A Comparative Study of Pre-Trained CNN Models with Transfer Learning for Content-Based Image Retrieval

Monica Palla
Department of EECE, GITAM (Deemed to be University), Visakhapatnam, Andhra Pradesh, India
mpalla3@gitam.in (corresponding author)

Renu Karra
Department of EECE, GITAM (Deemed to be University), Visakhapatnam,Andhra Pradesh, India
rkarra@gitam.edu (corresponding author)

Received: 15 April 2025 | Revised: 5 May 2025, 23 May 2025, 2 June 2025, and 9 June 2025 | Accepted: 14 June 2025

 DOI: https://doi.org/10.48084/etasr.11496

ABSTRACT

Content-Based Image Retrieval (CBIR) involves searching for images that are visually similar within large image databases.Convolutional Neural Networks (CNNs) are important in CBIR tasks and classifications. However, handling large-scale datasets with a high number of categories continues to be a significant challenge. This study investigated the effectiveness of four widely used pretrained models, EfficientNet-B0, DenseNet-201, VGG-16, and AlexNet, for CBIR on the Corel-10K dataset. The four models are fine-tuned using transfer learning and data augmentation techniques and evaluated using mean Average Precision at rank K (P@K) and across all queries (mAP@K) to enhance their retrieval performance. Experimental results prove that the DenseNet-201 architecture achieves the best retrieval performance among the models, with a P@10 of 92.1% and a mAP@10 of 94.7%. The VGG-16 model also performs well, while EfficientNet-B0, despite showing lower performance compared to others, maintainscomputational efficiency. DenseNet-201 achieved superior precision on the Corel-10K dataset compared to other methods, demonstrating its effectiveness in large-scale image retrieval. This study helps identify effective pre-trained models for CBIR and serves as a basis for future advances in retrieval systems, including better handling of image orientation changes.

Keywords-content-based image retrieval; pre-trained CNN models; transfer learning; Corel-10K dataset

I. INTRODUCTION

Content-Based Image Retrieval (CBIR) has become important in several fields, such as healthcare, security surveillance, and digital libraries, where correct image retrieval is important [1, 2]. CBIR aims to retrieve images based on visual content rather than textual metadata, because it is more accurate than textual information [3]. As digital image collections grow rapidly, traditional handcrafted feature methods struggle to handle complex image variations, including scale changes, background clutter, and semantic gaps [4], bridging between basic features and high-level image understanding. To address these challenges, recent advances in Deep Learning (DL), particularly Convolutional Neural Networks (CNNs), have gained importance in extracting reliable and high-level image features for retrieval tasks [5-9].Pre-trained CNN models, although primarily designed for large-scale image classification tasks, work well for CBIR due to their ability to capture hierarchical feature representations [10]. Several advanced models, including EfficientNet-B0[11],DenseNet-201[12],VGG-16[13], andAlexNet[14], have shownremarkable performance in various computer vision applications[15]. However, their effectiveness in CBIR, particularly in large-scale datasets such as Corel-10K, requires further evaluation.

Recently, the application of DL to CBIR tasks has gained more focus and exploration, with multiple studies examining its potential [16, 17].Several studies have examined how different CNN architectures perform on retrieval tasks, highlighting their strengths and limitations across diverse datasets. Although many models exist, detailed comparative research using several pretrained networks on large-scale datasets such as Corel-10K remains limited. To address this gap, this study conducts an in-depth comparison of EfficienNet-B0, DenseNet-201, VGG-16, and AlexNet for CBIR tasks. The models are fine-tuned using transfer learning and measured using Precision at rank K (P@K) and the mean Average Precision (mAP@K) [18] score across all queries to improve their effectiveness in image retrieval tasks. This study aims to find out which pre-trained model performs well in terms of both precision and computational efficiency for large-scale image searches.

