VISIONE: Redesigning an Interactive Retrieval System for Lifelog Search

Giuseppe Amato
CNR-ISTI
Pisa, Italy
giuseppe.amato@isti.cnr.it

Paolo Bolettieri
CNR-ISTI
Pisa, Italy
paolo.bolettieri@isti.cnr.it

Fabio Carrara
CNR-ISTI
Pisa, Italy
fabio.carrara@isti.cnr.it

Fabrizio Falchi
CNR-ISTI
Pisa, Italy
fabrizio.falchi@cnr.it

Claudio Gennaro
CNR-ISTI
Pisa, Italy
claudio.gennaro@isti.cnr.it

Nicola Messina
CNR-ISTI
Pisa, Italy
nicola.messina@isti.cnr.it

Lucia Vadicamo*
CNR-ISTI
Pisa, Italy
lucia.vadicamo@isti.cnr.it

Claudio Vairo
CNR-ISTI
Pisa, Italy
claudio.vairo@isti.cnr.it

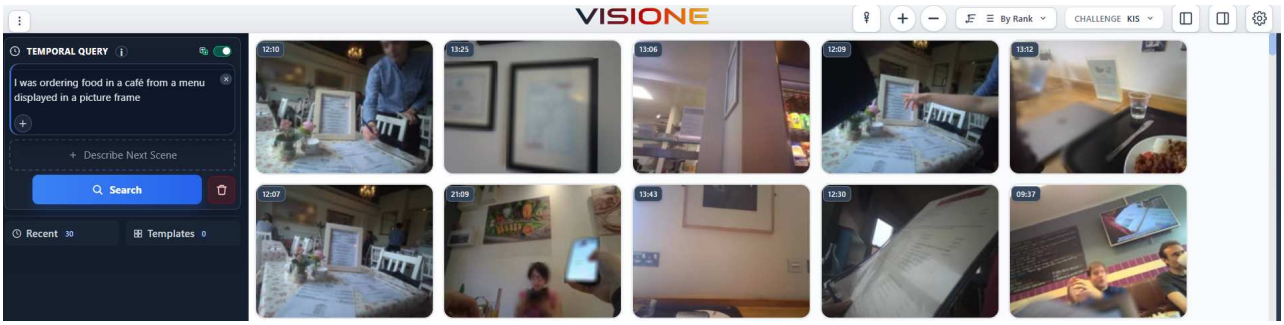


Figure 1: Example of an interactive lifelog search session in VISIONE, illustrating the query "*I was ordering food in a café from a menu displayed in a picture frame*".

Abstract

Finding specific moments in lifelog collections is often difficult, as users must navigate large streams of temporally ordered images while refining queries on the fly. In this paper, we present a redesigned version of VISIONE, an interactive retrieval system tailored to the Lifelog Search Challenge (LSC). Building on our previous participation, the system has been restructured to better support time-constrained search scenarios and iterative query refinement. The new version introduces a modular architecture that integrates multiple retrieval modalities, including multimodal embeddings and metadata-based filtering, within a unified framework. On the interaction side, the interface has been redesigned to simplify query formulation through sequence-based temporal queries, while still allowing expert users to control retrieval strategies and employed feature representations. Additional components, such as relevance feedback and LLM-assisted query processing, further support both exploratory and task-driven search. Taken together, these design choices aim to make the system more effective in practice, enabling users to quickly narrow down large lifelog collections and converge toward relevant moments under realistic LSC conditions.

CCS Concepts

• **Information systems** → **Information retrieval**; **Users and interactive retrieval**; **Retrieval models and ranking**; *Search engine architectures and scalability*; **Multimedia and multimodal retrieval**; **Video search**.

