E-LifeSeeker: An Interactive Lifelog Search Engine for LSC'23

Thao-Nhu Nguyen*
Tu-Khiem Le*
Van-Tu Ninh*
Dublin City University
Ireland

Minh-Triet Tran
Nguyen Thanh Binh
University of Science
Vietnam National University
Ho Chi Minh City
Vietnam

Graham Healy
Sinéad Smyth
Annalina Caputo
Cathal Gurrin
Dublin City University
Ireland

ABSTRACT

Lifelogging is referred to as the process of automatically capturing the everyday activities of an individual to ultimately create a digital diary for further sharing, which can be challenging to manage and retrieve due to its multimodal nature. Lifelog retrieval systems not only have the potential to transform the way people interact and understand their lives, but also provide insights into their behaviour, habits, and preferences. The Lifelog Search Challenge (LSC) is a live benchmarking challenge to evaluate the performance of lifelog search tools in a real-time. This paper describes the modifications made to the E-LifeSeeker retrieval system, which participates in the 6th LSC challenge. This year, we enhance not only the underlying core engine with the latest pre-trained embedding models but also the user interface to be more intuitive for novice users. Moreover, Differential Networks are implemented to address the new question-answering task this year. These new modalities are designed to provide users with a more intuitive and efficient search experience, easing the process of locating information needed from the huge collection of egocentric lifelog images.

CCS CONCEPTS

• **Information systems → Multimedia databases**; **Users and interactive retrieval**; *Search interfaces*; • **Human-centered computing → Interactive systems and tools**.

KEYWORDS

lifelog, interactive retrieval, information system

ACM Reference Format:
Thao-Nhu Nguyen, Tu-Khiem Le, Van-Tu Ninh, Minh-Triet Tran, Nguyen Thanh Binh, Graham Healy, Sinéad Smyth, Annalina Caputo, and Cathal Gurrin. 2023. E-LifeSeeker: An Interactive Lifelog Search Engine for LSC'23. In *6th Annual ACM Lifelog Search Challenge (LSC '23), June 12–15, 2023, Thessaloniki, Greece.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3592573.3593098

*Three authors contributed equally to this research.

LSC '23, June 12–15, 2023, Thessaloniki, Greece

ACM ISBN 979-8-4007-0188-7/23/06.
https://doi.org/10.1145/3592573.3593098

1 INTRODUCTION

With the widespread use of digital devices in everyday life, it is getting easier and easier to capture data generated by the daily activities of an individual, leading to a potentially huge collection of data. Lifelog retrieval systems have emerged as a result of the difficulty in organising and accessing those data which aims to provide users a convenient way to interact and retrieve specific moments from the collection. To benchmark for different search engines in indexing and retrieving multimodal lifelog data, there have been a number of interactive lifelog retrieval challenges such as NTCIR Lifelog [7, 8], ImageCLEF Lifelog [4, 5, 20], and Lifelog Search Challenge (LSC) [9, 10]. Inspired by an annual popular video search challenge named Video Browser Showdown (VBS), the LSC challenge has been organised since 2018, with the intention of benchmarking different search methods on lifelog dataset specifically. This challenge has attracted participants from all over the world to compete in interactive tasks designed to simulate real-world lifelog retrieval. Each team is tasked with developing an interactive retrieval tool capable of answering specific questions from the provided data collection within a time constraint. Particularly, the organisers introduce three different types of tasks for this year's challenge including Known-item Search (KIS), Ad-hoc Search (ADS), and Question Answering Search (QAS). Sharing the same objective of finding images from a given query, the KIS task seeks specific keyframes as fast as possible, while the ADS task seeks as many relevant items as possible for a given query. Meanwhile, the QAS tasks, newly introduced this year, require users to give textual responses answering the information needs which will be manually judged in real-time. All tasks will be measured by the score combining between correctness and solving time of the team's submissions.

Our retrieval system, LifeSeeker [15], was originally developed as a concept-based searching tool that matches the input queries with the pre-defined indices. After participating in the LSC through three research iterations [15, 16, 19], we have continued improving both the underlying engine and interface in response to the user feedback. Besides, we also identified the lack of context of the search queries when matching keywords standalone leading to a major change to the semantic-based system in LifeSeeker 4.0[18].

