

Research on the Application of Artificial Intelligence in Image Search

Zihao Wang

*School of Xuzhou, University of Technology, Xuzhou, China
Wangzihao031023@163.com*

Abstract. The era of information explosion has arrived, with vast amounts of data posing unprecedented challenges to traditional search engines. Efficiently extracting relevant content from complex and cluttered information has become a critical issue in technological development. Smart search technologies, particularly image search as a novel search method, have gradually become a vital component of modern search engines, playing a key role in improving search efficiency and user experience. To summarize the technologies supporting image search and analyze the performance of various AI models in this domain. This research aims to enhance the accuracy and efficiency of search engines, drive innovation in the information technology industry, and pave new paths for social progress and technological advancement. This study shows that deep-learning-based methods consistently surpass traditional content-based image-retrieval (CBIR) systems in precision and recall, yet their computational cost and interpretability remain open issues.

Keywords: Artificial Intelligence, Image Search, Deep Learning

1. Introduction

With the rapid advancement of information technology, the proportion of image data on the internet continues to grow year by year. Whether in social media, e-commerce, medical diagnostics, or security monitoring, images have become a crucial medium for information exchange. However, traditional image search methods primarily rely on manual labels, file names, or text descriptions. These approaches are not only inefficient but also highly prone to human subjectivity and language ambiguity, making it difficult to meet the demands of modern users for high accuracy, speed, and intelligence in image search.

In recent years, the emergence of artificial intelligence (AI) has introduced significant transformations to the image search domain. Powered by advancements in fields such as computer vision, deep learning, and natural language processing (NLP), image search has evolved beyond simple keyword-based matching. It has become an intelligent system capable of understanding the relationships between image content and semantics. Through the implementation of advanced algorithms like Convolutional Neural Networks (CNN), Generative Adversarial Networks (GANs), and Vision Transformers (ViT), these systems can automatically extract both low-level features

(such as texture, color, and shape) and high-level semantic information from images, enabling more precise matching and retrieval.

Moreover, intelligent image search has seen widespread application across various industries. In e-commerce, for instance, users can swiftly locate similar products by uploading images[1]. In healthcare, image retrieval systems assist doctors by enabling them to compare images of similar diseases to aid diagnosis[2]. In security systems, facial recognition technology has positioned image search as a key tool for tracking and identifying individuals[3]. These applications not only enhance the efficiency of information retrieval but also contribute to the digital and intelligent transformation of industries. Thus, the integration of AI in image search offers significant theoretical and practical value. This paper explores how AI can enhance the accuracy, intelligence, and scalability of image search systems, ultimately advancing the field of information retrieval technology.

2. Overview of image search technologies

2.1 Traditional image search methods

Traditional image search methods primarily depend on metadata such as manually assigned labels, file names, or textual descriptions for indexing and retrieving images[4]. While these methods laid the foundation for image retrieval systems, they come with several limitations.

In the early stages of image search, systems predominantly relied on textual descriptions associated with images. Users were required to input keywords or phrases related to the image they were searching for. The system would then match these queries with the available image metadata to deliver relevant results. For example, platforms like Google Images allowed users to retrieve images based on text-based queries. However, this method had notable drawbacks. First, it was highly subjective since text descriptions depended on human interpretation, leading to inconsistencies in labeling images. Additionally, words or phrases could have multiple meanings or interpretations, introducing ambiguity and causing inaccurate search results. Lastly, text-based search heavily relied on the availability of descriptive metadata, which was not always comprehensive or accurate, further limiting the effectiveness of this approach. File name-based search utilizes the file names or paths of images as keywords for indexing and retrieval. In this method, the filename serves as the basis for search queries. For instance, an image of a red apple named "red_apple.jpg" could be retrieved using its filename. Although this method offers a straightforward way to organize and retrieve images, it has significant limitations. The search scope is often restricted because filenames may not fully represent the content of the image. Moreover, filenames generally lack the context needed to describe complex or nuanced details, such as relationships between objects or hidden patterns within the image, resulting in less precise search outcomes.

