Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-15067-9

1228: ADVANCES IN LIFELOG RETRIEVAL AT THE LIFELOG SEARCH CHALLENGE 2021

Memento: a prototype search engine for LSC 2021

Naushad Alam[1] · **Yvette Graham**[2]

Received: 17 May 2022 / Revised: 27 January 2023 / Accepted: 2 March 2023

Abstract
In this extended paper, we describe our lifelog retrieval system called Memento which participated in the 2021 Lifelog Search Challenge in detail. Memento leverages semantic representations of images and textual queries projected into a common latent space to facilitate effective retrieval, aiming to bridge the existing semantic gap between complex visual scenes/events and user information needs expressed as textual and faceted queries. Our system also has a minimalist user interface which includes functionalities such as visual data filtering and temporal search. Finally, we include a comparative analysis of Memento's performance at LSC 2021 and suggest improvements for future iterations of the system.

Keywords Lifelogging · Information retrieval · Search and ranking model · Interactive user interface

1 Introduction

Lifelogging can be defined as the process of passively gathering, processing, and reflecting on life experience data collected by an individual using a variety of devices, such as wearable cameras, tracking devices, such as Fitbit, as well as other wearable sensor devices [28].

Vannevar Bush in his 1945 article 'As We May Think' [7] talks about Memex, a "future mechanised device" which can act as an "enlarged intimate supplement of an individual's memory" storing all his books, records, communications that can be consulted with "exceeding speed and flexibility". Bush further suggested that the items in Memex could be orgainised in the form of trails similar to how the human mind operates by association, such

✉ Naushad Alam
naushad.alam2@mail.dcu.ie

Yvette Graham
ygraham@tcd.ie

[1] Insight Centre for Data Analytics, Dublin City University, Dublin, Ireland

[2] School of Computer Science and Statistics, Trinity College, Dublin, Ireland

that with one item in its grasp it snaps instantly to the next that is suggested by the association of thought. The early 2000s saw the first effort of digitising one's life in the form of MyLifeBits project [21] which aimed at fulfilling the Memex vision of Bush. The goal was to digitize an entire lifetime capturing and storing storing sensory experience, images, books, call logs, TV and radio logs, emails, transactions etc.

Over the years lifelogging has grown from being an initial concept into an active research area finding its application in multiple research domains and use cases such as memory augmentation and reminiscence [5, 9, 28, 29], human activity recognition [10, 15, 17, 45], health monitoring for elderly and people suffering from chronic diseases [4, 34, 36, 51]. However, modern day lifelogging research is primarily focused on what can be termed as visual lifelogging, automatically capturing first-person images using a wearable camera apart from other data modalities such as location, activity, biometrics, sleep recorded using various wearable devices. Besides the numerous application areas and interesting use cases, the growth of this technology was propelled by the mass availability of low cost wearable cameras and sensor devices. Furthermore, the accessibility to cheap storage facilities has made recording oneself in a passive manner feasible as passive lifelogging can be quite a intensive task with respect to storage requirements.

Retrieving the desired information from the resulting large multimodal archive called lifelogs is very challenging. The passive capturing of the lifelog images makes it an extremely noisy and repetitive archive. Furthermore, the egocentric nature of the images might not convey the entire context at times unlike conventional third-person images. In addition to the images, the non-visual data modalities need to be taken into account when trying to retrieve a specific moment/event from the lifelog dataset. The Lifelog Search Challenge (LSC) is a comparative benchmarking workshop founded in 2018 [23] to foster advances in multimodal information retrieval similar to previous activities like NTCIR-Lifelog tasks [24–26, 55] and the ImageCLEF Lifelog tasks [12–14, 46]. However unlike NTCIR-Lifelog and ImageCLEF lifelog tasks, LSC is an interactive search challenge and poses a unique information retrieval problem to the participants, where the task is not given out entirely at the beginning to the participants but is rather revealed gradually at a 30 seconds interval; at times negating/correcting earlier revealed information. The structure of LSC evaluation queries is defined in a way as to mimic how humans recall memories from their daily life hence making the competition very challenging.

In this paper, we describe our prototype system, Memento, which participated in the 2021 edition of the Lifelog Search Challenge as well as analyse the system's performance in the competition. Our system was designed to address the challenge of interactive lifelog retrieval on two fronts; bridging the semantic gap between textual queries and lifelog images while also supporting the efficient searching/browsing of the lifelog data. We use the CLIP model [48] to derive generalised semantic representations from the lifelog images. The CLIP embeddings work well with Lifelogs without any fine-tuning as the model generates zero-shot transferable representation which can be transferred to most out-of-domain datasets and still perform better than supervised baselines. Moreover, the model supports instructions in natural language which is similar to how the evaluation queries in LSC are structured, allowing dictation of complex visual scenes efficiently. Furthermore, the user interface of our system has a minimalist layout supporting efficient navigation and functionalities such as visual data filtering, tagging important images during a run as well as temporal search and browsing.

The rest of the paper is structured as follows: Section 2 discusses the approaches and methodologies adopted by previous LSC participants to address the problem of lifelog

retrieval. Section 4 discusses the LSC dataset, the core components of Memento such as search engine, CLIP image-text embeddings, user interface as well as our adopted methodology to do temporal search and event segmentation in greater detail. In Section 5, we evaluate our search backend on multiple metrics using the evaluation queries from LSC 2019 and LSC 2021. Finally, in Section 6 we do a comparative performance analysis of our system using the final results from LSC 2021, perform a deep-dive analysis into the shortcomings and suggest improvements for future iterations of the system.

2 Related work

The Lifelog Search Challenge, since its inception in 2018, has attracted significant interest and active participation from the research community. Over the last 3 years, several innovative and novel ideas have been proposed by participants from across the globe to tackle this challenging competition.

Duane et al. [19] proposed a fully immersive virtual reality interface supporting multiple interaction modes such as gesture-based interaction, distance-based interaction etc. to query the lifelog data at the first LSC in 2018. An improved version of the system took part in LSC 2020 that integrated event-based data visualization as well as new search backend into their system. Hürst et al. [32] proposed another VR based approach to search and browse the data using geo-spatial information on a map interface. Several video retrieval systems that previously participated at the VBS Challenge have also taken part in the LSC over the years. vitrivr [31] which is a video retrieval system used a backend comprised of Cineast which is a feature extraction and query processing engine along with CottontailDB [20]. It also supports multiple query modes such as query-by-sketch, query-by-example, textual queries etc. VIRET [37] used dynamically computed self-organising maps adopting a hierarchical browsing structure to search the data. Similarly, SOMHunter [43] too leveraged dynamically computed self-organising maps with a user relevance feedback mechanism to improve the search result quality. LifeXplore [39] proposed a search mechanism based on feature maps browsing arranged as 2D hierarchical grids.

Exquisitor [35] proposed to use relevance feedback to build a model of the user's information needs without using any explicit query mechanism, while THUIR [40] employed user feedback to iteratively refine the retrieved results similar to SOMHunter [43]. MySceal [52] proposed a temporal query mechanism that allowed to search for up to 3 consecutive events simultaneously and introduced a novel concept weighing methodology to determine the importance of visual concepts in the image while LifeSeeker [38] approached the problem using a Bag-of-Words model with visual concept augmentation. Voxento [2] on the other hand proposed a voice-based interactive retrieval system leveraging speech-to-text APIs to convert voice commands as text input to the system.

