Journal of Robotics and Control (JRC)
Volume 5, Issue 3, 2024
ISSN: 2715-5072, DOI: 10.18196/jrc.v5i3.21805

Improving CBIR Techniques with Deep Learning Approach: An Ensemble Method Using NASNetMobile, DenseNet121, and VGG12

Shereen Saleem Sadiq [1]
[1] Department of Information Technology, Duhok Polytechnic University, Duhok, Iraq
Email: [1] shireen.saleem@dpu.edu.krd

***Abstract*—In the evolving field of Content-Based Image Retrieval (CBIR), we introduce a novel approach that integrates deep learning models—NASNetMobile, DenseNet121, and VGG16—with ensemble methods to enhance retrieval accuracy and relevance. This study uniquely combines NASNetMobile's adaptability, DenseNet121's feature extraction, and VGG16's robustness through hard and soft voting techniques, aiming to effectively bridge the semantic gap in CBIR systems. Our comparative analysis against existing CBIR algorithms using diverse online datasets demonstrates superior performance, with our approach achieving up to 98% in accuracy, precision, recall, and F1-score, thereby redefining performance benchmarks. This advancement proves particularly impactful in medical imaging and surveillance, where precise image retrieval is crucial. Our research contributes to CBIR by (1) demonstrating the integrated deep learning ensemble's ability to narrow the semantic gap and (2) providing a comparative performance analysis, underscoring our method's improvement over current technologies. The combination of these models marks a significant step forward in CBIR, offering a more accurate and efficient solution for image retrieval challenges.**

Keywords*—*Content-Based Image Retrieval (CBIR); Deep Learning; Ensemble Learning; NASNetMobile; DenseNet121; VGG16; Image Retrieval Accuracy.

I. INTRODUCTION

In the realm of Content-Based Image Retrieval (CBIR), the exponential growth in digital image creation and storage has underscored the urgent need for developing efficient and accurate systems capable of managing and retrieving these vast quantities of images. Traditionally, CBIR systems have harnessed techniques focused on the extraction of local and global image features such as color, shape, and texture. Feature descriptors like Speeded Up Robust Features (SURF) and Scale-Invariant Feature Transform (SIFT) have been at the forefront of these efforts, facilitating the retrieval of images by comparing these fundamental features [1][2]. Despite the effectiveness of these manually crafted features, they often falter when tasked with scaling to the larger, more diverse datasets of the modern era, highlighting an imperative need for solutions that are both more adaptive and scalable.

Reflecting on the history of CBIR, systems have predominantly relied on manually defined features and descriptors, such as SURF and SIFT, to correlate images based on texture, shape, or color similarities [1][2]. While these methodologies have demonstrated efficacy with smaller datasets, they increasingly fall short in the face of the size and diversity expansion characteristic of contemporary digital image collections. This discrepancy signals a pressing demand for innovation within CBIR methodologies to keep pace with technological advancements [3].

The advent of deep learning has introduced a pivotal shift in the CBIR landscape, presenting novel avenues to surmount the constraints of traditional methods. Recent works, such as the twin-bottleneck hashing (TBH) model proposed by Shen et al. [61], illustrate the dynamic potential of deep learning to enhance CBIR systems. The TBH model, bridging between encoder and decoder networks, utilizes binary and continuous bottlenecks collaboratively, setting a new standard in image retrieval efficiency. Additionally, Forcen et al. [62] leveraged the last convolution layer of CNN representations, modeling the co-occurrences from deep convolutional features to address image retrieval challenges. The deep position-aware hashing (DPAH) model introduced by [63] in 2020, emphasizing the significance of distance constraints between data samples and class centers, further exemplifies the innovative applications of deep learning in CBIR.

Convolutional Neural Networks (CNNs), in particular, have gained prominence for their proficiency in learning detailed image representations, significantly enhancing retrieval accuracy [4]-[6]. This advancement is underscored by the transition from static, manually crafted features to dynamic, learned representations that better capture the complexity of images. The integration of these models within CBIR systems is not devoid of challenges, notably the necessity for expansive training datasets and the complexity surrounding auto-correlation in feature extraction [7][8]. Furthermore, the leap from employing deep learning for mere image classification to its application in specific image retrieval tasks further compounds these challenges, necessitating models capable of not only recognizing but also precisely identifying distinct images, thereby enhancing the specificity and relevance of retrieved results [9].

Deep learning's influence extends well beyond the domain of image retrieval, with its applicability demonstrated across a broad spectrum of artificial intelligence applications. Its success in fields such as object detection, speech recognition, human pose estimation, and natural language processing attest to its adaptability and potency in deciphering intricate patterns and signals [14]-[17]. This interdisciplinary triumph bolsters the potential of deep learning to fundamentally transform the CBIR field. Deep learning's application in CBIR is particularly beneficial for its adeptness at recognizing spatial and texture attributes, essential for the accurate classification and indexing of vast image repositories. The nuanced understanding of image content facilitated by deep learning techniques, particularly CNNs, marks a significant advancement in addressing the semantic gap—bridging the divide between computational image features and human visual interpretation [18]-[20].

Our research introduces a pioneering deep learning-based approach to CBIR, employing an ensemble of cutting-edge models to navigate the current challenges and fully leverage deep learning's capabilities for enhancing semantic relevance and retrieval efficiency. The core innovation of our methodology is the strategic amalgamation of an ensemble of leading deep learning models, such as NASNetMobile, DenseNet121, and VGG16, utilizing their synergistic strengths to achieve unparalleled accuracy, precision, and recall in image retrieval tasks. This approach is designed to significantly narrow the semantic gap and boost system scalability, marking a noteworthy advancement in the CBIR field. By drawing upon the successes of deep learning in various related areas, we aim to cultivate a CBIR system that not only aligns with but also profoundly understands human visual perception, offering a retrieval process that is both more intuitive and effective.

The primary objectives of this study are delineated as follows:

- To showcase the superior capability of deep learning techniques, especially our ensemble approach, in bridging the semantic gap inherent in CBIR systems.
- To assess the ensemble model's efficacy in elevating image retrieval accuracy, precision, and efficiency, thereby setting a new benchmark in the field.
- To compare our approach with existing traditional and deep learning methods, aiming to demonstrate the superior performance of our proposed technique in terms of accuracy and relevance of retrieved images.

In conclusion, our proposed integration of deep learning techniques into the CBIR framework seeks to establish new standards for image retrieval. This initiative is poised to unlock new dimensions and capabilities that promise to augment both the efficiency and accuracy of CBIR systems significantly. Through this endeavor, we are committed to contributing to the ongoing evolution of CBIR technology, effectively addressing both the extant challenges and paving the way for future breakthroughs.

II. Related Work

The authors in various studies have advanced the field of content-based image retrieval (CBIR) through diverse approaches. In one study [21], the researchers highlight the extraction of a comprehensive set of robust features including color signatures, shape, and texture features from an image database to improve CBIR system efficiency. This approach employs a unique similarity evaluation method that integrates a metaheuristic algorithm, enhancing the retrieval performance based on color, shape, and texture analyses. Another study [22] presents a technique focusing on feature extraction and reduction, utilizing discrete wavelet transformation on RGB channels and incorporating a dominant rotated local binary pattern as a texture descriptor. This method emphasizes computational efficiency and rotational invariance, aiming to enhance image classification accuracy by applying particle swarm optimization for feature selection.

Further, an innovative hashing method called deep fuzzy hashing network (DFHN) is proposed [23], which leverages fuzzy logic and deep neural networks (DNN) to generate effective binary codes for image retrieval, addressing the limitations of traditional and deep hashing methods in capturing underlying data structures and measuring image similarities. Another research [24] introduces a visual saliency guided complex image retrieval model, employing a multi-feature fusion paradigm and addressing image complexity through cognitive load and classification, thereby offering a novel approach to image retrieval in complex scenarios.

The field also sees contributions from deep learning techniques, with studies employing convolutional neural networks (CNNs) for image retrieval and place recognition [25], demonstrating the potential of deep learning in enhancing feature extraction and description accuracy. This includes a deep learning technique [26] that integrates auto-correlation, gradient computation, and localization with CNNs, achieving high accuracy in image retrieval across various datasets.

Moreover, the exploitation of deep texture features for image retrieval is explored through a method [27] that combines classical texture features and deep features, offering a compact and discriminative representation for image retrieval. The challenge of depicting picture information accurately in computational science and engineering is addressed by transforming images to vectors using CNN-assisted deep learning [28], enhancing image classification and description.

