MemoriEase: An Interactive Lifelog Retrieval System for LSC'23

Quang-Linh Tran
linh.tran3@mail.dcu.ie
Dublin City University
Dublin, Ireland

Ly-Duyen Tran
ly.tran2@mail.dcu.ie
Dublin City University
Dublin, Ireland

Binh Nguyen
ngtbinh@hcmus.edu.vn
VNU-HCM University of Science
Ho Chi Minh City, Vietnam

Cathal Gurrin
cathal.gurrin@dcu.ie
Dublin City University
Dublin, Ireland

ABSTRACT

Lifelogging is an activity of recording all events that happen in the daily life of an individual. The events can contain images, audio, health index, etc which are collected through various devices such as wearable cameras, smartwatches, and other digital services. Exploiting lifelog data can bring significant benefits for lifeloggers from creating personalized healthcare plans to retrieving events in the past. In recent years, there has been a growing development of interactive lifelog retrieval systems, such as competitors at the annual Lifelog Search Challenge (LSC), to assist lifeloggers in finding events from the past. This paper introduces an interactive lifelog image retrieval called MemoriEase for the LSC'23 challenge. This system combines concept-based and embedding-based retrieval approaches to answer accurate images for LSC'23 queries. This system uses BLIP for the embedding-based retrieval approach to reduce the semantic gap between images and text queries. The concept-based retrieval approach uses full-text search in Elasticsearch to retrieve images having visual concepts similar to keywords in the query. Regarding the user interface, we make it as simple as possible to make novices users can use it with only a small effort. This is the first version of MemoriEase and we expect this can help users perform well in the LSC'23 competition.

CCS CONCEPTS

• **Information systems**; • **Information retrieval**; • **Human-centered computing**; • **Interactive systems**;

KEYWORDS

lifelog, interactive retrieval system, unified vision-language representation

ACM Reference Format:
Quang-Linh Tran, Ly-Duyen Tran, Binh Nguyen, and Cathal Gurrin. 2023. MemoriEase: An Interactive Lifelog Retrieval System for LSC'23. In *6th Annual ACM Lifelog Search Challenge (LSC '23), June 12–15, 2023, Thessaloniki, Greece.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3592573.3593101

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

ACM ISBN 979-8-4007-0188-7/23/06.
https://doi.org/10.1145/3592573.3593101

1 INTRODUCTION

The concept of Lifelogging involves collecting and storing data on all aspects of an individual's daily life, including actions through images and videos, sleep data, and health metrics. This data can be utilized for various beneficial purposes such as tracking one's health, aiding memory recall, and saving information as evidence when necessary [4, 10]. The idea of using lifelog data dates back to 1945 when Vannevar Bush proposed Memex [3] as a way to externalize human memories. However, it was not until the advent of the development of the internet and data capturing and storage devices such as wearable cameras, smartwatches, etc. These devices help to capture daily life data easily, store it efficiently and make lifelog gain popularity.

With the development of search engines such as Google, Bing, etc, which have revolutionized information retrieval, there is now a growing need for a search engine specifically designed for lifelog data. Such an engine would be helpful in assisting individuals in retrieving past moments, finding lost items, or identifying individuals encountered in daily life. Lifelog Search Challenge (LSC) was established to provide a benchmark evaluation for lifelog retrieval systems and a playground for researchers and developers to show their solutions. The first LSC was held in 2018[9] and we will participate in the 6th iteration in 2023.

In this paper, we explore the ability of state-of-the-art neural network models, such as CLIP[23] and BLIP[19], to bridge the gap between text and images. These models map images and texts to the same embedding space and compare their similarity to address various tasks such as image captioning, image-text retrieval, and text-image retrieval. They achieve excellent performance demonstrated on benchmark datasets, such as MS-COCO and Flickr. To utilize their power, we employ an embedding-based retrieval approach with BLIP as the main search engine in our retrieval system. To further improve accuracy, a concept-based retrieval approach is also employed to support the embedding-based approach.