Several systems have tried to address the issue of the existing semantic gap between textual query and images, as well as the poor contextual understanding of the data. FIRST [54] used an autoencoder like approach to map textual queries and images into a common semantic space in order to measure the similarity between them, LifeGraph [49] used a knowledge graph to represent the lifelog data aiming to capture the internal relationships of the various data modalities and then linked it to external static data sources for better semantic understanding. Chu et al. [11] extracted relation graphs from lifelog images to better describe the relationship between entities (subject-object) present within the image.

Fig. 1 Example images from the lifelog dataset. All images are redacted by blurring out people's faces as per GDPR norms

Our proposed system addresses the semantic gap issue by using generalised image-text representations derived from a pretrained model [48] and subsequently ranking the images based on cosine similarity scores. The model is not optimised for a specific task but can rather perform a variety of tasks such as object recognition, scene recognition, activity recognition, optical character recognition, etc., hence the generated image representations are zero-shot transferable to almost any out-of-domain dataset even beating supervised baselines by significant margins.

3 LSC dataset

The dataset for the Lifelog Search Challenge 2021 [23] consists of around ∼183K images captured using a wearable camera from a single lifelogger in 2015, 2016 and 2018 spanning a total of 114 days. The dataset is the same which was used for LSC 2020 but with ∼ 8K images removed due to data governance reasons (Fig. 1).

Along with the images, the dataset consists of the following two files:

- **Visual Concepts:** For every image in the dataset, this file contains scene description, object tags with confidence scores and object bounding boxes all derived from the non-redacted version of the images.
- **Metadata:** This file consists of information such as user's location, his current activity, elevation, and biometric data such as calories burnt, heart rate and step count all captured from a wearable device at 60 seconds interval throughout the day.

4 System overview

In this section, we present an overview of our proposed system and discuss its core components such as semantic image representation, search engine, user interface in detail, and the enhancements/modifications we did to the existing metadata to further improve it. Furthermore, we also elaborate on the system's temporal search and navigation functionality and its underlying algorithm.

4.1 Semantic image representation

The visual concepts data released by the LSC organisers provide a rich annotation of the lifelog image dataset where, for each image, it captures the objects detected along with their

corresponding bounding boxes and confidence scores. The data, although useful and comprehensive, fails to convey the semantic meaning of the image or the relationship between the detected objects, for example the detected objects could be 'person' and 'cup' but it cannot be inferred whether the person is drinking from the cup, buying a cup or doing something else entirely. The LSC evaluation topics, however, are in natural language and explicitly specify the task, thus demanding a semantic understanding of the image dataset at large as well as the relationship that exists among the entities within the image.

To bridge this existing semantic gap, we use image-text embeddings generated using the Vision Transformer [18] model (ViT-B/32) from CLIP [48] to facilitate semantic search over lifelog images. The model aims to leverage the supervision inherent in natural language texts and thus was trained in a contrastive manner (Fig. 2) on 400 million image-caption pairs gathered from the internet. The network is not directly optimised for a specific task but is trained on a proxy objective of matching the captions with their respective images, thus allowing it to learn generalised visual concepts which can further be applied to multiple downstream tasks such as information retrieval, object and activity detection, scene understanding, and optical character recognition using natural language input.

Besides accepting inputs in natural language which directly benefits use-cases like information retrieval, the model is also capable of zero-shot transfer to several out-of-domain datasets. The pre-trained CLIP model was able to beat supervised ResNet-50 [30] baselines on several benchmarks hence proving the robust transfer capability of the model under data distribution shift. This allowed our use of the model as-is on the lifelog dataset without any data specific fine-tuning. We evaluated the model's performance on several metrics (discussed in Section 5) and got encouraging results. The results prove the zero-shot transfer capability of the model on out-of-domain and complex datasets like lifelogs. However, it would also be worthwhile to experiment with fine-tuning the model to compare retrieval performance in future.

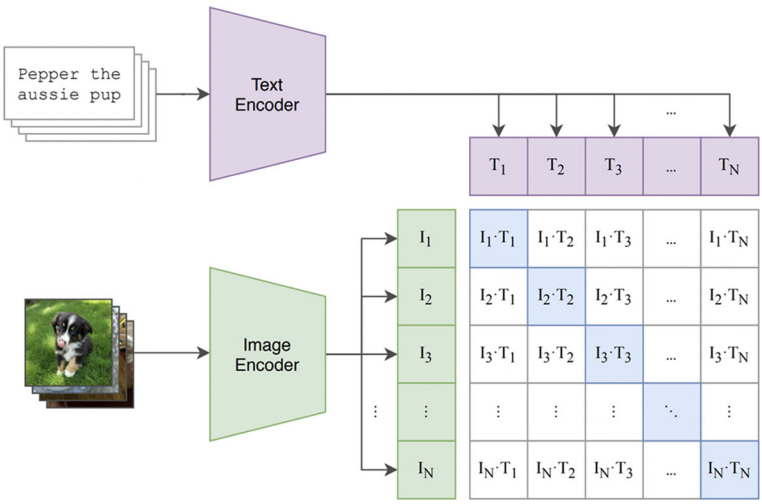


Fig. 2 CLIP's contrastive training methodology where for each batch of images and captions, the training objective is to match the correct caption with the correct image

4.2 Metadata enhancement

The LSC evaluation topics reveal information in a staged manner, where apart from a visual description of the target event, they also give out information like phases of the day (morning, night), date, and name of the location such as a cafe or a shopping mall. For example *Having coffee at Starbucks on a Sunday afternoon*

The metadata provided with the lifelog dataset contains visual concepts/annotations for the images and also has information like location, activity, date/time, etc., which is gathered from a wearable device. Our focus here was to enhance and enrich the specific part of the metadata dealing with location, activity type and date/time wherever possible as these play a crucial role in information retrieval given the fact that LSC evaluation queries explicitly reveal these bits of information during the search process.

- **Imputing Location Name:**
  Location name is a crucial piece of information given the queries explicitly specify it at times and hence it can directly be used for faceted filtering of results. It is also important to us from the event segmentation perspective given our approach (discussed in Section 4.3) uses this information to detect event boundaries from the lifelog images. We therefore try to impute the location name where location co-ordinates were available to us using a simple clustering approach.

  (i) Initially, from the existing metadata we create a dictionary with location names as **key** and their corresponding location co-ordinates as **value**

```
{location_1: (lat_1 ,lon_2),
 location_2: (lat_2 , lon_2),
   .
   .
 location_n: (lat_n , lon_n)}
```

  We also observed that few location names in the dataset are associated with multiple location co-ordinates, where the co-ordinates usually are from the same locality but slightly deviated from each other. In such scenarios, we consider the mode value of the co-ordinate corresponding to that location name.

  (ii) Next, we loop through the unnamed location co-ordinates in our dataset, assigning each of them a nearest location name using the lookup dictionary from step (i). This is done by calculating the distance of the co-ordinate in question with all co-ordinates from the dictionary.

  (iii) We finally discard those imputations where the closest location name has a distance of ≥ 3 kilometers which is an empirically derived threshold.

- **Identifying blurred images:**
  Passive lifelogging, i.e, continuous data capture at a regular interval without any human intervention generates a huge amount of data, a large chunk of which is images. Since the images are captured automatically at regular time intervals, a large volume of it tends to be noisy, blurred or occluded which is not very useful [27].

