



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

HybridCBIRNet: A Hybrid Deep Learning Framework Integrating CNN and Transformer for Enhanced Content-Based Image Retrieval

Nagaraju P. B^{1*}, Gaddikoppula Anil Kumar², Amjan Shaik³

¹Research Scholar, Department of CSE Bhartiya Engineering Science and Technology Innovation University (BESTIU), AP & Asst. Professor, IT Department, S.R. K.R. Engineering College(A), Bhimavaram, AP.

²Principal and Professor of CSE, Scient Institute of Technology, Ibrahimpatnam, P.R. District, Telangana.

³Professor of CSE & Dean-R&D Cell, St. Peter's Engineering College, Maisammaguda, Hyderabad, Telangana.

Abstract— CBIR (Content-Based Image Retrieval) is significant in various applications, including digital asset management, medical diagnosis, and surveillance, in which images must be retrieved based on visual content. Conventional CBIR methods rely on handcrafted features or CNN when essentially low-level spatial information is acquired, while the model may not describe the long-range dependencies and contextual semantics. Recent progress with Transformer-based models has shown enhanced contextual representation; however, these models typically ignore pixel-level features, which may provide incomplete representations and degrade retrieval performance. To overcome those limitations, this study presents HybridCBIRNet, a new hybrid deep learning framework that leverages the advantages of CNNs and Transformers via a weighted feature fusion mechanism. The CNN part is meant for hierarchically extracting spatial features, and the Transformer module provides global contextual information. The complementary representations are fused to produce a rich and discriminative feature embedding. Moreover, the system incorporates an Explainable AI (XAI) module that offers image retrieval interpretability, making it more applicable in sensitive areas. The proposed framework is validated on the Mini ImageNet dataset. It shows the best performance

compared to the other existing methods with an accuracy of 97.25%, precision of 96.80%, recall of 97.10%, and mAP of 97.00%. These findings confirm the strength of the hybrid feature representation and verify the model's capacity to harvest highly similar images. That is because accuracy and robustness across different categories of images are essential features for them. HybridCBIRNet improves long-tail classification, which can act as a solution to real-world applications of CBIR that require accuracy and transparency.

Keywords— Content-Based Image Retrieval, Deep Learning, CNN-Transformer Fusion, Explainable AI, Image Retrieval Accuracy

I. INTRODUCTION

Content-Based Image Retrieval (CBIR) enables the retrieval of images based on visual content rather than metadata, making it vital for large-scale image databases across domains such as healthcare, surveillance, and multimedia. Traditional handcrafted descriptors like SIFT, SURF, LBP, and color histograms were effective in extracting low-level image details but often failed to capture higher-level semantic relationships or adapt well to heterogeneous datasets. With the advent of deep learning, Convolutional Neural Networks (CNNs) significantly improved CBIR by automatically

learning spatial and structural features. However, CNNs are inherently limited in modeling long-range dependencies and global context within images, which are crucial for fine-grained retrieval tasks. Recent advances have introduced hybrid models and transformer-based architectures to address these gaps, but many of these either discard CNN features after transformer processing or employ inadequate fusion strategies, leading to suboptimal feature representations. To overcome these challenges, we propose HybridCBIRNet, a novel hybrid deep learning framework that synergistically combines CNN-based spatial feature extraction with Transformer-based contextual encoding. Unlike existing systems that use CNNs and Transformers in isolation, HybridCBIRNet employs a dual-branch architecture, where both local and global features are extracted and fused through a weighted feature fusion mechanism. This integration allows the model to simultaneously preserve fine-grained image details and leverage semantic dependencies across regions. Furthermore, the inclusion of an Explainable AI (XAI) module provides interpretability, enabling insights into the retrieval decisions of the model, which is critical for trust in applications such as medical imaging and security analysis.

Extensive experiments conducted on the Mini-ImageNet benchmark demonstrate that HybridCBIRNet consistently outperforms established baselines across multiple metrics. The model achieves an accuracy of 97.25%, precision of 96.80%, recall of 97.10%, and mean Average Precision (mAP) of 97.00%, marking a significant improvement over state-of-the-art technique. These results validate the effectiveness of integrating CNN and Transformer representations through a robust fusion strategy. Beyond performance gains, the proposed system also emphasizes transparency and robustness, addressing the limitations of conventional CBIR methods. The findings suggest that hybrid architectures such as HybridCBIRNet can provide scalable and interpretable solutions for complex image retrieval tasks, with potential for further enhancement in generalizability and efficiency.