Keywords

multimedia retrieval, image search, cross-modal search, interactive system, interactive retrieval, lifelog retrieval, egocentric images

ACM Reference Format:
Giuseppe Amato, Paolo Bolettieri, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, Nicola Messina, Lucia Vadicamo, and Claudio Vairo. 2026. VISIONE: Redesigning an Interactive Retrieval System for Lifelog Search. In *The 9th Annual ACM Workshop on the Lifelog Search Challenge (LSC '26), June 16–19, 2026, Amsterdam, Netherlands.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3810984.3819178

LSC '26, Amsterdam, Netherlands

ACM ISBN 979-8-4007-2697-2/26/06
https://doi.org/10.1145/3810984.3819178

1 Introduction

Lifelogging poses significant challenges for retrieval due to the scale and nature of the data, which typically consists of long-term, egocentric visual streams capturing everyday activities [9]. These collections are largely unstructured and exhibit strong temporal and contextual dependencies, making it difficult to retrieve specific moments, especially when the user's information need is vague or only partially specified. As a result, effective retrieval often relies on the user iteratively refining queries and exploring results, rather than issuing a single precise query. This highlights the need for effective interactive image and video retrieval systems, where users can iteratively refine queries and explore results to progressively converge toward the desired content. The Lifelog Search Challenge (LSC) [23] provides a well-established benchmark for evaluating such systems in realistic, time-constrained lifelogging tasks, including Known-Item Search (KIS), ad-hoc retrieval (AVS), and question answering (Q&A), within a live evaluation setting.

In recent years, lifelog retrieval systems have largely moved away from purely concept-based pipelines towards approaches that combine multimodal embeddings, often combined with higher-level reasoning components [23]. Top-performing systems in recent LSC editions, like MEMORIA [8], SnapSeek 3.0[10], and MemoriEase [25], rely mostly on CLIP [11, 20] and BLIP2 [14] as multimodal embeddings, increasingly integrate large language models (e.g., GPT [2], Qwen [6], Mistral7B [12]) to support query interpretation, reformulation, retrieval-augmented generation, or question answering. Alongside these developments, some systems have investigated alternative interaction strategies, including conversational interfaces [25] or dual-mode interfaces [5, 28], which balance simplicity for novices with parameterized control for experts, aiming to balance usability and search efficiency.

VISIONE is an interactive retrieval system that has been previously evaluated in the LSC and participated in several editions of the Video Browser Showdown (VBS) [21, 26]. Since its last participation in 2024 [4, 5], the system has undergone a substantial redesign. Earlier versions of the system offered a broad set of retrieval functionalities, including cross-modal search, similarity search, temporal search, concept-based retrieval, and spatial color and object search through a dedicated canvas interface [3–5]. However, observations from past participations indicated that some of these functionalities (particularly canvas-based interaction and concept-based search) could introduce interaction bottlenecks, especially for novice users. In practice, their use has progressively declined in favor of cross-modal search based on multimodal embeddings, which not only simplifies query formulation but also leads to more effective retrieval results. Based on these observations, the system has been substantially redesigned for LSC 2026. The new version adopts a modular architecture that separates user interaction, processing, and data management components, allowing for more flexible integration of retrieval methods. The backend has been unified around a PostgreSQL database with pgvector [19], replacing the previous dual-index setup (FAISS and Lucene) and simplifying data management. At the same time, the retrieval and analysis pipelines are now orchestrated by a customized LangChain middleware [13], enabling more effective search and system management, while supporting agentic interactions for question answering tasks. Furthermore, the user interface has been entirely rebuilt as a single-page web application, adopting a flexible multi-panel layout designed to streamline query formulation, result exploration, and relevance feedback.

The redesigned system was evaluated at LSC 2026 [24] with three independent users. VISIONE obtained strong results in the overall individual ranking, with two users placing 2nd and 3rd, and showed particularly strong performance in KIS, where VISIONE2 ranked 1st. These results suggest that the redesign effectively supported interactive search under realistic competition conditions.

In this paper, we present the updated VISIONE system and report the results achieved in LSC26. As with previous versions, the new VISIONE system will be released as open source at https://github.com/aimh-lab/visione/.

2 System Overview and Architecture

The redesigned VISIONE system adopts a modular and layered architecture aimed at supporting scalable and efficient interactive video retrieval. As illustrated in Figure 2, the system is organized into three main layers: the *User Interface Layer*, the *Processing Layer*, and the *Data Layer*. This design enables a clear separation of concerns while facilitating the integration of heterogeneous retrieval and analysis components.