  We tried to identify and tag blurred images in the dataset using an implementation of variance of the laplacian method [47] in the OpenCV [6] library. The algorithm converts the image into greyscale and convolve a 3x3 laplacian kernel over it to calculate a variance. A well focused image is expected to have a high variation in grey levels and vice-versa. We empirically choose a variance threshold of 50, below which we tag the

image as blurred. The objective of this exercise is to minimize less useful images during browsing of temporal events (discussed in Section 4.5) as we observe that more than 30% of the images in the corpus have some degree of blurring/occlusion.

- **Deriving specific fields from existing data:**
  LSC evaluation queries often explicitly specify key pieces of information such as name of city, day name, hour of the day etc. For example *At the london airport early morning on monday.* We therefore separated out specific data fields from the provided metadata in order to perform faceted filtering efficiently.

4.3 Segmentation of each day into events

It has been known for a long time that lifelog data is inherently sequential in nature where each day can be broken into coherent and meaningful chunks called 'events'. E.g. *driving in the car from home to the office* can be one event while *walking from office to the cafeteria to grab a coffee* can be another.

Since the last decade, several researchers have proposed innovative ideas to do event segmentation from lifelogs. Lin and Hauptmann [41] proposed an approach based on the K-means clustering algorithm using color features to do event segmentation. Doherty et al. [16] used MPEG-7 descriptors from images along with the metadata available from the camera sensors to address this problem. Similarly, Byrne et al. [8] proposed to use low-level MPEG-7 feature descriptors, along with light level sensor and motion sensor. More recently, approaches based on high-level image representations have been proposed to address event segmentation. Gupta et al. [22] used visual concepts derived from the Caffe framework [33] and semantic image categorization to form event boundaries while MySceal [52] approached the problem using SIFT [42] features along with features from pretrained VGG model [50].

In this work, we attempt to do event segmentation by adopting a crude approach leveraging only an individual's current activity along with their current location. We devised a simple algorithm which tracks how the activity and location of the individual changes chronologically to establish event boundaries. Our algorithm evaluates any changes in these two specific pieces of information (activity and location) to determine a change of event and further assigns an 'event number' to all events in the data which also serves as a unique ID. For instance, the Lifelogger is at his home where his current activity is specified as **'None'** while his location indicates **'Home'** in the metadata. As soon as he exits his home and starts to drive his car, the activity changes to **'Transport'** while the location updates to **'None'** indicating a change in event.

Our algorithm tracks such changes sequentially throughout the lifelog dataset to mark out the event boundaries taking into consideration the edge cases where the methodology will not work. As an example, in a situation when the lifelogger is driving across the city, the location name will continuously change since the car is moving while current activity will remain static, hence the algorithm handles the situation and considers it as a single 'car driving' event.

We adopt a coarse approach when compared with previous approaches discussed earlier where they use visual features from images to detect events, where two dissimilar images in a sequence suggest a change of event. Our approach hence is not suitable to do fine-grained event segmentation, for example when the individual is doing multiple things such eating, working, relaxing, all while sitting in his office, our algorithm will consider all these sub-events as part of a single large event since the location and activity is not changing. However, the proposed algorithm works best for our use case since the objective is to do data filtering

based on previous, next and current activity of the user as LSC queries explicitly specify user activity in temporal queries such as *Walking on a green footpath, to my car. I got into my car and drove to a lunch* which indicates the current activity as walking and the next activity as driving.

4.4 Search engine

The search and ranking functionality of our proposed system, Memento, is powered by the CLIP model [48]. The functionality is accessible to the end user via a Restful API endpoint built using the Flask framework. Figure 3 shows a high-level overview of the system architecture. Initially, all the lifelog images are encoded into 512-dimensional image embeddings using the Image Encoder of the Model and stored as a static numpy array on system's memory.

Furthermore, at run time our system takes the following steps to process the input and return the results back to the user:

1. The user inputs the search query on the user interface which hits the search API endpoint.
2. The search query is encoded into a 512-dimensional text embedding using the text encoder.
3. The generated text embedding from above is compared with all the image representations stored in memory using cosine similarity which gives us a ranked list of images.
4. Finally, the top 2000 images are sent to the user as a response. The metadata for these images having details like time, date, location etc. are also sent along with response.

The quality of the retrieved results will depend on how well the search query was 'engineered'. Since the model is trained on image-caption pairs, the search query besides being concise and compact should mimic the writing style of an image caption to get better results which the authors of [48] call 'prompt engineering'. LSC queries usually have information scattered across multiple sentences which should be rewritten in a compact way before a

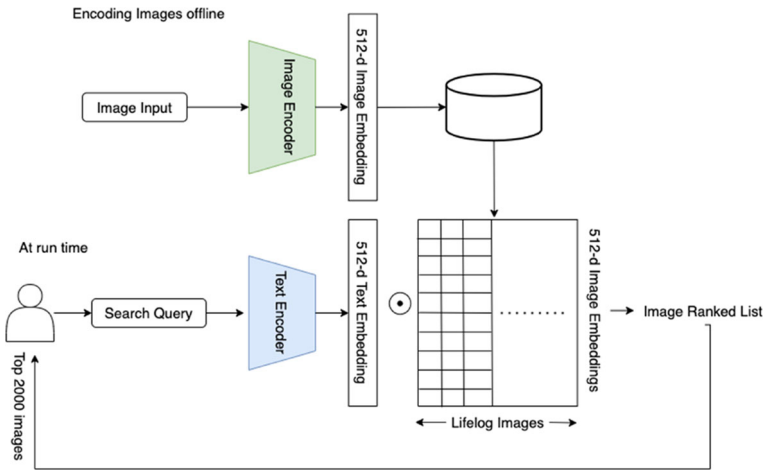


Fig. 3 System architecture - memento

query can be initiated. However, given the time sensitive nature of the competition, doing this iteratively can be a time consuming and cumbersome process. We discuss the impact prompt engineering can have on the system's performance in detail in Section 6. Furthermore, the design to store image embeddings and metadata as static files is by choice, to leverage the power of vectorised operations for a much faster turnaround time.

4.5 Temporal search and navigation

The LSC evaluation queries can be broadly categorised into two sub-categories, namely visual queries and temporal queries. Visual queries can be defined as those which puts more focus on the visual details of the target event revealing elaborate details about the event explicitly. For instance, *The Red House with lots of cars around that day and the weather was very nice with a beautiful blue sky*.

On the other hand, temporal queries might not delve too deep specifying visual details about the target but rather focuses more on specifying events which are in temporally closer vicinity of the target event. For example

- Target Event: *Buying a ticket for a train in Ireland from a ticket vending machine.*
- Future Event: *After the purchase, I walked up stairs to the platform and waited 8 minutes for the train to arrive.*
- Past Event: *I had walked (for 36 minutes) to the station after eating sushi and beer.*

The temporal search functionality hence allows the user to efficiently search such queries by specifying target and temporal events (past and future) as separate inputs to the system. Since the capability of searching and navigating the lifelog data using temporal context can be extremely crucial, several LSC systems in the past supported the functionality of searching using multiple temporal queries as input as well as navigating across the data temporally. VIRET [37] allowed the user to specify two temporally close scenes as input, while MySceal [52] allowed 3 temporal inputs to the system (past, present and future). LifeSeeker [38] used elastic sequencing to leverage temporal information of nearby moments while also supporting temporal browsing of past and future moments. Also SOMHunter [43] had the functionality to initialize a query with an optional temporal context and then continue to refine the search results with user relevance feedback.