In this paper, we introduce E-LifeSeeker, an enhanced version of our latest system [18], which maintains the semantic-driven approach to participate in the LSC'23 [11]. We integrate different state-of-the-art models for vision-language relations that were pre-trained on a huge dataset to further leverage the contextual meanings from visual scenes. Moreover, we will employ Differential Networks [28] to resolve the new coming task (QAS). Additionally,

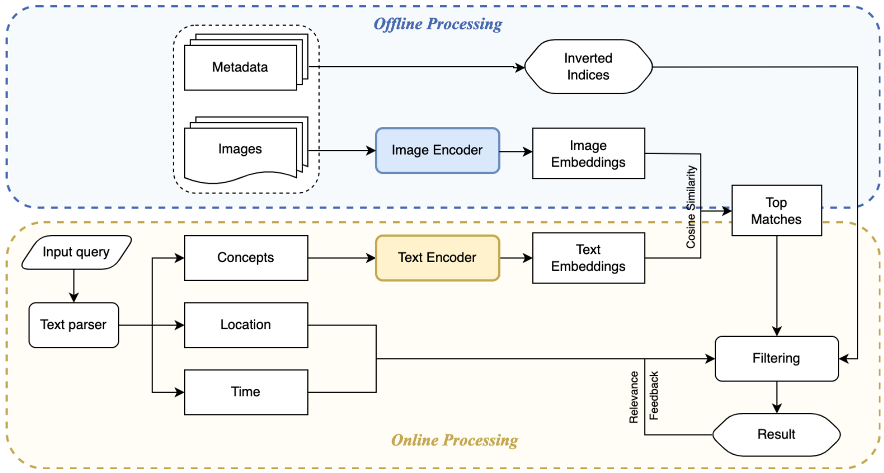

Figure 1: The System Architecture of E-LifeSeeker

the user interface (UI) was also redesigned to support easier user interaction, especially for novices.

2 RELATED RESEARCH

Participants have implemented different ranking models and search methods to solve the tasks posed in the live challenge. The most common mechanism that has been used during the first three years of the competition is concept-based search tool [3, 12, 19, 24, 27]. The key idea of this approach is to convert lifelogging images into a collection of visual annotations (including objects, text visible in the images or additional data) prior to matching them with the input keywords. While Vitrivr [13] provided multiple search modalities, including keywords, sketches, and audio, to facilitate the browsing process, Myscéal [27] incorporated users with the implementation of query expansion and expansion of location-based information in addition to text. Sharing the same core engine with Vitrivr [13], Vitrivr-VR [24] offered a unique user interaction experience by providing a virtual reality (VR) space for interactive retrieval. Meanwhile, LifeConcept [3] proposed a concept recommendation method by using ConceptNet alongside with a daily timeline view to aid the temporal navigation. Additionally, Lifegraph [22] leveraged a knowledge graph to match between text and vision, providing a more comprehensive understanding of the relationships between data for more accurate results. Another technique has been popular in constructing retrieval system is the vision-language-joint embedding approach. The popular pre-trained embedding model, CLIP [21] released in 2021 by open AI, has been adopted by many teams [1, 2, 18, 26]. Particularly, Memento [1] use ensemble models from two pre-trained Vision Transformers [6](ViT-L/14 and ResNet-50x64) as their backbone to search resulting in the better performance as compared to their previous version. Similarly, Voxento [2] provided a voice command experience while leveraging the same underlying model. Nevertheless, LifeSeeker [18] and Myscéal [26] also switched into semantic-driven systems as they made an effort to gain more insights into not only the visual content but the semantic context.

As we continue to employ semantic-based models on our new system E-LifeSeeker, we aim to investigate different models on encoding the visual content. Additionally, we facilitate the search process by enhancing the UI with additional functionalities including ranked list clustering, temporal search, and relevance feedback.

3 OVERVIEW OF E-LIFESEEKER

The E-LifeSeeker system has been developed specifically to provide fast and effective retrieval facilities over the LSC dataset from the challenge in 2023.