In this paper, we introduce a new interactive lifelog image retrieval system, MemoriEase, designed for participation in the LSC'23 competition [12]. MemoriEase employs a combination of embedding-based and concept-based retrieval approaches, implemented through a search engine on Elastic Search. The system also uses Edge detection to eliminate blurred and meaningless images. Additionally, the user interface of MemoriEase is designed to be simple and user-friendly, providing expert and novice users with sufficient information to retrieve images. The integration of these features is

expected to result in enhanced performance, with low latency and high accuracy, for MemoriEase in the LSC'23 competition.

2 RELATED WORKS

The field of information retrieval in general and image-text retrieval in particular have received significant attention from researchers thanks to their broad range of applications. There are various existing methods for retrieving images based on a textual query. One such approach proposed by Duyen et al. [25, 27] involves the extraction of visual concepts from images and calculating the similarity of visual concepts and keywords. Their approach employs a variation of the IF-IDF method to consider the region of visual concepts. Recently, the development of multi-modal models such as CLIP [23] and BLIP [18, 19] has helped to bridge the gap between text and images, and lots of image retrieval systems [1, 21] have adopted these models as the central component of their system.

There are various challenges and workshops have been organized in the field of lifelog image retrieval [5, 6, 22]. However, the LSC [7, 8, 11] is the pioneer in creating an interactive benchmark evaluation in lifelong search and it has attracted a significant number of researchers in building lifelog image retrieval systems. In the LSC'22 [11], nine teams participated. E-Myscéal [26] continued winning the competition by implementing an embedding-based approach that leverages CLIP. They used Elasticsearch as the search engine to retrieve images based on the embedding of images and enhanced text query. LifeSeeker 4.0 [21] also used CLIP to extract embedding the lifelog images and query. In addition, they enhanced their system by employing music metadata and event clustering techniques to enhance the user experience. Momento 2.0 [1] combined two CLIP models (ViT-L/14 and ResNet-50x64) to calculate the similarity score between images and queries, resulting in a significant improvement over their previous system. Voxento 3.0 [2] introduced a voice search system that used CLIP as an embedding extractor, but their main advantage was using voice recognition to input queries, which was faster than texting. Both vitrivr [14] and vitrivr-VR [24] used Cottontail DB as the database and Cineast as the retrieval engine. However, vitrivr used vitrivr-ng for the front-end which is "responsible for query formulation, result presentation, browsing, and filtering. Users can combine various query modalities, for example, first filtering by time and then making a color sketch" [13], while vitrivr-VR used a Virtual Reality (VR) interface for browsing and searching lifelog images. Memoria, a new competitor in LSC22, used a concept-based approach instead of an embedding-based approach like other competitors. They utilized several computer vision methods to process visual lifelogs. Finally, LifeXplore [17] presented an enhanced version of their previous system [16] with improvements to the system's interface.

In the field of lifelog image retrieval, the dominant approach is currently the embedding-based method, with nearly half of the LSC competitors [1, 2, 21, 26] using the CLIP model to extract visual and textual embeddings. However, the performance of CLIP has been surpassed by BLIP recently [18, 19], so we will use this model to extract feature embeddings in our system. Several databases have been used to store and query data, including MongoDB and Cottontail DB, but Elasticsearch is the most favored due to its efficient storing and searching engine. In addition, Elasticsearch has integrated vector search, which is useful for combining full-text search and vector search for embedding-based and concept-based approaches. We also learned from previous competitors of LSC to integrate event segmentation, blurred image removal, and provide an intuitive user interface into our system.

3 LSC'23 DATASET

3.1 Dataset Description

The LSC'23 dataset [12] comprises approximately 725K fully redacted lifelog images (1024 x 768 resolution) captured by a Narrative Clip camera worn by a lifelogger from January 2019 to June 2020. In addition to the visual data, the organizers of LSC'23 have also provided metadata relating to temporal, location, semantic, and music-related features, as well as biometric indices associated with the images. Tags, OCR, and captions are also provided as visual concepts. For the purpose of retrieving images through query only, we eliminate the health and music-related data to simplify the system and enhance querying time, as in Tran et. al [25].