Our proposed temporal search algorithm allows the user to input a target event along with option to specify a past event, a future event or both events as input. We leverage semantic image representations to search temporally similar to how normally searching works across the system as discussed in Section 4.4. We also define the search space of our algorithm using 'Event Numbers' (discussed in Section 4.3) as opposed to choosing a static time duration because we felt it would help reduce the noise in our final result set. The temporal search functionality of our system has the following execution steps:

1. The user initiates a query to search for the main event (discussed in Section 4.4).
2. Once the user has the ranked result set displayed on the screen, a temporal search can be initiated by specifying either a past event or a future event, or both as input.
3. The algorithm iterates through the initial result set, looking at the past and future context for every image in a predefined search space. For past context, the algorithm considers 2 events prior to the current event *(current event - 2)*, while for the future context it considers 2 events after the current *(current event + 2)* (Every image in our dataset has an event number associated to it already as discussed in Section 4.3).

4. The algorithm assigns a past temporal score and a future temporal score to every image in the initial result set, which is the maximum cosine similarity score within their respective search spaces.
5. The final score of each image is then computed as the sum of temporal scores (past and future) and initial cosine similarity score, based on which the images are re-ranked and rendered on screen.

The efficacy of this algorithm depends on how well the system locates and ranks the main event. Searching for temporal events when the main event is not within our initial result set, is futile. However, as discussed in [1], there is a very high probability of the target image being in the top-2000 results given that the query string is 'engineered' well. The system also supports sequential browsing of previous and next non-blurred images around a probable target image as in some scenarios browsing is fast and sufficient to arrive at a decision.

4.6 User interface

The user interface of Memento was built using ReactJS which is a free and open-source front-end JavaScript library developed by MetaAI. The system has a very minimalist user interface where most of the system functionalities are not on the main home screen but are accessible via separate overlay windows. The intention to choose overlay windows to deliver the functionalities was to keep the home screen clutter-free hence maximising the viewport area enabling efficient browsing of search results. Memento's user interface primarily consists of following user interfaces:

- **Primary Search Interface:** This interface consists of two components, one is the primary navigation bar on top of the window which embeds the search box as well as the buttons to access system functionalities such as data filtering and image starring while the other one displays the search results. Figure 4 shows a snapshot of the primary search interface of our system showing results for the query *"eating sushi"*.
- **Data Filtering Component:** The aim of designing the data filtering interface the way it is was twofold: one is providing faceted filtering functionality over the result set

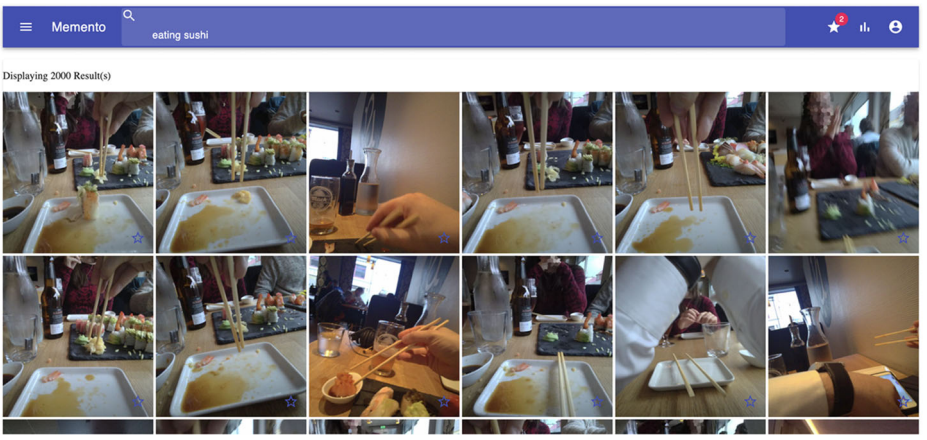


Fig. 4 Primary search interface

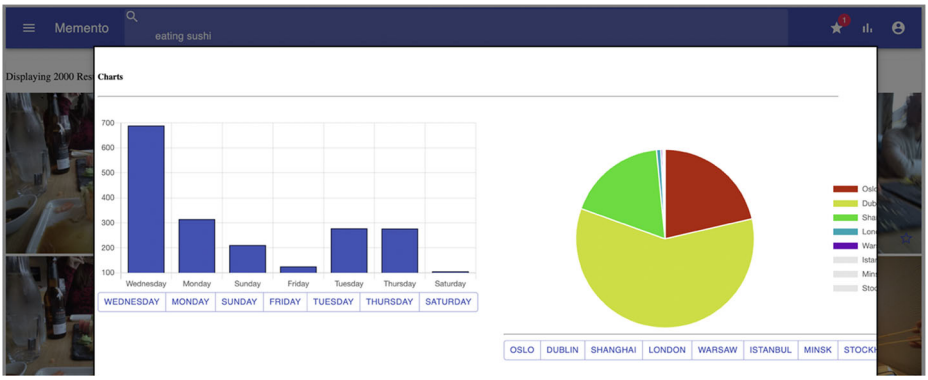


Fig. 5 Data filtering interface

and at the same time serving as a data visualisation component. The need of a faceted filtering component in a LSC system which can filter the data on the basis of time, day, location etc. cannot be stressed enough. However, adding a dynamic data visualisation functionality on top of it makes it extremely handy in a time-sensitive competition such as LSC. The result set visualisation is intended to help the user form a mental picture of the data quickly and aid in better decision making. Figure 5 shows a snapshot of the data filtering interface of Memento.

- **Starred Images:** Since the LSC queries reveal hints gradually, during an ongoing search the user would sometimes want to save an image which he thinks could be relevant and would like to revisit it later once subsequent hints are revealed. The starring interface provides this exact functionality helping the user revisit tagged images without having to reinitiate the search process all over again. The images can be added to the starred list by clicking on the star icon corresponding to that image from the primary search interface. The starring interface displays the starred images as well as the relevant metadata associated with it. The user can choose to submit the image to the evaluation server using the 'Submit' button or initiate temporal browsing to view

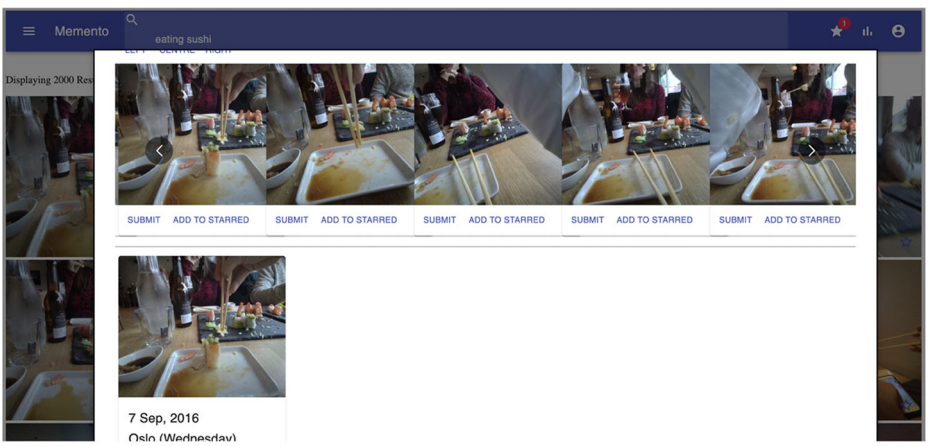


Fig. 6 Starring interface

Table 1 First 3 hints (revealed at 0, 30, and 60 seconds) for a LSC query and their corresponding rephrased version

Time	Original query text	Modified query text
0 seconds	I remember a doll's house	a doll's house
30 seconds	I remember a doll's house, a white dolls house	a white doll's house
60 seconds	I remember a doll's house, a white dolls house. There were other people there	a white doll's house with other people around

previous and next images in sequence using 'Inspect' functionality. Figure 6 shows a snapshot of the starring interface displaying a single starred image in the bottom, while the temporal context for that image is displayed on the top with left-right arrow buttons to navigate through the past or the future images in sequence.