2.2 Modern image search methods

Over the years, the field of image retrieval has undergone significant transformation, transitioning from basic text-based and file name-based methods to more advanced content-based and deep learning-driven techniques. These advancements aim to overcome the limitations of earlier approaches and enhance the accuracy, efficiency, and semantic understanding of image search systems.

Early image retrieval systems, such as Content-Based Image Retrieval (CBIR), focused on analyzing the actual content of images, including color, texture, and shape. CBIR systems extract low-level features from images and compare them with the features of the query image to identify

similar images in a database[5]. The primary advantage of CBIR is its ability to retrieve images based on visual content, reducing reliance on potentially subjective text descriptions or metadata. Furthermore, since CBIR does not require descriptive metadata, it proves particularly effective in scenarios where metadata is incomplete or unavailable. However, this approach also has limitations. Extracting meaningful low-level features can be computationally demanding and often requires substantial processing power. Additionally, CBIR systems typically struggle with understanding high-level semantics, such as object relationships or the context of the image, leading to less accurate searches when semantic understanding is necessary.

The emergence of deep learning techniques, especially Convolutional Neural Networks (CNNs), has transformed the field of image retrieval. Deep learning-based image search methods enable the automatic extraction of complex features from images, moving beyond low-level features to capture high-level semantic information. CNNs learn hierarchical representations of images, allowing them to capture detailed information such as object categories, scene context, and even abstract concepts within images. This results in improved accuracy and a deeper understanding of images compared to traditional CBIR methods. The main strength of deep learning models lies in their ability to automatically learn features from large datasets, eliminating the need for manual feature extraction[4][6]. However, deep learning methods also present challenges. They require large labeled datasets for training, which can be a significant barrier for certain applications. Moreover, training deep models demands considerable computational resources, and inference times may be longer compared to traditional techniques.

To address the limitations of both CBIR and deep learning-based methods, recent research has focused on hybrid approaches that combine traditional techniques with deep learning models. Hybrid image retrieval systems aim to leverage the strengths of both methods by integrating low-level feature extraction with deep learning-based semantic understanding[4]. These hybrid systems typically offer enhanced accuracy, as they can capture both detailed visual features and high-level semantic context. Additionally, hybrid models are more adaptable and can be tailored to specific applications, such as medical imaging, e-commerce, or security surveillance. However, combining multiple techniques can increase the overall complexity of the system, making it more challenging to implement and manage. Furthermore, hybrid methods often require higher computational resources due to the integration of different algorithms and processing layers.

3. Comparison of AI models for image searching

In the study of image search engines, performance comparison is an important criterion to measure the efficiency of image retrieval systems. This paper evaluates the performance of different image retrieval models by comparing them, especially the difference between content-based image retrieval (CBIR) methods and deep learning models. Through the analysis of multiple datasets, the performance of each model in terms of callback rate, precision and accuracy is compared, and the advantages and disadvantages of different models in image retrieval tasks are demonstrated.

Table 1: The performance comparison of models [7]

Author	Document	Backbone	Main Challenge	Recall	Precision	Accuracy
M. A. Rahman et al.	The Development of an Image Searching Method	CBIR	Blurred image recovery, edge preservation	Not Specified	Not Specified	89.7%
Sumiaya, Md. Amamuzzaman	ERISE	CBIR	High response time, memory usage	94.9%	89.9%	Not Specified
Shilpa Marathe et al.	Search Engine with GWO-NN	CBIR+GWO-NN	Feature extraction selection	95.59%	95.93%	95.59%
Arosh, Tamal Mondal	CNN-based Reverse Search Engine	Deep Learning	Feature dimensionality, CNN selection	97%	97.4%	97%
I Gede Susarma et al.	Reverse Search using CNN	Deep Learning	Image orientation, efficiency	18%	36%	Not Specified
C-H Chang et al.	Image Deblurring for Smartphones	Deep Learning	Motion blur removal	45%	Not Specified	Not Specified
Sumiaya & Md. Amamuzzaman	DURISE	Not Specified	Blur and haze, computational cost	85.89%	90.01%	Not Specified
K. Sashi Rekha et al.	Search Goals using SVM	Not Specified	Ambiguity in user preferences	91%	90%	91%
Che-Wei Chang & Chung-Ming Lo	CNN-based Commodity Search	Deep Learning	Semantic gap, product image diversity	Not Specified	91%	84%
Roland Szabo	Web-based Search Platform	Not Specified	Personalized search challenges	58.33%	77.78%	76.67%

