Transportation Research Record
2020, Vol. XX(X) 1–16
©National Academy of Sciences:
Transportation Research Board 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/ToBeAssigned
journals.sagepub.com/home/trr
SAGE

Human-Oriented Image Retrieval System (HORSE): A Neuro-Symbolic Approach to Optimizing Retrieval of Previewed Images

Abraham Itzhak Weinberg[1]

Abstract

Image retrieval remains a challenging task due to the complex interaction between human visual perception, memory, and computational processes. Current image search engines often struggle to efficiently retrieve images based on natural language descriptions, as they rely on time-consuming preprocessing, tagging, and machine learning pipelines. This paper introduces the Human-Oriented Retrieval Search Engine for Images (HORSE), a novel approach that leverages neuro-symbolic indexing to improve image retrieval by focusing on human-oriented indexing. By integrating cognitive science insights with advanced computational techniques, HORSE enhances the retrieval process, making it more aligned with how humans perceive, store, and recall visual information. The neuro-symbolic framework combines the strengths of neural networks and symbolic reasoning, mitigating their individual limitations. The proposed system optimizes image retrieval, offering a more intuitive and efficient solution for users. We discuss the design and implementation of HORSE, highlight its potential applications in fields such as design error detection and knowledge management, and suggest future directions for research to further refine the system's metrics and capabilities.

Keywords

Image Retrieval, Neuro-Symbolic AI (NeSy), Natural Language Processing (NLP), Computer Vision

Introduction

At the intersection of human expertise and artificial intelligence lies a fundamental challenge: how to effectively bridge the gap between human knowledge and computational intelligence. Traditional Symbolic AI approaches, which attempt to codify human expertise into rule-based systems, have long struggled with the closed world assumption problem—they can only reason within the boundaries of explicitly defined knowledge, limiting their adaptability to novel situations. Conversely, modern Machine Learning (ML) approaches excel at pattern recognition but suffer from two critical limitations: biased outputs reflecting their training data and a lack of explainability that renders their decision-making processes opaque to human understanding.
This paper is organized as follows: first, we introduce the image retrieval tasks, their current solution approaches, and their drawbacks. Afterwards, we present the NeSy approach and explain how it can be integrated with image retrieval. Finally, we propose our HORSE algorithm and discuss it.

"Great living starts with a picture, held in your imagination, of what you would like to do or be" (Harry Emerson Fosdick, (*1*)). This concept of envisioning outcomes resonates with the Human-Oriented Image Retrieval System (HORSE), which aims to improve image retrieval by focusing on how humans mentally visualize and describe images. By leveraging neurosymbolic indexing, HORSE bridges the gap between human cognition and computational systems, enabling users to retrieve images based on natural language descriptions, much like imagining a desired outcome and bringing it into reality.
Recent research highlights the potential of Neuro-Symbolic (NeSy) to enhance interpretability and bridge neural computation with human logic, offering a compelling foundation for building transparent and human-aligned AI systems (*2*). Our proposed HORSE algorithm offers a novel solution through a unique NeSy integration approach. Unlike conventional NeSy systems where reasoning processes are merely layered atop neural network outputs, HORSE begins by extracting relevant human knowledge and then implements AI processes specifically aligned with this knowledge foundation. This human-oriented integration ensures that computational reasoning remains compatible with human logic while leveraging the pattern recognition strengths of neural networks. Retrieving visual information exemplifies the challenges this

[1]AI-WEINBERG, AI Experts, Tel Aviv, Israel

Corresponding author:
Abraham Itzhak Weinberg, aviw2010@gmail.com

approach addresses. Current image retrieval systems face significant limitations in terms of human usability, often employing complex pipelines of preprocessing, tagging, and ML algorithms that create redundant feature storage while failing to optimize for previously viewed images. These systems frequently return results that, while algorithmically relevant, do not align with human perception and memory processes.

This research explores an alternative paradigm where users can retrieve previously viewed images by describing them in their natural language, creating a more intuitive human-computer interaction model. By optimizing retrieval processes around human memory and description capabilities, we aim to develop systems that work in harmony with human cognitive processes rather than requiring humans to adapt to machine limitations. The research aims to achieve the following outcomes:

- A better understanding of the image retrieval process from both human and computer perspectives.
- Insights into the interaction between human memory, storage, description, and retrieval.
- Development of human-computer metrics for evaluating the end-to-end retrieval process.
- An optimized image retrieval solution that improves accessibility and retrieval efficiency.

Image Search Engines versus Image Retrieval and Indexing Engines

The domain of visual information discovery and management is divided primarily between two types of systems, each serving distinct purposes and user needs. Image search engines, exemplified by Google Images (*3*) and Bing Images (*4*), are designed for general users seeking to discover web-based images through text queries. These consumer-oriented platforms offer straightforward interfaces with basic filtering options for characteristics like size, color, and image type, prioritizing accessibility and breadth of coverage over technical sophistication.

In contrast, image retrieval and indexing engines serve technical and enterprise applications with specialized capabilities for managing curated image collections. Systems like Apache Solr with image extensions (*5*) and Elasticsearch with image plugins (*6*) support both text-based and content-based queries, enabling more precise access to visual assets.

These platforms incorporate advanced features including visual/image similarity matching[*](*7*) or Structural Similarity Index (SSIM)[†], automated feature extraction[‡], content classification[§], and comprehensive metadata indexing—capabilities that support professional workflows in fields ranging from digital asset management to medical imaging.

The fundamental differences between these system types lie in their scope, query methodologies, and technical depth as can be seen in Table 1. While search engines cast a wider net across the public web with primarily text-based queries, retrieval systems offer deeper analysis capabilities within defined collections using multiple query methods. This distinction reflects their different purposes: search engines connect users with previously unknown images, while retrieval systems help users efficiently locate and leverage known visual assets within managed repositories. The technical sophistication of retrieval systems comes with increased complexity but enables the precision and analytical capabilities required for professional and enterprise applications.

Background and Related Work

Our proposed approach begins by examining the human image retrieval process. Human memory is capable of rapidly encoding visual information and storing it compactly for long-term retrieval (*10*), a characteristic that our approach seeks to leverage for efficient indexing and storage in computer systems. The novelty of this work lies in its integration of human cognitive factors into image retrieval, which, to the best of our knowledge, has not been widely explored.

HORSE translates these characteristics into NeSy rules. These rules are then used as human knowledge, enhancing further ML and retrieval steps. As far as we know, this integration has not been done before. Additionally, the translation of human memory into NeSy rules is unique in the field.

Our algorithm addresses the growing need for accessible image retrieval systems in the context of the increasing volume of images stored and shared on social media platforms and personal devices. The findings of this research will also have significant implications for improving image accessibility, knowledge management, and aiding professionals like designers and draftsmen in detecting design errors.

---

[*] Image similarity can be thought of as a numerical representation of how alike two images are in terms of their visual content. There are several dimensions along which images can be similar, such as color, shape, texture, and composition.

[†] A commonly used metric that evaluates the structural similarity between two images. It takes into account luminance, contrast, and structure, providing a score ranging from -1 (completely dissimilar) to 1 (identical).

[‡] It employs specialized algorithms or deep networks to automatically extract features from signals or images, eliminating the need for human intervention (*8*).

[§] It classifies images into predefined categories based on their visual content (*9*).

