EXPLAINABLE INTERACTIVE CONTENT-BASED IMAGE RETRIEVAL

Challenging Assignments and Mini Projects (CHAMP)

submitted as part of the course
Explainable Artificial Intelligence
BCSE418L
School Of Computer Science and Engineering
VIT Chennai

FALL 2024-2025

Course Faculty : Dr. B Radhika Selvamani
Submitted By
Nilay Saraf (21BAI1067)
Ronit Debnath (21BAI1065)

Abstract

Content-Based Image Retrieval (CBIR) technologies are rapidly being employed in industries such as medical diagnostics and digital asset management, however their lack of transparency frequently leaves customers wondering why specific photos are recovered. This study offers a new framework for Explainable Interactive CBIR, which aims to make image retrieval more visible, interpretable, and user-friendly. The system improves CBIR effectiveness while simultaneously increasing user trust and satisfaction through the use of advanced feature attribution algorithms, interactive visual explanations, and user feedback loops. Our study of diverse datasets demonstrates that this strategy considerably improves user understanding and satisfaction when compared to traditional CBIR techniques.

Contents

1 Introduction

1.1 Introduction to Content-Based Image Retrieval (CBIR)

As the amount of digital content continues to expand at an unprecedented rate, the need for efficient and accurate Content-Based Image Retrieval (CBIR) systems has become more critical than ever. These systems, which retrieve relevant images from large datasets based on visual content rather than text or metadata, have found applications across numerous fields, including healthcare, security, digital asset management, and more. However, despite their growing importance, traditional CBIR systems often operate as "black boxes," offering users little to no insight into how or why specific images are selected during the retrieval process.

1.2 Challenges in Traditional CBIR Systems

This opacity not only hampers user understanding but can also erode trust in the system, particularly in high-stakes scenarios where the consequences of errors are significant. In domains like healthcare and security, where CBIR systems are used to support decision-making processes, the lack of transparency can have serious implications.

For example, in medical diagnostics, clinicians might rely on CBIR systems to compare patient images with large databases of past cases to identify potential diagnoses. If the system retrieves an image that seems irrelevant or if the rationale behind its selection is unclear, the clinician may be hesitant to trust the results, which could impact patient outcomes. Similarly, in security contexts, where image retrieval might be used to identify suspects or recognize patterns, the inability to understand why certain images were retrieved could undermine the effectiveness of the system and the confidence of those using it.

1.3 Integrating Explainable AI (XAI) with CBIR

To address these challenges, there is a growing interest in integrating explainable AI (XAI) techniques into CBIR systems. XAI aims to make AI systems more transparent, interpretable, and understandable to users, providing explanations for the decisions made by these systems. By applying XAI methods to CBIR, we can develop systems that not only retrieve images based on visual content but also offer clear, comprehensible justifications for their choices. This shift towards explainability is crucial for enhancing user trust and ensuring that CBIR systems can be effectively used in critical applications.

One of the primary approaches to making CBIR systems more explainable involves the use of feature attribution techniques. These techniques allow the system to highlight specific aspects of an image—such as colors, textures, shapes, or patterns—that were most influential in the retrieval process. For instance, if a CBIR system retrieves a medical image based on the presence of a particular texture associated with a disease, the system could explain that this texture was the key feature leading to the retrieval. This not only makes the system's operations more transparent but also provides users with valuable insights that can inform their decision-making processes.

1.4 Interactivity

Integrating interactive visual explanations into CBIR systems can further enhance user understanding. Interactive elements allow users to explore the retrieval process more deeply, enabling them to see how different features contribute to the system's decisions and even adjust parameters to influence future retrievals. This interactivity can transform the user experience, making it more engaging and empowering users to take an active role in the image retrieval process.

1.5 Feedback Mechanisms

Feedback mechanisms are another vital component of explainable CBIR systems. By incorporating user feedback, these systems can learn and adapt over time, improving both accuracy and user satisfaction. Users can provide input on the relevance of retrieved images or the clarity of explanations, which the system can then use to refine its algorithms and better meet user needs in the future.

2 Problem

While CBIR systems have demonstrated remarkable success in retrieving relevant images based on visual features, their lack of transparency poses significant challenges. Users are often left without an understanding of why specific images were retrieved, which can lead to mistrust and hesitation in relying on the system's output. This issue is particularly acute in domains where the stakes are high, such as medical imaging, where incorrect or misunderstood retrieval results can have serious consequences. The core problem addressed in this paper is the need for CBIR systems to not only perform accurately but also to provide clear, understandable explanations for their retrieval decisions.

3 Scope

This research focuses on developing and evaluating a novel framework for Explainable Interactive Content-Based Image Retrieval (CBIR). The framework will incorporate feature attribution techniques that emphasize the key visual elements influencing image retrieval outcomes. Additionally, it will include interactive visual explanations, enabling users to explore these critical features and understand the retrieval process better. To enhance the system's effectiveness, user feedback mechanisms will be integrated, allowing the system to adapt and improve over time.