The field of CBIR has transitioned from traditional methods that involve handcrafted features, such as color histograms and texture descriptors, to more robust DL-based methods that better manage the complexity of image data.Pretrained architectures, such as AlexNet and VGG-16, were some of the first models utilized for CBIR, showing strong results due to their effectiveness in capturing high-level semantic features [19].Later architectures, such as DenseNet-201, known for its densely connected layers, and EfficientNet-B0, with compound scaling, have further refined both the retrieval accuracy and the efficiency of image retrieval systems [20]. Several studies have compared these models for classification tasks,but only a few have focused on their evaluation for CBIR tasks using extensive datasets such as Corel-10K [21]. Recent studies have explored the use of Vision Transformers (ViTs)in image retrieval, using self-attention mechanisms that allow for capturing global connections within visual data. Although ViTs show potential, their application in CBIR remains an active area of research [22].Hybrid approaches that integrate CNNs with attention mechanisms have also been proposed to improve retrieval performance [23].To address a gap in existing research, this study compares four widely used pretrained CNNs for CBIR on the Corel-10K dataset to guide future retrieval system design.

In addition to DL models, several handcrafted or hybrid feature-based approaches have been proposed for CBIR. In [24], a method was based on the separation of texture and color regions using a probability annular histogram and weighted similarity matching.In [25], a multihierarchical clustering approach used the Davis-Bouldin index for robust image retrieval.In [26], an unsupervised CBIR framework was developed using pretrained CNN and PCNN feature extractors. These methods offer alternative perspectives on feature representation and similarity computation in CBIR systems.

II. PROPOSED WORK

Figure 1 illustrates the workflow of the proposed CBIR model.

A. Dataset Details

This study used the Corel-10K dataset, which comprises 10,000 images arranged into 100 categories, each with 100 images.The Corel-10K dataset [27] includes many types of images, such as landscapes, animals, people, and more. Its structureand diverse content serve as a good reference for testing CBIR systems.

B. Dataset Preprocessing

Dataset splitting is a standard approach inDL to improvethe training and evaluation processes of a model. The dataset is divided into 80% training, 10% validation, and 10% testing.Preprocessing helps prepare the images for model training, using image resizing and normalization.The original dimensions of images in the dataset vary in size, ranging from 120×80 to 384×256 pixels. They were resized to 224×224 pixels to align with the input size requirements of pretrained networks. However, for AlexNet, the images were resized to 227×227 pixels as per its specific input size requirement. Image resizing is mathematically represented as:

$$I'(x', y') = I\left\{\frac{x'.W}{W\prime}, \frac{y'.H}{H\prime}\right\} \quad (1)$$

where $I(x, y)$ is the original image, and $I'(x', y')$ is the resized image. In this context, W and H correspond to the width and height of the original images, while W' and H' denote the width and height of the resized image. The variables x' and y'denote the coordinates in the resized image.

After resizing, a normalization technique was applied to all images. Min-max scaling was applied to bring pixel values into the $[0, 1]$ range by dividing each pixel by 255. Pixel normalization is mathematically represented as:

$$NormalizedPixelValue = \frac{PixelValue - Min}{Max - Min} \quad (2)$$

where the $PixelValue$ is the original intensity, Min denotes the lowest value, and Max indicates the highest value in the image.

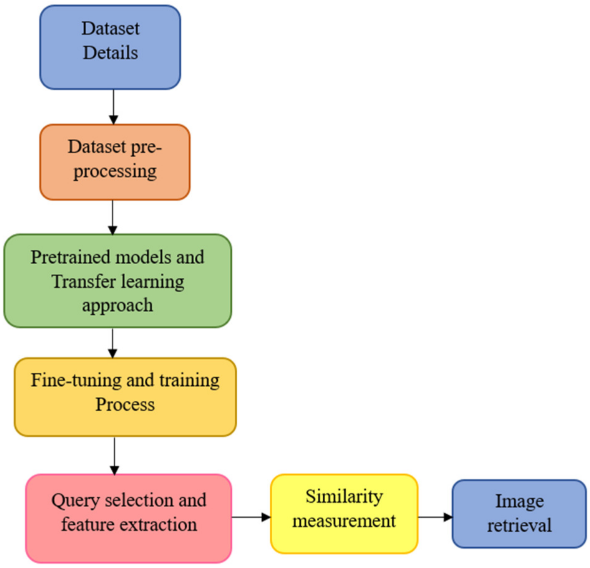

Fig. 1. Workflow of the proposed CBIR system using CNN models (EfficientNet-B0, DenseNet-201, VGG-16, and AlexNet) on the Corel-10K dataset.