Characteristic	Image Search Engines	Image Retrieval and Indexing Engines
Target Users	General consumers	Technical professionals and enterprises
Primary Purpose	Web-based image discovery	Management of curated image collections
Examples	Google Images, Bing Images	Apache Solr with image extensions, Elasticsearch with image plugins
Indexing Methodology	Web crawling with indexing based on surrounding text, file names, and basic visual features	Sophisticated multi-dimensional indexing structures (k-d trees, R-trees, locality-sensitive hashing)
Query Types	Primarily text-based queries	Multiple query paradigms: text-based, query-by-example, query-by-sketch, hybrid approaches
Query Processing	Keyword matching against indexed text	Complex similarity searches across feature vectors with semantic understanding
Feature Extraction	Basic image features and metadata	Advanced visual feature extraction with domain-specific optimizations
Data Sources	Public web content	Curated collections with structured metadata
Data Management	Continuous discovery via web crawling with limited dataset control	Carefully maintained collections with version control and access permissions
Analytical Capabilities	Basic filtering (size, color, type)	Advanced capabilities (object detection, scene understanding, facial recognition)
Computational Focus	Speed and relevance across massive datasets	Precision within defined domains with domain-specific algorithms
Integration Options	Limited API access	Robust APIs and integration frameworks for enterprise applications
Customization	Generalized algorithms across all content	Can be tailored for domain-specific visual patterns and use cases
Technical Complexity	Designed for ease of use	Higher complexity with more sophisticated controls
Scope	Broader coverage across public web	Deeper analysis within defined collections
Scalability Focus	Horizontal scaling for billions of images	Precise indexing and retrieval for domain-specific collections

Table 1. Comparison between Image Search Engines and Image Retrieval and Indexing Engines

Key Approaches and Algorithms in Image Retrieval and Indexing Systems

Image retrieval and indexing systems rely on various approaches and algorithms that help efficiently manage and search visual data. One key area is feature extraction (*8*), which involves extracting information from images in different forms. Color features, for example, are often used to capture the distribution of colors across an image, with techniques like color histograms, color moments, and color correlograms (*11*). Color histograms track the distribution of colors, while color moments capture the mean, variance, and skewness of color distributions, and the color correlogram represents the spatial correlation of colors within an image. Dominant color descriptors are also important for identifying key colors that characterize the overall appearance of an image.

Texture features focus on analyzing the surface properties of images (*12*), with methods such as the Gray Level Co-occurrence Matrix (GLCM), which studies the spatial relationships between pixel intensities (*13*). Gabor filters detect specific frequency content in various directions (*14*), and Local Binary Patterns (LBP) capture local texture patterns (*15*). Wavelet transforms offer a multi-resolution approach to texture analysis, which is crucial in many applications (*16*).

Shape features are another critical aspect, and methods like edge detection, including Canny and Sobel operators, highlight the edges of an image (*17*). Contour representations capture the shape outline (*18*), while moment invariants provide a way to describe shapes in a rotation- and scale-invariant manner (*19*). Shape contexts are also employed to capture the spatial relationships between points on the shape's boundary, offering a distinct representation of an object.

Deep learning-based approaches have gained significant attention, especially with Convolutional Neural Networks (CNNs), which are commonly used for feature extraction. Pre-trained models like Visual Geometry Group (VGG), Residual Network (ResNet), and Inception are often employed to generate feature vectors that serve as effective image descriptors (*20*). Siamese networks have become popular for learning similarity metrics between image pairs, improving the ability to distinguish between similar and dissimilar images (*21*). Auto-encoders are applied for dimensionality reduction, simplifying the data representation while preserving key features. Self-supervised learning methods, which allow for better representations of images without requiring labeled data, have also proven beneficial in enhancing retrieval performance.

Indexing structures are essential for efficient search and retrieval. Tree-based methods, such as KD-trees (K-Dimensional) for low-dimensional features and R-trees for spatial indexing, are commonly used (*22*). M-trees are suitable for indexing in metric spaces (*23*), and VP-trees (Vantage-Point) are particularly effective for high-dimensional data (*24*). Hashing methods, including Locality Sensitive Hashing (LSH), Semantic Hashing, Spectral Hashing, and Product Quantization, also play a vital role in reducing the complexity of retrieval tasks, allowing for faster searches by encoding data into more compact representations (*25*).

Similarity measures are used to assess the closeness between images based on their features (*26*). Euclidean distance is commonly used for comparing feature vectors, while cosine similarity is often employed in high-dimensional spaces. Earth Mover's Distance is useful for comparing histograms by measuring the cost of transforming one distribution into another (*27*). Hamming distance is applied to binary features (*28*), and Mahalanobis distance is used when the data involves correlated features, providing a more accurate similarity measure in such cases (*29*).

Content-Based Image Retrieval (CBIR) is one of the most popular retrieval techniques, allowing systems to retrieve images based on their content (*30*). Query by Example (QBE) and relevance feedback mechanisms are often used to refine searches, and multi-feature fusion strategies combine multiple types of features for improved retrieval accuracy (*31*). Cross-modal retrieval, which includes techniques such as text-to-image search and image-to-text mapping, allows for retrieving images based on textual descriptions and vice versa (*32*). These techniques rely on joint embedding spaces that bridge the gap between different modalities.

Modern optimization approaches like Approximate Nearest Neighbor (ANN) search are widely used to improve retrieval efficiency (*33*). Tools like FAISS (Facebook AI Similarity Search), Annoy (Spotify's ANN library), and HNSW (Hierarchical Navigable Small World) enable faster nearest neighbor searches, reducing computational overhead in large datasets (*34*). When evaluating retrieval systems, metrics such as precision and recall, Mean Average Precision (MAP), and Normalized Discounted Cumulative Gain (NDCG) are commonly used to assess the relevance and ranking of results (*35*). Other important factors include retrieval time and memory efficiency, which are critical in real-world applications.

For real-world implementation, several considerations must be addressed. Scalability is crucial to handle large datasets, and update mechanisms are required to accommodate dynamic collections. Storage optimization ensures efficient data storage, while query optimization improves the speed of query processing. Load balancing in distributed systems is also necessary to handle varying demands and ensure smooth operation.

Enhanced retrieval methods, such as using multiple query images and query expansion, can help improve retrieval accuracy (*36*). Semantic search techniques improve retrieval by understanding the meaning behind search queries (*37*), while attribute-based filtering allows for more fine-grained searches based on specific image attributes (*38*). Spatial verification is another technique that ensures the returned results align with the spatial characteristics of the query image (*39*).

Finally, advanced topics in image retrieval and indexing include fine-grained image retrieval, which focuses on retrieving images with subtle differences, and instance-level retrieval, which targets specific instances of objects. Cross-view image matching addresses challenges in matching images of the same object from different angles (*40*), and zero-shot image retrieval enables retrieval of images without prior examples or training data (41). Continuous learning systems, which adapt and improve over time through new data, are another emerging area in the field, ensuring that image retrieval systems remain effective as they evolve (42).