The *User Interface Layer* provides the main entry point to the system and is implemented as a web-based application using Svelte [1]. It supports query submission, result visualization, and interaction logging, as well as integration with the DRES evaluation framework [22] used in LSC. User queries are forwarded to the backend, while ranked results and intermediate feedback are presented through an interactive interface designed for rapid exploration.

The core functionality of the system is handled within the *Processing Layer*, which is centered around the *Search and Management Server* implemented using a customized version of LangChain [13]. This component orchestrates the retrieval workflow, including query encoding, selection of retrieval strategies, and aggregation of results from multiple sources. It also interacts with an LLM (Ministral-3 14B [16]) to support higher-level functionalities such as automatic query translation and retrieval-augmented question answering. In addition, it coordinates a GPU-enabled *Analysis Server*, responsible for computationally intensive multimedia processing, including feature extraction and embedding generation. This separation allows the system to scale efficiently by offloading heavy processing tasks to dedicated hardware resources. The system currently supports multiple feature extraction pipelines to enable different retrieval modalities. In particular, DINOv2 features are available for image similarity search, while OpenCLIP *ViT-H/14* embeddings [11, 20] and Qwen3-VL-Embedding 8B [15] are used for both cross-modal and similarity-based retrieval. Users can dynamically choose between OpenCLIP, Qwen3-VL, or a "*Smart*" retrieval mode, which combines the rankings produced by the two models to exploit their complementary strengths.

The extracted features are stored and indexed within the *Data Layer* using PostgreSQL with the pgvector extension [19], which manages both metadata and high-dimensional vector representations. It enables efficient vector similarity search alongside structured filtering (e.g., temporal constraints or metadata-based queries) within a single system. Access to multimedia assets is supported

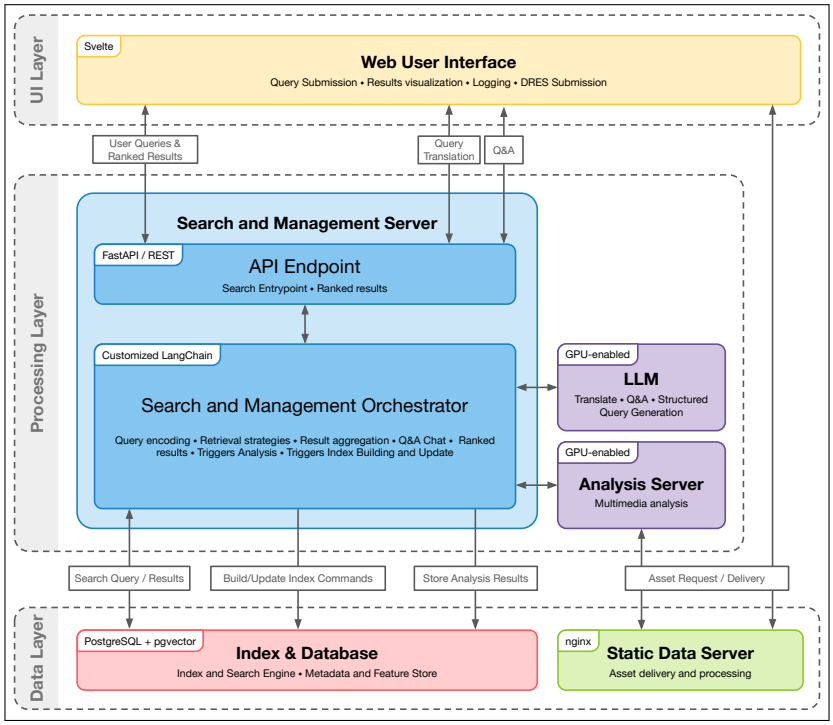


Figure 2: Overview of the VISIONE system architecture.

through lightweight HTTP services. The Data layer manages asset requests and delivery between the *Data Server* and both the user interface and the analysis components, ensuring efficient data transfer and enabling on-demand visualization of multimedia content.

Table 1 summarizes the main differences between the redesigned VISIONE system and its previous 2024 release.

3 User Interface and Interaction

The VISIONE user interface is designed as a unified environment for interactive image and video retrieval, where users can formulate multimodal queries over large-scale visual collections. It is implemented as a single-page web application and organized using a *multi-panel layout*, with a left sidebar for query formulation, a central area for result visualization, and a right sidebar for relevance feedback and submission tracking (Figure 3). A top toolbar provides quick access to navigation and display controls, while a persistent status bar at the bottom reports real-time information about the current session.