C. Pretrained Models and Transfer Learning Approach

This study uses four popular pre-trained models, EfficientNet-B0, DenseNet-201, VGG-16, and AlexNet, to classify and retrieve images from Corel-10 K. All were initially trained on ImageNet, which contains more than 14 million images spread across 21,000 categories. At this stage, the models learn to identify important visual elements that can be applied to many different image categories.These models were chosen for their strong results in image recognition and can extract meaningful patterns from images. EfficientNet and DenseNet were selected based on their strong past results, while VGG-16 and AlexNet were used because they are popular and well-established models. Their unique structures allow them to learn a wide range of features, making the image search process more accurate. Strong feature extraction by each model increases how well images can be retrieved based on how they look.

Transfer learning [28] helps adapt a model trained on one dataset to perform well in another similar task by reusing its features. This technique speeds up training and helps the model work better on new data by building on what it has already learned. This work applied transfer learning to customize each model to the Corel-10K dataset by updating its final layers and retraining the feature extraction layers.

1) EfficientNet-B0

EfficientNet-B0 is a compact CNN model, known for its strong performance in image retrieval tasks. It achieves this by efficiently balancing model depth, width, and input image resolution through a technique called compound scaling. The model incorporates a Mobile inverted Bottleneck Convolution (MBConv) layer and squeeze and excitation techniques to improve feature extraction. The model achieves a strong balance of performance and computational speed by adjusting the network's size, structure, and input image dimensions. This architecture enables strong performance in large-scale CBIR tasks while maintaining low computational costs.

In this study, EfficientNet-B0 began with weights from ImageNet and was modified by replacing the original classification layers with custom fully connected layers. The model was fine-tuned using the Corel-10K dataset to adapt to specific image categories. Feature vectors were extracted from the *global_pooling2d layer*, which captures compact and meaningful semantic information. These features are stored for use in the image retrieval process.

2) DenseNet-201

DenseNet-201 is aneural network with a total of 201 layers.It includes dense connections, where each layer takes input from all previous layers and passes its output to every following layer. This helps the model better capture complex features and reduces the number of parameters. This architecture allows information to flow easily between layers, helping the network learn better and avoid training problems. As a result, it achieves high performance in tasks such as image retrieval while using fewer parameters, making it more computationally efficient.

In this study, the original classification layers of DenseNet-201 were replaced with custom fully connected layers during fine-tuning and retrained on Corel-10K to improve its feature extraction capabilities for image retrieval. Feature vectors were extracted from the *avg_pool layer*, which captures deep and distinctive representations suitable for CIBR.

3) VGG-16

The VGG-16 architecture is a deep CNN consisting of 16 layers in total, including a sequence of 13 convolutional layers and three fully connected ones. This design helps to capture fine-grained features in images. It is commonly considered a strong baseline model due to its consistent performance across various classification and image retrieval tasks. Its deeper network allows it to extract more detailed feature patterns, although it requires more computational resources compared to other models. However, it is widely used in CBIR because it performs well and can effectively manage large datasets.

In this study, the pretrained VGG-16 model was tailored for the Corel-10K dataset by fine-tuning its final layers, which were replaced with custom fully connected layers and fine-tuned for the image retrieval task. Features were extracted from the *fc7* layer, which provides high-level semantic representations suitable for retrieving visually similar images.

4) AlexNet

AlexNet is an early and well-known CNN model that played a key role in advancing DL for image recognition. The model includes five convolutional and three fully connected layers, incorporating ReLu activations and dropout to manage overfitting. These design choices contribute to faster learning and more robust performance. Although less complex than modern deep networks, it efficiently extracts low- and mid-level features, making it effective for tasks such as image retrieval where speed and simplicity are important.

AlexNet was pretrainedonImageNet.Its final layers were fine-tuned and retrained for the Corel-10K dataset to better adapt to its specific classes. The original classification layers were replaced with custom fully connected layers, and features were extracted from the *fc7* layer, capturing abstract image characteristics useful for image retrieval.

D. Fine-Tuning and Training Process

This study employed transfer learning by fine-tuning each pretrained model to adjust to the distinct characteristics of the Corel-10K dataset. The following fine-tuning steps were applied consistently:

- Freezing Initial Layers: The initial convolutional layers of all models' weights were frozen, retaining the general feature extraction capabilities learned from the ImageNet dataset, ensuring that the learned low-level features, such as edges, textures, and colors, remained intact.
- Replacing Final Classification Layers: The original fully connected layers, which were specific to ImageNet classification, were replaced with new layers tailored to match the 100 categories of the Corel-10K dataset.
- Retraining Top Layers: Only the newly added layers and the last few convolutional layers were retrained using the Corel-10K dataset to allow specialization for image retrieval and classification tasks without losing general features.
- Data Augmentation: To improve the generalization and robustness of the model, data augmentation techniques such as random rotation, scaling, translation, and reflection were applied during the training. This helped the model handle variations in image appearance and reduce overfitting.