Notable Image Retrieval and Indexing Engines

There are various notable image retrieval and indexing engines, which can be categorized into commercial, open source, and deep learning-based solutions. Commercial solutions include Elastic Image Search (43), which offers visual search capabilities within the Elasticsearch ecosystem, and Amazon Rekognition (44), an AWS service designed for image analysis and similarity search. Google Cloud Vision Product Search enables visual product search and cataloging (45), while Microsoft Azure Computer Vision provides image analysis and visual search capabilities (46).
Additionally, Algolia Visual Search serves as a visual search add-on for their search platform (47). Open-source solutions include Milvus, a distributed vector database that handles image feature vectors (48), and FAISS (Facebook AI Similarity Search), a high-performance similarity search library (49). Qdrant is another vector similarity search engine that offers extended filtering support (50), while Vearch is a high-performance vector similarity search engine developed by Jina AI (51). Vespa is a real-time big data serving engine that also supports image search (50), and the ImageHash Python library allows simple image matching using perceptual hashes (52).
LIRE (Lucene Image REtrieval) is a Java library designed for image retrieval based on Lucene (53). On the deep learning front, CLIP (Contrastive Language Image Pre-Training) is a model developed by OpenAI for text-to-image search (54), and DupDetector uses neural networks to identify duplicate or similar images (55). DeepSight is another deep learning-based visual search engine (56). These solutions typically offer several key features such as CBIR, reverse image search, near-duplicate detection, feature vector indexing, perceptual hashing, and multimodal search, which combines text and images, all with scalable distributed indexing capabilities. A comparison of these image search solutions and their technical approaches is provided in Table 2, which summarizes their key approaches and core algorithms.

Google's Various Image Search and Retrieval Technologies Google's innovations in image search and retrieval technologies play a pivotal role in reshaping how users interact with and access visual information online. Google has developed various advanced image search and retrieval technologies. Google Images, the standard consumer-facing search engine, supports reverse image search using advanced computer vision and ML techniques, and integrates with Google Lens technology (57).
Google Lens is a mobile-first visual search tool that can identify objects, text, and landmarks in real-time and provides shopping capabilities for visual product search, along with text extraction and translation features (58). It is available as a standalone app and is integrated into other Google products. For enterprise applications, Google Cloud Vision AI offers multiple services, including Product Search for building retail catalog search systems, Vision AI for general-purpose image analysis and classification, AutoML Vision for custom model training (59), and the Video Intelligence API for analyzing video content (60).
Google's internal technologies include PlaNet, a model for geographic location estimation from images (61), DELF (DEep Local Features) for landmark recognition (62), SwAV (Swapping Assignments between Views) for self-supervised learning in image understanding (63), and MUM (Multimodal Unified Model), which processes both text and images simultaneously (64).

Neuro-Symbolic Approaches to Image Information Retrieval

NeSy approaches to image information retrieval combine the strengths of neural networks and symbolic reasoning to enhance the effectiveness of retrieval systems (65). Traditional information retrieval models typically rely on either neural networks or symbolic approaches, each with distinct advantages and limitations. Neural networks excel at processing high-dimensional data and recognizing patterns within visual content, while symbolic reasoning offers interpretability, logical consistency, and the integration of structured human knowledge. The integration of these complementary strengths in NeSy systems provides a promising framework to address the complex challenges of image retrieval, bridging the semantic gap between low-level image features and high-level human understanding.
NeSy approaches offer several key advantages for image retrieval systems (66). One major benefit is the ability to maintain interpretability while leveraging the pattern recognition capabilities of neural networks (2). Unlike pure neural models, which often function as "black boxes," NeSy systems retain transparency through their symbolic components, making the reasoning behind their decisions more understandable. Additionally, these approaches can explicitly integrate domain knowledge, human expertise, and cognitive principles that are difficult to capture through data alone. This knowledge integration is particularly valuable in specialized fields where labeled data may be scarce or incomplete.
Moreover, the incorporation of symbolic reasoning allows NeSy systems to perform logical inferences about spatial relationships, object properties, and semantic contexts—key aspects of human-like image understanding. This ability to reason about images adds a layer of depth that pure neural systems cannot match. The inclusion of symbolic knowledge also reduces the dependency on massive datasets, making

Solution	Key Approaches	Core Algorithms
Elastic Image Search	Vector-based similarity search with CNN feature extraction	HNSW graphs Approximate k-NN Dense vector scoring Cosine similarity computation
Amazon Rekognition	Multi-model approach combining detection and recognition	Deep CNNs Cascade classifiers Siamese networks YOLO variants
Google Cloud Vision	AutoML-based custom model training with efficient feature extraction	EfficientNet backbone Contrastive learning Deep metric learning BERT text-image matching
Microsoft Azure Vision	Transformer-based architecture with multi-task capabilities	ResNet extraction Scene graph generation Faster R-CNN Few-shot learning
Milvus	Distributed vector search with multiple index support	IVF/HNSW/ANNOY indexing GPU acceleration Dynamic quantization SIMD optimization
FAISS	High-performance similarity search with compression	Product Quantization Inverted File Index Multi-probe LSH Cluster-based indexing
Qdrant	Graph-based vector search with filtering	HNSW indexing Payload filtering Segment storage Query optimization
CLIP	Contrastive learning between image and text	Vision Transformer Zero-shot classification Cross-modal attention Temperature scaling
ImageHash	Perceptual hash-based image matching	Average hashing Difference hashing Wavelet hashing Color moment hashing
LIRE	Traditional computer vision features	SIFT/SURF descriptors Color/Edge detection Gabor textures Edge histograms

Table 2. Comparison of Image Search Solutions and Their Technical Approaches

NeSy approaches more viable in scenarios where training data is limited. Finally, the symbolic component acts as a safeguard, providing robustness against the common failure modes seen in purely neural approaches and improving overall system consistency and reliability.

The unique challenges of image retrieval make NeSy approaches particularly well-suited for this task. Visual information is inherently hierarchical and relational, with meaning emerging not only from individual objects but also from their spatial arrangement, context, and relationships (*67*). This complexity aligns well with the capabilities of NeSy systems. One of the key challenges in image retrieval is the substantial gap between pixel-level representations and semantic understanding. While neural networks can efficiently process raw visual data, they often struggle to capture the abstract relationships and contextual knowledge that humans rely on when describing or searching for images. Symbolic components help bridge this gap by explicitly representing higher-level concepts, such as spatial relationships or object properties, which are crucial for understanding and interpreting images.

Moreover, natural language queries for image retrieval often involve imprecise, subjective, or context-dependent terms that require interpretation beyond simple keyword matching (*68*). NeSy systems are particularly adept at handling such queries by mapping linguistic descriptions to visual features through symbolic representations of concepts like "above," "larger than," or "similar to." This ability to handle nuanced language makes NeSy systems more effective in real-world applications, where queries are rarely straightforward.

Another reason NeSy approaches are well-suited for image retrieval is that human memory for images operates on multiple levels of abstraction—from broad impressions to specific details. This mirrors the complementary processing of neural and symbolic components in NeSy systems, making them more aligned with human cognitive processes (*69*). This alignment allows NeSy systems to create retrieval mechanisms that feel intuitive and natural to users, further enhancing the user experience.

Furthermore, NeSy systems can leverage prior knowledge about common object relationships and spatial configurations to make inferences about images (*70*). This ability reduces the need for large amounts of training data, which is often required for pure neural approaches to learn these relationships. In specialized domains, where training data may be limited but domain knowledge is rich, this advantage becomes particularly important. By combining the strengths of neural and symbolic approaches, these systems offer a powerful solution for overcoming the challenges of image retrieval, making them an ideal choice for applications that require both high accuracy and interpretability.