The main entry point is the temporal search view, where users construct queries as a sequence of descriptive steps. Each step represents a scene or event within a temporally ordered lifelog image stream, and can be individually enabled, disabled, reordered, or removed, allowing users to model complex temporal queries by simply describing the different moments of interest. Queries can be expressed in natural language and optionally enriched with visual inputs (e.g., images from previous results, local files, or external URLs), supporting combined text–image retrieval. For each step, users can select different embedding models for both textual and visual components, enabling fine-grained control over the retrieval process. In addition, each step can be combined with metadata-based filters (e.g., time, date, or location), allowing users to further refine the search space. The interface also supports multilingual queries through automatic translation into English.

To support iterative search, the system provides a history of recent queries and reusable templates (bottom part of left panel in Figure 3), enabling users to quickly revisit and refine previous search attempts.

The central area displays results in a navigable and customizable grid, where items can be grouped according to different criteria, including relevance, temporal attributes (e.g., date and time), and other metadata associated with the lifelog data. Each result is presented as an image and provides contextual interactions, including a slideshow-based exploration of temporally related images (e.g., within the same time window, Figure 4a) and access to a summary view that aggregates related content (Figure 4b). Users can also click on a result to open a dedicated modal view showing a larger version of the selected keyframe along with some additional information such as video ID, frame ID, and the associated metadata (Figure 4c). Additional interactions include image similarity search, relevance feedback through labelling the image as a positive or negative example, and submission of the selected result.

Relevance feedback is supported through the selection of positive and negative examples, displayed in a dedicated side panel that the user can expand or collapse (Figure 4d). Two main relevance feedback strategies, namely Rocchio and SVM-based approaches [27], allow expert users to configure the underlying algorithm through system settings according to their preferences. Submitted items

Table 1: Technological Evolution of the VISIONE System (2024 vs. 2026)

Feature / Module	VISIONE 2024 [4, 5]	VISIONE 2026 (this paper)
System Architecture	Tightly coupled architecture with limited separation between components	Modular layered architecture (UI, Processing, Storage and Retrieval, Data Access)
Backend Index and Storage	Dual-index setup: FAISS for CLIP embeddings and Apache Lucene for metadata and surrogate text representations	Unified index of embeddings and metadata using PostgreSQL with `pgvector` extension.
Retrieval Orchestration	Fixed retrieval pipeline with late fusion of results	Dynamic orchestration via a customized LangChain-based middleware
Multimodal Models	OpenCLIP ViT-L/14 [11, 20], CLIP2Video [7], ALADIN [17], and DINOv2 [18]	OpenCLIP ViT-H/14 and ViT-L/14 [11, 20], Qwen3-VL-Embedding 8B [15], and DINOv2 [18]
Temporal Search	Pairwise temporal matching within a predefined time window	Sequence-based queries with ordered multimodal steps representing event chains
User Interface	Dual-mode interface (novice/advanced) with canvas-based spatial search, text boxes for natural language queries, and fixed grid-based result visualization	Single-page application with a configurable multi-panel layout, supporting user-specific preferences beyond the expert/novice paradigm
QA Capabilities	Manual inspection of retrieved results	LLM-assisted question answering with retrieval-augmented generation
Query Processing	Single-modality queries with optional automatic text translation (SeamlessM4T)	Composed multi-modal, multi-step queries with LLM-assisted translation and search orchestration