There are three main types of topics in the LSC'23, namely known-item, ad-hoc, and question-answering topics. The known-item topic is the most straightforward topic, where the query asks for a specific moment in the lifelog dataset and users only need to provide images that illustrate the moment. Meanwhile, the ad-hoc topic contains queries asking about general moments in the dataset. This means that there are several images at different times that can be the answer to this topic. The last topic is question answering, in which users need to answer a textual answer instead of images like the previous topics. This topic asks for information such as the brand of a car or the number of a room in the lifelog dataset. This can be seen as the most difficult topic because it requires users to highly interact with the system to find the answers.

3.2 Challenges

During our investigation, we encounter numerous challenges with respect to image quality and missing metadata in the dataset. Because this is a lifelog dataset, the images are not taken with the intent to capture with high-quality but rather are automatically taken, sometimes resulting in blurry or meaningless images. To assess the degree of detail of the images, we have implemented an edge detection method. Images that are blurred or obstructed by hands or clothing typically have few edges, resulting in a small sum of edge weights. Our findings indicate that a considerable number of images have a sum of edge weights below 350,000, as illustrated in Figure 2, thereby indicating that they are likely to be blurred.

Another obstacle that we have faced is missing data in the metadata. Some important attributes for retrieving such as latitude, longitude, time zone, and semantic name have a missing rate of nearly 67%, posing difficulty in retrieving images with exact time and location. To address these problems, we propose some methods which are illustrated in section 4.1.

4 MEMORIEASE SYSTEM

This section provides general information about the MemoriEase system. The system utilizes ElasticSearch as the search engine to retrieve images and the BLIP model as the main feature extractor to obtain textual and visual embedding. The process of extracting visual

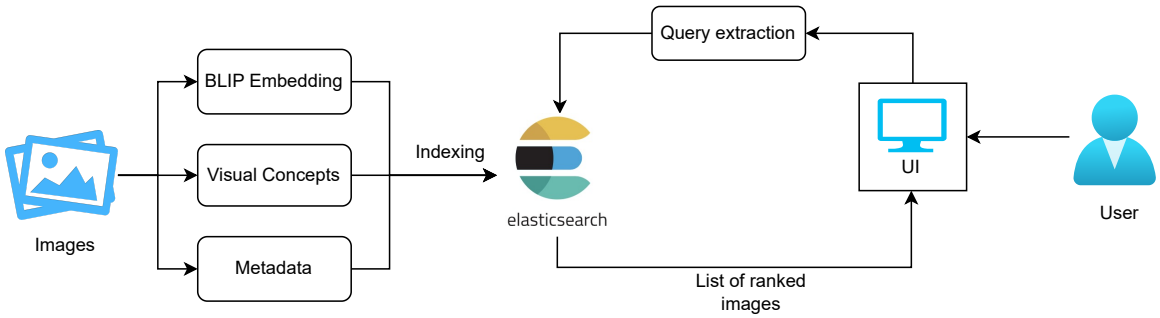

Figure 1: MemoriEase Overview

embedding involves removing blurred images and grouping meaningful images into segments, which are then passed through the BLIP model. The metadata is imputed and cleaned before combining with visual concepts and embedding to index to the ElasticSeach. The details of this data processing and indexing stage are presented in section 4.1. In the user interface (UI), when a user performs a search, the query is processed and decomposed to time and location for filtering, visual concepts, and embedding for calculating similarity. We combine both concept-based and embedding-based retrieval approaches in the system and propose a scoring mechanism to balance the relevance score from the two approaches. Section 4.3 provides a detailed explanation of this stage. An overview of the MemoriEase system is shown in figure 1.

4.1 Data processing and indexing

As mentioned in section 3.2 previously, the LSC'23 dataset contains a significant number of blurred and meaningless images. To address this issue, we have implemented a method to remove such images by filtering out those with a summation of edge weights below a certain threshold. We found that a threshold of 350,000 results in the removal of a large proportion of blurred or covered images. Specifically, we removed a total of 94,450 images, which accounts for nearly 13% of the original dataset. This approach helps to significantly reduce the storage requirements for the system.

Regarding the issue of missing metadata, we have implemented an imputation strategy to fill in missing latitude and longitude values in the dataset. Specifically, we leverage GPS trajectories and images captured from wearable cameras to infer missing values between two existing values based on image activity, inspired by the work in [15, 28]. We then use the location information to calculate the timezone and impute local time accordingly. We also extract time-related information such as time of day (morning, afternoon, etc), and day of the week (Monday, Tuesday, etc) to enhance the metadata.