Key Disadvantages and Limitations of Traditional Major Approaches to Image Retrieval

Although various image retrieval and indexing techniques offer promising solutions, each approach is accompanied by inherent limitations that impact their performance in different contexts. Table 3 provides a comprehensive overview of the key disadvantages associated with the most widely used methods in this field.

Traditional feature extraction methods, such as color, texture, and shape features, exhibit several drawbacks. Color features, for instance, are highly sensitive to illumination changes and perform poorly with grayscale images (*11*). Texture features, while useful in many scenarios, are computationally expensive and sensitive to rotation and scaling. Additionally, shape features often struggle with occlusions and complex or deformable objects, and they require clean segmentation for optimal performance. These limitations underscore the challenges faced by traditional approaches in real-world applications.

In contrast, deep learning-based approaches, such as CNNs and Siamese networks, have shown significant promise in image retrieval tasks (*20, 21*). However, these methods come with their own set of challenges. CNNs, for example, require large amounts of training data and are computationally intensive, which can make them impractical in resource-limited environments. Moreover, both CNNs and Siamese networks suffer from poor interpretability, and the risk of overfitting is a constant concern. Additionally, these methods typically have high memory usage.

Indexing structures, such as tree-based methods and hashing techniques, also present significant disadvantages (*22, 25*). Tree-based indexing approaches degrade in high-dimensional spaces and suffer from issues related to unbalanced trees and high memory overhead. Hashing methods, while efficient in some cases, are prone to information loss due to quantization and require complex handling of collisions. These issues can lead to poor retrieval performance, especially in dynamic or high-dimensional datasets.

Similarity measures, such as Euclidean distance and cosine similarity, are fundamental in many image retrieval systems but have limitations that hinder their effectiveness (*26*). Euclidean distance, for example, is highly sensitive to outliers and assumes equal feature weights, which may not always hold true in real-world data. Similarly, cosine similarity fails to account for magnitude differences, which can be a critical factor in certain applications. Both methods also struggle with high-dimensional data and lack the ability to understand semantic relationships between features.

Finally, modern optimization techniques, such as ANN search and cross-modal retrieval, offer advanced capabilities but are not without their challenges (*33*). ANN search faces a trade-off between accuracy and speed, and it requires careful tuning of parameters to balance performance. Cross-modal retrieval, which involves the alignment of different types of data such as images and text, suffers from semantic gaps between modalities and challenges related to paired data requirements. Moreover, methods like fine-grained retrieval require extensive annotations and can be computationally expensive, which limits their practicality for large-scale applications.

Table 3. Popular Image Retrieval and Indexing Approaches and Disadvatanges

Approach	Category	Key Disadvantages
Color Features	Traditional Feature Extraction	• Sensitive to illumination changes • Ignores spatial relationships • Fails with similar color distributions • Poor with grayscale images • High storage overhead
Texture Features	Traditional Feature Extraction	• Computationally expensive • Sensitive to rotation/scale • Limited for non-textured images • Poor with viewpoint changes • Requires multiple descriptors
Shape Features	Traditional Feature Extraction	• Highly sensitive to occlusion • Struggles with complex shapes • Poor with deformable objects • Computationally intensive • Requires clean segmentation
CNN Features	Deep Learning	• Large training data requirement • High computational cost • Poor interpretability • Overfitting risks • Heavy memory usage
Siamese Networks	Deep Learning	• Complex training pair selection • Training instability • Limited multi-class handling • Domain-specific retraining • High memory requirements
Tree-based Indexing	Indexing Structures	• Degrades in high dimensions • Unbalanced tree issues • High memory overhead • Expensive updates • Poor for dynamic data
Hashing Methods	Indexing Structures	• Information loss from quantization • Hash function dependency • Complex collision handling • Multiple table requirements • Precision-recall trade-off

Continued on next page

Table 3 continued

Approach	Category	Key Disadvantages
Euclidean Distance	Similarity Measures	• Outlier sensitivity • Equal feature weight assumption • Poor in high dimensions • Fixed-length feature requirement • No semantic understanding
Cosine Similarity	Similarity Measures	• Ignores magnitude differences • Poor with sparse data • High-dimension sensitivity • Vector space limitation • Simple relationship modeling
Query by Example	CBIR	• Query quality dependency • Semantic gap issues • Feature representation limits • No abstract query support • Multiple example needs
Relevance Feedback	CBIR	• Required user interaction • Time-consuming process • Convergence issues • User fatigue • Implementation complexity
ANN Search	Modern Optimization	• Accuracy-speed trade-off • Complex parameter tuning • High memory overhead • Missed relevant results • Data distribution dependency
Cross-Modal Retrieval	Advanced Methods	• Semantic gap between modalities • Paired data requirement • Vocabulary limitations • Modal alignment issues • Training complexity
Fine-grained Retrieval	Advanced Methods	• Detailed annotation needs • High computation costs • Visual sensitivity • Domain-specific features • Limited generalization

Neuro-Symbolic Integration

Nesy approaches aim to overcome the limitations of both by integrating neural networks for pattern recognition with symbolic reasoning mechanisms for logic and interpretability. This integration results in a more robust system capable of reasoning about images and queries in a way that pure neural networks or symbolic systems alone cannot.

Neural Components Neural components in a NeSy image retrieval system primarily handle unstructured data, such as images, by learning representations that capture the underlying semantics and visual relationships. In this section, we will describe the components of Embedding Generation (*71*), Feature Learning (*72*), and Natural Language Understanding (*73*), which are essential for processing both visual and textual data in a neuro-symbolic image retrieval system.
These components work together to process both visual and textual data in a meaningful way. Embedding generation is the first key step, where neural networks transform raw image data into dense vector representations (*71*). CNNs and Vision Transformers (ViTs) are commonly used in this process, enabling the extraction of image features while capturing contextual relationships between objects. Following this, feature learning allows neural models to automatically discover relevant patterns from raw image data, creating hierarchical representations at various levels of abstraction (*72*). This adaptability helps the system recognize domain-specific features. Finally, natural language understanding plays a crucial role in processing text queries, enabling the system to interpret natural language inputs, resolve ambiguity, and effectively match visual content with textual descriptions (*73*). Through these interconnected processes, the system is able to bridge the gap between visual and textual data, providing a robust and efficient retrieval mechanism.

Symbolic Components Symbolic components in a NeSy system add a layer of reasoning and structure, ensuring that the retrieved information is logically coherent and interpretable. This section will describe the components of Knowledge Representation (*74*), Reasoning Mechanisms (*75*), and Constraint Management (*76*).
Knowledge Representation involves the use of formal ontologies and taxonomies to define concepts such as “dog,” “car,” and “building,” along with their relationships (*74*). Symbolic representations, such as scene graphs and spatial relationships such as ”above,” ”below,” ”contains”, help to provide a deeper understanding of the image context. Reasoning Mechanisms utilize methods like forward and backward chaining to apply logical rules over the learned representations, clarifying relationships between different objects or concepts within an image and enhancing the retrieval process (*75*). Finally, Constraint Management ensures that the retrieved images meet specific criteria by enforcing logical constraints, such as matching certain attributes or fitting within predefined categories, thus ensuring the relevance and accuracy of the retrieval results (*76*).