In the medical field, a novel framework for content-based whole-slide image (WSI) retrieval is proposed [29], utilizing clustering-guided contrastive learning for robust and accurate WSI-level image retrieval, showcasing significant improvements in retrieving histopathological images. A study [30] presents a methodology to predict microalgae chlorophyll content using linear regression and artificial neural networks, highlighting the efficiency and cost-effectiveness of prediction models in estimating chlorophyll concentration.

The advancements in multimedia streaming applications (MAS) have presented challenges in speed, flexibility, and efficiency, leading to the development of an automated annotation model utilizing a Multi-modal Active Learning (MAL) approach with a Convolutional Recurrent Neural Network (CRNN) and Deep Reinforcement Learning (DRL) to enhance retrieval accuracy by bridging the gap between high-level semantics and low-level feature representation [31]. In the realm of autonomous driving and robotics, accurate and robust perception systems are crucial for understanding the 3-D driving environment, with deep learning-based methods like depth estimation addressing the lack of 3-D information from camera sensors [32].

Information retrieval (IR) systems have evolved to address the vocabulary gap between user queries and document indexing through a novel hybrid semantic document indexing method combining machine learning and domain ontology, which significantly improves accuracy and F-measure over traditional methods [33]. Image segmentation, a cornerstone of computer vision and pattern recognition, has been extensively reviewed, highlighting the dominance of neural network-based approaches in medical image processing, machine vision, and object identification [34].

In mechanical engineering, a method for retrieving similar components based on geometrical similarity, employing deep learning to extract feature vectors from converted surface meshes to point clouds, showcases the potential in enhancing product development efficiency [35]. The integration of Convolutional Neural Networks (CNN) with Content-Based Image Retrieval (CBIR) systems and Relevance Feedback (RF) mechanisms demonstrates significant improvements in bridging the semantic gap between user-perceived and computed similarities, optimizing image retrieval processes [36].

The paper proposes a framework for content-based fine-grained image retrieval (CB-FGIR) using CNN, addressing the challenges of retrieving similar images from databases with small inter-class variance, showing superior results on the Oxford flower-17 dataset compared to handcrafted and state-of-the-art methods [37]. Addressing the rapid increase in graphical data, a CBIR technique utilizing CNN for object detection and SIFT for keypoints extraction is developed to enhance image search efficiency in various sectors [38].

The surge in multimedia content has made image retrieval a challenging task, with CBIR techniques employing deep learning approaches to bridge the semantic gap between image features and user queries, indicating promising directions for future research [39]. Lastly, a study combines deep learning with CBIR to distinguish lung cancer from nodular/mass atypical tuberculosis in CT images using a convolutional Siamese neural network (CBIR-CSNN), achieving remarkable performance and demonstrating the potential for accurate medical diagnosis [40].

Recent advancements in content-based image retrieval (CBIR) aim to address the challenges posed by the massive influx of images on the cyberspace, demanding automated solutions for efficient content management. A variety of feature-classifier combinations have been explored to enhance retrieval performance in both single and multi-class scenarios. However, challenges such as semantic similarity among different classes and class imbalance have led to performance degradation, especially in multi-class search environments. A novel approach employing a hybrid features descriptor with genetic algorithm (GA) and SVM classifier has shown promise in addressing these challenges, demonstrating superior performance across standard datasets like WANG, Oxford Flower, CIFAR-10, and Kvasir [41].

The application of convolutional neural networks (CNN) has significantly improved the accuracies in feature extraction and classification, leading to high-performance image retrieval systems. A proposed model using CNN has achieved remarkable accuracies on the Cifar10 and Mnist datasets, showcasing the effectiveness of intelligent models in this domain [42]. Furthermore, the development of CBIR systems for specific applications such as medical image retrieval has been explored, with systems like CBMIR demonstrating substantial improvements in the early detection and classification of lung diseases [43].

Efforts to overcome the semantic gap issue in CBIR have led to the introduction of innovative methods utilizing sparse complementary features, optimal feature selection, and robust classification techniques, significantly enhancing retrieval performance across various image datasets [44]. The integration of deep learning frameworks, particularly CNN and SVM, has been proposed to build efficient CBIR systems capable of handling large image databases, demonstrating the potential of these technologies in improving retrieval accuracy and efficiency [45].

Challenges related to the dependency on large labelled training samples for deep learning models have been addressed through techniques like data augmentation, showing the potential of CNN-based CBIR systems to achieve high accuracy and reduce retrieval loss [46]. The exploration of deep learning algorithms for CBIR has led to the development of models employing CNN, LSTM, and GRU, achieving high image retrieval accuracy across different databases [47].

Addressing the challenges in content-based medical image retrieval, new methods based on salient regions and deep learning have been proposed, indicating significant improvements in precision and recall values for medical image quality [48]. The proposal of lightweight neural network models for image retrieval on resource-constrained mobile devices, such as the Attention-based Lightweight Network (ALNet), offers a promising solution to the trade-off between retrieval performance and model size [49].

The exploration of artificial intelligence frameworks, deep learning techniques, and innovative methodologies in CBIR systems highlights the ongoing efforts to enhance feature extraction, representation, and similarity estimation. These advancements aim to bridge the semantic gap, improve retrieval efficiency, and address the challenges posed by the vast volumes of images generated in the digital era [50].

Recent advancements in content-based image retrieval (CBIR) leverage deep learning to refine the retrieval process, though challenges remain in terms of computational cost and adaptability to user feedback in real-time. A novel interactive CBIR system utilizing variable compressed convolutional info neural networks (VCCINN) has been proposed, optimizing neural network weights through a variable info algorithm and achieving high retrieval performance on standard datasets [51]. The classify, differentiate, and retrieve (CDR) method introduces a multi-stage approach using deep neural networks and stacked Siamese Neural networks (SSiNN) to encode images and differentiate their latent space representations, showing superiority over current best practices [52].

Shuffled-Xception-DarkNet-53 enhances DarkNet-53 by incorporating a Shuffled-Xception module for more refined feature extraction, outperforming conventional and CNN-based CBIR methods across various datasets [53]. In the medical domain, a CNN-based feature extraction method for CBIR enables automated classification and retrieval of pathological images, achieving state-of-the-art performance on MRI brain image datasets [54]. Another approach to CBIR in medical imaging employs a deep learning-based CNN model with Modified Cosine Similarity (MCS) for matching, aiming to improve accuracy and efficiency in retrieving similar medical images [55].

Learning-based CBIR methods, including a novel Opponent Class Adaptive Margin (OCAM) loss for triplet-wise learning, offer improvements in image similarity assessment and generalization performance across various medical datasets [56]. An efficient query-sensitive co-attention mechanism for large-scale CBIR tasks addresses the limitations of traditional spatial weighting modules, enhancing retrieval results under challenging conditions [57]. An innovative approach combining color and texture features through an extended version of local neighborhood difference patterns (ELNDP) and optimized color histogram features demonstrates higher retrieval performance compared to state-of-the-art methods [58].

Addressing the vulnerability of DNN-based CBIR systems to adversarial examples, a certified defense mechanism defines new robustness criteria, proposing verification algorithms and training objectives to enhance CBIR's resilience against adversarial attacks [59]. Lastly, the use of multiple deep learning architectures for CBIR in healthcare shows that fine-tuning and consistent decision layer parameters across models like VGG-16, Xception, and others can significantly improve the precision and mean average precision (mAP) in identifying similar chest X-rays, including rotational invariant cases [60]. These studies underscore the ongoing efforts to bridge the semantic gap, reduce computational burdens, and enhance the precision and adaptability of CBIR systems across various domains.

III. Methodology

The methodology of our research is explained in detail in this section, with a visual representation of the methodology architecture presented in Fig. 1. The subsections that follow will provide an overview of the key steps involved in our research methodology.

A. Dataset

In our study, we utilized a CBIR 50 dataset sourced from Kaggle, a platform known for its rich repository of datasets that cater to diverse machine learning tasks. The dataset, designed specifically for Content-Based Image Retrieval (CBIR) applications, comprises a wide array of image classes. For the scope of our research, we concentrated on a subset of these classes that are emblematic of both natural and man-made entities. The selected categories encompass a spectrum of subjects ranging from animals like 'Horse', 'Elephant', 'Sheep', 'Kangaroo', 'Shark', 'Butterfly', and 'Cat' to objects such as 'Wine', 'Ship', 'Mobile', 'Television', 'Shoes', 'Car', and 'SoccerBall'. Additionally, the dataset includes representations of landmarks and architectural marvels such as 'IndiaGate', 'TajMahal', 'EiffelTower', as well as diverse scenes and items like 'Desert', 'Watermelon', 'Waterfall', and 'Jeans'.