We use the BLIP model to extract embeddings from images and cluster them into segments using the segmentation algorithm described in algorithm 1. This results in a total of 173,269 main events, which are then indexed in ElasticSearch for later retrieval. The indexing includes 17 attributes, including tags, captions, and OCR for visual concept search, as well as time and location-related attributes for filtering, and a BLIP embedding attribute for embedding-based retrieval.

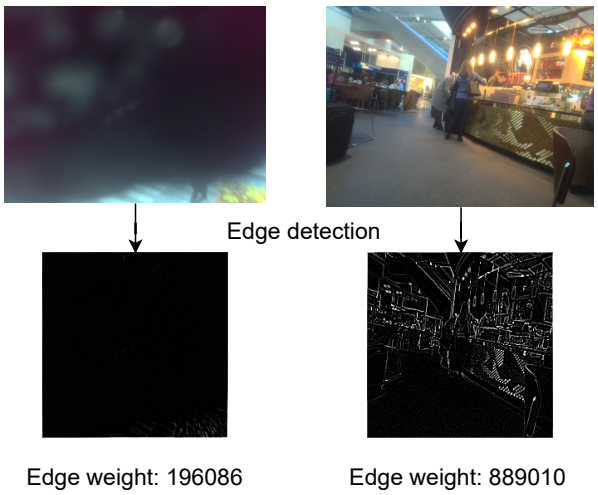

Figure 2: Examples of edge detection.

Algorithm 1 The event segmentation algorithm

```
I ← image embeddings ordered by time
E ← events list
id ← event id in the list E
α ← threshold
for i in I do
    if Cosine(i − 1, i) < α then
        if Cosine(i, i + 1) > α then
            E.append({i : id})
        else
            id ← id + 1
            E.append({i : id})
        end if
    else
        E.append({i : id})
    end if
end for
```

4.2 Visual and Textual Embedding by BLIP

Thanks to the zero-shot capacity of BLIP [19], it shows robustness in several tasks without the need for retraining or fine-tuning. This

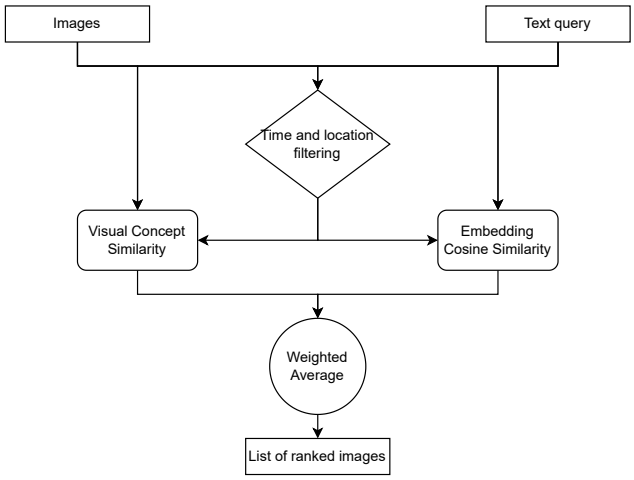

Figure 3: ElasticSearch retrieving process

advantage combined with the model's state-of-the-art performance in image retrieval made it a good choice as the embedding extractor for our system.

To extract image embedding, we use a pre-trained BLIP model provided by the authors, which was trained on the COCO dataset [20] for retrieval tasks. We use the base version of the model, rather than the larger version, due to resource constraints. For image embedding extraction, we need to resize the image from 768x768 to fit the model input and we receive an array of 256 dimensions.

The text query is padded to the max length of 60 because the maximum length of the previous competition is smaller than 60 words. After processing and padding, the query is converted into an array of 256 dimensions, similar to the image embeddings, to allow for cosine similarity calculations.