1) Training Process

All four models were trained using Stochastic Gradient Descent with Momentum (SGDM), with a consistent set of hyperparameters to ensure a fair comparison. The training process was carried out for 50 epochs, with periodic validation based on the dataset size.The following hyperparameters were applied uniformly across all models:

- Mini-batch size: 32,

- Maximum epochs: 50,
- Initial learning rate:0.0001,
- Optimizer: SGDM,
- Validation frequency: Adjusted based on dataset size.

E. Query Selection and Feature Extraction

Feature extraction plays a vital role in representing images within a high-dimensional feature space by identifying key patterns and essential information that make each image distinct from others. The extracted features serve as numerical descriptors that capture essential features, such as texture, shape, and color, enabling effective image retrieval.

1) Feature Extraction

The EfficientNet-B0, DenseNet-201, AlexNet, and VGG-16 models were utilized to extract robust and discriminative features. The feature extraction layers were carefully selected as follows:

- EfficientNet-B0: Features are extracted from the *global_average_pooling* layer.
- DenseNet-201: Features are taken from the *avg_pool* layer.
- AlexNet: Features are obtained from the *fc7* layer.
- VGG-16: Features are collected from the *fc7* layer.

These layers capture high-level abstract information from the images, ensuring that the retrieved images are semantically close to the query image. The feature vectors obtained from the model are kept in a database to support futureimage retrieval tasks.

2) Query Selection

In the retrieval process, a query image is selected from the dataset or provided by the user. The query images underwent the same preprocessing and feature extraction steps to ensure compatibility. Once the query feature vector is obtained, it is used in the similarity computation process to retrieve relevant images.

F. Similarity Measurement

Euclidean distance evaluates how similar the query image is to the stored features. It is calculated as the straight-line distance between two points in the feature space, where smaller distances indicate higher similarity. The retrieved images are ranked based on the calculated Euclidean distances, with the closest matches presented as output. The Euclidean distance between two feature vectors is given by:

$$EuclideanDistance(a,b) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2} \quad (3)$$

where a and b are feature vectors, a_i and b_i are the i^{th} components, and n is the number of dimensions.

G. Image Retrieval

The system ranks the retrieved images according to how similar they are to the query, showing the most relevant ones first. This method works by extracting deep features and measuring similarity with Euclidean distance for better results.

III. RESULTS AND DISCUSSION

A. Experimental Setup

All experiments were carried out using MATLAB R2024a on a Windows 11 DELL machine equipped with an Intel Core i7-13700 2.10 GHz processor and an NVIDIA GeForce RTX 3050 OEM GPU.

B. Classification Results

Before feature extraction for retrieval, classification performance was evaluated for all four models. Each pretrained model was fine-tuned for 50 epochs, and the corresponding validation metrics are shown in Table I.

TABLE I. CLASSIFICATION METRICS OF PRE-TRAINED MODELS ON COREL-10K DATASET

Model	Validation accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
EfficientNet-B0	88.9	88.90	90.18	88.67
DenseNet-201	92.2	92.20	92.82	92.13
VGG-16	91.0	91.00	91.87	90.91
AlexNet	88.2	88.20	89.70	88.02

C. Computational Efficiency of the Models

Computational metrics were used to further analyze the trade-off between performance and efficiency. These metrics, including the number of parameters, average inference time, and available memory usage for each model, are critical in evaluating the practical feasibility of deploying these models for image retrieval tasks. Table II summarizes the computational details of each model used. EfficientNet-B0 provides a reasonable balance, but DenseNet-201 and VGG-16 offer better performance in terms of faster inference times, despite having more parameters. The results provide a comprehensive view of the trade-off between model complexity and computational efficiency.