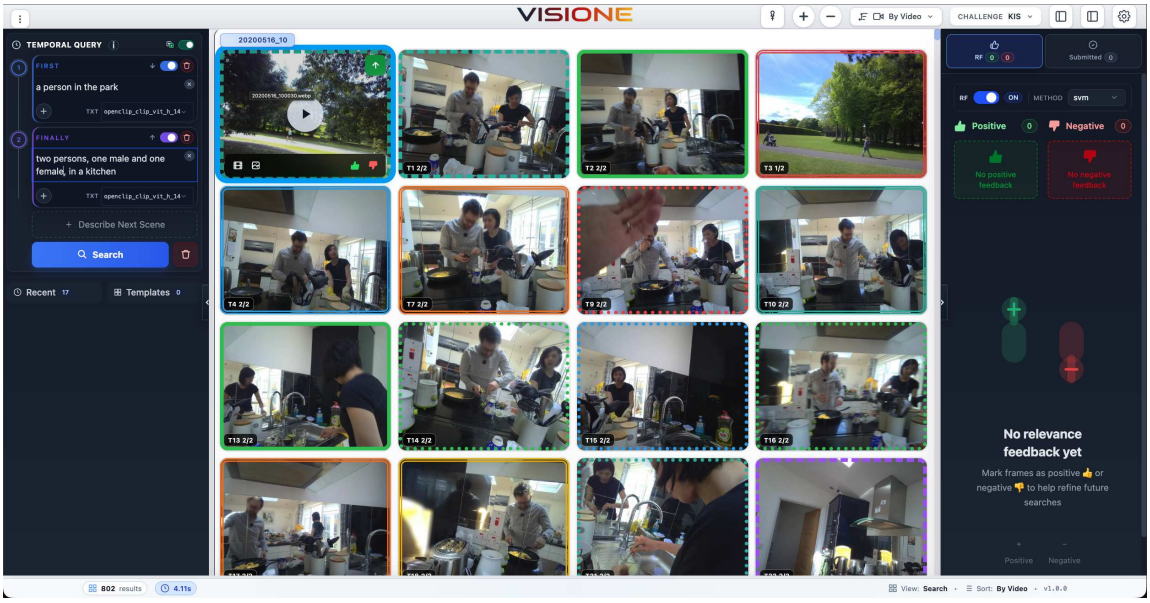


Figure 3: VISIONE UI with multi-panel view for search (left), results (center), and relevance feedback/submissions (right).

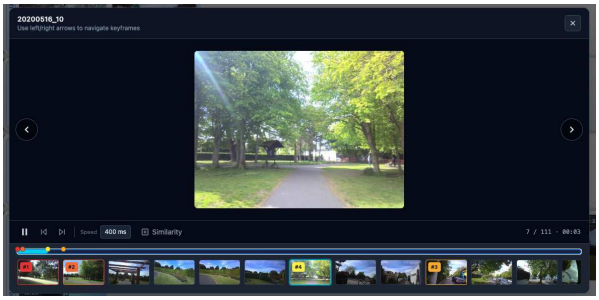
are tracked in a dedicated tab within the same side panel, allowing users to monitor their submission history and maintain awareness of previously selected results.

The interface is designed to support both exploratory search and task-oriented scenarios, such as those in the LSC. In particular, for the Q&A task, the system integrates an LLM-based chatbot that, given a user question, performs retrieval over the collection and analyzes the retrieved results to generate a direct answer. This component complements the interactive search workflow by allowing users to obtain immediate responses while still having access to the underlying visual evidence and retrieval results.

Finally, all user interactions are logged according to standard evaluation protocols, enabling reproducibility and offline analysis.

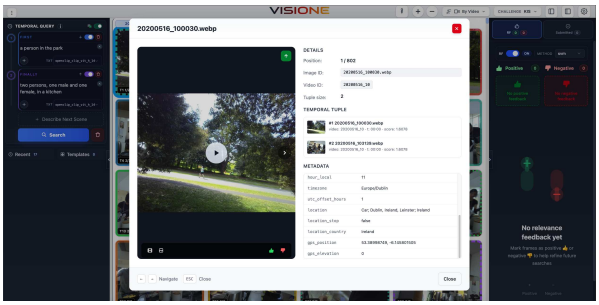
4 Performance at LSC26

The VISIONE system participated in LSC 2026 [24] with three independent users, denoted as VISIONE1, VISIONE2, and VISIONE3. The results show that the redesigned system was competitive, while

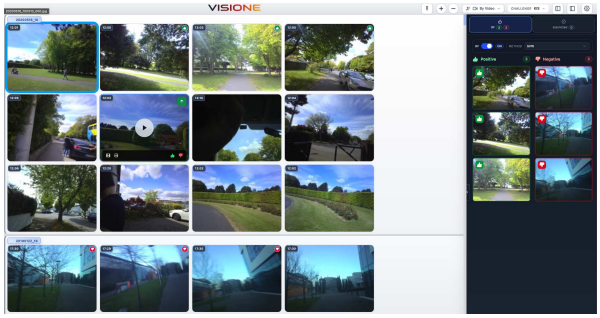


(a) Slideshow-based exploration of temporally related images.