3.1 LSC dataset

The dataset used for this year's competition is almost identical to the Lifelog dataset[1] in LSC'22, which includes the personal data of one active lifelogger for the period of 18 months (527 days). The multimodal dataset comprises the following three components:

- **Egocentric images**: 725,226 images captured by the Narrative Clip device[2] with a resolution of 1024 x 768. To protect the privacy of both the lifelogger and people's surroundings, these images are fully anonymised with blurred faces and censored text;
- **Metadata**: descriptive, spatial, and temporal information corresponding with the images;
- **Visual Concepts**: visual objects, scenes, captions, and text are extracted from the non-redacted version of the visual dataset.

In 2023, some additional location annotations have been provided by the Myscéal team.

3.2 E-LifeSeeker System Architecture

Similar to other retrieval systems [25], the system architecture of E-LifeSeeker consists of both offline and online stages, as illustrated in Figure 1. In general, the offline (indexing) stage is responsible for creating a searchable database of features from the dataset. Initially,

[1] http://lifelogsearch.org/lsc/lsc_data/
[2] http://getnarrative.com/

while provided metadata (including spatial, temporal, and other information) was indexed into inverted files, the egocentric images were transferred to high-dimensional representations for further matching in the next stage. In the online (searching) stage, the system processes user queries and performs similarity matching on the pre-processed database in real-time, returning the most relevant results measured by the cosine similarity score. The higher the score, the more similar the content is likely to be. Furthermore, users have options for filters adding more information related to time, location, objects, or text visible in the images (like part of the day, at work, etc.) to the current query.

3.3 Embedding Model Selection for Semantic Search

Semantic-based models have gained popularity across different research areas due to their efficiency and versatility. Their applications in information retrieval, in general, and lifelog search in particular, have shown a significant increase in recent years. From the last competition and our previous user study [18], we observed that semantic-based models, OpenAI CLIP models with a Vision Transformer [6] pre-trained at 336-pixel resolution (ViT-L/14@336p), outperformed the conventional concept-matching method in terms of both accuracy and searching time. Building on this success, we aim to enhance the underlying models used in LifeSeeker 4.0 by leveraging larger scale and more powerful pre-trained models such as CoCa [29], CLIP [21] pre-trained on LAION-5B dataset [23], BLIP [17], and ALIGN [14]. These state-of-the-art models, trained on massive amounts of data, can have the potential to capture more complex and abstract visual-textual relations, making them feasibly better suited for lifelog data. The E-LifeSeeker system is designed to integrate all these models, and it is possible to select any of them as the core embedding model for the search engine in the LSC'23 competition.

3.4 Question Answering task

The Question-Answering task is a notable new task in this year's LSC challenge. Unlike the known-item search tasks in previous years, where users are asked to locate target images for a provided description, the question-answering task in 2023 requires the submission of a textual answer to a set of previously unprecedented questions concerning the given personal collection, such as questions about quantities, activities, or durations. It is a real-world scenario, yet a challenging task to solve since the queries will be highly personalised and may require the searcher to quickly develop a deep understanding of the user's individual experiences and activities. E-LifeSeeker integrates latest embedding models for retrieval and utilises differential networks to support Q&A queries. Firstly, we input the initial query and get the ranked list by taking advantage of the above embedding models. Once the system retrieves a set of relevant images, users can use the Differential Networks based Fusion model [28], a plug-and-play module that enables differences between pairwise feature elements, or involves judging which image contains the information necessary to answer the question. For example, the task "*What did I have for breakfast on Christmas Day 2019*" cannot be answered by submitting a single image, it requires users to navigate throughout that event and then specify the objects or information needed from those images.

3.5 User Interface Improvements

In our efforts to offer users a better search experience, we modify our UI to meet the demands of both expert and novice users, with the rearrangement depicted in Figure 2. The improved UI is composed of four main components: the free-text search and filter box (**A**), the automatic question generation display (**B**), the search progress bar (**C**), and vertically-scrollable panel displaying the retrieved result in groups (**D**). We recognise that the intuitive interface is critical for enhancing overall system performance as well as user experience. Those modifications include ranked list clustering, user relevance feedback, and temporal search.