TABLE II. COMPUTATIONAL EFFICIENCY OF PRETRAINED MODELS FOR CBIR

Model	Total parameters	Average inference time (s)	Used GPU memory (MB)
EfficientNet-B0	4,114,640	0.226	7,152.71
DenseNet-201	18,072,996	0.0717	7,152.71
VGG-16	134,670,244	0.0657	6,633.28
AlexNet	57,277,924	0.0395	6,411.46

D. Retrieval Performance Evaluation

Model performance in retrieval tasks was tested using Precision at K (P@K), Average Precision at K (AP@K), and mean Average Precision at K (mAP@K) to determine how relevant the top retrieved images are.

1) Precision at Rank K (P@K)

P@K calculates the ratio of relevant images in the top K retrieved results and is defined as:

$$P@K = \frac{Number\ of\ relevant\ images\ in\ top\ K}{K} \quad (4)$$

where K is the number of top retrieved images, and P@K represents the precision among the top K results.

2) Average Precision at K (AP@K)

AP@K quantifies the precision of a retrieval system by averaging the precision scores at positions where relevant images appear, normalized by the minimum of total relevant images and K. This ensures that AP@K accounts for varying numbers of relevant images across different queries. It can be calculated using the following expression.

$$AP@K = \frac{1}{min(TotalRelevantImages, K)} \sum_{i=1}^{K} \left(\frac{relevantCount(i)}{i}\right) \quad (5)$$

where $relevantCount(i)$ refers to the number of relevant images in the top i retrieved results, and $TotalRelevantImages$ denotes the count of ground truth relevant images for the query.

3) Mean Average Precision at K (mAP@K)

This is a metric used to evaluate the performance of information retrieval systems, particularly where the relevance of retrieved items is crucial. It computes the mean of the AP@K scores over all queries as.

$$mAP@K = \frac{1}{N} \sum_{n=1}^{N} AP@K(n) \quad (6)$$

where N indicates the total number of queries, and $AP@K(n)$ refers to the AP@K for the n-th query.

Table III presents the P@K and mAP@k scores, calculated for all 100 classes in the Corel-10K dataset, for different values of K, providing insight into how effectively each pretrained CNN model ranks relevant images within the top retrieved results.Figures 2 and 3 present the P@K and mAP@k across various K values to compare the retrieval effectiveness of each pretrained CNN model, allowing a clearer comparison of model performance and highlighting variations in retrieval effectiveness across different architectures.

TABLE III. RETRIEVAL PERFORMANCE AT DIFFERENT K VALUES

Models	Metrics	Top-5	Top-10	Top-15	Top-20
EfficientNet-B0	P@K	87.4	83.1	79.6	77.1
	mAP@K	92.7	88.5	86.0	84.0
DenseNet-201	P@K	94.0	92.1	90.6	89.1
	mAP@K	96.3	94.7	93.4	92.5
VGG-16	P@K	91.0	85.1	81.4	79.5
	mAP@K	94.3	90.5	87.9	86.0
AlexNet	P@K	82.8	78.3	74.2	70.7
	mAP@K	89.8	85.6	82.0	79.6

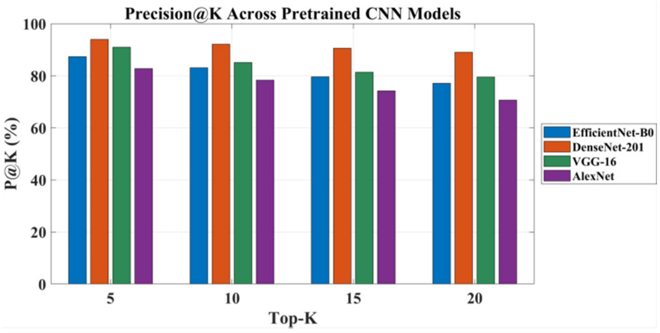

Fig. 2. Comparison of P@K values for different pretrained models.

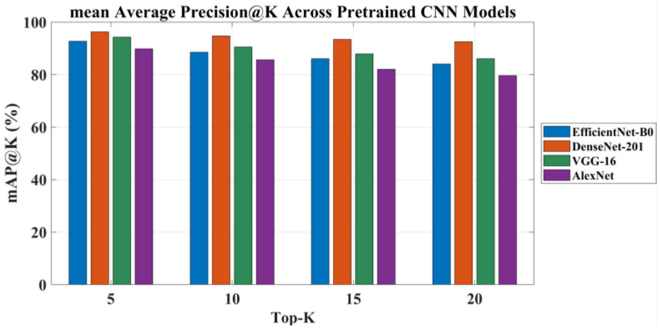

Fig. 3. Comparison ofmAP@K values for different pretrained models.