5 System evaluation

The LSC evaluation topics reveal information sequentially in parts at a time interval of 30 seconds giving out the 1st hint at t = 0 seconds and then every 30 seconds till t = 150 seconds. The queries usually reveal visually descriptive information about the target event early on while more explicit information like time and date, place, etc., are revealed towards the later stages.

We, however, only consider the information available to us by t = 60 seconds to evaluate our system as the CLIP model which powers the search engine being trained on image-caption pairs cannot make sense of explicit information such as exact time, day, month. The performance of the model also depends a lot on the input prompt provided to it where compact prompts mimicking the writing style of an image caption works almost always better than other prompts. We, therefore, manually rephrased the LSC evaluation queries in a compact form to suit the requirements of the CLIP model. Table 1 shows the first 3 hints and their corresponding modified version of a query from LSC 2019.

We evaluated our proposed system on the evaluation topics from LSC 2019 and LSC 2021 on the following metrics:

1. Hit@K: For a given topic, Hit@K is defined as finding at least one target image among top-K images in the result set;
2. Precision@K;
3. Recall@k.

Tables 2 and 3 shows the results of Hit@K for LSC 2019 and LSC 2021 respectively at different amounts of elapsed time, t and K values.

Table 2 Hit@K calculated at different amounts of elapsed times, t and K values across 24 evaluation topics for LSC'19

t	@1	@3	@5	@10	@20	@50	@75	@100
0s	8.33	25.00	29.17	29.17	37.50	50.00	54.17	62.50
30s	8.33	25.00	25.00	33.33	33.33	54.17	54.17	58.30
60s	12.50	29.17	29.17	41.67	54.17	75.00	75.00	**79.20**

Table 3 Hit@K calculated at different amounts of elapsed times, t and K values across 23 evaluation topics for LSC'21

t	@1	@3	@5	@10	@20	@50	@75	@100
0s	8.70	17.39	26.09	26.09	34.78	47.83	56.52	56.52
30s	17.39	21.74	26.09	26.09	34.78	39.13	56.52	60.87
60s	21.74	26.09	30.43	34.78	47.83	56.52	60.87	**65.22**

At t=60 seconds and K=50, the system is able to find at least one target image in top-100 results for 79.4% of LSC 2019 and 65.2% of LSC 2021 queries. The hit rate observed is similar at lower values of K for both the set of queries, however the performance of the model diverges at higher K values for LSC 2021 queries when compared with the performance on LSC 2019 queries.

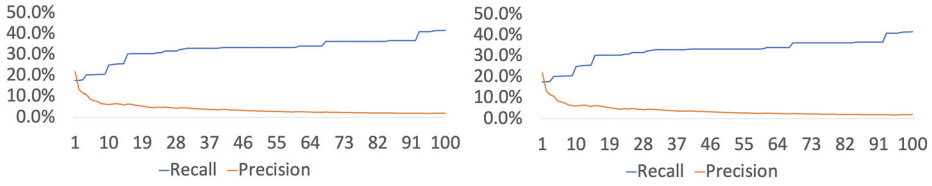
We further evaluated our proposed system on precision and recall metrics at multiple K values. Figure 7 shows the precision versus recall curve at K = 1 to 100 averaged across 24 and 23 evaluation topics from LSC 2019 and LSC 2021 respectively by only considering the information available to us by t=60 seconds.

6 Memento at the lifelog search challenge 2021

Our system, Memento participated in the 2021 Lifelog Search Challenge and successfully submitted correct response for 16 out the 23 LSC queries. It secured an overall 6th position on the final leader board competing against 16 other participating teams. Figure 8 shows the count of correct and incorrect submissions made by all 17 participating teams at LSC 2021 with Memento making 16 correct and 11 incorrect submissions.

LifeSeeker [44] which stood at position 3 on the leader board submitted the most number of correct responses **(20)** and hence achieved the highest recall **(0.86)** among all participating teams. Voxento [3] on the other hand submitted the least number of wrong answers **(3)** achieving the highest precision **(0.85)** overall. We define Precision as the *number of correct submissions* upon the *total number of submissions (correct and incorrect)* made by a team while Recall is defined as *total number of correct submissions* upon the *total number of queries*. Memento's performance in terms of precision was worst when compared with the top-5 systems on the leaderboard hence pulling down its overall score despite performing better than CVHunter in terms of recall (Table 4).

Voxento [3] which shared the same backend with Memento [1] performed better in terms of both recall and precision. However, interestingly the 16 queries solved by Memento are



(a) Precision versus Recall - LSC'19 (b) Precision versus Recall - LSC'21

Fig. 7 Precision versus Recall curves at K = 1 to 100 averaged across 24 and 23 evaluation topics from LSC'19 and LSC'21 respectively using all information available by t= 60 seconds

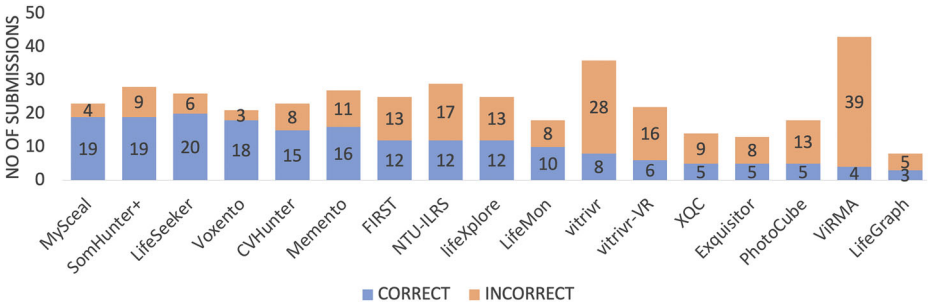


Fig. 8 Number of incorrect and correct submissions plotted for each participating system. Systems are ordered left to right according to their position in the final leaderboard

not a subset of the 18 correct responses from Voxento. As shown in Table 5, apart from the 13 queries for which both systems submitted correct responses within the stipulated time, Voxento managed to solve an additional 5 queries which Memento couldn't while Memento similarly submitted correct responses for 3 additional queries which Voxento failed to locate. Both these system hence collectively submitted correct response for 21 out of 23 LSC 2021 queries surpassing LifeSeeker's tally of 20. This shows the robust transfer capability of the CLIP model to data distribution shift and how well it generalises to such a complex dataset like lifelog. The difference in performance of both these systems despite using the same backend is largely due to human factors as well as the differences in their respective user interfaces which adopts different search and browsing mechanism.

There can be a large number of human characteristics which can impact the outcome of such a time-sensitive and competitive event like the Lifelog Search Challenge. However, one crucial aspect is the ability to write well 'engineered' prompts which go as input to the backend CLIP model given the limited amount of time participants have during the challenge. The performance of the model depends a lot on how well the input was structured, its conciseness, the keywords being used or simply the failure to remove keywords which can unnecessarily confuse the model.