Integration Mechanisms The integration of neural and symbolic components is a key aspect of NeSy systems, enabling the combination of learning-based and reasoning-based approaches for enhanced performance. This section will describe the following integration mechanisms: Neural-to-Symbolic Translation (*77*), Symbolic-to-Neural Guidance (*77*), and Hybrid Reasoning Paths (*78*).
One way integration occurs is through Neural-to-Symbolic Translation, where learned neural representations are transformed into symbolic forms that can be reasoned with. Visual features are mapped to symbols that are understood within the context of a predefined ontology (*77*). Another important mechanism is Symbolic-to-Neural Guidance, where symbolic knowledge, such as logical rules, influences and refines the neural learning process. Constraints from the symbolic layer can guide the neural network outputs, ensuring the generation of more accurate or contextually appropriate features (*77*). Additionally, Hybrid Reasoning Paths combine neural pattern recognition and symbolic inference in a synergistic manner. In this approach, the neural component is responsible for initial image feature extraction, while symbolic reasoning applies logical rules to filter, refine, and enhance the results, providing a more comprehensive retrieval mechanism.

Applications to Image Information Retrieval

In the context of image retrieval, NeSy approaches offer several key advantages. These systems excel in handling complex queries that require both pattern recognition and logical reasoning (*66*). For example, they can process queries like ”find images with a red car,” which involves pattern recognition, as well as more complex requests like ”find images where the car is in front of a building,” which necessitate logical reasoning.
Additionally, symbolic reasoning enables semantic search, where images are retrieved not only based on visual similarity but also by considering logical constraints and the semantic relationships between objects. NeSy systems also support multi-modal search, allowing users to query with a combination of text and images. For instance, users might upload a reference image and ask the system to find similar images, considering both visual features and the semantic meanings behind them.
Finally, the symbolic layer enhances explainability by providing interpretable results, allowing users to understand why certain images were retrieved based on explicit logical rules and relationships, rather than relying solely on the opaque decision-making processes of deep learning models.

Example Architecture for Image Retrieval

A typical architecture for a NeSy image retrieval system consists of several stages that work together to process and interpret both visual and textual data (*79*). The first stage, image processing, involves CNNs or ViTs, which process the image data to extract important features such as object recognition, localization, and contextual understanding.
Following this, the query parsing stage employs a Natural Language Processing (NLP) module to interpret user queries, converting text input into symbolic representations. For example, a query like ”Find images with a red ball” would be mapped to a search for red-colored objects within images. Once the query is parsed, symbolic filtering is applied. In this stage, symbolic reasoning methods are used to refine the retrieved images based on logical constraints.
For instance, a constraint such as ”only images where the ball is on a table” would narrow down the results. Finally, in the retrieval stage, both the neural and symbolic components collaborate to rank and return the most relevant images, considering not only visual similarity but also the logical constraints provided by the symbolic reasoning.

Algorithm 1 Human-Oriented Image Retrieval (HORSE)

```
Input: Image dataset
Output: Matching images based on query

Step 1: Extracting Human Retrieval Patterns
Step 2: Translate the Retrieval Patterns into NeuroSymbolic Meta Rules

for each image do
    Step 3.1: Detect Meaningful Objects (using OCR, for instance)
    Step 3.2: Extract properties of each Meaningful Object
    Step 3.3: According to Step 2, find relations between Meaningful Objects
end for
Step 4: Index the images using their meaningful objects and relations

Step 5: Enable NLP search query
Step 5.1: Extract the query objects and relations
Step 5.2: Search the image indexed database for matching images

Output: View the matching images
```

Human-Oriented Image Retrieval System (HORSE) Methodology and Problem Definition

As mentioned earlier, NeSy starts from human knowledge. Since the image retrieval target is to serve human users it has to be designed to their usage profile. We can diagnose the way that human approach for image retrieval. The transformation from retrieving image by using key words to a free language increase the challenge.
Analyzing the description of images shows that the human memory plays a main role in this task. As far as we researched, the human memory retrieval of images consists of several characteristics: image objects spacial relations, their size and characteristics such as color and if there are human also body and facial gesture.
Human memory plays a crucial role in image analysis and description, as research shows that when humans recall and describe images, they typically encode and retrieve various aspects of visual information. In this paper, we define several meta-rules based on human cognition and map them to the NeSy. These include spatial relationships between objects, often referred to as the ”where” information, which helps to understand the positioning of elements within an image (*80, 81*).
Additionally, we consider object properties like size, color, and shape, which contribute to the detailed recognition of individual items (*82, 83*). Finally, the semantic meaning and contextual relationships of objects within the image play an essential role in comprehending the broader narrative and interpretation of the visual content.

Human memory plays a crucial role in image analysis and description, as research shows that when humans recall and describe images, they typically encode and retrieve various aspects of visual information. This process is often hierarchical and gist-based, with people typically remembering the overall meaning or impression of an image first, followed by specific details. Additionally, emotionally significant elements tend to be remembered more vividly, highlighting the role of emotional salience in memory retrieval (*84, 85*).
Moreover, memory retrieval for images involves both bottom-up (feature-driven) and top-down (knowledge/expectation-driven) processes working together, rather than relying on feature extraction alone (*86, 87*). HORSE algorithm proposes the NeSy approach, which combines insights from human memory with AI techniques for image retrieval. The proposed approach exemplifies the extraction of rules from human knowledge, supported by psychological insights and the human memory retrieval system. These rules can be based on organizational data and generalized beyond past data, incorporating research on human memory, decision-making, and brain characteristics.
The HORSE algorithm 1 follows several steps: recognizing objects in the image and their names, as well as human emotions; extracting object characteristics like color; normalizing object sizes and ranking them by relative size; mapping relationships between objects using 2D and 3D

relations (such as A being above B, or to the left/right, or behind/in front); and learning relations based on an image corpus. For instance, the system learns that, in 99% of cases, a car is on the ground, the sky is in the upper part of images, and a house is bigger than a human. The system can then identify the uniqueness of an image by comparing it to other images and a 'normal' baseline. To develop a solution that bridges the human visual system and memory with computational systems, it is essential to take an interdisciplinary approach. Traditional algorithms often fail to account for human cognitive processes, which can limit the reliability and efficiency of image retrieval. Thus, our methodology examines the image retrieval process from both the human and computer perspectives.

Key parameters for evaluating this solution include human-side factors such as memory functionality, psychological aspects, and linguistic description capabilities, as well as computer-side factors like algorithm complexity, running time, and storage compactness. By assessing the retrieval process from these angles, we aim to provide a solution that optimally balances human needs and computational efficiency.

Discussion

The proposed HORSE algorithm represents a significant advancement in the field of image retrieval by addressing fundamental limitations of traditional approaches. By incorporating NeSy principles that mirror human cognitive processes, HORSE offers several advantages that warrant further discussion.

Alignment with Human Cognitive Processes

Our approach deliberately mirrors the hierarchical and relational nature of human visual memory. Traditional computer vision systems often emphasize raw feature extraction or pure statistical learning, which can create a mismatch between how machines index images and how humans naturally recall them. HORSE bridges this gap by establishing a framework that captures spatial relationships, object properties, and semantic meanings—the three key dimensions that characterize human image recall as identified in our research.

The incorporation of both bottom-up (feature-driven) and top-down (knowledge-based) processes in our algorithm acknowledges the bidirectional nature of human image processing (*86*, *87*). This dual-process approach allows the system to balance concrete visual features with contextual understanding, making it particularly effective for natural language queries that may contain imprecise or subjective descriptions.