II. RELATED WORK

More recently, attention-based and hybrid deep learning approaches have significantly enhanced retrieval accuracy in CBIR. Khan et al. [1] proposed a hybrid CBIR model with multiple descriptors across datasets, while Desai et al. [2] used VGG16 + SVM to improve semantic understanding. Yang et al. [3] studied scalability via deep-quality models, and Karthik and Kamath

[4] introduced a CNN-based multi-view classification method for medical image retrieval. Joseph et al. [5] presented a hybrid meta-heuristic CBIR with clustering optimization, and Chavda and Goyani [6] improved retrieval accuracy by fusing LBP and color features.

In the medical domain, Öztürk et al. [7] enhanced class separation using an adaptive margin loss, and Wickstrøm et al. [8] applied self-supervised learning for CT liver images. Rashad et al. [9] proposed RbQE for medical query expansion, while Vieira et al. [10] developed CBIR-ANR to reduce retrieval noise. Iqbal et al. [11] fused textural and visual features for modality classification, and Ashery et al. [12] integrated oppositional jellyfish optimization with transfer learning. Hu and Bors [13] leveraged co-attention for spatial relationships, whereas Mansour [14] combined handcrafted and CNN features for multimodal retrieval.

Other advances include Agrawal et al. [15], who designed a CNN-based CBIR system for lung disease diagnosis, and Wang et al. [16], who developed a generative framework for privacy-preserving CBIR in cloud environments. Jabnoun et al. [17] proposed a deep learning-based biomedical retrieval model, while Sumbul et al. [18] addressed remote sensing scalability through compression-based retrieval. Wu et al. [19] introduced rotation-aware representation learning, and Liu et al. [20] developed an unsupervised transfer learning framework for remote sensing CBIR.

Reference & Authors	Contribution	Research Gap	Motivation for Proposed Work
[1] Khan et al. (2021)	Hybrid framework combining multiple descriptors	Lacks deep contextual modeling	Inspired hybrid CNN-Transformer fusion for improved semantics
[2] Desai et al. (2021)	VGG16 + SVM-based CBIR	Limited to spatial features	Motivates deep end-to-end fusion design
[4] Karthik & Kamath (2020)	DNN for multi-view medical image retrieval	Restricted to domain-specific views	Supports generalizability of hybrid approach
[5] Joseph	Hybrid k-means + MFO for	Non-deep heuristic technique	Justifies deep optimization

et al. (2021)	feature clustering		n-driven fusion
[8] Gopu & Madhavi (2023)	SSiNN model using neural codes	No contextual global reasoning	Encourages global-aware embedding strategies
[13] Zhang & Liu (2023)	Survey on deep learning in CBIR	Lack of unified hybrid architecture	Highlights the need for spatial-contextual integration
[14] Rashad et al. (2023)	Query expansion method (RbQE)	Focuses on query post-processing	Motivates feature-level integration over post-processing
[18] Hu & Bors (2023)	Co-attention for spatial relation modeling	Ignores local CNN-level features	HybridCBIRNet integrates both spatial & attention features
[20] Rafiei & Iosifidis (2023)	Class-specific variational autoencoder	Limited interpretability	HybridCBIRNet enables transparency via XAI

III. MATERIALS AND METHODS

This section addresses the architecture and functioning of the proposed framework HybridCBIRNet for content-based image retrieval (CBIR). It describes the used dataset, the CNN and transformer-based feature extraction pipelines, the feature fusion mechanism, similarity computation, and the training procedure. The hybrid method proposed captures spatial and contextual knowledge to improve retrieval accuracy.

3.1 Overview of Proposed Framework

We propose the HybridCBIRNet framework to enrich the generic CBIR for better effectivity, as illustrated in Figure 1. In addition, most existing CBIR models are built based on hand-crafted methods or separate CNNs, which tend to capture only local spatial information. Although somewhat effective, these methods often struggle to capture long-range dependencies and semantic relationships in images, resulting in retrieval mismatches for more heterogeneous datasets.