In Figures 2 and 3, different colors were assigned to each model to ensure clear differentiation and improve visual clarity and accessibility. It is evident that DenseNet-201 consistently achieves the best performance across all metrics, clearly outperforming the other models in both P@K and mAP@K. It is followed by VGG-16, which, despite being an older architecture, shows strong retrieval performance, particularly in top-5 and top-10 retrievals. EfficientNet-B0, although it has a relatively lightweight architecture, demonstrates competitive performance, particularly at lower K values. This suggests that EfficientNet-B0 effectively captures high-quality features while maintaining computational efficiency. AlexNet shows the lowest performance among the four, particularly as K increases, indicating its limitations in feature extraction for complex image retrieval tasks. The declining trends in P@K and mAP@K with increasing K indicate that higher values of K introduce more non-relevant images in the retrieved set, which is expected in CBIR systems.

In addition to the quantitative evaluation metrics, the retrieval performance of the pretrained CNN models was also qualitatively analyzed.For each model, the top 10 results retrieved for several query images were examined. The relevance of the retrieved images was assessed based on class similarity, as defined by the Corel-10K dataset labels.DenseNet-201 and EfficientNet-B0 show strong visual and semantic relevance in retrieval, especially for detailed categories.VGG-16 and AlexNet also perform reasonably well, although with occasional inconsistencies in precision for certain classes. In this study, a retrieved image is considered correct if it belongs to the same class as the query image, based on the Corel-10K dataset labels. Precision is calculated as the ratio of correct images in the Top-K results. All evaluated models demonstrate robustness to image orientation by effectively retrieving the correct images even when using rotated versions of the query images, although AlexNet shows comparatively lower retrieval precision. These results visually supported the quantitative findings.

E. Ablation Study

To assess the contribution of each model individually, an ablation study was performed of the proposed models separately on the Corel-10K dataset. The results showed that DenseNet-201 achieved the highest classification accuracy (92.2%) and retrieval performance (P@5=94.0%, mAP@5=96.3%), outperforming the other models. EfficientNet-B0, while offering efficient computational performance, achieved a classification accuracy of 88.9% and retrieval metrics of P@5=87.4% and mAP@5=92.7%. VGG-16 also performed well with 91.0% classification accuracy and strong retrieval results (P@5=91.0%,mAP@5=94.3%). However, AlexNet showed relatively lower performance in both classification (88.2% accuracy) and retrieval tasks (P@5=82.8%, mAP@5=89.8%). Theseresults highlight the superior performance of DenseNet-201 in both classification and retrieval tasks, while EfficientNet-B0 offers a more efficient solution with slightly lower performance.

F. Comparison Study

TheDenseNet-201 model, which achieved the highest precision among the tested architectures, was compared with recent CBIR methods. As shown in Table IV, DenseNet-201 achieved a Top-10 precision of 92.1% on the Corel-10K dataset.This result exceeds the methods in [24-26] in top-10 precision values. This comparison indicates that DenseNet-201 provides superior feature representation and retrieval precision compared to existing approaches. All methods were testedon the same dataset, ensuring a fair comparison in terms of retrieval performance at top-10 precision.

TABLE IV. TOP-10 PRECISION COMPARISON OF CBIR APPROACHES ON THE COREL-10K DATASET

Reference	Method	Dataset	Precision@10 (%)
[24]	Texture and Color Region Separation	Corel-10K	76.45
[25]	CBIR using multi-Hierarchical clustering	Corel-10K	64.0
[26]	Unsupervised CBIR using PCNN features	Corel-10K	70.2
Proposed method	DenseNet-201	Corel-10K	92.1