Technical Implications and Advantages

The NeSy foundation of HORSE offers several technical advantages over purely neural or purely symbolic approaches. By extracting meta-rules from human cognitive patterns, the system can operate with greater interpretability than black-box deep learning models, while maintaining more flexibility than rigid rule-based systems. This middle-ground approach is particularly valuable for debugging, system refinement, and trustworthiness.

Our indexing strategy based on meaningful objects and their relationships represents a more efficient computational approach than exhaustive feature extraction. By focusing on elements that would be salient to human memory, we potentially reduce the dimensionality of the search space without sacrificing retrieval accuracy. This efficiency becomes increasingly important as image databases grow in size.

The normalization of object sizes and relative positioning in three-dimensional space allows HORSE to generalize across images with different perspectives and scales. This capability addresses a common limitation in traditional image retrieval systems, which often struggle with viewpoint invariance.

Limitations and Future Work

Despite its advantages, HORSE faces several challenges that require further research. First, the extraction of accurate spatial relationships depends on reliable object detection and scene understanding, which remain active research areas. Errors in object recognition can propagate through the system, potentially affecting retrieval accuracy. Second, while our approach aims to mimic human memory patterns, individual differences in visual perception and memory remain a challenge. Future iterations of HORSE could benefit from personalization mechanisms that learn individual users' recall patterns and adjust accordingly.

Third, the current implementation primarily focuses on static images. Extending the framework to video retrieval would require additional considerations for temporal relationships and motion patterns, which are crucial aspects of human memory for dynamic visual content. Future work should address these limitations while exploring several promising directions:

1. Multimodal Integration: Incorporating audio descriptions, text captions, and other contextual metadata could enhance retrieval accuracy, especially for ambiguous queries.

2. Adaptive Learning: Developing mechanisms for HORSE to continuously refine its understanding of human memory patterns based on user interactions and feedback.

3. Cross-Cultural Validation: Testing the system across diverse cultural contexts to ensure that the extracted meta-rules generalize across different user populations.

4. Computational Optimization: Further refining the indexing structures to balance comprehensiveness with computational efficiency, particularly for large-scale image collections.

Broader Applications

The principles underlying HORSE extend beyond simple image retrieval. The same NeSy approach could be applied to several adjacent domains:

- Visual Anomaly Detection: By establishing normative relationships between objects (e.g., ”cars are typically on roads”), the system could identify unusual or incorrect images.
- Accessibility Tools: HORSE could facilitate image descriptions for visually impaired users by focusing on the aspects of images that sighted humans find most memorable.
- Educational Applications: The system could support visual learning by helping students locate relevant images based on conceptual descriptions rather than keywords alone.
- Design Assistance: Creative professionals could use natural language descriptions to retrieve inspirational images that match their conceptual vision.

Conclusion

In this paper, we proposed a novel approach for human-oriented image retrieval, utilizing neuro-symbolic indexing. The method takes into account the user's cognitive and linguistic abilities, in addition to standard computational parameters, to optimize the image retrieval process. Future work will focus on refining the metrics and exploring potential applications in areas like design error detection and knowledge management.

HORSE represents a promising step toward more human-centric image retrieval systems. By grounding computational approaches in cognitive science research, we aim to create systems that feel more intuitive and accessible to users. The NeSy framework offers a balanced approach that maintains the advantages of both neural networks and symbolic reasoning while mitigating their respective limitations.

As visual content continues to proliferate across digital platforms, the need for effective retrieval systems becomes increasingly critical. HORSE demonstrates that by better understanding how humans process, store, and recall visual information, we can design more effective computational systems that serve human needs. Future research should continue exploring this intersection of cognitive science and computer vision to further enhance human-computer interaction in visual domains.

References

References

1. Maltz, M. *New psycho-cybernetics*. Penguin, 2002.
2. Weinberg, A. I. Neurosymbolic AI and Mechanistic Interpretability: Can They Align in the Artificial General Intelligence Era?, 2025.
3. Fergus, R., P. Perona, and A. Zisserman. A visual category filter for google images. In *Computer Vision-ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part I 8*. Springer, 2004, pp. 242–256.
4. Hu, H., Y. Wang, L. Yang, P. Komlev, L. Huang, X. Chen, J. Huang, Y. Wu, M. Merchant, and A. Sacheti. Web-scale responsive visual search at bing. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 2018, pp. 359–367.
5. Shahi, D. *Apache Solr: a practical approach to enterprise search*. Springer, 2015.
6. Dixit, B. *Elasticsearch essentials*. Packt Publishing Ltd, 2016.
7. Santini, S. and R. Jain. Similarity measures. *IEEE Transactions on pattern analysis and machine Intelligence*, Vol. 21, No. 9, 1999, pp. 871–883.
8. McKeown, D. M., T. Bulwinkle, S. Cochran, W. Harvey, C. McGlone, and J. A. Shufelt. Performance evaluation for automatic feature extraction. *International Archives of Photogrammetry and Remote Sensing*, Vol. 33, No. B2; PART 2, 2000, pp. 379–394.
9. Park, S. B., J. W. Lee, and S. K. Kim. Content-based image classification using a neural network. *Pattern Recognition Letters*, Vol. 25, No. 3, 2004, pp. 287–300.
10. Wang, L., X. Zhao, Y. Si, L. Cao, and Y. Liu. Context-associative hierarchical memory model for human activity recognition and prediction. *IEEE Transactions on Multimedia*, Vol. 19, No. 3, 2016, pp. 646–659.
11. Singh, S. M. and K. Hemachandran. Image retrieval based on the combination of color histogram and color moment. *International journal of computer applications*, Vol. 58, No. 3.
12. Armi, L. and S. Fekri-Ershad. Texture image analysis and texture classification methods-A review. *arXiv preprint arXiv:1904.06554*.
13. Park, Y. and J.-M. Guldmann. Measuring continuous landscape patterns with Gray-Level Co-Occurrence Matrix (GLCM) indices: An alternative to patch metrics? *Ecological Indicators*, Vol. 109, 2020, p. 105802.
14. Kamarainen, J.-K., V. Kyrki, and H. Kalviainen. Invariance properties of Gabor filter-based features-overview and applications. *IEEE Transactions on image processing*, Vol. 15, No. 5, 2006, pp. 1088–1099.
15. Pietikäinen, M. and G. Zhao. Two decades of local binary patterns: A survey. In *Advances in independent component analysis and learning machines*. Elsevier, 2015, pp. 175–210.
16. Busch, A. W. *Wavelet transform for texture analysis with application to document analysis*. Ph.D. thesis, Queensland

University of Technology, 2004.