To investigate the impact of well engineered prompts on the performance of the CLIP model, we evaluate it on both rephrased queries as well as original queries from LSC 2021 and observe the differences. Figure 9 shows the performance comparison of the original queries (blue bars) versus the modified queries (orange bars) from LSC 2021 on 'Hit@k' metric for multiple values of K at t = 0, 30 and 60 seconds. The modified queries almost always beat the raw original queries. The difference in performance further widens at t= 60 seconds because the original queries by this time usually span multiple sentences while the modified queries have the information consolidated into a single sentence.

Figure 10 shows the comparison of original and modified queries from LSC 2021 on precision and recall averaged across 23 queries at t= 60 seconds. The modified queries beat the raw original queries by a significant margin in terms of recall, however in terms

Table 4 Count of correct submissions, overall precision and recall for our system Memento (on the left) compared with overall best figures from LSC 2021 (on the right)

Count of correct submissions	16	Max correct submission (LifeSeeker)	20
Precision	0.59	Max Precision (Voxento)	0.85
Recall	0.69	Max Recall (LifeSeeker)	0.86

Table 5 Overview of correct submissions made by Memento and Voxento

Description	Count
Count of queries solved by both Memento and Voxento (Memento $\bigcup$ Voxento)	13
Count of queries solved by Voxento but not Memento ($Memento^c \bigcap$ Voxento)	5
Count of queries solved by Memento but not Voxento ($Voxento^c \bigcap$ Memento)	3

of precision the performance of both queries is neck-to-neck. The results provide an insight into the impact input queries can have on the performance of the CLIP model.

Besides input prompt, the user interface of the system has a very important role to play at the Lifelog Search Challenge. Factors like the granularity of faceted filtering functionality enabling the user to apply filters across a large number of data dimensions, the support of temporally searching and browsing the data, or simply the layout of the buttons and widgets on the screen can significantly impact the performance of a system in the competition. Figure 11 shows the correct and incorrect submissions made by Memento for all 23 queries from LSC 2021. Out of the total 16 correct submissions by Memento, 6 submissions were made even before the 3rd hint was revealed, other 6 were made without any prior incorrect submission while the remaining 4 were done after making 1 or 2 incorrect submissions. The system, however, was unable to submit a correct response for 7 out of 23 queries.

To do a deep dive analysis on the queries which our system failed to correctly answer, we try to leverage our local system logs generated from the Flask Framework. Table 6 lists out 7 unsolved queries from LSC 2021 while Table 7 shows a high-level report for the 7 queries derived from the local system logs.

Analysis of topic 1: For this topic, despite initiating 4 different queries with different keywords and language style we could not manage to locate a single target image within the top-2000 range. Surprisingly, Voxento managed to submit a correct response for this query indicating that the 'prompt engineering' factor might have played a role in this case.

Analysis of topic 4: Topic 4 was a challenging Optical Character recognition task where text was written on a t-shirt using a mix of alphabets and symbols. We managed to initiate 8 different queries trying out different combinations of the keywords available and managed to locate the target at position 489 but failed to submit the correct response. Our system although having a month and year filter lacked the date filtering functionality which might have come handy given the last hint in the topic revealed the date information explicitly.

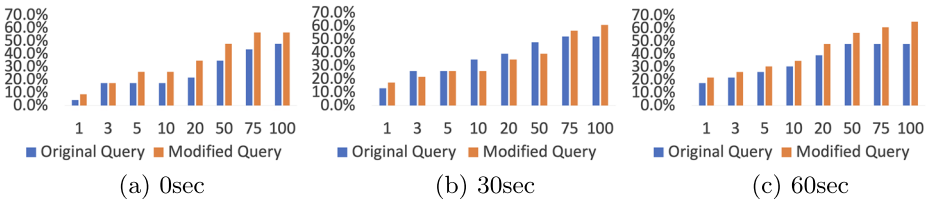


Fig. 9 Performance comparison of original queries (blue bars) with modified queries (orange bars) on Hit@K metric using LSC 2021 queries

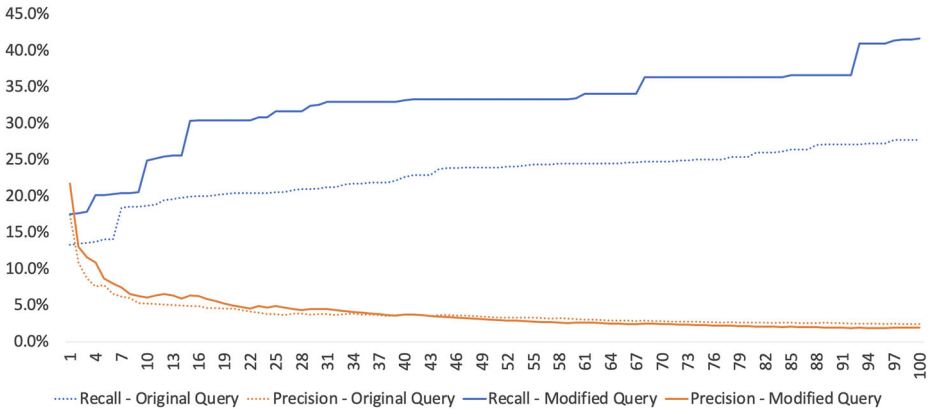


Fig. 10 Precision versus Recall curve for both original and modified queries from LSC 2021. The modified queries significantly outperforms the original queries in terms of recall while being neck-to-neck on precision

Analysis of topic 9: This topic again could be categorised as an OCR task where the ask was to locate a restaurant given its name. The other hints in this topic were too vague and hence were not very useful. We managed to locate the image but very far down the order and hence couldn't submit the correct response within the stipulated time. Voxento on the other hand correctly submitted the response for this topic which again shows the impact of input prompt given to the system.

Analysis of topic 11: This topic by far was the toughest of all as despite trying out multiple different prompts the system was unable to locate any target image within the 1-2000 range. The system was not able to comprehend the finer visual details provided in the topic such as objects being seen in the mirror reflection and not directly.

Analysis of topic 13: We were able to locate multiple target images for this topic in the top-50 results but still could not capitalise on this opportunity. The 'orange suitcase' occupied a very small area of the image with 'ride-on-suitcase' written over it in even smaller font size making it hard to spot quickly among a large number of similar looking images on the home screen. The lack of an image zooming feature on our home screen resulted in the target image being overlooked in this case. **Analysis of topic 15**: The issue

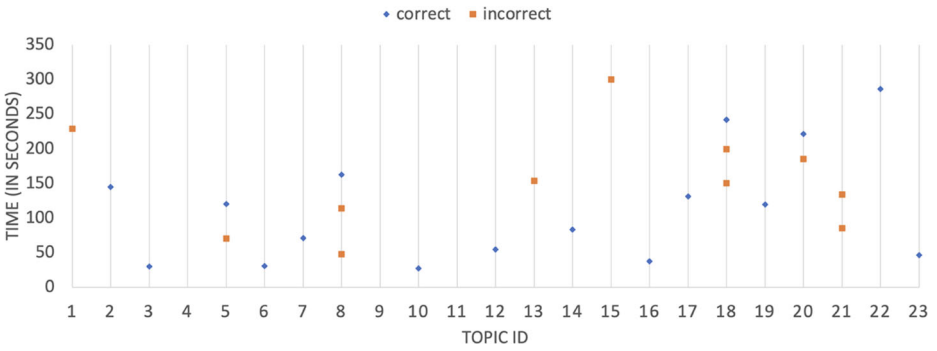


Fig. 11 Correct and incorrect submissions by Memento plotted for all 23 queries against time of submission (in seconds). Empty line indicates no submission made by the system for that topic