(b) Summary view showing chronologically ordered images.



(c) Detailed modal view of a selected image.



(d) Search results with relevance feedback panel.

Figure 4: Views of the main components of the new VISIONE user interface.

Table 2: Overall individual ranking.

Rank	User	Score
1	MyEachtra	2714.4
2	**VISIONE2**	**2706.7**
3	**VISIONE3**	**2688.4**
4	ExploringLife2	2545.8
...		
10	**VISIONE1**	**2028.9**

Table 3: Ranking positions achieved by the VISIONE users in each task type.

User	AVS rank	KIS rank	QA rank
VISIONE1	13	19	10
VISIONE2	4	1	5
VISIONE3	2	3	6

also highlighting the role of the user in interactive retrieval scenarios. In the overall individual ranking (see Table 2), VISIONE2 and VISIONE3 achieved the 2nd and 3rd positions, respectively, with scores close to the top-ranked system, while VISIONE1 ranked 10th. This variability suggests that, as expected in interactive search settings, performance depends not only on the system capabilities but also on user strategy, speed, and familiarity with the interface. At task level, particularly strong results were obtained in the KIS tasks, where VISIONE2 ranked 1st overall and and VISIONE3 ranked 3rd. VISIONE3 also achieved 2nd place in AVS, while all three VISIONE users ranked within the top 10 for QA. Table 3 reports the rankings of all VISIONE users by task type.

Table 4 summarises the number of successfully solved tasks by user and task type. Overall, the three VISIONE users solved most of the attempted tasks, with an average of 16.3 solved tasks out of 21, corresponding to a success rate of 77.8%. The results also show some variability across users: VISIONE3 achieved the highest success rate, solving 18 tasks, followed by VISIONE2 with 17 and VISIONE1 with 14.

At task-type level, AVS obtained the highest average success rate, with 5.7 solved tasks out of 6. However, this result should be interpreted in light of the AVS task design, where multiple relevant submissions can be submitted and contribute to a positive score. KIS also showed strong results, with an average of 4.7 solved tasks out of 6, while QA achieved the lowest average success rate, suggesting that question answering remains the task type with the largest margin for improvement in future versions of the system.

5 Conclusions

In this paper, we presented the redesigned VISIONE system for lifelog retrieval, focusing on both architectural improvements and interaction design. The new version departs from earlier implementations by adopting a more modular backend and a simplified, yet flexible, user interface that better supports iterative and time-constrained search. In particular, we emphasized the role of sequence-based query formulation, the integration of multimodal

Table 4: Successfully solved tasks by user and task type.

User	Task type	Attempted	Solved	Success rate
VISIONE1	AVS	6	5	83.3%
	KIS	6	4	66.7%
	QA	9	5	55.6%
	All	21	14	66.7%
VISIONE2	AVS	6	6	100.0%
	KIS	6	5	83.3%
	QA	9	6	66.7%
	All	21	17	81.0%
VISIONE3	AVS	6	6	100.0%
	KIS	6	5	83.3%
	QA	9	7	77.8%
	All	21	18	85.7%
Average user	AVS	6	5.7	94.4%
	KIS	6	4.7	77.8%
	QA	9	6.0	66.7%
	All	21	16.3	77.8%

representations and LLM-based components, and the use of relevance feedback to guide the retrieval process. The interface maintains a balance between flexibility and usability by allowing advanced configurations while providing simplified interaction modes for less experienced users.

Future work will focus on extending the question answering functionalities, which are currently in a prototype stage. This will likely involve revisiting the interaction paradigm, for instance by introducing a chat-based interface to support more natural query formulation and exploration.

References

[1] Svelte Contributors. 2026. *Svelte.* https://svelte.dev/ Accessed: 2026-04-23.

[2] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).