17. Jing, J., S. Liu, G. Wang, W. Zhang, and C. Sun. Recent advances on image edge detection: A comprehensive review. *Neurocomputing*, Vol. 503, 2022, pp. 259–271.
18. Yang, D., B. Peng, Z. Al-Huda, A. Malik, and D. Zhai. An overview of edge and object contour detection. *Neurocomputing*, Vol. 488, 2022, pp. 470–493.
19. Qi, S., Y. Zhang, C. Wang, J. Zhou, and X. Cao. A survey of orthogonal moments for image representation: Theory, implementation, and evaluation. *ACM Computing Surveys (CSUR)*, Vol. 55, No. 1, 2021, pp. 1–35.
20. Gkelios, S., A. Sophokleous, S. Plakias, Y. Boutalis, and S. A. Chatzichristofis. Deep convolutional features for image retrieval. *Expert Systems with Applications*, Vol. 177, 2021, p. 114940.
21. Melekhov, I., J. Kannala, and E. Rahtu. Siamese network features for image matching. In *2016 23rd international conference on pattern recognition (ICPR)*. IEEE, 2016, pp. 378–383.
22. Li, M., H. Wang, H. Dai, M. Li, C. Chai, R. Gu, F. Chen, Z. Chen, S. Li, Q. Liu, et al. A survey of multi-dimensional indexes: past and future trends. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 36, No. 8, 2024, pp. 3635–3655.
23. Chen, L., Y. Gao, X. Song, Z. Li, Y. Zhu, X. Miao, and C. S. Jensen. Indexing metric spaces for exact similarity search. *ACM Computing Surveys*, Vol. 55, No. 6, 2022, pp. 1–39.
24. Ukey, N., Z. Yang, B. Li, G. Zhang, Y. Hu, and W. Zhang. Survey on exact knn queries over high-dimensional data space. *Sensors*, Vol. 23, No. 2, 2023, p. 629.
25. Han, L., M. E. Paoletti, X. Tao, Z. Wu, J. M. Haut, P. Li, R. Pastor-Vargas, and A. Plaza. Hash-based remote sensing image retrieval. *IEEE Transactions on Geoscience and Remote Sensing*.
26. Zhang, L., L. Zhang, X. Mou, and D. Zhang. FSIM: A feature similarity index for image quality assessment. *IEEE transactions on Image Processing*, Vol. 20, No. 8, 2011, pp. 2378–2386.
27. Tang, Y., L. H. U, Y. Cai, N. Mamoulis, and R. Cheng. Earth mover's distance based similarity search at scale. *Proceedings of the VLDB Endowment*, Vol. 7, No. 4, 2013, pp. 313–324.
28. Ionescu, M. and A. Ralescu. Fuzzy hamming distance in a content-based image retrieval system. In *2004 IEEE International Conference on Fuzzy Systems (IEEE Cat. No. 04CH37542)*, Vol. 3. IEEE, 2004, pp. 1721–1726.
29. Akyilmaz, E. and U. M. Leloglu. Similarity ratio based adaptive Mahalanobis distance algorithm to generate SAR superpixels. *Canadian Journal of Remote Sensing*, Vol. 43, No. 6, 2017, pp. 569–581.
30. Hameed, I. M., S. H. Abdulhussain, and B. M. Mahmmod. Content-based image retrieval: A review of recent trends. *Cogent Engineering*, Vol. 8, No. 1, 2021, p. 1927469.
31. Mei, T., Y. Rui, S. Li, and Q. Tian. Multimedia search reranking: A literature survey. *ACM Computing Surveys (CSUR)*, Vol. 46, No. 3, 2014, pp. 1–38.
32. Cheng, Q., Y. Zhou, P. Fu, Y. Xu, and L. Zhang. A deep semantic alignment network for the cross-modal image-text retrieval in remote sensing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 14, 2021, pp. 4284–4297.
33. Tu, Z., W. Yang, Z. Fu, Y. Xie, L. Tan, K. Xiong, M. Li, and J. Lin. Approximate nearest neighbor search and lightweight dense vector reranking in multi-stage retrieval architectures. In *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*. 2020, pp. 97–100.
34. Rahman, M., S. M. E. Rabbi, and M. M. Rashid. Optimizing Domain-Specific Image Retrieval: A Benchmark of FAISS and Annoy with Fine-Tuned Features. *arXiv preprint arXiv:2412.01555*.
35. Moffat, A. Computing maximized effectiveness distance for recall-based metrics. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 30, No. 1, 2017, pp. 198–203.
36. Xie, H., Y. Zhang, J. Tan, L. Guo, and J. Li. Contextual query expansion for image retrieval. *IEEE Transactions on Multimedia*, Vol. 16, No. 4, 2014, pp. 1104–1114.
37. Fernández, M., I. Cantador, V. López, D. Vallet, P. Castells, and E. Motta. Semantically enhanced information retrieval: An ontology-based approach. *Journal of Web Semantics*, Vol. 9, No. 4, 2011, pp. 434–452.
38. Hossen, M. B., Z. Ye, A. Abdussalam, and S. U. Hassan. Attribute-driven filtering: A new attributes predicting approach for fine-grained image captioning. *Engineering Applications of Artificial Intelligence*, Vol. 137, 2024, p. 109134.
39. Zhou, W., H. Li, Y. Lu, and Q. Tian. SIFT match verification by geometric coding for large-scale partial-duplicate web image search. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, Vol. 9, No. 1, 2013, pp. 1–18.
40. Shi, Y., X. Yu, D. Campbell, and H. Li. Where am i looking at? joint location and orientation estimation by cross-view matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 4064–4072.
41. Kan, S., Y. Cen, Y. Cen, M. Vladimir, Y. Li, and Z. He. Zero-shot learning to index on semantic trees for scalable image retrieval. *IEEE Transactions on Image Processing*, Vol. 30, 2020, pp. 501–516.
42. Yang, R., S. Wang, H. Zhang, S. Xu, Y. Guo, X. Ye, B. Hou, and L. Jiao. Knowledge decomposition and replay: A novel cross-modal image-text retrieval continual learning method. In *Proceedings of the 31st ACM International Conference on Multimedia*. 2023, pp. 6510–6519.
43. Amato, G., P. Bolettieri, F. Carrara, F. Falchi, and C. Gennaro. Large-scale image retrieval with elasticsearch. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 2018, pp. 925–928.