Ranked List Clustering. Similarly to almost all conventional lifelog retrieval systems, our previous UI [18] displayed results as all keyframes individually ranked by their relevance score for the input query. However, this approach made it difficult for users to navigate to find the desired moments when a sequence of duplicated images happened in one event. For example, when querying for "the lifelogger was watching the TV", it would return hundreds of images of this activity, many of which are temporally sequential. To overcome this problem, we introduce a new way of displaying the results (**D**) by clustering them based on temporal features. They will be clustered together and displayed in the top-3 with the highest score, with the option to view all images in the group if desired. To be more precise, the re-ranking algorithm has the execution steps below:

(1) Get the ranked list when entering a new query,
(2) Convert the ranked list into sub-lists by grouping images in the same part of the day (each day will be divided into 3 parts: morning, afternoon, and evening),
(3) Calculate the average confidence score of the top-3 highest ranked images in each sub-list,
(4) Re-rank sub-lists based on their new average confidence score scores.

Temporal Search. This functionality provides users with the option to navigate through the selected moment and its temporally-related images by adjusting the temporal range between them. By allowing users to explore these data, they will get a better understanding of the activities and context surrounding that event or during a day.

Relevance Feedback. To better support users when interacting with the system, we incorporate relevance feedback (**B**), where users are enabled to provide more information to adjust the current result. In particular, after inputting a query, the system will automatically generate a question related to the visual content of their desired images, and users have the option to choose to include or eliminate those. By doing so, we are able to improve the ranking algorithm and ensure that more relevant and less irrelevant images will be displayed in future searches. In addition, a progress bar (**C**) can also be very helpful for users, especially when performing complex search queries that involve multiple steps. By showing the results of the current step and allowing users to go back to any previous step, they can easily correct the queries or modify their search criteria in case if necessary, further enhancing the usability of the system.