[3] Giuseppe Amato, Paolo Bolettieri, Fabio Carrara, Franca Debole, Fabrizio Falchi, Claudio Gennaro, Lucia Vadicamo, and Claudio Vairo. 2021. The VISIONE video search system: exploiting off-the-shelf text search engines for large-scale video retrieval. *Journal of Imaging* 7, 5 (2021), 76.

[4] Giuseppe Amato, Paolo Bolettieri, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, Nicola Messina, Lucia Vadicamo, and Claudio Vairo. 2024. VISIONE 5.0: Enhanced User Interface and AI Models for VBS2024. In *International Conference on Multimedia Modeling*. Springer, 332–339.

[5] Giuseppe Amato, Paolo Bolettieri, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, Nicola Messina, Lucia Vadicamo, and Claudio Vairo. 2024. Will VISIONE Remain Competitive in Lifelog Image Search?. In *Proceedings of the 7th Annual ACM Workshop on the Lifelog Search Challenge* (Phuket, Thailand) *(LSC '24)*. Association for Computing Machinery, New York, NY, USA, 58–63. doi:10.1145/3643489.3661122

[6] Shuai Bai, Yuxuan Cai, Ruizhe Chen, Keqin Chen, Xionghui Chen, Zesen Cheng, Lianghao Deng, Wei Ding, Chang Gao, Chunjiang Ge, et al. 2025. Qwen3-vl technical report. *arXiv preprint arXiv:2511.21631* (2025).

[7] Han Fang, Pengfei Xiong, Luhui Xu, and Yu Chen. 2021. Clip2video: Mastering video-text retrieval via image clip. *arXiv preprint arXiv:2106.11097* (2021).

[8] Alexandre Gago, Bernardo Kaluza, and António J. R. Neves. 2025. MEMORIA: A Memory Enhancement and MOment RetrIeval Application at the LSC2025. In *Proceedings of the 8th Annual ACM Workshop on the Lifelog Search Challenge (LSC '25)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3729459.3748693

[9] Cathal Gurrin, Alan F Smeaton, and Aiden R Doherty. 2014. Lifelogging: Personal big data. *Foundations and Trends® in information retrieval* 8, 1 (2014), 1–125.

[10] Minh-Quan Ho-Le, Duy-Khang Ho, Van-Tu Ninh, Trong-Thuan Nguyen, Cathal Gurrin, and Minh-Triet Tran. 2025. SnapSeek 3.0: Interactive Lifelog Search System with Enhanced Scene Understanding. In *Proceedings of the 8th Annual ACM Workshop on the Lifelog Search Challenge (LSC '25)*. Association for Computing Machinery, New York, NY, USA, 7–14. doi:10.1145/3729459.3748697

[11] Gabriel Ilharco, Mitchell Wortsman, Ross Wightman, Cade Gordon, Nicholas Carlini, Rohan Taori, Achal Dave, Vaishaal Shankar, Hongseok Namkoong, John Miller, Hannaneh Hajishirzi, Ali Farhadi, and Ludwig Schmidt. 2021. *OpenCLIP.* doi:10.5281/zenodo.5143773

[12] Albert Q Jiang, A Sablayrolles, A Mensch, C Bamford, D Singh Chaplot, Ddl Casas, F Bressand, G Lengyel, G Lample, L Saulnier, et al. 2023. Mistral 7b. *arXiv preprint arXiv:2310.06825* 10 (2023), 3.

[13] LangChain Team. 2024. LangChain. https://github.com/langchain-ai/langchain. Accessed: 2026-04-23.

[14] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International conference on machine learning*. PMLR, 19730–19742.

[15] Mingxin Li, Yanzhao Zhang, Dingkun Long, Keqin Chen, Sibo Song, Shuai Bai, Zhibo Yang, Pengjun Xie, An Yang, Dayiheng Liu, et al. 2026. Qwen3-VL-Embedding and Qwen3-VL-Reranker: A Unified Framework for State-of-the-Art Multimodal Retrieval and Ranking. *arXiv preprint arXiv:2601.04720* (2026).