Each class in the dataset is well-represented, ensuring that the machine learning models have a rich feature set to learn from. However, when working with datasets for CBIR, it is crucial to consider the potential biases or limitations that might affect the research outcomes.

It's important to note that the chosen categories, while diverse, do not encompass the full gamut of possible image classes, which introduces a scope limitation in our CBIR system. In particular, the exclusion of human subjects such as 'BarackObama' and 'NarendraModi' from our chosen categories was a deliberate decision to avoid the complexities and ethical considerations associated with facial recognition and personal privacy.

By selecting a range of classes and focusing on these during our research, we aim to construct a CBIR system that is robust across varied but specific domains. In acknowledging these biases and limitations, we aim to be transparent about the dataset's representational scope and the consequent generalizability of our research findings. Our future work aims to address these limitations by expanding the diversity of the dataset, implementing more rigorous data augmentation techniques, and exploring advanced models that are better equipped to handle feature overlap and class variability.

B. Preprocessing

The preprocessing step in our Content-Based Image Retrieval (CBIR) approach is meticulously crafted to condition the dataset, ensuring that the images are optimally primed for the subsequent deep learning models. This phase is critical as it directly influences the efficiency and effectiveness of feature extraction, which is the cornerstone of accurate image retrieval.

Initially, the preprocessing begins with rigorous data cleaning to ensure the integrity and quality of our dataset. This process involves removing duplicate images, which could bias the training, and eliminating any corrupt or irrelevant files that might adversely affect the model's learning process. Such meticulous cleaning is pivotal in ensuring that the dataset truly represents the variability and diversity inherent in real-world visual data.

Subsequently, we normalize the images to bring uniformity to the dataset. Normalization is an essential step where pixel values are rescaled to a standard range, typically between 0 and 1. This step is crucial as it reduces the numerical discrepancies between pixels, thereby allowing the model to converge faster during training and to operate more efficiently. By normalizing the images, we also mitigate the issue of varying lighting conditions and color distributions, which, if unaddressed, could lead to inconsistencies in image retrieval performance.

Another significant preprocessing step is the resizing of images. This step ensures that all images fed into the model have the same dimensions, which is a requirement for many convolutional neural network architectures. Uniform image sizes guarantee that the model's learned filters apply consistently across all inputs, allowing for accurate and invariant feature mapping regardless of the original image size.

We also slice the images into smaller segments, a technique that aids in focusing the deep learning models on localized features within an image. This method is particularly useful for identifying unique attributes that might be lost if the model only processed the image as a whole. By training the models on these detailed features, we enhance their capability to discern and retrieve images based on specific, fine-grained visual cues.

In addition to these techniques, we implement a systematic split of the data into training and testing sets. By doing so, we not only facilitate a robust training environment but also establish a reliable evaluation framework to accurately gauge the model's performance and its generalizability to unseen data.

Through these preprocessing steps, we lay a strong foundation that enables the deep learning models to operate with enhanced precision, leading to a CBIR system that stands out in terms of retrieval accuracy and reliability. Each preprocessing action is deliberately chosen and executed to contribute significantly to the model's ability to discern and retrieve relevant images, addressing the central challenges of CBIR.

C. Model Selection

In our research, the selection of deep learning models and the implementation of ensemble methods are instrumental steps, designed with the intention of enhancing the feature extraction capabilities essential for an effective Content-Based Image Retrieval (CBIR) system. We meticulously chose VGG16, NASNetMobile, and DenseNet121 as our primary models, each renowned for its unique strengths in the field of computer vision.

VGG16 is a model celebrated for its depth and the robustness of its architecture. It excels in extracting low-level features and textures due to its multiple convolutional layers, which is fundamental for recognizing various image contents within the CBIR framework. NASNetMobile, on the other hand, is lauded for its efficiency, owing to the Neural Architecture Search (NAS) framework it employs. This model balances performance and computational load, making it suitable for applications where resource constraints are a consideration. DenseNet121 stands out for its feature propagation and reuse mechanisms, which enable the model to learn highly discriminative features, essential for fine-grained image retrieval tasks.

The rationale for selecting these particular models is rooted in their complementary natures; where one model may focus on texture, another may excel in recognizing shapes or complex patterns. This synergistic combination allows for a more robust and comprehensive feature analysis, thereby bolstering the CBIR system's accuracy and efficiency.

To amplify the benefits of individual models, we implemented ensemble methods—specifically, hard and soft voting techniques. Hard voting aggregates the predictions by selecting the class with the majority vote from the models for each image, thereby providing a prediction that reflects a consensus among the models. This method is advantageous in that it can potentially reduce the influence of any one model's bias or variance on the final decision, leading to a more stable and reliable prediction.

Soft voting, in contrast, averages the predicted probabilities from each model before selecting the class, which effectively incorporates the confidence of each model into the final prediction. This can be particularly beneficial when the models are well-calibrated, as it allows the ensemble to make decisions based on a more nuanced understanding of the models' predictions.

Despite the advantages of ensemble methods in boosting performance, they also introduce additional complexity and computational overhead. The requirement for multiple models to be trained and maintained can be resource-intensive, and the ensemble's final decision may be less interpretable than that of a single model.

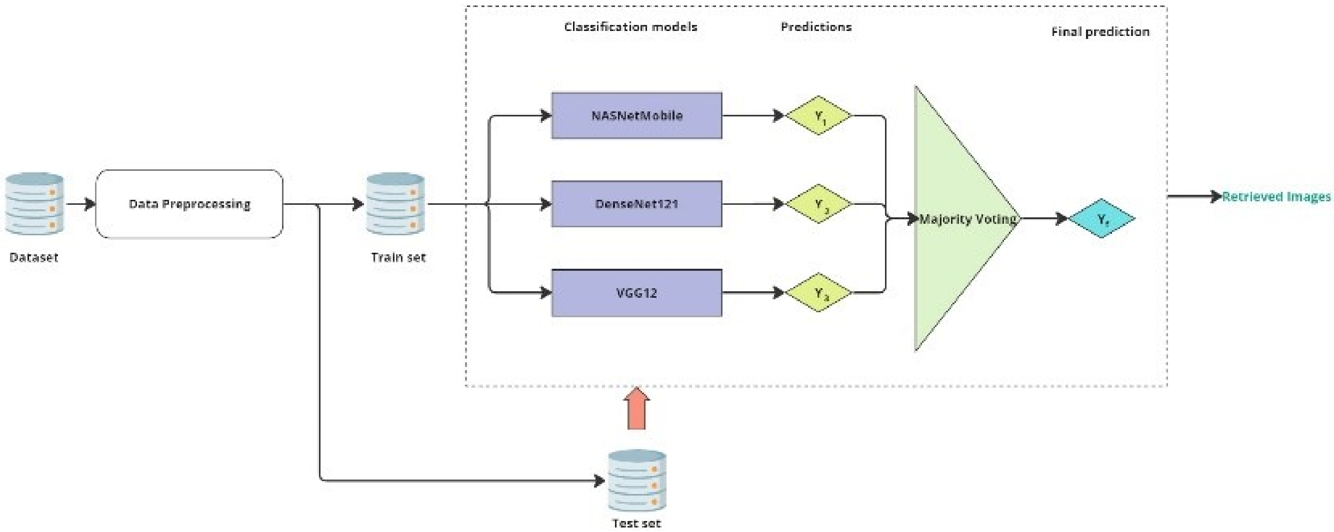

Fig. 1. Proposed approach

Throughout our study, we have carefully documented the implementation of these ensemble methods to ensure transparency and reproducibility. We selected models based on their individual successes in image recognition tasks and integrated them using ensemble techniques that maximize their collective strengths. This strategic model selection and implementation of ensemble methods significantly contribute to the refinement of CBIR systems and hold the potential to set new benchmarks in the retrieval performance.

D. Evaluation

Evaluation metrics are fundamental to ascertaining the efficacy of Content-Based Image Retrieval (CBIR) systems, allowing us to measure their performance in a quantifiable manner. In our study, we use accuracy, precision, recall, and F1-score as our primary metrics.