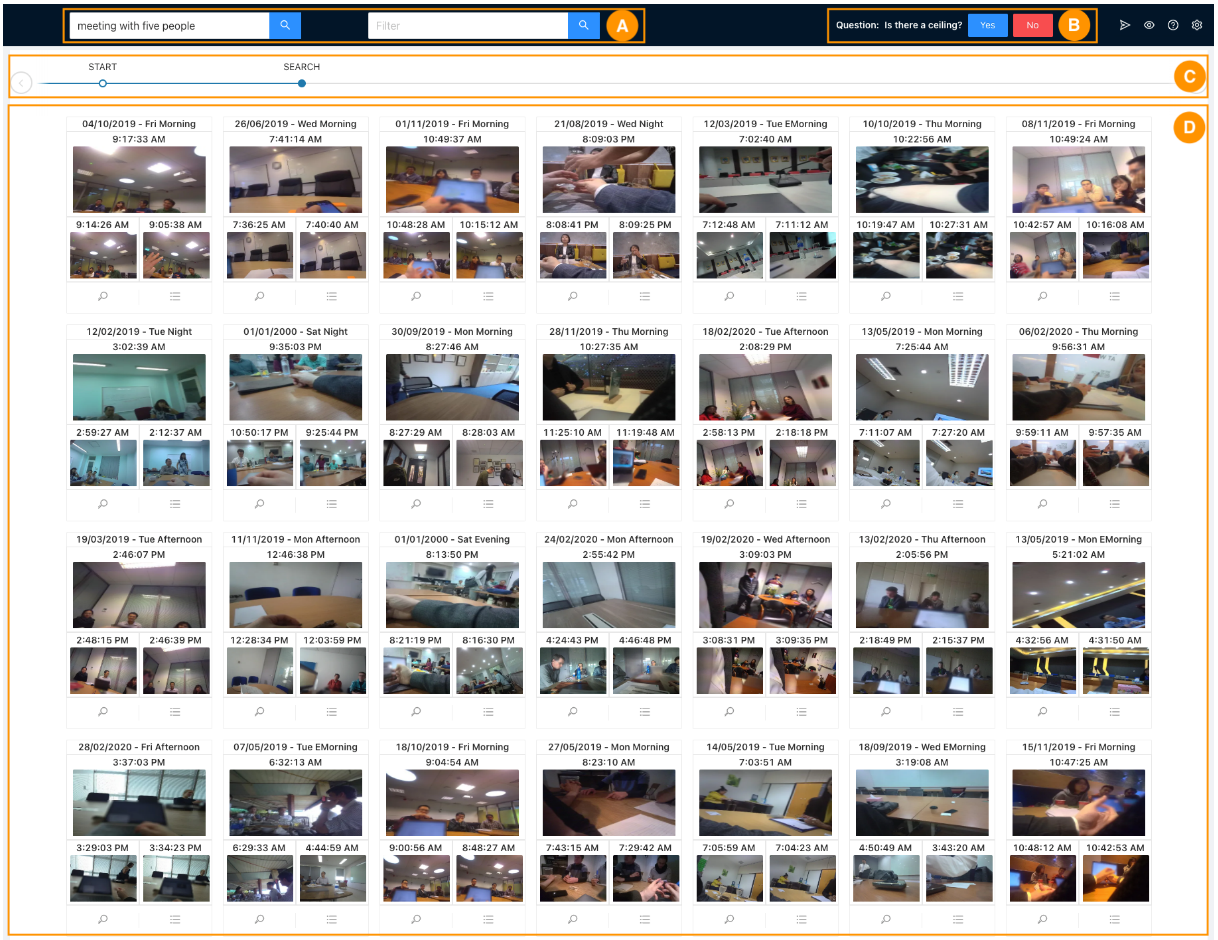

Figure 2: The User Interface of E-LifeSeeker. The screen displayed the result of an example query "meeting with five people".

4 CONCLUSION

This paper presents the improvements made to our retrieval system, E-LifeSeeker, for the 6th Lifelog Search Challenge. The image representations were enhanced by incorporating state-of-the-art embedding models, while the newly released task (QAS) was resolved by the use of Differential Networks. We further made improvements in terms of the UI with better visualisation and additional functionalities in order to make the user experience more efficient. With these modifications, we hope that our system will achieve competitive results, as it did in previous challenges.

ACKNOWLEDGMENTS

This publication has emanated from research supported in part by research grants from Irish Research Council (IRC) under Grant Number GOIPG/2016/741, Science Foundation Ireland under grant numbers SFI/12/RC/2289_P2, 18/CRT/6223, and SFI/13/RC/2106_P2 and co-funded by the European Regional Development Fund.

REFERENCES

[1] Naushad Alam, Yvette Graham, and Cathal Gurrin. 2022. Memento 2.0: An Improved Lifelog Search Engine for LSC'22. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 2–7. https://doi.org/10.1145/3512729.3533006

[2] Ahmed Alateeq, Mark Roantree, and Cathal Gurrin. 2022. Voxento 3.0: A Prototype Voice-Controlled Interactive Search Engine for Lifelog. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 43–47. https://doi.org/10.1145/3512729.3533009

[3] Wei-Hong Ang, An-Zi Yen, Tai-Te Chu, Hen-Hsen Huang, and Hsin-Hsi Chen. 2021. LifeConcept: An Interactive Approach for Multimodal Lifelog Retrieval through Concept Recommendation. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 47–51. https://doi.org/10.1145/3463948.3469070

[4] Duc-Tien Dang-Nguyen, Luca Piras, Michael Riegler, Liting Zhou, Mathias Lux, and Cathal Gurrin. 2018. Overview of ImageCLEFlifelog 2018: Daily Living Understanding and Lifelog Moment Retrieval.. In *CLEF (working notes)*.

[5] Duc Tien Dang Nguyen, Luca Piras, Michael Riegler, Liting Zhou, Mathias Lux, Minh Triet Tran, Tu-Khiem Le, Van-Tu Ninh, and Cathal Gurrin. 2019. Overview of ImageCLEFlifelog 2019: solve my life puzzle and lifelog moment retrieval. CEUR Workshop Proceedings.