[16] Alexander H Liu, Kartik Khandelwal, Sandeep Subramanian, Victor Jouault, Abhinav Rastogi, Adrien Sadé, Alan Jeffares, Albert Jiang, Alexandre Cahill, Alexandre Gavaudan, et al. 2026. Ministral 3. *arXiv preprint arXiv:2601.08584* (2026).

[17] Nicola Messina, Matteo Stefanini, Marcella Cornia, Lorenzo Baraldi, Fabrizio Falchi, Giuseppe Amato, and Rita Cucchiara. 2022. ALADIN: Distilling Fine-grained Alignment Scores for Efficient Image-Text Matching and Retrieval. *arXiv preprint arXiv:2207.14757* (2022).

[18] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. 2023. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193* (2023).

[19] pgvector. 2024. *pgvector 0.7.0 Released!* https://www.postgresql.org/about/news/pgvector-070-released-2852/ PostgreSQL News.

[20] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 8748–8763.

[21] Luca Rossetto, Klaus Schoeffmann, Cathal Gurrin, Jakub Lokoč, and Werner Bailer. 2025. Results of the 2025 Video Browser Showdown. *arXiv preprint arXiv:2509.12000* (2025).

[22] Loris Sauter, Ralph Gasser, Heiko Schuldt, Abraham Bernstein, and Luca Rossetto. 2025. Performance Evaluation in Multimedia Retrieval. *ACM Trans. Multim. Comput. Commun. Appl.* 21, 1, 23:1–23:23. doi:10.1145/3678881

[23] Allie Tran, Werner Bailer, Duc-Tien Dang-Nguyen, Graham Healy, Steve Hodges, Björn Þór Jónsson, Luca Rossetto, Klaus Schoeffmann, Minh-Triet Tran, Lucia Vadicamo, et al. 2025. The State-of-the-Art in Lifelog Retrieval: A Review of Progress at the ACM Lifelog Search Challenge Workshop 2022-2024. *IEEE Access* 13 (2025), 216340–216363.

[24] Allie Tran, Werner Bailer, Duc-Tien Dang-Nguyen, Graham Healy, Steve Hodges, Björn Þór Jónsson, Wolfgang Hürst, Luca Rossetto, Klaus Schoeffmann, Minh-Triet Tran, Liting Zhou, and Cathal Gurrin. 2026. Introduction to the 9th Annual Lifelog Search Challenge, LSC'26. In *Proceedings of the 2026 International Conference on Multimedia Retrieval (ICMR '26)*. Association for Computing Machinery, New York, NY, USA, 2904–2905. doi:10.1145/3805622.3811234

[25] Quang-Linh Tran, Binh Nguyen, Gareth J F Jones, and Cathal Gurrin. 2025. MemoriEase 3.0: A RAG-Enhanced Conversational Lifelog Retrieval System at LSC'25. In *Proceedings of the 8th Annual ACM Workshop on the Lifelog Search Challenge (LSC '25)*. Association for Computing Machinery, New York, NY, USA, 52–57. doi:10.1145/3729459.3748689

[26] Lucia Vadicamo, Rahel Arnold, Werner Bailer, Fabio Carrara, Cathal Gurrin, Nico Hezel, Xinghan Li, Jakub Lokoc, Sebastian Lubos, Zhixin Ma, et al. 2024. Evaluating Performance and Trends in Interactive Video Retrieval: Insights from the 12th VBS Competition. *IEEE Access* (2024).

[27] Lucia Vadicamo, Francesca Scotti, Alan Dearle, and Richard Connor. 2025. Comparative analysis of relevance feedback techniques for image retrieval. In *International Conference on Multimedia Modeling*. Springer, 206–219.

[28] Michela Wilhelmsen, Linnea Sannes Rønning, Duc-Tuan Luu, Minh-Quan Ho-Le, Thanh-Hai Tran, Cathal Gurrin, Minh-Triet Tran, and Duc Tien Dang Nguyen. 2025. VitaChronicle 2.0: A Dual-Mode UX-Centered Approach to Lifelog Image Retrieval for Novice and Expert Users. In *Proceedings of the 8th Annual ACM Workshop on the Lifelog Search Challenge (LSC '25)*. Association for Computing Machinery, New York, NY, USA, 63–69. doi:10.1145/3729459.3748692