IV. CONCLUSION

This study presented a comparative analysis of four pretrained CNN architectures for CBIR using the Corel-10K dataset. All models were fine-tuned using transfer learning, and their retrieval effectiveness was evaluated using P@K and mAP@k at various top-K levels (5,10,15, 20). DenseNet-201 consistently achieved the highest retrieval performance, with a P@5 of 94.0% and mAP@5 of 96.3%, demonstrating its strong ability to retrieve relevant images. VGG-16 and EfficientNet-B0 also showed competitive results, particularly at lower K values, while AlexNet exhibited lower performance across all top-K metrics, indicating its limited representational power compared to deeper architectures. The results highlight the effectiveness of deeper and more modern CNNs, such as DenseNet and EfficientNet, in capturing discriminative features for image retrieval, providing a strong foundation for selecting pretrained CNNs in CBIR systems, especially for retrieval tasks requiring high precision at various top-K levels. DenseNet-201 achieved the highest performance,with superior precision on the Corel-10K dataset compared to existing methods, highlighting its effectiveness for large-scale image retrieval tasks. In addition, using rotation-based data augmentation helps the system handle rotated images better, improving its overall performance in real-world retrieval tasks.

REFERENCES

[1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years,"*IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, Sep. 2000, https://doi.org/10.1109/34.895972.

[2] M. Yasmin, S. Mohsin, and M. Sharif, "Intelligent Image Retrieval Techniques: A Survey,"*Journal of Applied Research and Technology*, vol. 12, no. 1, pp. 87–103, Feb. 2014, https://doi.org/10.1016/S1665-6423(14)71609-8.

[3] E. Yildizer, A. M. Balci, M. Hassan, and R. Alhajj, "Efficient content-based image retrieval using Multiple Support Vector Machines Ensemble,"*Expert Systems with Applications*, vol. 39, no. 3, pp. 2385–2396, Feb. 2012, https://doi.org/10.1016/j.eswa.2011.08.086.

[4] J. Wan *et al.*, "Deep Learning for Content-Based Image Retrieval: A Comprehensive Study," in *Proceedings of the 22nd ACM international conference on Multimedia*, Orlando, FL, USA, Nov. 2014, pp. 157–166, https://doi.org/10.1145/2647868.2654948.

[5] R. Gorle and A. Guttavelli, "Enhanced Image Tampering Detection using Error Level Analysis and CNN,"*Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19683–19689, Feb. 2025, https://doi.org/10.48084/etasr.9593.

[6] M. Shaikh, I. F. Siddiqui, Q. Arain, J. Koo, M. Ali Unar, and N. M. F. Qureshi, "MDEV Model: A Novel Ensemble-Based Transfer Learning Approach for Pneumonia Classification Using CXR Images,"*Computer Systems Science and Engineering*, vol. 46, no. 1, pp. 287–302, 2023, https://doi.org/10.32604/csse.2023.035311.

[7] K. Sanjeevaiah., T. S. Reddy, S. Karthik, M. Kumar, and D. Vivek, "Content-Based Image Retrieval Using Hybrid Densenet121-Bilstm and Harris Hawks Optimization Algorithm,"*International Journal of Software Innovation*, vol. 11, no. 1, pp. 1–15, Dec. 2022, https://doi.org/10.4018/ijsi.315661.

[8] S. R. Dubey, "A Decade Survey of Content Based Image Retrieval Using Deep Learning,"*IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 5, pp. 2687–2704, May 2022, https://doi.org/10.1109/tcsvt.2021.3080920.

[9] S. Chopparapu and J. B. Seventline, "An Efficient Multi-modal Facial Gesture-based Ensemble Classification and Reaction to Sound Framework for Large Video Sequences,"*Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11263–11270, Aug. 2023, https://doi.org/10.48084/etasr.6087.

[10] X. Li, J. Yang, and J. Ma, "Recent developments of content-based image retrieval (CBIR),"*Neurocomputing*, vol. 452, pp. 675–689, Sep. 2021, https://doi.org/10.1016/j.neucom.2020.07.139.

[11] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of the 36th International Conference on Machine Learning*, May 2019, pp. 6105–6114.

[12] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, Jul. 2017, https://doi.org/10.1109/cvpr.2017.243.

[13] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition." arXiv, Apr. 10, 2015, https://doi.org/10.48550/arXiv.1409.1556.

[14] M. S. Sayed, A. A. A. Gad-Elrab, K. A. Fathy, and K. R. Raslan, "A deep learning content-based image retrieval approach using cloud computing,"*Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 3, Mar. 2023, Art. no. 1577, https://doi.org/10.11591/ijeecs.v29.i3.pp1577-1589.