4.3 Scoring Mechanism

In our system, we adopt two approaches to retrieve images and propose a scoring mechanism to balance their advantages and disadvantages. Figure 3 depicts the process for generating a list of candidates, which are ordered based on the final score. Specifically, the images and text query are processed to extract time, location, visual concepts, and embedding information. We use time and location information to filter out irrelevant images before the similarity computation begins. Subsequently, the embedding of the filtered images and the query are used to calculate the cosine similarity score.

For the concept-based approach, we use the default BM25 algorithm of ElasticSearch[1] for calculating the relevance score of visual concepts in the query and in the metadata.

However, we also apply a normalization stage to reduce the range of BM25 score from 0 to 1, as the range for Cosine similarity for the Embedding-based approach. We apply the sigmoid function to normalize the BM25 score if the score is higher than 0.5, else we keep the original BM25 score. Then we use the weighted average to get the final relevance score between the query and the images.

[1]https://www.elastic.co/blog/practical-bm25-part-2-the-bm25-algorithm-and-its-variables

4.4 User Interface

To make MemoriEase accessible to novice users, the user interface is designed with only two main parts, query, and results, as illustrated in Figure 4. The query panel can be expanded or shrunk by clicking on Show more or Show less, respectively, and it supports two types of search: moment search and temporal search. For the moment search, users can enter the query in the Find box and optionally filter by time or location to retrieve relevant images. For temporal search, users can search for images before and after the main event in the Before and After boxes with a specific hour gap.

The result part of the user interface is optimized to display as many images as possible, with up to 28 images for the moment search and up to 12 image triplets for the temporal search. Figure 4 shows an example of temporal search, where the main image is displayed in the center of the triplet, with the left image for the previous event and the right image for the later event. For each image, we display its location and time information (day of week, hour, and date) to help users make informed decisions.

For each returned image, we also display similar images of it to help users easily look at them to compare with the query and make a decision. When users click on the returned images, a pop-up will appear to show all similar images of it and users can submit all these images. Figure 5 depicts the similar images pop-up in moment search.

Both novice and expert users can easily search for queries through the query panel and save potential images by clicking on the Save button on each image. Users can even submit images immediately by clicking on the Submit button on each image or submit all saved images in the Save window. For the question-answering topic, users can browse through all returned images and find the answer and submit text query in the special submit box for question answering topic.

5 EVALUATION

To measure the performance of the system, we carried out this experiment instead of other novice users because this experiment only test the efficiency of the system without any further actions such as temporal browsing, or similar image checking. We use 14 queries from the LSC'22 [11] in the Know-Item topic to perform this experiment. This topic has specific answers for each query instead of the queries in Ad-hoc and Question Answering topics, so it is easier to measure the accuracy of the system through Recall at K (R@k). However, it should be noted that this measurement does not consider user interactions like filtering or image browsing. Additionally, users may have to divide the query into Before, Find, and After based on their interpretation to fit the input for temporal search. For example, in the query "Meeting friends outside a bar called the Brazen Head, before walking to another bar for drinks.", users need to choose "Meeting friends outside a bar called the Brazen Head" as the query for the Find box as the main event, and "walking to another bar for drinks." as the previous event to put into the Before box thanks to the word "before".

Figure 4: MemoriEase main user interface

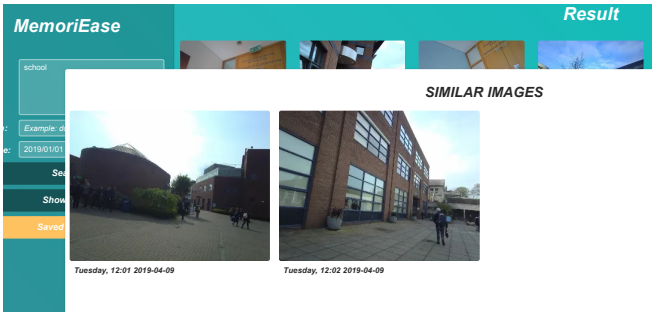

Figure 5: Similar images pop up.

It is worth noting that the LSC competition provides hints after 30 seconds, and there are a total of 6 hints. In addition, for each hint, if we submit the true answer, we do not need to perform more searches in the later hints so as to make our evaluation as realistic in the real challenge as possible, we will calculate a modified version of R@k in which R@k(i) is the Recall at k at hint i and $R@k(i) = Max(R@k(i), R@k(i-1))$. Table 1 illustrates the performance of MemoriEase in LSC'22 queries.