[6] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. https://doi.org/10.48550/ARXIV.2010.11929

[7] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, and Rami Albatal. 2016. NTCIR Lifelog: The First Test Collection for Lifelog Research. 705–708. https://doi.org/10.1145/2911451.2914680

[8] Cathal Gurrin, H. Joho, Frank Hopfgartner, Liting Zhou, Tu Ninh, Tu-Khiem Le, Rami Albatal, Duc-Tien Dang-Nguyen, and Graham Healy. 2019. Overview of the NTCIR-14 Lifelog-3 task.

[9] Cathal Gurrin, Björn Þór Jónsson, Klaus Schöffmann, Duc-Tien Dang-Nguyen, Jakub Lokoč, Minh-Triet Tran, Wolfgang Hürst, Luca Rossetto, and Graham Healy. 2021. Introduction to the Fourth Annual Lifelog Search Challenge, LSC'21. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3460426.3470945

[10] Cathal Gurrin, Tu-Khiem Le, Van-Tu Ninh, Duc-Tien Dang-Nguyen, Björn Þór Jónsson, Jakub Lokoč, Wolfgang Hürst, Minh-Triet Tran, and Klaus Schöffmann. 2020. *Introduction to the Third Annual Lifelog Search Challenge (LSC'20)*. Association for Computing Machinery, New York, NY, USA, 584–585. https://doi.org/10.1145/3372278.3388043

[11] Cathal Gurrin, Björn Þór Jónsson, Duc Tien Dang Nguyen, Graham Healy, Jakub Lokoc, Liting Zhou, Luca Rossetto, Minh-Triet Tran, Wolfgang Hürst, Werner Bailer, and Klaus Schoeffmann. 2023. Introduction to the Sixth Annual Lifelog Search Challenge, LSC'23. In *Proceedings of the 2023 International Conference on Multimedia Retrieval* (Thessaloniki, Greece) *(ICMR '23)*. New York, NY, USA.

[12] Silvan Heller, Ralph Gasser, Mahnaz Parian-Scherb, Sanja Popovic, Luca Rossetto, Loris Sauter, Florian Spiess, and Heiko Schuldt. 2021. Interactive Multimodal Lifelog Retrieval with Vitrivr at LSC 2021. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 35–39. https://doi.org/10.1145/3463948.3469062

[13] Silvan Heller, Ralph Gasser, Mahnaz Parian-Scherb, Sanja Popovic, Luca Rossetto, Loris Sauter, Florian Spiess, and Heiko Schuldt. 2021. Interactive Multimodal Lifelog Retrieval with Vitrivr at LSC 2021. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 35–39. https://doi.org/10.1145/3463948.3469062

[14] Chao Jia, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc V. Le, Yun-Hsuan Sung, Zhen Li, and Tom Duerig. 2021. Scaling Up Visual and Vision-Language Representation Learning With Noisy Text Supervision. *CoRR* abs/2102.05918 (2021). arXiv:2102.05918 https://arxiv.org/abs/2102.05918

[15] Tu-Khiem Le, Van-Tu Ninh, Duc-Tien Dang-Nguyen, Minh-Triet Tran, Liting Zhou, Pablo Redondo, Sinead Smyth, and Cathal Gurrin. 2019. LifeSeeker: Interactive Lifelog Search Engine at LSC 2019. In *Proceedings of the ACM Workshop on Lifelog Search Challenge* (Ottawa ON, Canada) *(LSC '19)*. Association for Computing Machinery, New York, NY, USA, 37–40. https://doi.org/10.1145/3326460.3329162

[16] Tu-Khiem Le, Van-Tu Ninh, Minh-Triet Tran, Thanh-An Nguyen, Hai-Dang Nguyen, Liting Zhou, Graham Healy, and Cathal Gurrin. 2020. *LifeSeeker 2.0: Interactive Lifelog Search Engine at LSC 2020*. Association for Computing Machinery, New York, NY, USA, 57–62. https://doi.org/10.1145/3379172.3391724

[17] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. 2022. BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. In *ICML*.