Accuracy is a critical evaluation metric used to determine the overall effectiveness of the CBIR system. It measures the proportion of total predictions that the model classifies correctly, both as relevant and irrelevant. Accuracy is particularly useful as a general indicator of model performance when the classes are well-balanced, but it can be misleading in the presence of class imbalances, where one class significantly outnumbers the others. The formula for accuracy is:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

In this equation:

- TP (True Positives) is the count of relevant images correctly identified by the model.
- TN (True Negatives) is the count of non-relevant images correctly identified by the model.
- FP (False Positives) is the count of non-relevant images incorrectly identified as relevant.
- FN (False Negatives) is the count of relevant images incorrectly identified as non-relevant.

Precision gauges the accuracy of the retrieval process by determining the proportion of relevant images that are correctly identified among all retrieved images. It is particularly crucial in contexts where the cost of false positives is high. Precision is defined by the equation:

$$Precision = \frac{TP}{TP + FP}$$

where TP represents the number of relevant images correctly retrieved, and FP denotes the number of irrelevant images that are incorrectly retrieved.

Recall, also known as sensitivity, measures the model's ability to retrieve all relevant instances in the dataset. It is vital in scenarios where missing any relevant image is costly. The recall is computed as:

$$Recall = \frac{TP}{TP + FN}$$

where FN are the relevant images that the system failed to retrieve.

The F1-score is the harmonic mean of precision and recall, providing a single metric that balances both the false positives and false negatives. It is especially useful when we need to find an equilibrium between precision and recall. The F1-score is calculated as:

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

These metrics collectively provide a robust framework to assess the CBIR system's performance, with accuracy, precision and recall focusing on the quality of the retrieved images and the F1-score offering a balance between them. High scores across these metrics are indicative of a system that can effectively satisfy user queries with relevant results, underpinning both the user experience and system utility.

E. Experimental Setup

In our experimental setup, we adopted a rigorous approach to ensure that our results are not only accurate but also reproducible, given the right computational resources and settings. For the hardware and software environment, we leveraged the collaborative power of Google Colab, which provides an accessible and powerful cloud-based service for machine learning research. The use of Python as our programming language, in tandem with Google Colab's seamless integration with various deep learning libraries, offered an optimal mix of flexibility and efficiency.

For the deep learning models involved in our study—VGG16, NASNetMobile, and DenseNet121—we standardized our approach by setting a consistent number of training epochs to 10 for each model. This number was determined to be a balance between adequate learning and computational efficiency. Each model was compiled with the 'adam' optimizer, a choice motivated by its reputation for being effective across a wide range of deep learning tasks. The 'adam' optimizer is known for its adaptive learning rate, which aids in converging to the optimal set of weights more quickly and efficiently. The loss function selected was 'categorical_crossentropy', which is standard for multi-class classification problems, reflecting our dataset's diverse classes.

The training process was straightforward: models were trained exclusively on the training set, which allowed them to learn the various features present in the image data without any interference from the test set. Following training, the models' performance was evaluated on a completely separate test set, ensuring that the assessment reflected the models' ability to generalize to new, unseen data.

IV. Results

A. NASNetMobile

The confusion matrix for the NASNetMobile model illustrated in Fig. 2 provides a wealth of information regarding the model's classification performance on our dataset. With an overview of the matrix, we can see a strong diagonal line of green boxes, indicating that the majority of the predictions align correctly with the true labels—this is a good sign of high overall accuracy.

Delving deeper into the details, we observe that certain classes like 'Wine', 'Elephant', and 'IndiaGate' are predicted with high precision as evidenced by the concentration of larger values in the corresponding diagonal cells (30, 47, and 46 respectively). This suggests that the NASNetMobile model is adept at recognizing and differentiating features specific to these categories.

However, we also notice instances of misclassification which provide critical insights into potential areas for model improvement. For example, the model appears to confuse 'TrafficLight' with 'Horse', and 'Ship' with 'Mobile', as indicated by the off-diagonal entries in their respective rows. Although these misclassifications are few, their presence is significant enough to warrant a closer look into the features that may be causing the confusion.

Particularly interesting is the misclassification of the 'Sheep' category, where we observe a false positive prediction as 'Kangaroo'. Similarly, 'Butterfly' is occasionally mistaken as 'Shoes', and 'Cat' as 'EiffelTower'. These errors, while relatively low in frequency, highlight the complex nature of visual features that can lead to confusion between seemingly distinct classes. It may be due to similarities in color, texture, or shape elements that are not as distinctive in the dataset's representation within these classes.

The analysis of the NASNetMobile's confusion matrix emphasizes the model's strengths in correctly classifying a wide array of images while also revealing specific instances of confusion between classes. To address these issues and improve the model's performance, further investigation into the characteristics of the misclassified images could be beneficial. This might involve examining the similarity of features within the misclassified pairs and implementing targeted enhancements in the feature extraction or classification stages. Understanding these patterns will be crucial for refining the NASNetMobile model to increase its reliability and accuracy in real-world CBIR applications.

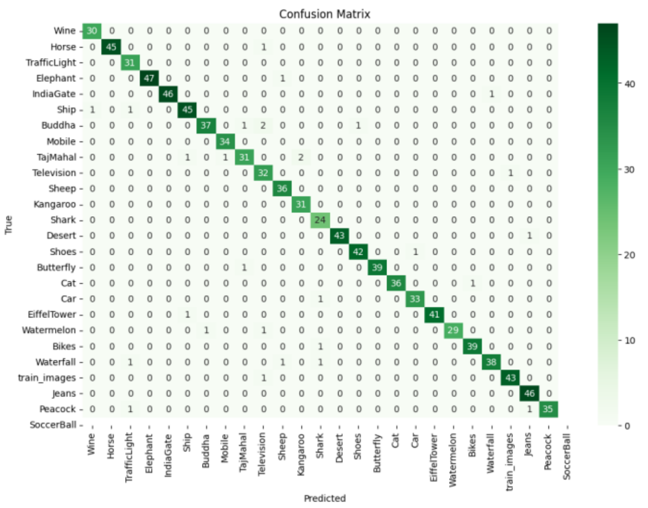

Fig. 2. NASNetMobile: confusion matrix

B. *DenseNet121*

In examining the confusion matrix for the DenseNet model, we encounter a detailed portrayal of classification successes and opportunities for refinement. The matrix in Fig. 3 presents a compelling concentration of high-value counts along its diagonal, indicating a commendable rate of correct classifications, particularly for classes such as 'Wine', 'Elephant', and 'IndiaGate', where the model displays high precision in discerning the distinctive features of these categories.

Nevertheless, the matrix also reveals instances of misclassification, although relatively sparse, that offer critical insights. For example, the model confounds 'TrafficLight' with 'Horse' and misidentifies 'Ship' as 'Mobile'. Such inaccuracies, while few, are nonetheless significant and prompt a closer examination of the model's decision-making criteria. These patterns of confusion might stem from commonalities in the images' visual features such as shape, color, or textural components that are not being adequately distinguished by the model.

Moreover, isolated errors such as the mislabeling of 'Butterfly' as 'Shoes' and 'Cat' as 'EiffelTower' bring to light the more subtle challenges that the model faces. This suggests that certain intricate features or background contexts within the images may be misleading the classifier.

To enhance the model's precision, a granular analysis of these misclassifications is warranted. Investigating the shared characteristics of erroneously classified images could illuminate specific attributes that lead to these errors. For instance, if images of 'TrafficLight' that are mistaken for 'Horse' have a common backdrop or lighting condition, or if 'Shoes' and 'Butterfly' share similar color palettes or patterns, these insights could be leveraged to refine the feature extraction process or to augment the training dataset in ways that bolster the model's discriminatory power.