Traditional content-based image retrieval (CBIR) relies on low-level features of images, such as color, texture, and shape, to retrieve images[8]. In contrast, in recent years, the introduction of deep learning methods has enabled image retrieval systems to improve retrieval accuracy by learning high-level features in images. By comparing different retrieval models, the results show that deep learning-based models are generally able to provide higher accuracy and precision. According to the data in Table 1, the accuracy of traditional CBIR methods is generally between 89% and 95%, while the accuracy of deep learning models can exceed 97%, showing a significant improvement.

The recall rate is another key performance indicator that reflects how many relevant images the retrieval system can find. Deep learning models often outperform traditional methods on this metric[8]. For example, convolutional neural network (CNN)-based models can achieve a callback rate of 97%, while traditional CBIR methods are typically lower than that. Especially when dealing with complex scenes, deep learning models are better able to extract features from images, thereby improving the coverage and accuracy of retrieval.

Precision is also a very important evaluation criterion in image retrieval, indicating the accuracy of the search results. Deep learning models also show advantages in this metric. The data in Table 1 show that the accuracy of CNN-based models is generally over 95% and can reach more than 97%

at most, while the accuracy of traditional CBIR methods is generally 90% to 96%. This shows that the deep learning model can not only find more relevant images, but also effectively reduce false detections and improve the quality of search results.

However, deep learning models also have certain limitations. While they perform well in terms of accuracy, callback rate, and precision, they are computationally expensive. Deep learning models often require a lot of computational resources and time, especially when working with large-scale datasets. Compared with this, the traditional CBIR method is more computationally efficient and can perform image retrieval in a shorter time, but its performance is less than that when processing complex images. Therefore, although deep learning methods can provide higher retrieval accuracy, traditional CBIR methods still have their advantages in some applications with high real-time requirements.

4. Challenges and solving of AI in image search

In complex backgrounds, object recognition can be affected by ambient noise, resulting in reduced accuracy. Especially when faced with visually intricate scenes—such as crowded public spaces, cluttered retail environments, or medical imaging with overlapping structures—it becomes more difficult to extract effective features. With the continuous advancement of deep learning technology, many innovative neural network architectures have been proposed, achieving remarkable results in image recognition tasks across real-world applications. Among them, ResNet and DenseNet are two of the most widely adopted and efficient deep learning models.

ResNet (Residual Network) addresses the vanishing gradient problem in deep networks by introducing residual connections[9]. As the number of layers in deep neural networks increases, gradients tend to either vanish or explode, making training difficult. ResNet overcomes this by adding "shortcuts" between layers, allowing information to bypass multiple layers and maintain strong gradient flow. This innovation enables the training of very deep networks, improving both the performance and stability of image recognition systems. For instance, in facial recognition used in airport security systems, ResNet's deep feature extraction enhances the model's ability to distinguish between individuals under varying lighting and occlusion conditions.

DenseNet, on the other hand, improves feature reuse by connecting each layer to all preceding layers[10]. Unlike traditional architectures where each layer is only connected to its immediate neighbors, DenseNet ensures that every layer can access the collective knowledge of all earlier layers. This design reduces the number of parameters, enhances gradient flow, and improves both the accuracy and robustness of recognition. A practical example is in medical diagnostics, such as detecting tumors in MRI scans, where DenseNet's ability to retain fine-grained details helps in identifying small or hidden abnormalities more precisely.