[15] T. L. Durga Likhitha, M. Noushika, V. S. Deepika, and V. M. Manikandan, "A Detailed Review on CBIR and Its Importance in Current Era," in *2021 International Conference on Data Science and Its Applications (ICoDSA)*, Bandung, Indonesia, Oct. 2021, pp. 124–128, https://doi.org/10.1109/icodsa53588.2021.9617481.

[16] S. Maji and S. Bose, "CBIR Using Features Derived by Deep Learning,"*ACM/IMS Transactions on Data Science*, vol. 2, no. 3, pp. 1–24, Aug. 2021, https://doi.org/10.1145/3470568.

[17] G. Gautam and A. Khanna, "Content Based Image Retrieval System Using CNN based Deep Learning Models,"*Procedia Computer Science*, vol. 235, pp. 3131–3141, 2024, https://doi.org/10.1016/j.procs.2024.04.296.

[18] G. V. R. M. Kumar and D. Madhavi, "Stacked Siamese Neural Network (SSiNN) on Neural Codes for Content-Based Image Retrieval,"*IEEE Access*, vol. 11, pp. 77452–77463, 2023, https://doi.org/10.1109/access.2023.3298216.

[19] R. Vishraj, S. Gupta, and S. Singh, "A comprehensive review of content-based image retrieval systems using deep learning and hand-crafted features in medical imaging: Research challenges and future directions,"*Computers and Electrical Engineering*, vol. 104, Dec. 2022, Art. no. 108450, https://doi.org/10.1016/j.compeleceng.2022.108450.

[20] P. Miguel, A. Cansian, G. Rozendo, G. Medalha, M. Zanchetta Do Nascimento, and L. Neves, "An Investigation of Deep-Learned Features for Classifying Radiographic Images of COVID-19," in *Proceedings of the 25th International Conference on Enterprise Information Systems*, Prague, Czech Republic, 2023, pp. 675–682, https://doi.org/10.5220/0012038500003467.

[21] Y. Yang *et al.*, "A comparative analysis of eleven neural networks architectures for small datasets of lung images of COVID-19 patients toward improved clinical decisions,"*Computers in Biology and Medicine*, vol. 139, Dec. 2021, Art. no. 104887, https://doi.org/10.1016/j.compbiomed.2021.104887.

[22] G. A. Pereira and M. Hussain, "A Review of Transformer-Based Models for Computer Vision Tasks: Capturing Global Context and Spatial Relationships." arXiv, Aug. 27, 2024, https://doi.org/10.48550/arXiv.2408.15178.

[23] K. Karthik and S. S. Kamath, "A deep neural network model for content-based medical image retrieval with multi-view classification,"*The Visual Computer*, vol. 37, no. 7, pp. 1837–1850, Jul. 2021, https://doi.org/10.1007/s00371-020-01941-2.

[24] J. Pradhan, S. Kumar, A. K. Pal, and H. Banka, "Texture and colour region separation based image retrieval using probability annular histogram and weighted similarity matching scheme,"*IET Image Processing*, vol. 14, no. 7, pp. 1303–1315, 2020, https://doi.org/10.1049/iet-ipr.2018.6619.

[25] R. Hidayat, A. Harjoko, and A. Musdholifah, "A Robust Image Retrieval Method Using Multi-Hierarchical Agglomerative Clustering and Davis-Bouldin Index,"*International Journal of Intelligent Engineering and Systems*, vol. 15, no. 2, pp. 441–453, Apr. 2022, https://doi.org/10.22266/ijies2022.0430.40.

[26] M. S. Sayed, A. A. A. Gad-Elrab, K. A. Fathy, and K. R. Raslan, "Unsupervised Content Based Image Retrieval Using Pre-Trained CNN and PCNN Features Extractors,"*International Journal of Intelligent Engineering and Systems*, vol. 16, no. 1, pp. 584–596, Feb. 2023, https://doi.org/10.22266/ijies2023.0228.50.

[27] M. Wilson, "Corel-10K." Kaggle, [Online]. Available: https://www.kaggle.com/datasets/michelwilson/corel10k.

[28] D. Chen *et al.*, "An Ensemble Deep Neural Network for Footprint Image Retrieval Based on Transfer Learning,"*Journal of Sensors*, vol. 2021, no. 1, Jan. 2021, https://doi.org/10.1155/2021/6631029.