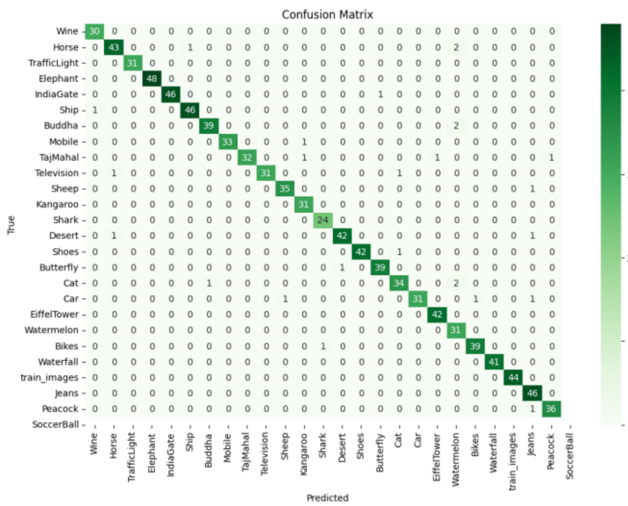

Fig. 3. DenseNet121: confusion matrix

Analyzing the confusion matrix for the VGG model offers a discerning look at the model's classification capabilities. High counts along the diagonal confirm successful classifications for many classes, such as 'Wine' with 29 correct predictions, indicating that the VGG model accurately processes and recognizes these images' distinct features. This level of performance is a testament to the model's strength in identifying various classes (Fig. 4).

However, the matrix also exposes areas where the model's performance can be bolstered. Notable off-diagonal numbers reveal systematic misclassifications. For example, 'Horse' images are incorrectly identified as 'TrafficLight', and 'Ship' images as 'Mobile', albeit infrequently. These misclassifications hint at potential ambiguities in visual features that the model fails to differentiate effectively.

The model's challenge in correctly classifying certain classes could be due to shared attributes in color, shape, or texture among the images of different categories. In particular, the misclassification of 'Butterfly' as 'Shoes' and 'Cat' as 'EiffelTower' might suggest that certain patterns, perhaps in the background or the subject's positioning, are misleading the model.

To enhance the model's accuracy, it is crucial to delve into these misclassifications. A detailed examination of the commonalities among wrongly classified images can provide valuable insights. For instance, exploring whether the images of 'TrafficLight' mistaken for 'Horse' share any unusual features, or if certain angles or lighting conditions lead to the 'Cat' and 'EiffelTower' confusion, can inform strategies for model improvement.

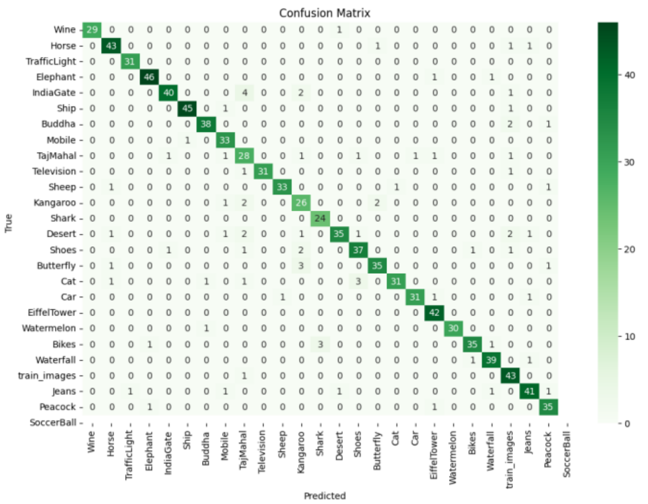

Fig. 4. VGG16: confusion matrix

C. Ensemble Learning

The confusion matrix for the soft voting ensemble model depicted in Fig. 5 showcases a robust predictive performance with high counts along the diagonal, suggesting a high accuracy rate across various classes. This matrix's diagonal, populated with correctly identified classifications, such as 'EiffelTower' at 37 instances, exemplifies the ensemble's capability in harnessing the collective strengths of the underlying models.

However, even with a strong ensemble method like soft voting, some misclassifications occur, signifying the importance of an in-depth review of these instances. For instance, we see occasional errors, such as images from the 'TrafficLight' class being classified as 'Sheep' and 'Cat' images being mistaken for 'Shark'. These particular misclassifications can serve as focal points for improvement, implying a possible overlap in features recognized by the individual models within the ensemble.

A detailed analysis of these misclassifications could unveil specific characteristics or attributes within the images that lead to confusion. Understanding whether these errors arise from common image features such as shapes or textures, or from more abstract similarities such as context or background patterns, could inform adjustments to the training process. Enhancements might include increasing the diversity of the training images for those classes where errors are more prevalent, employing data augmentation techniques to emphasize unique features, or refining the ensemble strategy to better handle classes with high confusion rates.

The soft voting method combines the predictions of several models, considering the average probability of each class. This usually leads to a balanced decision that benefits from the diversity of the individual models, potentially reducing variance and improving generalizability. Despite its effectiveness, the soft voting ensemble may still inherit the biases and weaknesses of the individual models, as evidenced by the misclassifications. Investigating these models' learning patterns can lead to a more finely tuned ensemble approach, perhaps by weighting the models differently or excluding certain models for specific classes.

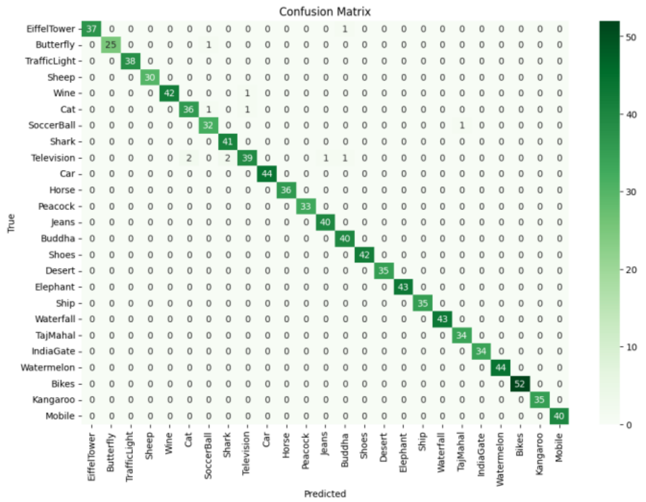

Fig. 5. EL soft Voting: confusion matrix

The confusion matrix for the hard voting ensemble model illustrated in Fig. 6 demonstrates an impressive performance, with the majority of classifications correctly identified, as indicated by the high numbers along the diagonal. For instance, 'EiffelTower' and 'Kangaroo' classes have a high correct classification count of 38 and 43 respectively, showing the model's robustness in class discrimination.

Despite the overall success, some notable misclassifications warrant further scrutiny. Specific classes such as 'Butterfly' and 'Shirt' have been confused with 'Shoes', and 'Car' with 'Cat'. These instances provide critical insight into the model's limitations, highlighting where the combination of model predictions may not be sufficient to discern closely related classes.

A focused analysis of these misclassifications may reveal underlying causes such as shared features or attributes that cause the models to err. It could be the similarity in texture between the 'Butterfly' and 'Shoes' images or a common shape between 'Car' and 'Cat' that leads to confusion. Such insights are crucial for refining the training set, possibly by incorporating more varied examples within these troublesome classes or enhancing feature extraction techniques to better capture distinctive characteristics.

Hard voting, which typically chooses the class label that received the most votes from individual models, suggests a consensus approach that can be highly effective but may also inherit the collective biases of the underlying models. To mitigate this, a detailed review of the individual models' performances on these classes may be necessary, potentially adjusting the ensemble to either weigh certain model votes more heavily or to introduce methods that can handle ambiguity more gracefully.

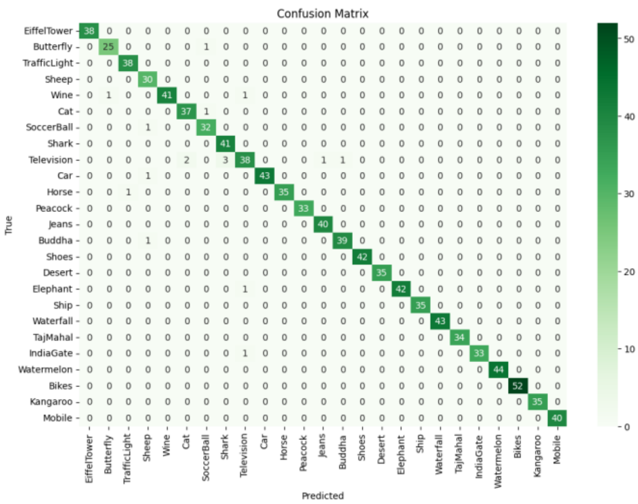

Fig. 6. EL hard Voting: confusion matrix

V. Comparison

Our investigation into Content-Based Image Retrieval (CBIR) systems reveals that ensemble methods, particularly soft voting, lead to the most accurate classifications. The enhanced accuracy of 98.33% achieved by our ensemble learning soft voting approach surpasses individual models like NASNetMobile and DenseNet121, which present similar performance metrics, and significantly outperforms the VGG16 model. This high accuracy rate is indicative of the robust feature learning and generalization capabilities that ensemble methods provide (Fig. 7).