Both models have demonstrated significant improvements in performance when applied to complex image recognition tasks in various domains such as autonomous driving, intelligent retail systems, healthcare, and urban surveillance. Their success in handling noisy or cluttered visual data makes them essential components in modern AI-powered image search and recognition applications.

5. Conclusion

With the continuous development of artificial intelligence, image search technologies have undergone a fundamental transformation—from relying on metadata and simple visual features to adopting advanced deep learning models capable of understanding complex semantic information. Traditional methods such as keyword-based and file name-based search, while simple and easy to

implement, are limited by subjectivity and lack of semantic depth. Content-Based Image Retrieval (CBIR) introduced the use of visual features but struggled with high-level understanding. The rise of deep learning, particularly Convolutional Neural Networks (CNNs), has significantly improved the accuracy, recall, and precision of image retrieval by enabling automatic feature extraction and semantic interpretation. Furthermore, innovative architectures such as ResNet and DenseNet have addressed technical challenges like gradient vanishing and limited feature reuse, further enhancing the performance of image search systems.

Despite the remarkable progress, challenges remain—especially regarding computational cost, data requirements, and performance in complex environments. Hybrid approaches that combine traditional and AI-based methods have emerged as a promising direction, balancing performance and efficiency. In conclusion, the integration of AI into image search not only meets the growing demands for intelligent information retrieval but also paves the way for broader applications across fields such as e-commerce, healthcare, and security. Future advancements are expected to focus on improving real-time capabilities, reducing computational overhead, and enhancing the interpretability and adaptability of image retrieval systems.

References

- [1] Pallathadka, H., Ramirez-Asis, E. H., Loli-Poma, T. P., Kaliyaperumal, K., Ventayen, R. J. M., & Naved, M. (2023). Applications of artificial intelligence in business management, e-commerce and finance. *Materials Today: Proceedings*, 80, 2610–2613. <https://doi.org/10.1016/j.matpr.2021.06.419>
- [2] Mennella, C., Maniscalco, U., De Pietro, G., & Esposito, M. (2023). The role of artificial intelligence in future rehabilitation services: A systematic literature review. **IEEE Access**, 11, 11024–11043. <https://doi.org/10.1109/ACCESS.2023.3236084>
- [3] Irene, S., John Prakash, A., & Rhymend Uthariaraj, V. (2024). Person search over security video surveillance systems using deep learning methods: A review. *Image and Vision Computing*, 143. <https://doi.org/10.1016/j.imavis.2024.104930>
- [4] Zga, A., & Zair, N. (2022). L'interprétation sémantique des extractions du contenu d'une image. University of Kasdi Merbah Ouargla. Retrieved from <https://dspace.univ-ouargla.dz/jspui/handle/123456789/31330>
- [5] Latif, A., Rasheed, A., Sajid, U., Ahmed, J., Ali, N., Ratyal, N. I., Zafar, B., Dar, S. H., Sajid, M., & Khalil, T. (2019). Content-based image retrieval and feature extraction: A comprehensive review. *Computational Intelligence and Neuroscience*, 2019, 9658350. <https://doi.org/10.1155/2019/9658350>
- [6] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. <https://doi.org/10.1145/3065386>
- [7] Singh, M. K., Chakraverti, A., & Gupta, S. (2024). Comprehensive analysis on image search engines. *Proceedings of the 2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)*, 1438–1442. <https://doi.org/10.1109/ICTACS62700.2024.10841130>
- [8] Manjula, & Kumar, S. (2021). A comprehensive study on deep learning approach for CBIR. *Proceedings of the 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)*, 560–564. <https://doi.org/10.1109/CSNT51715.2021.9509633>
- [9] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- [10] Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>