Table 6 Unsolved queries by Memento in LSC 2021

Topic 1	*"I was building a computer alone in the early morning on a Friday at a desk with a blue background. Sometimes I needed to refer to the manual. I remember some Chinese posters on the desk background. I was in Dublin City University in 2015"*
Topic 4	*"There was a while t-shirt for sale. I remember it said I love bicycle. It was in a bicycle and parts store. A big sale, bicycles were half price. It was in the afternoon on the 15th May 2015"*
Topic 9	*"I was lost and looking for directions on a street, close to an asian restaurant called Maple Leaf. It was in the late afternoon or evening and it was in Wexford. I had driven there in 2015"*
Topic 11	*"Telescope in the mirror before going to the airport. I was able to see my telescope and a red flower vase in the mirror in my bedroom. I also remember a white violin, flowers and a nice painting. I was playing with my computer and phone at the time. It was in 2015"*
Topic 13	*"It was an orange suitcase, a ride-on suitcase, one with a face. I remember it was in a store selling car and bicycle equipment. It was in the afternoon... before I drove home via a supermarket"*
Topic 15	*"Boarding pass for PVG. I was going to Shanghai and queuing at the airport gate... in Dublin on a day with a nice blue sky. My name is clearly visible on the boarding pass which was for the second leg of my flight (from FRA to PVG) in 2015"*
Topic 21	*"Eating mandarins and an apple, while working on a paper on my laptop. The paper is about Quantified Self. It was in 2016 in early September"*

we faced with this topic was quite similar to the one we faced with topic 13. We again managed to locate the target image within the top-50 range but the details like boarding pass and the name written over it was tough to spot given there were a lot of similar looking images within the 1-50 range. Again a simple image zooming feature might have helped us locate the correct response for this topic within the prescribed time.

Analysis of topic 21: The topic 21 was a easy take as the visual description was not very complex. It, however, resulted in multiple similar looking images (containing mandarins and apples) captured at different points in time with the task being to locate the image

Table 7 A high-level report generated from the local system logs for the 7 unsolved queries at LSC 2021

Topic ID	Count of queries initiated	Found in top-2000	Position of target image within top-2000	Solved by Voxento
1	4	No	–	Yes
4	8	Yes	489	No
9	7	Yes	1976	Yes
11	12	No	–	No
13	7	Yes	23	Yes
15	6	Yes	12	Yes
21	3	Yes	47	Yes

Topic ID refers to the unique ID of the query. **Count of queries initiated** refers to the number of different rephrased texts given as input to the system during the run. **Found in top-2000** is a boolean variable indicating whether any one of the target image was found within the 1-2000 range for any of the 'n' queries initiated by the user as well as specifying the lowest position it was found at. Finally, **Solved by Voxento** indicates whether the system Voxento could solve that particular query

from 'early September'. Here again, a date filtering functionality in our user interface could have helped locate the correct image as the month filter alone was not able to reduce the similar images from the result set.

Figure 12 shows the time distribution for correct responses across all participating teams and evaluation topics. MySceal [53] has the best median time among the top 6 systems on the leaderboard while Memento [1] stands at 5th position in terms of median time of submission, better only to LifeSeeker's [44] tally. Figure 13 shows the count of occurences when the system was in top 3 to submit a correct response for a query. Memento managed to be in the top 3 for 9 out of 16 queries it correctly submitted demonstrating competitive performance.

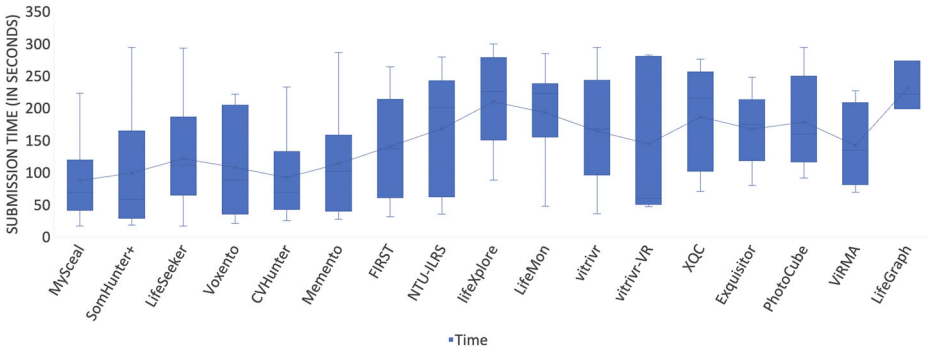


Fig. 12 Time distribution (in sec) for correct submissions across all teams. Teams are ordered left-to-right as per their position on the final leaderboard

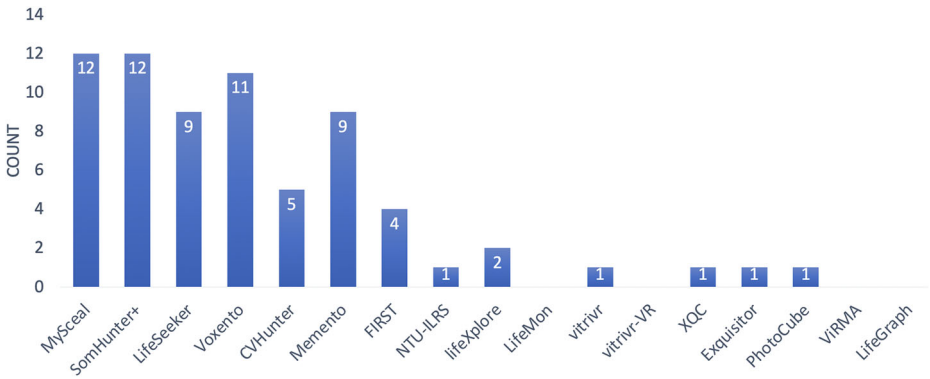


Fig. 13 Count of occurrences when the system was in top 3 to submit a correct for any given query

7 Conclusion and future work

In this extended work, we describe our lifelog retrieval system Memento in further detail. Our proposed system uses a pretrained model to generate semantic image-text representations which is used to search and rank relevant images based on cosine similarity score. We evaluate our system on multiple metrics using queries from LSC 2019 and LSC 2021 and got encouraging results despite testing in a constrained environment which proves the generalisation capability of the model to unseen and complex datasets like lifelogs. The comparative performance analysis of our system sheds light into some of the shortcomings of the CLIP model as well as highlights the crucial role human factors play in the Lifelog Search Challenge. It is quite evident that manually engineering good prompts for the model is tough given there are no set rules to do so and is largely dependent on the individual's own understanding of what should work and what shouldn't. Iteratively trying out multiple prompts could work but is not a viable approach for LSC given it is a very time sensitive competition demanding quicker turn around time. An ensembling approach where the embeddings of multiple search queries can be combined together and sent as a consolidated input to the model could be a feasible solution as the consolidated vector will encompass multiple contexts and is likely to perform better than the individual prompts. Manually writing multiple prompts is again a time consuming task and an automated solution would be required to generate multiple prompts given a raw LSC query.

The future iteration of Memento is supposed to participate in the 2022 Lifelog Search Challenges with an improved user interface incorporating all the findings of our analysis and will incorporate embeddings derived from the subsequently released larger CLIP models demonstrating superior search and ranking capability.