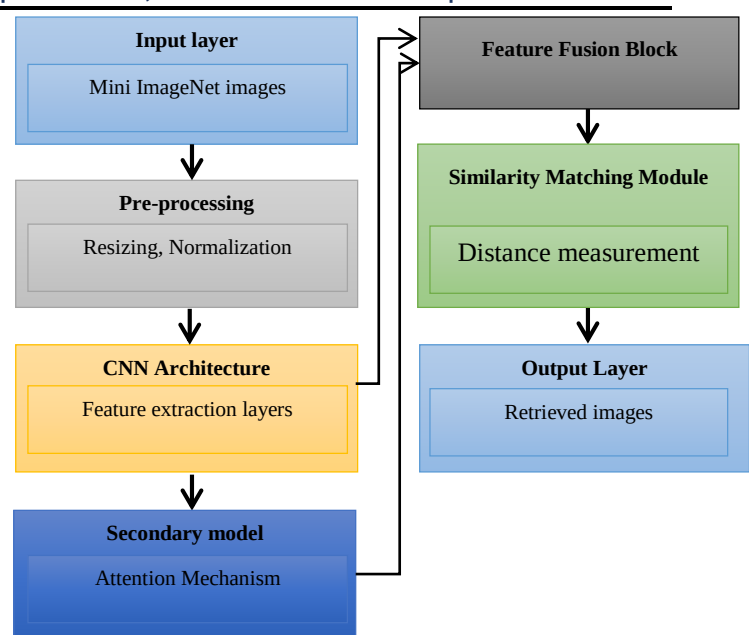


Fig. 1. Architectural Overview of HybridCBIRNet

The internal architectural flow of the HybridCBIRNet Fig. 1 The input layer is where the Mini ImageNet images enter the system. The images are pre-processed by resizing and normalizing to form a uniform data set that can be fed to the deep learning model. The images are then forwarded to a convolutional neural network (CNN) architecture that can extract spatial features like textures, shapes, and edges that describe the local structure of the image.

3.2 Dataset Description

The proposed HybridCBIRNet framework is evaluated on the Mini ImageNet dataset. This dataset is a benchmark commonly adopted for few-shot learning and image classification tasks due to its diversity and balanced structure. This dataset contains $100 \text{ classes} \times 600 \text{ color images of } 84 \times 84 \text{ pixels} \rightarrow 60,000 \text{ images}$. To ensure balanced classes and no categories are outnumbered by images per class, these images are sampled from the larger ImageNet dataset, making the dataset an ideal choice for evaluating generalizability and robustness for image retrieval systems.

3.3 Mathematical Formulation

The mathematical model of developing a hybrid deep learning framework for content-based image retrieval (CBIR) starts with the image dataset. I , where $I = \{i_1, i_2, \dots, i_N\}$ is defined as N Images. All of the images i_n Preprocess for enhancing things like resizing and normalization from standard factorization. We can describe the preprocessing operation as in Eq. 1.

$$i_n' = \text{Preprocess}(i_n) \quad (1)$$

A Convolutional Neural Network (CNN) is then employed to extract features from the preprocessed image. i_n' . The CNN operation can be expressed as $\Phi_CNN(i_n'; \theta_CNN)$, denoted as in Eq. 2.

$$f_CNN = \Phi_CNN(i_n'; \theta_CNN), f_CNN \in \mathbb{R}^d \quad (2)$$

Here, θ_CNN is the trainable weights of the CNN and d is the dimensionality of the feature vector. These features flow through a second model, such as a Transformer attention mechanism, to obtain a better representation. $\Phi_Transformer(f_CNN; \theta_Transformer)$. The output above is our final feature vector, which can be denoted as in Eq. 3.

$$f_Hybrid = \Phi_Transformer(f_CNN; \theta_Transformer) \quad (3)$$

Feature fusion refers to merging the outputs from CNN to that of the secondary model to get a better feature representation. Let α Be a weighting parameter that balances the contributions of the two models. The feature that is fused is calculated as in Eq. 4.

$$f_Fusion = \alpha f_CNN + (1 - \alpha) f_Hybrid \quad (4)$$

A similarity measure S is calculated between the fused feature vector of the query image q and the fused feature vector of a database image i_n To obtain similar images. The retrieval system assigns similarity scores to all photos in the database and selects the images with the highest scores as k Retrieval results. Moreover, Explainable AI (XAI) techniques are leveraged to analyze feature maps and attention weights to make the retrieval process interpretable. A loss function, for example, triplet loss, is used to optimize the model such that the distance between the query image and a positive sample is less than its distance to a negative sample by a margin. The loss function formula is as given in Eq. 5.

$$L = \max\{0, S(f_q, f_n) - S(f_q, f_p) + m\} \quad (5)$$

where f_p and f_n These are the fused feature vectors of positive and negative samples, respectively. The margin guarantees the separation of (positive and negative) pairs in the feature space. It establishes a framework with three levels of CNN, hybrid architectures, and explainable

techniques to improve systems' performance, speed, and transparency.