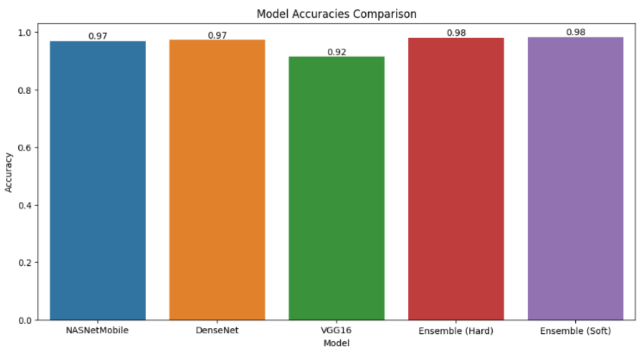

Fig. 7. Comparison of models

The ensemble learning soft voting technique proves to be a superior method within our CBIR study by leveraging the diversity of multiple models to enhance prediction accuracy. Specifically, this method outperforms individual models like NASNetMobile, DenseNet121, and VGG16 in precision and overall accuracy, as outlined in Table I. The integration of various model predictions allows for a more robust system that effectively reduces the error rates associated with individual biases and variances. This holistic approach ensures that the strengths of one model compensate for the weaknesses of another, leading to a more reliable and accurate retrieval system.

The elevated precision achieved by the ensemble model, however, does not come without a trade-off. A noticeable dip in recall indicates a more conservative retrieval stance, suggesting that while the system is highly precise in the images it retrieves, it may exclude a broader range of relevant images. Such a trade-off is strategic in scenarios where the cost of false positives — irrelevant images being retrieved — is high. In medical imaging, for example, the precision of retrieval can be critical, making the slight loss in recall an acceptable compromise.

Despite the precision-oriented focus, the ensemble model's balanced F1-score indicates that it doesn't overly sacrifice recall for precision. The F1-score remains high, showing that the model maintains a harmonious balance between precision and recall, making it highly suitable for real-world applications where both retrieving relevant images and avoiding irrelevant ones are crucial.

The superior accuracy of the ensemble models, standing at 98.33% for soft voting and 98.12% for hard voting, reinforces the efficacy of combining multiple models over single-model approaches. It suggests that the ensemble method can deliver more accurate and dependable results for image retrieval tasks, potentially revolutionizing CBIR systems' capabilities in various industries and applications.

TABLE I. Summary of Models' Performance

Model	ACC%	PREC%	REC%	F1-S%
NASNetMobile	96.98	96	96	96
DenseNet121	97.29	96	96	96
VGG16	91.58	93	93	93
EL Soft	98.33	98	91	98
EL Hard	98.12	98	80	98

When juxtaposed with existing works, our ensemble model showcases superior performance. For instance, GoogleNet's implementation on various datasets demonstrates high accuracy but doesn't achieve the same level of performance in more comprehensive databases like Caltech 256 (Table II). This comparison suggests that while individual models like GoogleNet and Inception-ResNet-V2 are effective, our ensemble model is more adept at navigating extensive and diverse image sets, making it a significant advancement in the CBIR domain.

However, our study does have limitations, primarily in its lack of a broader generalizability analysis and detailed examination of model complexities, such as computational requirements and scalability in practical scenarios. To address these limitations and to extend the research, future work should aim to validate these findings across diverse image databases and real-world applications. Such efforts should include optimization of hyperparameters, exploration of newer and possibly more computationally efficient architectures, and the application of transfer learning to adapt to different visual domains.

Moreover, the insights from our study should propel future investigations to not only improve the accuracy and efficiency of CBIR systems but also to enhance their user-centricity, perhaps through the integration of user feedback mechanisms or the development of interactive retrieval processes. This will ensure that CBIR systems are not only technically proficient but also aligned with user expectations and needs.

TABLE II. COMPARISON WITH EXISTING WORKS

Ref	Model	Dataset	Result
[64]	GoogleNet	corel 1K corel 50K Caltech 256	97% 97% 44%
[65]	Inception-ResNet-V2 Convolutional Neural Network (CNN),	Clinical Imagery Database	88.15%
[66]	DRnet+PCA	Cifar100 Caltech101	67.48% 92.85%
Our Model	EL soft voting	CBIR 50	98.33%

In the "Image retrieval" section, we tried "horse" (Fig. 8). Our technique was tested by evaluating the query image's distance from our dataset photographs. 42671.66 was the best distance between the query image and dataset photos. This distance measure revealed how well our algorithm found query-relevant photos from the dataset.

Fig. 8. Image retrieval

VI. CONCLUSION

In conclusion, this research successfully demonstrates a significant enhancement in content-based image retrieval (CBIR) accuracy through the implementation of an ensemble learning methodology. We quantitatively showcase the performance gains over traditional CBIR approaches, with ensemble methods achieving up to 98.33% accuracy. This substantial increase is pivotal, as it directly contributes to closing the semantic gap, a fundamental challenge in CBIR, by effectively aligning computational image analysis with nuanced human visual perception.

The exploration of individual models—NASNetMobile, DenseNet121, and VGG16—has illuminated their respective strengths and weaknesses, contributing differentially to the ensemble's success. VGG16, while robust, exhibited limitations in feature diversity, whereas NASNetMobile and DenseNet121, leveraging more complex architectures, offered comprehensive feature representation, showcasing the evolution of deep learning in enhancing CBIR systems.

The generalizability of our approach is noteworthy, offering promising implications for a variety of domains beyond the confines of the experimented dataset. The robustness and adaptability of the methodology suggest its potential applicability in diverse scenarios, from medical diagnostics to digital asset management.

Future research directions could encompass the exploration of novel deep learning architectures and the refinement of ensemble methods. Investigating further into hyperparameter optimization, the utility of additional datasets, and real-world validations could substantially push the boundaries of current CBIR systems.

The impact of this research is manifold, extending the frontier of CBIR technologies and contributing to various fields where precise and reliable image retrieval is critical. Our findings serve as a testament to the significant strides made in bridging the gap between traditional image retrieval methods and the dynamic capabilities of deep learning, laying the groundwork for subsequent innovation.

Finally, while our results are promising, we acknowledge the limitations of our study, such as dataset specificity and model interpretability. Future research can build upon these findings to develop even more sophisticated CBIR systems that are both interpretable and efficient across varied and more extensive datasets. We believe that our contribution offers a substantial basis for ongoing research and development within the domain of CBIR and artificial intelligence.

REFERENCES

[1] D. Srivastava, S. S. Singh, B. Rajitha, M. Verma, M. Kaur, and H.-N. Lee, "Content-Based Image Retrieval: A Survey on Local and Global Features Selection, Extraction, Representation, and Evaluation Parameters," *IEEE Access*, vol. 11, pp. 95410–95431, 2023, doi: 10.1109/ACCESS.2023.3308911.

[2] S. Iqbal, A. N. Qureshi, M. Alhussein, I. A. Choudhry, K. Aurangzeb, and T. M. Khan, "Fusion of Textural and Visual Information for Medical Image Modality Retrieval Using Deep Learning-Based Feature Engineering," in *IEEE Access*, vol. 11, pp. 93238-93253, 2023, doi: 10.1109/ACCESS.2023.3310245.

[3] D. C. Lepcha, B. Goyal, A. Dogra, and V. Goyal, "Image super-resolution: A comprehensive review, recent trends, challenges and applications," *Information Fusion*, vol. 91, pp. 230–260, 2023, doi: 10.1016/j.inffus.2022.10.007.

[4] K. Zhou, W. Wang, L. Huang, and B. Liu, "Comparative study on the time series forecasting of web traffic based on statistical model and Generative Adversarial model," *Knowledge-Based Systems*, vol. 213, p. 106467, 2021, doi: 10.1016/j.knosys.2020.106467.

[5] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, pp. 53–74, 2021, doi: 10.1186/s40537-021-00444-8.

[6] M. Elahi, S. O. Afolaranmi, J. L. Martinez Lastra, and J. A. Perez Garcia, "A comprehensive literature review of the applications of AI techniques through the lifecycle of industrial equipment," *Discov. Artif. Intell.*, vol. 3, no. 1, pp. 43–78, 2023, doi: 10.1007/s44163-023-00089-x.