44. Sharma, V. Object detection and recognition using Amazon Rekognition with Boto3. In *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2022, pp. 727–732.
45. Zhong, Y., P. J. Garrigues, and J. P. Bigham. Real time object scanning using a mobile phone and cloud-based visual search engine. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. 2013, pp. 1–8.
46. Meena, M., A. R. Singh, and V. A. Bharadi. Architecture for software as a service (SaaS) model of CBIR on hybrid cloud of microsoft azure. *Procedia Computer Science*, Vol. 79, 2016, pp. 569–578.
47. Dasic, P., J. Dasic, and J. Crvenkovic. Applications of the search as a service (SaaS). *Bulletin of the Transilvania University of Brasov. Series I-Engineering Sciences*, 2016, pp. 91–98.
48. Wang, J., X. Yi, R. Guo, H. Jin, P. Xu, S. Li, X. Wang, X. Guo, C. Li, X. Xu, et al. Milvus: A purpose-built vector data management system. In *Proceedings of the 2021 International Conference on Management of Data*. 2021, pp. 2614–2627.
49. Danopoulos, D., C. Kachris, and D. Soudris. Approximate similarity search with faiss framework using fpgas on the cloud. In *International Conference on Embedded Computer Systems*. Springer, 2019, pp. 373–386.
50. Singh, P. N., S. Talasila, and S. V. Banakar. Analyzing embedding models for embedding vectors in vector databases. In *2023 IEEE International Conference on ICT in Business Industry & Government (ICTBIG)*. IEEE, 2023, pp. 1–7.
51. Sun, C., L. Bin Song, and L. Ying. Product Re-identification System in Fully Automated Defect Detection. In *International Conference on Smart Multimedia*. Springer, 2022, pp. 144–156.
52. Diyasa, I. G. S. M., A. D. Alhajir, A. M. Hakim, and M. F. Rohman. Reverse image search analysis based on pre-trained convolutional neural network model. In *2020 6th Information Technology International Seminar (ITIS)*. IEEE, 2020, pp. 1–6.
53. Lux, M. Lire: Open source image retrieval in java. In *Proceedings of the 21st ACM international conference on Multimedia*. 2013, pp. 843–846.
54. Gao, P., S. Geng, R. Zhang, T. Ma, R. Fang, Y. Zhang, H. Li, and Y. Qiao. Clip-adapter: Better vision-language models with feature adapters. *International Journal of Computer Vision*, Vol. 132, No. 2, 2024, pp. 581–595.
55. Vanam, D. and V. R. Pulipati. Identifying duplicate questions in community question answering forums using machine learning approaches. In *Machine Learning Technologies and Applications: Proceedings of ICACECS 2020*. Springer, 2021, pp. 131–140.
56. Zhang, Y., S. Qiao, S. Ji, and Y. Li. DeepSite: bidirectional LSTM and CNN models for predicting DNA–protein binding. *International Journal of Machine Learning and Cybernetics*, Vol. 11, 2020, pp. 841–851.
57. Al-Lohibi, H., T. Alkhamisi, M. Assagran, A. Aljohani, and A. O. Aljahdali. Awjedni: a reverse-image-search application. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, Vol. 9, No. 3, 2020, p. 49.
58. Maurya, A., A. Kumar, and D. Alimohammadi. Application of google lens to promote information services beyond the traditional techniques. *Qualitative and Quantitative Methods in Libraries*, Vol. 12, No. 1, 2023, pp. 111–136.
59. Bisong, E. Google automl: cloud vision. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*. Springer, 2019, pp. 581–598.
60. KC, A. and S. Aravind. Integrating Google AI: Enhancing Library Services for the Digital Age, 2024.
61. Xu, Y., L. Pan, C. Du, J. Li, N. Jing, and J. Wu. Vision-based UAVs aerial image localization: A survey. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*. 2018, pp. 9–18.
62. Noh, H., A. Araujo, J. Sim, T. Weyand, and B. Han. Large-scale image retrieval with attentive deep local features. In *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 3456–3465.
63. Jaiswal, A., A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon. A survey on contrastive self-supervised learning. *Technologies*, Vol. 9, No. 1, 2020, p. 2.
64. Fei, N., H. Jiang, H. Lu, J. Long, Y. Dai, T. Fan, Z. Cao, and Z. Lu. VEMO: A Versatile Elastic Multi-modal Model for Search-Oriented Multi-task Learning. In *European Conference on Information Retrieval*. Springer, 2024, pp. 56–72.
65. Dietz, L., H. Bast, S. Chatterjee, J. Dalton, J.-Y. Nie, and R. Nogueira. Neuro-symbolic representations for information retrieval. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2023, pp. 3436–3439.
66. Bouneffouf, D. and C. C. Aggarwal. Survey on applications of neurosymbolic artificial intelligence. *arXiv preprint arXiv:2209.12618*.
67. Silva, I. C. S. d. Visualization of intensional and extensional levels of ontologies, 2014.
68. Datta, R., D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (Csur)*, Vol. 40, No. 2, 2008, pp. 1–60.
69. Alford, S. *A neurosymbolic approach to abstraction and reasoning*. Ph.D. thesis, Massachusetts Institute of Technology, 2021.
70. Khan, M. J., F. Ilievski, J. G. Breslin, and E. Curry. A survey of neurosymbolic visual reasoning with scene graphs and common sense knowledge. *Neurosymbolic Artificial Intelligence*, Vol. 1, 2025, pp. NAI–240719.
71. Bhuyan, B. P., A. Ramdane-Cherif, T. P. Singh, and R. Tomar. Graph Neural Networks in Neural-Symbolic Computing. In *Neuro-Symbolic Artificial Intelligence: Bridging Logic and Learning*. Springer, 2024, pp. 231–253.
72. Gong, N., W. Ying, D. Wang, and Y. Fu. Neuro-symbolic embedding for short and effective feature selection via autoregressive generation. *ACM Transactions on Intelligent*

Systems and Technology, Vol. 16, No. 2, 2025, pp. 1–21.
73. Hamilton, K., A. Nayak, B. Božić, and L. Longo. Is neuro-symbolic AI meeting its promises in natural language processing? A structured review. *Semantic Web*, Vol. 15, No. 4, 2024, pp. 1265–1306.
74. Di Maio, P. Towards a Web Standard for Neuro-Symbolic Integration and Knowledge Representation Using Model Cards. In *Data Science with Semantic Technologies*. CRC Press, 2023, pp. 173–193.
75. Prentzas, J. and I. Hatzilygeroudis. Exploring aspects regarding reasoning in neuro-symbolic rules and connectionist expert systems. In *2021 12th International Conference on Information, Intelligence, Systems & Applications (IISA)*. IEEE, 2021, pp. 1–8.
76. Liu, J. Knowledge Completion Method Based on Interference Principle. In *2024 6th International Conference on Frontier Technologies of Information and Computer (ICFTIC)*. IEEE, 2024, pp. 398–401.
77. Pulicharla, M. R. Neurosymbolic AI: Bridging neural networks and symbolic reasoning, 2025.
78. Confalonieri, R. and G. Guizzardi. On the multiple roles of ontologies in explanations for neuro-symbolic AI. *Neurosymbolic Artificial Intelligence*, Vol. 1, 2025, pp. NAI–240754.
79. Oltramari, A., J. Francis, C. Henson, K. Ma, and R. Wickramarachchi. Neuro-symbolic architectures for context understanding. In *Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges*. IOS Press, 2020, pp. 143–160.
80. Rosman, B. and S. Ramamoorthy. Learning spatial relationships between objects. *The International Journal of Robotics Research*, Vol. 30, No. 11, 2011, pp. 1328–1342.
81. Hollingworth, A., C. C. Williams, and J. M. Henderson. To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review*, Vol. 8, No. 4, 2001, pp. 761–768.
82. Hu, Z. and J. Yang. Effects of memory cue and interest in remembering and forgetting of gist and details. *Frontiers in Psychology*, Vol. 14, 2023, p. 1244288.
83. Greene, N. R. and M. Naveh-Benjamin. The effects of divided attention at encoding on specific and gist-based associative episodic memory. *Memory & Cognition*, Vol. 50, No. 1, 2022, pp. 59–76.
84. Cooper, R. A., E. A. Kensinger, and M. Ritchey. Memories fade: The relationship between memory vividness and remembered visual salience. *Psychological Science*, Vol. 30, No. 5, 2019, pp. 657–668.
85. Baker, A. L., M. Kim, and J. E. Hoffman. Searching for emotional salience. *Cognition*, Vol. 214, 2021, p. 104730.
86. Ochsner, K. N., R. R. Ray, B. Hughes, K. McRae, J. C. Cooper, J. Weber, J. D. Gabrieli, and J. J. Gross. Bottom-up and top-down processes in emotion generation: common and distinct neural mechanisms. *Psychological science*, Vol. 20, No. 11, 2009, pp. 1322–1331.
87. Shelton, J. T. and M. K. Scullin. The dynamic interplay between bottom-up and top-down processes supporting prospective remembering. *Current Directions in Psychological Science*, Vol. 26, No. 4, 2017, pp. 352–358.

Copyright Statement