[18] Thao-Nhu Nguyen, Tu-Khiem Le, Van-Tu Ninh, Minh-Triet Tran, Thanh Binh Nguyen, Graham Healy, Sinéad Smyth, Annalina Caputo, and Cathal Gurrin. 2022. LifeSeeker 4.0: An Interactive Lifelog Search Engine for LSC'22. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 14–19. https://doi.org/10.1145/3512729.3533014

[19] Thao-Nhu Nguyen, Tu-Khiem Le, Van-Tu Ninh, Minh-Triet Tran, Nguyen Thanh Binh, Graham Healy, Annalina Caputo, and Cathal Gurrin. 2021. LifeSeeker 3.0: An Interactive Lifelog Search Engine for LSC'21. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 41–46. https://doi.org/10.1145/3463948.3469065

[20] Van-Tu Ninh, Tu-Khiem Le, Liting Zhou, Luca Piras, Michael Alexander Riegler, Pål Halvorsen, Mathias Lux, Minh-Triet Tran, Cathal Gurrin, and Duc Tien Dang Nguyen. 2020. Overview of imageclef lifelog 2020: lifelog moment retrieval and sport performance lifelog. (2020).

[21] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. *CoRR* abs/2103.00020 (2021). arXiv:2103.00020 https://arxiv.org/abs/2103.00020

[22] Luca Rossetto, Matthias Baumgartner, Ralph Gasser, Lucien Heitz, Ruijie Wang, and Abraham Bernstein. 2021. Exploring Graph-Querying Approaches in LifeGraph. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 7–10. https://doi.org/10.1145/3463948.3469068

[23] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade W Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, Patrick Schramowski, Srivatsa R Kundurthy, Katherine Crowson, Ludwig Schmidt, Robert Kaczmarczyk, and Jenia Jitsev. 2022. LAION-5B: An open large-scale dataset for training next generation image-text models. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*. https://openreview.net/forum?id=M3Y74vmsMcY

[24] Florian Spiess and Heiko Schuldt. 2022. Multimodal Interactive Lifelog Retrieval with Vitrivr-VR. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 38–42. https://doi.org/10.1145/3512729.3533008

[25] Ly-Duyen Tran, Manh-Duy Nguyen, Duc-Tien Dang-Nguyen, Silvan Heller, Florian Spiess, Jakub Lokoč, Ladislav Peška, Thao-Nhu Nguyen, Omar Shahbaz Khan, Aaron Duane, Björn Þór Jónsson, Luca Rossetto, An-Zi Yen, Ahmed Alateeq, Naushad Alam, Minh-Triet Tran, Graham Healy, Klaus Schoeffmann, and Cathal Gurrin. 2023. Comparing Interactive Retrieval Approaches at the Lifelog Search Challenge 2021. *IEEE Access* (2023), 1–1. https://doi.org/10.1109/ACCESS.2023.3248284

[26] Ly-Duyen Tran, Manh-Duy Nguyen, Binh Nguyen, Hyowon Lee, Liting Zhou, and Cathal Gurrin. 2022. E-Myscéal: Embedding-Based Interactive Lifelog Retrieval System for LSC'22. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 32–37. https://doi.org/10.1145/3512729.3533012

[27] Ly-Duyen Tran, Manh-Duy Nguyen, Nguyen Thanh Binh, Hyowon Lee, and Cathal Gurrin. 2021. Myscéal 2.0: A Revised Experimental Interactive Lifelog Retrieval System for LSC'21. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 11–16. https://doi.org/10.1145/3463948.3469064

[28] Chenfei Wu, Jinlai Liu, Xiaojie Wang, and Ruifan Li. 2019. Differential Networks for Visual Question Answering. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence* (Honolulu, Hawaii, USA) *(AAAI'19/IAAI'19/EAAI'19)*. AAAI Press, Article 1104, 8 pages. https://doi.org/10.1609/aaai.v33i01.33018997

[29] Jiahui Yu, Zirui Wang, Vijay Vasudevan, Legg Yeung, Mojtaba Seyedhosseini, and Yonghui Wu. 2022. CoCa: Contrastive Captioners are Image-Text Foundation Models.