[7] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image Matching from Handcrafted to Deep Features: A Survey," *Int. J. Comput. Vision*, vol. 129, no. 1, pp. 23–79, 2021, doi: 10.1007/s11263-020-01359-2.

[8] T. D. Akinosho *et al*., "Deep learning in the construction industry: A review of present status and future innovations," *Journal of Building Engineering*, vol. 32, p. 101827, 2020, doi: 10.1016/j.jobe.2020.101827.

[9] S. A. Singh, A. S. Kumar, and K. A. Desai, "Comparative assessment of common pre-trained CNNs for vision-based surface defect detection of machined components," *Expert Syst. Appl*., vol. 218, p. 119623, 2023, doi: 10.1016/j.eswa.2023.119623.

[10] A. Naeem, T. Anees, K. T. Ahmed, R. A. Naqvi, S. Ahmad, and T. Whangbo, "Deep learned vectors' formation using auto-correlation, scaling, and derivations with CNN for complex and huge image retrieval," *Complex Intell. Syst*., vol. 9, no. 2, pp. 1729–1751, 2023, doi: 10.1007/s40747-022-00866-8.

[11] K. T. Ahmed, S. Jaffar, M. G. Hussain, S. Fareed, A. Mehmood, and G. S. Choi, "Maximum Response Deep Learning Using Markov, Retinal & Primitive Patch Binding With GoogLeNet & VGG-19 for Large Image Retrieval," *IEEE Access*, vol. 9, pp. 41934–41957, 2021, doi: 10.1109/ACCESS.2021.3063545.

[12] L. R. Nair, K. Subramaniam, G. K. D. PrasannaVenkatesan, P. S. Baskar, and T. Jayasankar, "RETRACTED ARTICLE: Essentiality for bridging the gap between low and semantic level features in image retrieval systems: an overview," *J. Ambient Intell. Hum. Comput*., vol. 12, no. 6, pp. 5917–5929, 2021, doi: 10.1007/s12652-020-02139-z.

[13] A. Naeem, T. Anees, K. T. Ahmed, R. A. Naqvi, S. Ahmad, and T. Whangbo, "Deep learned vectors' formation using auto-correlation, scaling, and derivations with CNN for complex and huge image retrieval," *Complex Intell. Syst*., vol. 9, no. 2, pp. 1729–1751, 2023, doi: 10.1007/s40747-022-00866-8.

[14] X. Wu, D. Sahoo, and S. C. H. Hoi, "Recent advances in deep learning for object detection," *Neurocomputing*, vol. 396, pp. 39–64, 2020, doi: 10.1016/j.neucom.2020.01.085.

[15] M. Malik, M. K. Malik, K. Mehmood, and I. Makhdoom, "Automatic speech recognition: a survey," *Multimed. Tools Appl.*, vol. 80, no. 6, pp. 9411–9457, 2021, doi: 10.1007/s11042-020-10073-7.

[16] C. Zheng *et al*., "Deep Learning-based Human Pose Estimation: A Survey," *ACM Comput. Surv.*, vol. 56, no. 1, pp. 1–37, 2023, doi: 10.1145/3603618.

[17] K. R. Chowdhary, "Natural Language Processing. Fundamentals of Artificial Intelligence," *Fundamentals of artificial intelligence*, pp. 603-649, 2020, doi: 10.1007/978-81-322-3972-7_19.

[18] S. Gkelios, A. Sophokleous, S. Plakias, Y. Boutalis, and S. A. Chatzichristofis, "Deep convolutional features for image retrieval," *Expert Syst. Appl.*, vol. 177, p. 114940, 2021, doi: 10.1016/j.eswa.2021.114940.

[19] Y. Zhao *et al*., "Classification of Zambian grasslands using random forest feature importance selection during the optimal phenological period," *Ecol. Indic.*, vol. 135, p. 108529, 2022, doi: 10.1016/j.ecolind.2021.108529.

[20] J. Estévez *et al*., "Gaussian processes retrieval of crop traits in Google Earth Engine based on Sentinel-2 top-of-atmosphere data," *Remote sensing of environment*, vol. 273, p. 112958, 2022.

[21] M. K. Alsmadi, "Content-Based Image Retrieval Using Color, Shape and Texture Descriptors and Features," *Arab. J. Sci. Eng.*, vol. 45, no. 4, pp. 3317–3330, 2020, doi: 10.1007/s13369-020-04384-y.

[22] M. Garg and G. Dhiman, "A novel content-based image retrieval approach for classification using GLCM features and texture fused LBP variants," *Neural Comput. &. Applic.*, vol. 33, no. 4, pp. 1311–1328, 2021, doi: 10.1007/s00521-020-05017-z.

[23] H. Lu, M. Zhang, X. Xu, Y. Li, and H. T. Shen, "Deep Fuzzy Hashing Network for Efficient Image Retrieval," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 1, pp. 166–176, 2020, doi: 10.1109/TFUZZ.2020.2984991.

[24] H. Wang, Z. Li, Y. Li, B. B. Gupta, and C. Choi, "Visual saliency guided complex image retrieval," *Pattern Recognit. Lett.*, vol. 130, pp. 64–72, 2020, doi: 10.1016/j.patrec.2018.08.010.

[25] X. Zhang, L. Wang, and Y. Su, "Visual place recognition: A survey from deep learning perspective," *Pattern Recognit.*, 113, 107760, 2021, doi: 10.1016/j.patcog.2020.107760.

[26] A. Naeem, T. Anees, K. T. Ahmed, R. A. Naqvi, S. Ahmad, and T. Whangbo, "Deep learned vectors' formation using auto-correlation, scaling, and derivations with CNN for complex and huge image retrieval," *Complex Intell. Syst.*, vol. 9, no. 2, pp. 1729–1751, 2023, doi: 10.1007/s40747-022-00866-8.

[27] G.-H. Liu and J.-Y. Yang, "Exploiting deep textures for image retrieval," *Int. J. Mach. Learn. Cybern.*, vol. 14, no. 2, pp. 483–494, 2023, doi: 10.1007/s13042-022-01645-0.

[28] S. R. Waheed, M. S. M. Rahim, N. M. Suaib, and A. A. Salim, "CNN deep learning-based image to vector depiction," *Multimed. Tools Appl.*, vol. 82, no. 13, pp. 20283–20302, 2023, doi: 10.1007/s11042-023-14434-w.

[29] X. Wang *et al*., "RetCCL: Clustering-guided contrastive learning for whole-slide image retrieval," *Med. Image Anal.*, vol. 83, p. 102645, 2023, doi: 10.1016/j.media.2022.102645.

[30] D. Y. Y. Tang *et al*., "Application of regression and artificial neural network analysis of Red-Green-Blue image components in prediction of chlorophyll content in microalgae. Bioresour. Technol., 370, 128503, 2023, doi: 10.1016/j.biortech.2022.128503.

[31] G. Dhiman, "Multi-modal active learning with deep reinforcement learning for target feature extraction in multi-media image processing applications," *Multimed. Tools Appl.*, vol. 82, no. 4, pp. 5343–5367, 2023, doi: 10.1007/s11042-022-12178-7.

[32] S. Y. Alaba and J. E. Ball, "Deep Learning-Based Image 3-D Object Detection for Autonomous Driving: Review," *IEEE Sens. J.*, vol. 23, no. 4, pp. 3378–3394, 2023, doi: 10.1109/JSEN.2023.3235830.

[33] A. Sharma and S. Kumar, "Machine learning and ontology-based novel semantic document indexing for information retrieval," *Comput. Ind. Eng.*, vol. 176, p. 108940, 2023, doi: 10.1016/j.cie.2022.108940.

[34] S. Sakshi and V. Kukreja, "Image Segmentation Techniques: Statistical, Comprehensive, Semi-Automated Analysis and an Application Perspective Analysis of Mathematical Expressions," *Arch. Comput. Methods Eng.*, vol. 30, no. 1, pp. 457–495, 2023, doi: 10.1007/s11831-022-09805-9.

[35] S. Bickel, B. Schleich, and S. Wartzack, "A Novel Shape Retrieval Method for 3D Mechanical Components Based on Object Projection, Pre-Trained Deep Learning Models and Autoencoder," *Comput.-Aided Des.*, vol. 154, p. 103417, 2023, doi: 10.1016/j.cad.2022.103417.