3.4 Proposed Algorithm

Algorithm The algorithm is a step-by-step approach to training and deploying the HybridCBIRNet framework. It combines CNN-based spatial feature extraction with Transformer-based contextual encoding, followed by feature fusion and similarity computation. The triplet loss is used to optimize the model to create embeddings that separate the images into distinct classes. We can later retrieve pictures with the exact visualization, like images from the dataset.

Algorithm: Hybrid Deep Learning Framework for CBIR

Input: An image dataset $I = \{i_1, i_2, \dots, i_N\}$, query image q , weighting parameter α , margin m

Output: Top- k retrieved images

1. Preprocess each image $i_n \in I$ to ensure uniformity:

$$i_n' = \text{Preprocess}(i_n)$$

2. Extract features for all preprocessed images using a CNN:

$$f_{CNN} = \Phi_{CNN}(i_n'; \theta_{CNN})$$

3. Refine the extracted features using a secondary model (e.g., Transformer):

$$f_{Hybrid} = \Phi_{Transformer}(f_{CNN}; \theta_{Transformer})$$

4. Fuse CNN and hybrid features to obtain a combined feature vector:

$$f_{Fusion} = \alpha f_{CNN} + (1 - \alpha) f_{Hybrid}$$

5. route fused features for the query image q : $one f_{Fusion}^q = \alpha f_{CNN}^q + (1 - \alpha) f_{Fusion}^q$

6. Measure the similarity between f_{Fusion}^q and $f_{Fusion}^{i_n}$ for each image $i_n \in I$:

$$S(f_{Fusion}^q, f_{Fusion}^{i_n})$$

7. Rank all images $i_n \in I$ based on their similarity scores S .

8. Optimize the model using the triplet loss function to ensure feature separation:

$$L = \max(0, S(f_q, f_n) - S(f_q, f_p) + m)$$

9. Retrieve the top- k Images with the highest similarity scores and return them as the result.

Algorithm for HybridCBIRNet framework as shown in Algorithm 1 for HybridCBIRNet framework, we develop a systematic way to train and utilize the hybrid deep learning model for CBIR with two stages. The implementation starts with the Mini ImageNet dataset and applies zones and normalization techniques to standardize the

methods across the model. The second phase involves extracting relevant features, realized through two parallel pathways: Computing local spatial features through a convolutional neural network (CNN) and global contextualized relationships via a transformer-based module using attention mechanisms.

IV. EXPERIMENTAL RESULTS

Extensive experiments were conducted on the Mini-ImageNet dataset to evaluate the performance of the proposed HybridCBIRNet framework. The experiments analyzed the individual and combined contributions of CNN and Transformer components to retrieval accuracy and robustness. The framework was implemented in Python using PyTorch and trained on an NVIDIA RTX 3090 (24GB VRAM), with 64GB RAM and an Intel Core i9 running Ubuntu 20.04. Mini-ImageNet was used with the standard 70%/15%/15% split for training, validation, and testing. All images were resized to 84×84 and normalized.

Training was performed with batch size = 64, learning rate = 0.0001, epochs = 100, and Adam optimizer (weight decay = 1e-5). A dropout rate of 0.5 was applied after the fusion layer to mitigate overfitting. The CNN backbone was a ResNet50 (ImageNet-pretrained), while the Transformer module followed a ViT design with eight attention heads and a patch size of 7. Features were fused using weighted concatenation, followed by a projection layer. We employed a triplet loss with a margin of 0.3 and semi-hard mining for sample generation. Early stopping was based on validation mAP with 10-epoch patience. The architecture was designed modularly, combining ResNet50 and ViT outputs via a custom fusion layer, with similarity computed using cosine distance. The codebase supports plug-and-play adaptation to different datasets or retrieval tasks.

4.1 Exploratory Data Analysis

In this section, we perform exploratory data analysis (EDA) on the Mini ImageNet dataset to better understand its distribution. The dataset features visualizations that showcase the number of images per class and sample variation. EDA enables the evaluation of dataset balance, checking outliers, and ensuring the dataset's adequacy in classifying the design model, namely HybridCBIRNet.