Table 1: Modified Mean R@k for LSC'22 queries

Hint	R@1	R@3	R@5	R@10	R@20	R@50
1	0.07	0.14	0.14	0.29	0.36	0.43
2	0.29	0.36	0.43	0.57	0.64	0.71
3	0.50	0.50	0.57	0.64	0.71	0.79
4	0.50	0.57	0.71	0.71	0.79	0.86
5	0.57	0.64	0.71	0.71	0.79	0.86
6	0.57	0.71	0.79	0.79	0.79	0.93

The results indicate that the Recall at K after the first hint is relatively low, with an average of 23%. However, there is a significant improvement in performance after the 2^{nd} hint, with an average of 50% across k. This is because 1^{st} hint usually contains very little information, while 2^{nd} hint provides more useful information for retrieval. The performance of our MemoriEase system continues to improve after the 3^{rd} hint, reaching its highest point at the 6^{th} hint, where R@1 is 57%, and R@50 is 93%.

The low R@1 score at all hints suggests that the correct answers are not among the top returned images. However, after the 3^{rd} hint, 50% of the queries are resolved in the first returned image. For the top 20 returned images that appear on the first page of the MemoriEase interface, R@20 is 36% in the first hint and this doubles to 79% in the fourth hint. However, the R@20 metric does not increase after the fourth hint since all the answered images have been found, and no further useful information is available in the later hints. The maximum R@k score is 93% in the top 50 returned images after the 6^{th} hint, as this hint provides information about date, time, and location, making it easier to filter and obtain accurate answers.

6 CONCLUSION

In this paper, we introduce the MemoriEase system for the LSC'23 competition. This system comprises embedding-based and concept-based approaches for retrieving lifelog images from textual queries. The embedding-based approach facilitates BLIP to extract visual and textual embedding and compute cosine similarity for ranking images. The concept-based approach uses ElasticSearch to calculate the similarity of keywords in the query and visual concepts of images. These two approach helps users to get the results quickly without the effort to modify input queries. The evaluation of the system on the previous competition queries shows a potential result.

The system does not only need the query, but this also needs users to perform searching, browsing, and submitting the results so we propose a simple and user-friendly user interface for MemoriEase. The system is expected to help both expert and novice users perform well in the LSC'23 competition.

ACKNOWLEDGMENTS

This research was conducted with the financial support of Science Foundation Ireland at ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology at Dublin City University [13/RC/2106_P2]. For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

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] Vannevar Bush. 1996. As We May Think. *Interactions* 3, 2 (mar 1996), 35–46. https://doi.org/10.1145/227181.227186

[4] Mariona Carós, Maite Garolera, Petia Radeva, and Xavier Giro-i Nieto. 2020. Automatic Reminiscence Therapy for Dementia. In *Proceedings of the 2020 International Conference on Multimedia Retrieval* (Dublin, Ireland) *(ICMR '20)*. Association for Computing Machinery, New York, NY, USA, 383–387. https://doi.org/10.1145/3372278.3391927

[5] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, and Rami Albatal. 2016. NTCIR Lifelog: The First Test Collection for Lifelog Research. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Pisa, Italy) *(SIGIR '16)*. Association for Computing Machinery, New York, NY, USA, 705–708. https://doi.org/10.1145/2911451.2914680

[6] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, Duc-Tien Dang-Nguyen, Rashmi Gupta, and Rami Albatal. 2017. Overview of NTCIR-13 Lifelog-2 Task. In *NTCIR Conference on Evaluation of Information Access Technologies*.