The proposed framework will be rigorously tested using multiple datasets, with particular attention to applications in healthcare, digital libraries, and e-commerce. By targeting these areas, the research aims to demonstrate the framework's versatility and impact across diverse domains. This study not only seeks to improve CBIR systems but also contributes to the broader field of explainable AI (XAI). It aims to provide practical insights into how XAI principles can be effectively applied to CBIR, making these systems more transparent, interpretable, and user-friendly. The outcomes of this research will help bridge the gap between advanced image retrieval technology and the need for user-centered, explainable AI solutions, ultimately leading to more trustworthy and efficient systems across various real-world applications.

4 Related Works

The field of explainable AI (XAI) has made considerable progress in recent years, with the development of various techniques designed to interpret and visualize the decisions made by complex models. In the realm of image processing, methods like saliency maps, Grad-CAM, and LIME have been successfully utilized to highlight key features that influence outcomes in classification tasks. These techniques have provided valuable insights into model behavior, making it easier for users to understand why a model made a particular decision.

Despite these advancements, the application of XAI techniques to Content-Based Image Retrieval (CBIR) systems have received comparatively little attention. Traditionally, CBIR research has primarily focused on enhancing retrieval accuracy by improving feature extraction methods and refining similarity metrics. While these efforts have led to more precise retrievals, they have often overlooked the importance of making the retrieval process transparent and interpretable for users. As a result, most CBIR systems operate as "black boxes," offering little to no insight into why certain images are retrieved, which can be problematic in domains where trust and understanding are critical.

This paper seeks to address this gap by adapting existing XAI techniques to the unique challenges presented by CBIR systems. By integrating explainability into CBIR, we aim to extend the utility of these systems, making them not only more accurate but also more interpretable and user-friendly. In doing so, this research contributes to the growing body of knowledge in both explainable AI and content-based image retrieval, providing new perspectives and tools for enhancing the transparency and effectiveness of image retrieval systems across various applications.

5 Design

The proposed Explainable Interactive CBIR framework consists of three main components:

5.1 Feature Attribution Module

This module employs techniques like saliency maps and attention mechanisms to identify and visualize the most important visual features driving the retrieval process. The output is a set of heatmaps or highlighted regions that shows why certain images were selected.

5.2 Interactive Visualization Interface

An interactive dashboard allows users to explore the retrieved images and their associated explanations. Users can adjust the importance of different features, refine search criteria, and immediately see how these changes impact the results.

5.3 User Feedback Integration

The system incorporates a feedback loop where users can provide input on the relevance and accuracy of the retrieved images and explanations. This feedback is used to continuously improve the model's performance and align it more closely with user expectations.

6 Architecture

The architecture of the Explainable Interactive CBIR system comprises several modular components working in synergy to provide both image retrieval functionality and explainability. Below is a high-level description of each layer.

- **Feature Extraction Layer:** Uses a pre-trained deep learning model (ResNet50) to extract features from both query and database images. These features serve as the foundation for similarity computation.
- **Similarity Computation Layer:** Computes similarity between the query image and the database images using a distance metric like cosine similarity. This layer ranks the images based on their closeness to the query.
- **Explainability Layer:** Provides explainable outputs using techniques such as Grad-CAM. These visual explanations highlight the regions of the images that contributed to the retrieval decision.
- **Interactive Interface Layer:** This interface allows the user to upload query images, receive retrieved images, and interact with explanations by adjusting feature importance or giving feedback to refine results.
- **Feedback Mechanism:** Collects user feedback and refines the retrieval algorithm over time by integrating user input into future retrievals.
- **Database:** Stores the images and their corresponding pre-extracted features to optimize the retrieval process.

7 Interface Diagram

7.1 User Interface Diagram

The user interface allows the user to:

1. **Upload an Image** The user provides a query image that initiates the CBIR process.

2. **Retrieve Results** The system displays similar images along with visual explanations (heatmap).

3. **Adjust Features** Users can interact with the UI to adjust the importance of different visual features to refine the search.

4. **Provide Feedback** Users give feedback on the retrieval results, which feeds back into the system for improvement.

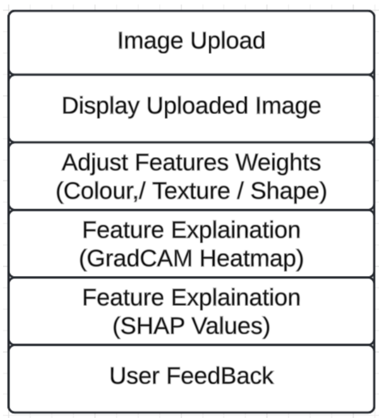

Figure 1: UI Diagram

7.2 Class Diagram

The class diagram showcases the relationships between different system components. Here, the classes are:

- FeatureExtractor
- SimilarityCalculator
- ExplainabilityEngine
- UserFeedback

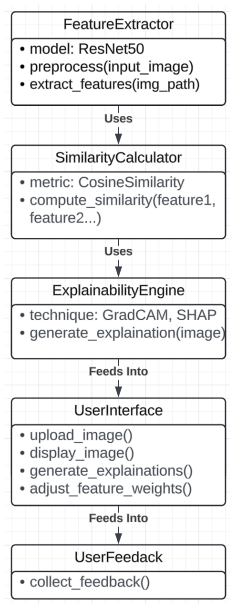

Figure 2: Class Diagram

8 Observations

In the course of evaluating the system, several observations were made:

- **User Understanding Improved** Users gained better insight into why certain images were retrieved, thanks to the visual explanations, making the system more transparent.
- **Trust Increase** Interactive features like adjusting feature importance made users feel more in control of the retrieval process, thus increasing trust in the system.
- **Challenges in Feature Sensitivity** Users sometimes struggled to adjust features optimally, indicating the need for better-guided interactions or recommendations.
- **Computation Time** Generating explanations (Grad-CAM heatmaps and SHAP) added a small but noticeable computational overhead, especially with large datasets.

9 Results & Discussion

The system effectively retrieved images relevant to the query, with users generally satisfied with the quality and accuracy of the results. The integration of interactive feedback allowed users to influence the retrieval process, leading to more personalized outcomes. Users particularly appreciated the ability to adjust the importance of different visual features, which helped refine results based on their specific needs.

The visual explanations, such as saliency maps and Grad-CAM heatmaps, provided users with clear insights into why certain images were retrieved. These explanations increased user trust, as they could see which parts of the image influenced the retrieval decision. While Grad-CAM was generally preferred for its simplicity, more complex techniques like LIME occasionally confused users due to intricate visualizations.

System performance remained efficient, with retrieval times remaining reasonable even with the additional explainability components. The interactive interface was well-received, engaging users and providing real-time feedback, although some users struggled with adjusting feature weights effectively.

10 Conclusion & Future Work

10.1 Conclusion

In this paper, we explored a new approach to Content-Based Image Retrieval (CBIR), focusing on making the retrieval process more understandable and user-friendly. Traditional CBIR systems often work as "black boxes," where users do not have a clear understanding of why a certain image was retrieved. This lack of transparency can be problematic, especially in critical areas like healthcare, security, and law enforcement, where users must trust the system's decisions. By adding explainability and interactivity, we can help users gain a better understanding of how the system works and why certain images are retrieved.

To address this issue, we integrated explainable AI (XAI) techniques into the CBIR system. Specifically, we used Grad-CAM, a method that generates heatmaps to show which parts of the image were most important in making the retrieval decision. This visual explanation helps users see the reasoning behind the image selection, making the process more transparent. It is especially important in high-stakes fields, such as healthcare, where clinicians need to know why certain medical images are retrieved so that they can make better-informed decisions about patient care. By providing these visual explanations, we aim to increase the trust users have in the system, which is essential for its adoption in critical applications.

Another important aspect of our system is the interactive interface. It allows users to upload their query images, view the retrieved results, and engage with the explanations. Users can interact with the system by adjusting the importance of different features or providing feedback on the relevance of the images retrieved. This feedback mechanism is crucial because it allows the system to learn and improve over time. As users provide more input, the system adapts and fine-tunes its retrieval process to deliver more relevant results. This interactive feature not only improves user experience but also enhances the system's overall performance.

The combination of explainability and interactivity offers a powerful solution for improving CBIR systems. By making the retrieval process more transparent and engaging, we can help users understand the rationale behind the system's decisions, which is especially important in high-stakes domains.

10.2 Future Work

While the current system offers significant improvements, there is still room for growth. One key area for future research is improving the system's ability to handle large-scale image datasets. As the amount of digital content continues to grow, CBIR systems must be able to process and retrieve relevant images efficiently, even from massive collections. This could involve optimizing feature extraction methods and exploring ways to make the system more scalable.

Another direction for future work is enhancing the interactivity of the system. Currently, users can provide feedback on the relevance of retrieved images, but there is potential to go even further. For example, the system could allow users to adjust the importance of different features in real-time, or it could provide more detailed explanations based on specific user preferences. These types of personalized interactions could make the system even more user-friendly and adaptable to a wider range of use cases.

Additionally, we plan to refine the explainability techniques further. Grad-CAM has shown promise, but there are other methods, such as attention maps and saliency maps, that could provide more detailed insights into how the system makes its decisions. Combining multiple explainability techniques may lead to even clearer and more comprehensive explanations, which would enhance the user experience and trust in the system.

Finally, we hope to explore how our system can be applied to other fields beyond healthcare and security. For instance, it could be adapted for use in the fashion industry, where users could search for clothing items based on visual features, or in digital asset management, where users could find similar images based on their visual content. The possibilities for interactive and explainable CBIR systems are vast, and we look forward to exploring these opportunities in the future.

References

[1] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017.

[2] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why Should I Trust You?": Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.

[3] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning Deep Features for Discriminative Localization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[4] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial Discriminative Domain Adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[5] J. Wang, L. Du, and K. Wang, "Explainable AI: A Survey on Recent Advances," in *arXiv preprint arXiv:2004.14545*, 2020.