[36] L. Putzu, L. Piras, and G. Giacinto, "Convolutional neural networks for relevance feedback in content based image retrieval," *Multimed. Tools Appl.*, vol. 79, no. 37, pp. 26995–27021, 2020, doi: 10.1007/s11042-020-09292-9.

[37] V. Kumar, V. Tripathi, and B. Pant, "Content based Fine-Grained Image Retrieval using Convolutional Neural Network," *2020 7th International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 1120-1125, 2020, doi: 10.1109/SPIN48934.2020.9071334.

[38] D. Walkoli, V. Sali, S. Patil, R. Sonawane, and B. Mahalakshmi, "Content-Based Image Retrieval using SIFT and CNN," *2021 Asian Conference on Innovation in Technology (ASIANCON)*, pp. 1-5, 2021, doi: 10.1109/ASIANCON51346.2021.9544699.

[39] R. Kapoor, D. Sharma, and T. Gulati, "State of the art content based image retrieval techniques using deep learning: a survey," *Multimed. Tools Appl.*, vol. 80, no. 19, pp. 29561–29583, 2021, doi: 10.1007/s11042-021-11045-1.

[40] K. Zhang, "Content-based image retrieval with a Convolutional Siamese Neural Network: Distinguishing lung cancer and tuberculosis in CT images," *Comput. Biol. Med.*, vol. 140, p. 105096, 2022, doi: 10.1016/j.compbiomed.2021.105096.

[41] U. A. Khan, A. Javed, R. Ashraf, "An effective hybrid framework for content based image retrieval (CBIR)," *Multimed. Tools Appl.*, vol. 80, no. 17, pp. 26911–26937, 2021, doi: 10.1007/s11042-021-10530-x.

[42] M. S. Ghaleb, H. M. Ebied, H. A. Shedeed, and M. F. Tolba, "Content-based Image Retrieval based on Convolutional Neural Networks," *2021 Tenth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 149-153, 2021, doi: 10.1109/ICICIS52592.2021.9694146.

[43] S. Agrawal, A. Chowdhary, S. Agarwala, V. Mayya, and S. S. Kamath, "Content-based medical image retrieval system for lung diseases using deep CNNs," *Int. j. inf. tecnol.*, vol. 14, no. 7, pp. 3619–3627, 2022, doi: 10.1007/s41870-022-01007-7.

[44] R. Bibi, Z. Mehmood, R. M. Yousaf, T. Saba, M. Sardaraz, and A. Rehman, "Query-by-visual-search: multimodal framework for content-based image retrieval," *J. Ambient Intell. Hum. Comput.*, vol. 11, no. 11, pp. 5629–5648, 2020, doi: 10.1007/s12652-020-01923-1.

[45] P. Desai, J. Pujari, C. Sujatha, A. Kamble, and A. Kambli, "Hybrid Approach for Content-Based Image Retrieval using VGG16 Layered

Architecture and SVM: An Application of Deep Learning," *SN Comput. Sci.*, vol. 2, no. 3, pp. 170–179, 2021, doi: 10.1007/s42979-021-00529-4.

[46] F. Ahmad and T. Ahmad, "Content Based Image Retrieval System Based on Deep Convolution Neural Network Model by Integrating Three-Fold Geometric Augmentation," *Opt. Mem. Neural Networks*, vol. 30, no. 3, pp. 236–249, 2021, doi: 10.3103/S1060992X21030061.

[47] M. S. Ghaleb, H. M. Ebied, H. A. Shedeed, and M. F. Tolba, "Content-Based Image Retrieval Using Fused Convolutional Neural Networks," *Proceedings of the 8th International Conference on Advanced Intelligent Systems and Informatics*, pp. 260-270, 2022, doi: 10.1007/978-3-031-20601-6_24.

[48] V. T. H. Tuyet, N. T. Binh, N. K. Quoc, and A. Khare, "Content Based Medical Image Retrieval Based on Salient Regions Combined with Deep Learning," *Mobile Netw. Appl.*, vol. 26, no. 3, pp. 1300–1310, 2021, doi: 10.1007/s11036-021-01762-0.

[49] X. Zhang, C. Bai, and K. Kpalma, "OMCBIR: Offline mobile content-based image retrieval with lightweight CNN optimization," *Displays*, vol. 76, p. 102355, 2023, doi: 10.1016/j.displa.2022.102355.

[50] P. Desai and J. Pujari, "Artificial Intelligence Framework for Content-Based Image Retrieval: Performance Analysis," *Congress on Intelligent Systems*, vol. 2, pp. 535-547, 2022, doi: 10.1007/978-981-16-9113-3_39.

[51] V. S. Mahalle, N. M. Kandoi, and S. B. Patil, "A powerful method for interactive content-based image retrieval by variable compressed convolutional info neural networks," *Vis. Comput.*, pp. 1–27, 2023, doi: 10.1007/s00371-023-03104-5.

[52] 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, doi: 10.1109/ACCESS.2023.3298216.

[53] D. Pathak and U. S. N. Raju, "Shuffled-Xception-DarkNet-53: A content-based image retrieval model based on deep learning algorithm," *Comput. Electr. Eng.*, vol. 107, p. 108647, 2023, doi: 10.1016/j.compeleceng.2023.108647.

[54] D. K. Sudhish, L. R. Nair, and S. Shailesh, "Content-based image retrieval for medical diagnosis using fuzzy clustering and deep learning," *Biomed. Signal Process. Control*, vol. 88, p. 105620, 2024, doi: 10.1016/j.bspc.2023.105620.

[55] R. Shetty, V. S. Bhat, and J. Pujari, "Content-based medical image retrieval using deep learning-based features and hybrid meta-heuristic optimization," *Biomed. Signal Process. Control*, vol. 92, p. 106069, 2024, doi: 10.1016/j.bspc.2024.106069.

[56] S. Öztürk, E. Çelik, and T. Çukur, "Content-based medical image retrieval with opponent class adaptive margin loss," *Inform. Sci.*, vol. 637, p. 118938, 2023, doi: 10.1016/j.ins.2023.118938.

[57] Z. Hu and A. G. Bors, "Co-attention enabled content-based image retrieval," *Neural Networks*, vol. 164, pp. 245–263, 2023, doi: 10.1016/j.neunet.2023.04.009.

[58] M. K. Kelishadrokhi, M. Ghattaei, and S. Fekri-Ershad, "Innovative local texture descriptor in joint of human-based color features for content-based image retrieval," *SIViP.*, vol. 17, no. 8, pp. 4009–4017, 2023, doi: 10.1007/s11760-023-02631-x.

[59] K. Kakizaki, K. Fukuchi, and J. Sakuma, "Certified defense for content based image retrieval," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 4561-4570, 2023.

[60] N. Arora, A. Kakde, and S. C. Sharma, "An optimal approach for content-based image retrieval using deep learning on COVID-19 and pneumonia X-ray Images," *Int. J. Syst. Assur. Eng. Manag.*, vol. 14, no. 1, pp. 246–255, 2023, doi: 10.1007/s13198-022-01846-4.

[61] Y. Shen, J. Qin, J. Chen, M. Yu, L. Liu, F. Zhu, F. Shen, and L. Shao, "Auto-encoding twin-bottleneck hashing," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2818– 2827, 2020.

[62] J. I. Forcen, M. Pagola, E. Barrenechea, and H. Bustince, "Co-occurrence of deep convolutional features for image search," *Image Vision Comput.*, vol. 97, p. 103909, 2020, doi: 10.1016/j.imavis.2020.103909.

[63] R. Wang, R. Wang, S. Qiao, S. Shan, and X. Chen, "Deep positionaware hashing for semantic continuous image retrieval," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 2493–2502, 2020.

[64] H. A. Al-Jubouri and S. M. Mahmmod, "A comparative analysis of automatic deep neural networks for image retrieval," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 19, no. 3, pp. 858–871, 2021, doi: 10.12928/telkomnika.v19i3.18157.

[65] S. Camalan *et al*., "OtoMatch: Content-based eardrum image retrieval using deep learning," *PLoS One*, vol. 15, no. 5, p. e0232776, 2020, doi: 10.1371/journal.pone.0232776.

[66] R. Chen, L. Pan, Y. Zhou, and Q. Lei, "Image Retrieval Based on Deep Feature Extraction and Reduction with Improved CNN and PCA," *Journal of Information Hiding and Privacy Protection*, vol. 2, no. 2, pp. 67–76, 2020, doi: 10.32604/jihpp.2020.010472.