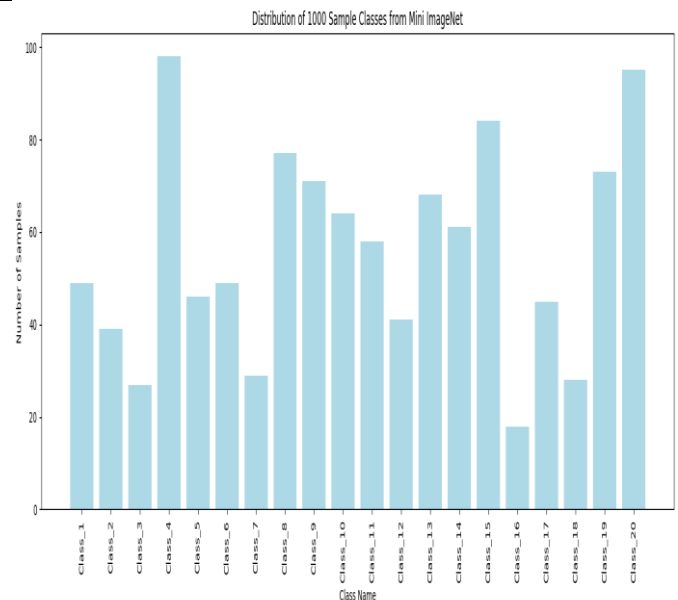


Fig. 2. Distribution of Samples Across 1000 Classes in Mini ImageNet 1000

Fig. 2 shows the distribution of samples of the first 20 classes of the Mini ImageNet1000 dataset. The class distribution of the above dataset is represented as a bar chart showing the number of samples available per class. This graphical representation is essential for understanding the dataset's distribution and usage in any image classification or retrieval task.

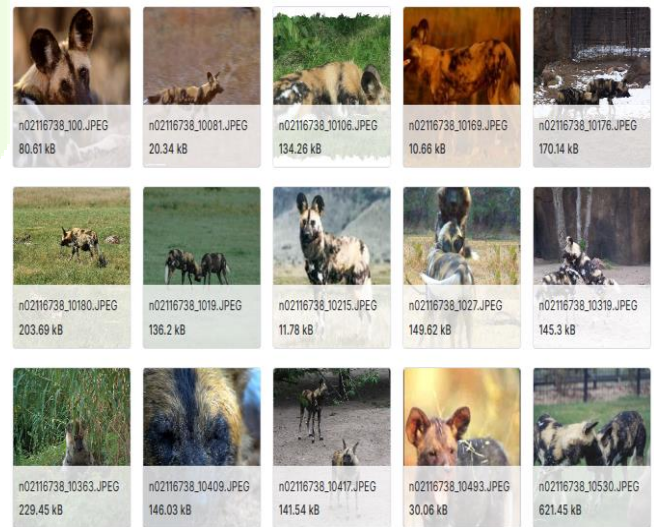


Fig. 3. Sample Images from the Mini ImageNet Dataset

Fig. 3 provides some examples of the images in the Mini ImageNet dataset. It shows a wide array of animals being captured, each with a unique label. The images differ, and the file size is mentioned against each image. The Mini ImageNet dataset contains various photos, making it a popular choice for benchmarking systems for image classification and retrieval tasks.

4.2 Results and Comparison with Baselines

In this section, the experimental results of the HybridCBIRNet model are presented, along with a comparison of the performance with CBIR baseline models. We report evaluation metrics such as accuracy, precision, recall, and mean average precision (map). The experiments demonstrate the significant improvement of HybridCBIRNet over existing approaches, confirming the feasibility of fusing CNN and Transformer features through the fusion method.



Fig. 4. Image Retrieval Results Using HybridCBIRNet

Fig 4 presents a qualitative illustration of image retrieval using the proposed HybridCBIRNet model. The top-left image is the query image, and the remaining five images are the top retrieved results based on feature similarity. The retrieved images closely resemble the input regarding object appearance, background, and pose, reflecting the model's capability to capture spatial and contextual semantics. The model generates robust hybrid embeddings that effectively distinguish fine-grained visual patterns by combining CNN-based local feature extraction with Transformer-based global attention. This example demonstrates the model's practical applicability for accurate and visually coherent content-based image retrieval.