[7] 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. In *Proceedings of the 2021 International Conference on Multimedia Retrieval* (Taipei, Taiwan) *(ICMR '21)*. Association for Computing Machinery, New York, NY, USA, 690–691. https://doi.org/10.1145/3460426.3470945

[8] 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). In *Proceedings of the 2020 International Conference on Multimedia Retrieval* (Dublin, Ireland) *(ICMR '20)*. Association for Computing Machinery, New York, NY, USA, 584–585. https://doi.org/10.1145/3372278.3388043

[9] Cathal Gurrin, Klaus Schoeffmann, Hideo Joho, Andreas Leibetseder, Liting Zhou, Aaron Duane, Duc-Tien Dang-Nguyen, Michael Riegler, Luca Piras, Minh-Triet Tran, Jakub Lokoč, and Wolfgang Hürst. 2019. Comparing Approaches to Interactive Lifelog Search at the Lifelog Search Challenge (LSC2018). *ITE Transactions on Media Technology and Applications* 7 (04/2019 2019), 46–59. https://doi.org/10.3169/mta.7.46

[10] Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. 2014. LifeLogging: Personal Big Data. *Found. Trends Inf. Retr.* 8, 1 (jun 2014), 1–125. https://doi.org/10.1561/1500000033

[11] Cathal Gurrin, Liting Zhou, Graham Healy, Björn Þór Jónsson, Duc-Tien Dang-Nguyen, Jakub Lokoč, Minh-Triet Tran, Wolfgang Hürst, Luca Rossetto, and Klaus Schöffmann. 2022. Introduction to the Fifth Annual Lifelog Search Challenge, LSC'22. In *Proceedings of the 2022 International Conference on Multimedia Retrieval* (Newark, NJ, USA) *(ICMR '22)*. Association for Computing Machinery, New York, NY, USA, 685–687. https://doi.org/10.1145/3512527.3531439

[12] 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.

[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] Silvan Heller, Luca Rossetto, Loris Sauter, and Heiko Schuldt. 2022. Vitrivr at the Lifelog Search Challenge 2022. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 27–31. https://doi.org/10.1145/3512729.3533003

[15] Sungsoon Hwang, Christian Evans, and Timothy Hanke. 2017. *Detecting Stop Episodes from GPS Trajectories with Gaps*. Springer International Publishing, Cham, 427–439. https://doi.org/10.1007/978-3-319-40902-3_23

[16] Andreas Leibetseder and Klaus Schoeffmann. 2021. LifeXplore at the Lifelog Search Challenge 2021. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 23–28. https://doi.org/10.1145/3463948.3469060

[17] Andreas Leibetseder, Daniela Stefanics, and Klaus Schoeffmann. 2022. LifeXplore at the Lifelog Search Challenge 2022. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 48–52. https://doi.org/10.1145/3512729.3533005

[18] 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. https://doi.org/10.48550/ARXIV.2301.12597

[19] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. 2022. BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. https://doi.org/10.48550/ARXIV.2201.12086

[20] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. 2014. Microsoft COCO: Common Objects in Context. http://arxiv.org/abs/1405.0312 cite arxiv:1405.0312Comment: 1) updated annotation pipeline description and figures; 2) added new section describing datasets splits; 3) updated author list.

[21] 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

[22] Van-Tu Ninh, Tu-Khiem Le, Liting Zhou, Luca Piras, Michael Riegler, Pål Halvorsen, Minh-Triet Tran, Mathias Lux, Cathal Gurrin, and Duc-Tien Dang-Nguyen. 2020. Overview of ImageCLEF Lifelog 2020:Lifelog Moment Retrieval and Sport Performance Lifelog. In *CLEF2020 Working Notes (CEUR Workshop Proceedings)*. CEUR-WS.org <http://ceur-ws.org>, Thessaloniki, Greece.

[23] 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. https://doi.org/10.48550/ARXIV.2103.00020

[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, Nguyen Thanh Binh, Hyowon Lee, and Cathal Gurrin. 2020. Myscéal: An Experimental Interactive Lifelog Retrieval System for LSC'20. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge* (Dublin, Ireland) *(LSC '20)*. Association for Computing Machinery, New York, NY, USA, 23–28. https://doi.org/10.1145/3379172.3391719

[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] Ly-Duyen Tran, Dongyun Nie, Liting Zhou, Binh Nguyen, and Cathal Gurrin. 2023. VAISL: Visual-Aware Identification of Semantic Locations in Lifelog. In *MultiMedia Modeling: 29th International Conference, MMM 2023, Bergen, Norway, January 9–12, 2023, Proceedings, Part II*. Springer, 659–670.