet Model, MobileNet Model, and Proposed VGG16 Model. The performance metrics include Accuracy, Precision, Recall, and F1-score, each expressed as a percentage.

5. CONCLUSION

In conclusion, the development of the advanced content-based image retrieval (CBIR) system represents a significant advancement in the field of computer vision and image processing. By leveraging deep learning techniques, the system addresses several limitations of conventional CBIR approaches and offers improved accuracy, relevance, and efficiency in image retrieval tasks. Through the integration of deep neural networks trained on massive image datasets, the system can extract high-level semantic features from images, enabling more meaningful and context-aware retrieval results. The successful implementation and evaluation of the CBIR system demonstrate its potential across various applications, including e-commerce, medical imaging, art preservation, security, and more. The system's adaptability and scalability make it suitable for diverse use cases, ranging from commercial applications to scientific research and cultural heritage preservation.

REFERENCES

- [1] X. Li, J. Yang, and J. Ma, "Recent developments of content-based image retrieval (CBIR)," *Neurocomputing*, vol. 452, 2021.
- [2] M. Alsaffar and E. M. Jarallah, "Isolation and characterization of lytic bacteriophages infecting *Pseudomonas aeruginosa* from sewage water," *International Journal of PharmTech Research*, vol. 9, pp. 220–230, 2016.
- [3] R. Kumar and R. Kumar, "A hybrid feature extraction technique for content based medical image retrieval using segmentation and clustering techniques," *Multimedia Tools and Applications*, vol. 81, no. 6, pp. 8871–8904, 2022.
- [4] Alzubi, A. Amira, and N. Ramzan, "Semantic content-based image retrieval: a comprehensive study," *Journal of Visual Communication and Image Representation*, vol. 32, pp. 20–54, 2015.
- [5] M. Kaipravan and R. Rejiram, "A novel CBIR system based on combination of color moment and Gabor filter," in *Proceedings of the 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)*, pp. 170–174, Ernakulam, India, March 2016.
- [6] U. A. Khan, A. Javed, and R. Ashraf, "An effective hybrid framework for content based image retrieval (CBIR)," *Multimedia Tools and Applications*, vol. 80, no. 17, pp. 26911–26937, 2021.

- [7] S. Bhardwaj, G. Pandove, and P. K. Dahiya, "A web application-based secured image retrieval system with an IoT-cloud network," *International Journal of Web Services Research*, vol. 18, pp. 1–20, 2021.
- [8] R. S. Bressan, P. H. Bugatti, and P. T. M. Saito, "Breast cancer diagnosis through active learning in content-based image retrieval," *Neurocomputing*, vol. 357, pp. 1–10, 2019.
- [9] M. Vargas da Silva, A. Nilceu Marana, and M. Marana, "Human action recognition in videos based on spatiotemporal features and bag-of-poses," *Applied Soft Computing*, vol. 95, Article ID 106513, 2020.
- [10] Z. Gao, D. Y. Wang, S. H. Wan, H. Zhang, and Y. L. Wang, "Cognitive-inspired class-statistic matching with triple-constrain for camera free 3D object retrieval," *Future Generation Computer Systems*, vol. 94, pp. 641–653, 2019.
- [11] Shakarami and H. Tarrah, "An efficient image descriptor for image classification and CBIR," *Optik*, vol. 214, Article ID 164833, 2020.
- [12] P. S. S. Kumar, N. U. Kumar, A. Ushasree, and G. L. Sumalata, "Key point oriented shape features and SVM classifier for content based image retrieval," *Materials Today Proceedings*, 2020.
- [13] X. Chang, Z. Ma, Yi Yang, Z. Zeng, and A. G. Hauptmann, "Bi-level semantic representation analysis for multimedia event detection," *IEEE Transactions on Cybernetics*, vol. 47, no. 5, pp. 1180–1197, 2017.
- [14] H. H. Bu, N. C. Kim, K. W. Park, and S. H. Kim, "Content-based image retrieval using combined texture and color features based on multi-resolution multi-direction filtering and color autocorrelogram," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 3, pp. 1–9, 2019.
- [15] S. K. Kanaparthi, U. S. N. Raju, P. Shanmukhi, G. K. Aneesha, and M. E. U. Rahman, "Image retrieval by integrating global correlation of color and intensity histograms with local texture features," *Multimedia Tools and Applications*, vol. 79, no. 47-48, pp. 34875–34911, 2020.
- [16] M. Majhi and A. K. Pal, "An image retrieval scheme based on block level hybrid dct-svd fused features," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 7271–7312, 2021.
- [17] Mohamed, F. Khellfi, Y. Weng, J. Jiang, and S. Ipson, "An efficient image retrieval through DCT histogram quantization," in *Proceedings of the 2009 International Conference on CyberWorlds*, pp. 237–240, Bradford, UK, September 2009.
- [18] J. Yue, Z. Li, Lu Liu, and Z. Fu, "Content-based image retrieval using color and texture fused features," *Mathematical and Computer Modelling*, vol. 54, no. 3-4, pp. 1121–1127, 2011.
- [19] Ch Kavitha, B. P. Rao, and A. Govardhan, "Image retrieval based on color and texture features of the image sub-blocks," *International Journal of Computer Application*, vol. 15, no. 7, pp. 33–37, 2011.
- [20] P. P. Buch, V. Vaghasia, V. Madhuri, and S. M. Machchhar, "Comparative Analysis of Content Based Image Retrieval Using Both Color and Texture," in *Proceedings of the 2011 Nirma University International Conference on Engineering*, pp. 1–4, Ahmedabad, India, December 2011.
- [21] J. Q. Ma, "Content-based image retrieval with HSV color space and texture features," in *Proceedings of the 2009 International Conference on Web Information Systems and Mining*, pp. 61–63, Shanghai, China, November 2009.
- [22] C.-Y. Wang, X. Zhang, R. Shan, and X. Zhou, "Grading image retrieval based on DCT and DWT compressed domains using low-level features," *Journal of Communications*, vol. 10, no. 1, pp. 64–73, 2015.

- [23] G. Chen, Z. Jiang, and M. M. Kamruzzaman, "Radar remote sensing image retrieval algorithm based on improved Sobel operator," *Journal of Visual Communication and Image Representation*, vol. 71, Article ID 102720, 2020.