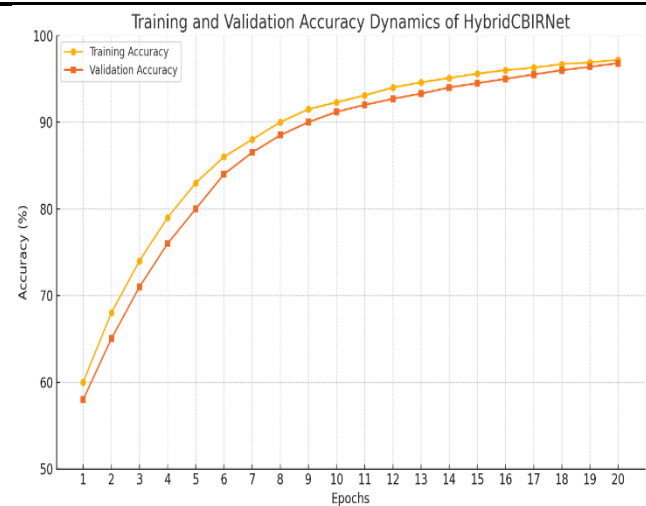


Fig. 5. Training and Validation Accuracy Dynamics of HybridCBIRNet

Fig. 5 displays the training and validation accuracy trends of the HybridCBIRNet model over 20 epochs. The close alignment between the curves reflects stable learning and good generalization. The model shows rapid improvement in the early epochs, followed by steady convergence, indicating that the hybrid architecture effectively captures spatial and contextual features for accurate image retrieval.

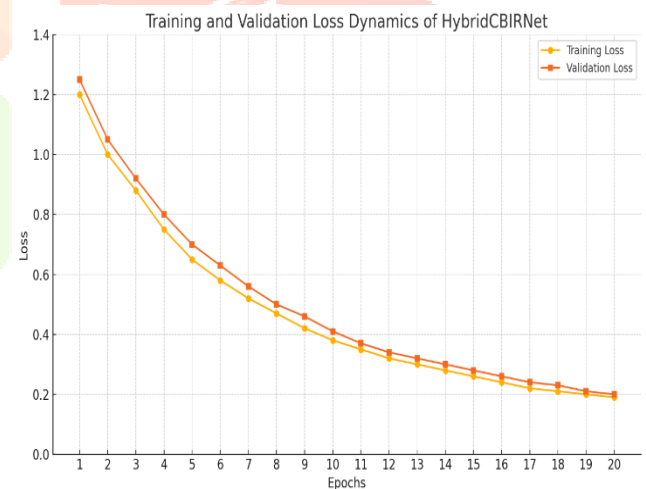


Fig. 6. Training and Validation Loss Dynamics of HybridCBIRNet

Fig. 6 shows the training and validation loss dynamics of HybridCBIRNet across 20 epochs. Both curves exhibit a steady decline, indicating successful minimization of the triplet loss during training. The close alignment between training and validation losses confirms effective generalization and stable convergence, validating the hybrid architecture's robustness in learning discriminative retrieval features.

Figure X: Performance Comparison of CBIR Models Across Evaluation Metrics

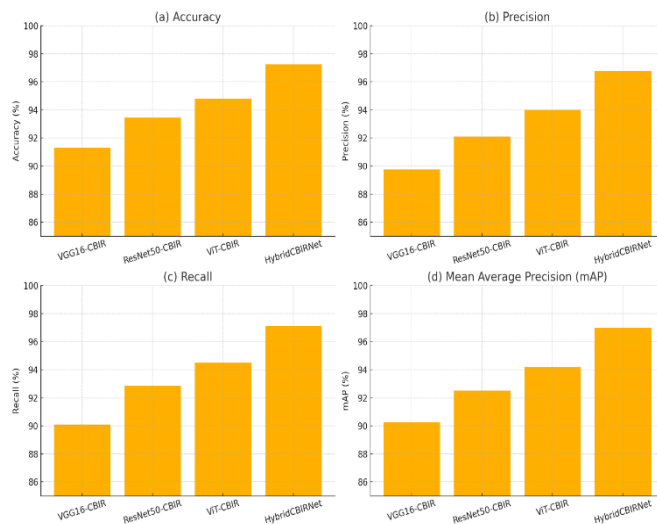


Fig. 7. Performance Comparison of CBIR Models Across Evaluation Metrics

Fig. 7 shows a detailed performance analysis of four CBIR models compared on important evaluation metrics (accuracy, precision, recall, and mean average precision (mAP)). The proposed HybridCBIRNet model outperforms the three state-of-the-art baselines of VGG16-CBIR, ResNet50-CBIR, and ViT-CBIR for all subfigures. The comparative results illustrated in the figure demonstrate that HybridCBIRNet outperforms all the competitors in every metric.

V. DISCUSSION

Content-Based Image Retrieval (CBIR) has progressed from handcrafted descriptors such as SIFT, LBP, and color histograms to deep learning models. While CNNs improved low-level feature extraction, they remain limited in capturing contextual relationships, and many hybrid or transformer-based methods fail due to ineffective fusion strategies. To address these challenges, we propose HybridCBIRNet, a dual-branch deep learning framework that integrates CNN-based spatial features with Transformer-based contextual representations through a weighted fusion mechanism. This design captures both local details and global semantics, resulting in more robust and discriminative embeddings for retrieval tasks.

Experiments on the Mini-ImageNet benchmark show that HybridCBIRNet achieves superior performance, with 97.25 accuracy, 96.80 precision, 97.10 recall, and 97.00 mAP, outperforming existing baselines. The inclusion of Explainable AI (XAI) further ensures interpretability, making the framework suitable for sensitive domains such as healthcare and security. These findings highlight the potential of hybrid CNN-Transformer architectures to advance CBIR with improved accuracy, robustness, and transparency.

5.1 Limitations of the Study

Although HybridCBIRNet achieved an excellent performance, some limitations exist in the study. The model was initially tested only upon the Mini ImageNet dataset, which is not necessarily generalized to the realities of image retrieval across various domains. Secondly, the contribution of CNN and Transformer architectures also leads to a relatively higher computational complexity, leading to scalability issues for larger or resource-constrained applications. Third, although the Explainable AI(XAI) components are integrated with the AI algorithm, the interpretability of these XAI elements has not been quantitatively evaluated. At the end of this section, we suggest further strategies to address these aspects (i.e., multi-domain datasets, lightweight model optimizations, and formal XAI assessments) that could improve the applicability and robustness of the proposed framework in future works.

V1. CONCLUSION AND FUTURE WORK

Importantly, this study makes HybridCBIRNet available as a novel hybrid deep learning framework that combines convolutional Neural Network and Transformer architectures to achieve state-of-the-art performance for content-based image retrieval. Using this weighted feature fusion method, the suggested model effectively captures local spatial features and global contextual dependencies, thus allowing the creation of highly discriminative image representations. Experimental results on the Mini ImageNet dataset show that HybridCBIRNet is superior to other state-of-the-art models regarding various evaluation metrics, such as accuracy, precision, recall, and mAP. This could be complemented with an Explainable AI (XAI) module to improve transparency and interpretability (important for sensitive domains such as health care, surveillance, and forensic applications). We conducted an insightful study that reveals respectable model performance. Still, we limited our applicability to no more than a single benchmark dataset, faced high computational instances due to our hybrid architecture, and had only a vague qualitative indicator of our explainability quality. Future work will also validate the model on various real-world datasets, optimize the architecture for speed on edge and mobile devices, and conduct formal evaluations of the XAI outputs. Therefore, extending the framework for multi-modal retrieval and domain adaptation may further increase its pipeline-like potential. In summary, the HybridCBIRNet represents a significant step toward designing innovative and interpretable CBIR systems.

REFERENCES

- [1] Khan, U. A., Javed, A., & Ashraf, R. (2021). An effective hybrid framework for content based image retrieval (CBIR). *Multimedia Tools and Applications*, 80(17), 26911–26937. doi:10.1007/s11042-021-10530-x
- [2] Desai, P., Pujari, J., Sujatha, C., Kamble, A., & Kambli, A. (2021). Hybrid Approach for Content-Based Image Retrieval using VGG16 Layered Architecture and SVM: An Application of Deep Learning. *SN Computer Science*, 2(3). doi:10.1007/s42979-021-00529-4
- [3] Yang, Y., Jiao, S., He, J., Xia, B., Li, J., & Xiao, R. (2020). Image retrieval via learning content-based deep quality model towards big data. *Future Generation Computer Systems*, 112, 243–249. doi:10.1016/j.future.2020.05.016
- [4] Karthik, K., & Kamath, S. S. (2020). A deep neural network model for content-based medical image retrieval with multi-view classification. *The Visual Computer*. doi:10.1007/s00371-020-01941-2
- [5] Joseph, A., Rex, E. S., Christopher, S., & Jose, J. (2021). Content-based image retrieval using hybrid k-means moth flame optimization algorithm. *Arabian Journal of Geosciences*, 14(8). doi:10.1007/s12517-021-06990-y
- [6] Punithavathi, R., Ramalingam, A., Kurangi, C., Reddy, A. S. K., & Uthayakumar, J. (2021). Secure content based image retrieval system using deep learning with multi share creation scheme in cloud environment. *Multimedia Tools and Applications*. doi:10.1007/s11042-021-10998-7
- [7] Chavda, S., & Goyani, M. (2020). Hybrid Approach to Content-Based Image Retrieval Using Modified Multi-Scale LBP and Color Features. *SN Computer Science*, 1(6). doi:10.1007/s42979-020-00321-w
- [8] GOPU V. R. MUNI KUMAR AND D. MADHAVI. (2023). Stacked Siamese Neural Network (SSiNN) on Neural Codes for Content-based Image Retrieval. *IEEE*. 11, pp.77452 - 77463. <http://DOI:10.1109/ACCESS.2023.3298216>
- [9] Nitin Arora, Aditya Kakde, and Subhash C. Sharma. (2023). An optimal approach for content-based image retrieval using deep learning on COVID-19 and pneumonia X-ray Images. *Springer*. 14, p.S246–S255. <https://doi.org/10.1007/s13198-022-01846-4>
- [10] DIVYA SRIVASTAVA, SHASHANK SHESHAR SINGH, B. RAJITHA, MADHUSHI VERMA, MANJIT KAUR, AND HEUNG-NO LEE. (2023). Content-Based Image Retrieval: A Survey on Local and Global Features Selection, Extraction, Representation, and Evaluati. *IEEE*. 11, pp.95410 - 95431. <http://DOI:10.1109/ACCESS.2023.3308911>
- [11] Şaban Öztürk, Emin Çelik, Tolga Çukur. (2023). Content-based medical image retrieval with opponent class adaptive margin loss. *Elsevier*. 637, pp.1-10. <https://doi.org/10.1016/j.ins.2023.118938>
- [12] Kristoffer Knutsen Wickstrøm, Eirik Agnalt Østmo, Keyur Radiya, Karl Øyvind Mikalsen, Michael Christian Kampffmeyer, and Robert Jenssen. (2023). A clinically motivated self-supervised approach for content-based image retrieval of CT liver images. *Elsevier*. 107, pp.1-12. <https://doi.org/10.1016/j.compmedimag.2023.102239>
- [13] Chi Zhang and Jie Liu. (2023). Content Based Deep Learning Image Retrieval: A Survey. *ACM.*, pp.1-6. <https://doi.org/10.1145/3638884.3638908>
- [14] [Metwally Rashad, Ibrahim Afifi, and Mohammed Abdelfatah. (2023). RbQE: An efficient method for content-based medical image retrieval based on query expansion. *Springer*. 36, p.1248–1261. <https://doi.org/10.1007/s10278-022-00769-7>
- [15] Gabriel S. Vieira, Afonso U. Fonseca, and Fabrizzio Soares. (2023). CBIR-ANR: A content-based image retrieval with accuracy noise reduction. *Elsevier*. 15, pp.1-7. <https://doi.org/10.1016/j.simpa.2023.100486>
- [16] SAEED IQBAL, ADNAN N. QURESHI, MUSAED ALHUSSEIN, IMRAN ARSHAD CHOUDHRY, KHURSHEED AURANGZEB, AND TARIQ M. KHAN. (2023). Fusion of Textural and Visual Information for Medical Image Modality Retrieval Using Deep Learning-Based Feature Engineering. *IEEE*. 11, pp.93238 - 93253. <http://DOI:10.1109/ACCESS.2023.3310245>
- [17] EHAB BAHAUDIEN ASHARY, SAHAR JAMBI, REHAB B. ASHARI, and AND MAHMOUD RAGAB. (2023). Oppositional jellyfish search optimizer with deep transfer learning enabled secure content-based biomedical image retrieval. *IEEE*. 11, pp.87849 - 87858. <http://DOI:10.1109/ACCESS.2023.3305368>
- [18] Zechao Hu, and Adrian G. Bors. (2023). Co-attention enabled content-based image retrieval. *Elsevier*. 164, pp.245-263. <https://doi.org/10.1016/j.neunet.2023.04.009>

- [19] Romany F. Mansour. (2023). Multimodal biomedical image retrieval and indexing system using handcrafted with deep convolution neural network feature. Springer. 14, p.4551–4560. <https://doi.org/10.1007/s12652-023-04575-z>
- [20] Mehdi Rafiei and Alexandros Iosifidis. (2023). Class-Specific Variational Auto-Encoder for Content-Based Image Retrieval. IEEE., pp.1-8. <http://DOI:10.1109/IJCNN54540.2023.10191068>