Funding Open Access funding provided by the IReL Consortium. This work has emanated from research supported by Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289_P2, co-funded by the European Regional Development Fund.

Data Availability Authors declare that the dataset used in this work can be accessed by following the instructions available at http://lsc.dcu.ie/lsc_data/

References

1. Alam N, Graham Y, Gurrin C (2021) Memento: a prototype lifelog search engine for lsc'21. In: Proceedings of the 4th annual on lifelog search challenge. Association for Computing Machinery, New York, pp 53–58. https://doi.org/10.1145/3463948.3469069
2. Alateeq A, Roantree M, Gurrin C (2020) Voxento: a prototype voice-controlled interactive search engine for lifelogs. In: Proceedings of the third annual workshop on lifelog search challenge. ACM, Dublin, pp 77–81. https://doi.org/10.1145/3379172.3391728
3. Alateeq A, Roantree M, Gurrin C (2021) Voxento 2.0: a prototype voice-controlled interactive search engine for lifelogs. In: Proceedings of the 4th annual on lifelog search challenge. Association for Computing Machinery, New York, pp 65–70. https://doi.org/10.1145/3463948.3469071
4. Amin MB, Banos O, Khan WA, Muhammad Bilal HS, Gong J, Bui D-M, Cho SH, Hussain S, Ali T, Akhtar U, Chung TC, Lee S (2016) On curating multimodal sensory data for health and wellness platforms. Sensors (Basel, Switzerland) 16:7. https://doi.org/10.3390/s16070980. Accessed 2021-04-13
5. Bahrainian SA, Crestani F (2018) Augmentation of human memory: anticipating topics that continue in the next meeting, 10
6. Bradski G (2000) The OpenCV library. Dr. Dobb's Journal of Software Tools
7. Bush V (1945) As we may think. Section: technology. https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/. Accessed 24 Apr 2021
8. Byrne D, Lavelle B, Doherty AR, Jones GJF, Smeaton AF (2007) Using bluetooth & gps metadata to measure event similarity in sensecam images. In: Information sciences 2007. World Scientific, pp 1454–1460
9. Carós M, Garolera M, Radeva P, Giro-i Nieto X (2020) Automatic reminiscence therapy for dementia. In: Proceedings of the 2020 international conference on multimedia retrieval. ACM, Dublin, pp 383–387. https://doi.org/10.1145/3372278.3391927. Accessed 11 May 2022
10. Cartas A, Marín J, Radeva P, Dimiccoli M (2017) Recognizing activities of daily living from egocentric images. arXiv:1704.04097 [cs]. Accessed 24 Apr 2021
11. Chu T-T, Chang C-C, Yen A-Z, Huang H-H, Chen H-H (2020) Multimodal retrieval through relations between subjects and objects in lifelog images. In: Proceedings of the third annual workshop on lifelog search challenge. ACM, Dublin, pp 51–55. https://doi.org/10.1145/3379172.3391723. Accessed 2021-03-31
12. Dang-Nguyen D-T, Piras L, Riegler M, Boato G, Zhou L, Gurrin C (September 2017) Overview of ImageCLEF lifelog 2017: lifelog retrieval and summarization. In: Dang-Nguyen, Duc-Tien ORCID: 0000-0002-2761-2213 <https://orcid.org/0000-0002-2761-2213>, Piras, Luca, Riegler, Michael, Boato, Giulia, Zhou, Liting ORCID: 0000-0002-7778-8743 <https://orcid.org/0000-0002-7778-8743> and Gurrin, Cathal ORCID: 0000-0003-2903-3968 <https://orcid.org/0000-0003-2903-3968> (2017) Overview of ImageCLEF lifelog 2017: lifelog retrieval and summarization. In: ImageCLEF 2017, 11-13 Sept 2017, Dublin. ISBN ISSN 1613-0073, vol 1866. CEUR-WS, Dublin. http://ceur-ws.org/Vol-1866/invited_paper_10.pdf. Accessed 2022-05-12
13. Dang-Nguyen D-T, Piras L, Riegler M, Zhou L, Lux M, Gurrin C Overview of ImageCLEFlifelog 2018: daily living understanding and lifelog moment retrieval, 19
14. Dang-Nguyen D-T, Piras L, Riegler M, Zhou L, Lux M, Tran M-T, Le T-K, Ninh V-T, Gurrin C Overview of ImageCLEFlifelog 2019: solve my life puzzle and lifelog moment retrieval, 17
15. Dobbins C, Rawassizadeh R, Momeni E Detecting physical activity within lifelogs towards preventing obesity and aiding ambient assisted living | Elsevier Enhanced Reader
16. Doherty AR, Smeaton AF (2008) Automatically Segmenting LifeLog Data into Events. In: 2008 Ninth international workshop on image analysis for multimedia interactive services, pp 20–23. ISSN: 2158-5881
17. Doherty AR, Kelly P, Kerr J, Marshall S, Oliver M, Badland H, Hamilton A, Foster C (2013) Using wearable cameras to categorise type and context of accelerometer-identified episodes of physical activity. Int J Behav Nutr Phys Act 10(1):22. https://doi.org/10.1186/1479-5868-10-22. Accessed 2022-05-11

18. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N (2021) An image is worth 16x16 words: transformers for image recognition at scale. arXiv:2010.11929 [cs]. Accessed 2022-03-20
19. Duane A, Gurrin C, Huerst W (June 2018) Virtual reality lifelog explorer: lifelog search challenge at acM ICMR 2018. In: Proceedings of the 2018 ACM workshop on the lifelog search challenge. ACM, Yokohama, pp 20–23. https://doi.org/10.1145/3210539.3210544. Accessed 2021-04-13
20. Gasser R, Rossetto L, Heller S, Schuldt H (2020) Cottontail db: an open source database system for multimedia retrieval and analysis. In: Proceedings of the 28th ACM international conference on multimedia. Association for Computing Machinery, New York, pp 4465–4468
21. Gemmell J, Bell G, Lueder R (2006) MyLifeBits: a personal database for everything. Commun ACM 49(1):88–95. https://doi.org/10.1145/1107458.1107460. Accessed 2021-04-12
22. Gupta R, Gurrin C (2018) Approaches for event segmentation of visual lifelog data. In: Schoeffmann K, Chalidabhongse TH, Ngo CW, Aramvith S, O'Connor NE, Ho Y-S, Gabbouj M, Elgammal A (eds) MultiMedia modeling, vol 10704. Springer International Publishing, Cham, pp 581–593
23. Gurrin C, Jónsson BT, Schöffmann K, Dang-Nguyen D-T, Lokoč J, Tran M-T, Hürst W, Rossetto L, Healy G (2021) Introduction to the fourth annual lifelog search challenge, lsc'21. In: Proc. International conference on multimedia retrieval (ICM'1). ACM, Taipein
24. Gurrin C, Joho H, Hopfgartner F (2016) Overview of NTCIR-12 lifelog task, 7
25. Gurrin C, Joho H, Hopfgartner F, Zhou L, Gupta R, Albatal R, Dang-Nguyen D-T (2017) Overview of NTCIR-13 lifelog-2 task, 6
26. Gurrin C, Joho H, Hopfgartner F, Zhou L, Ninh V-T, Le T-K, Albatal R, Dang-Nguyen D-T, Healy G (2019) Overview of the NTCIR-14 lifelog-3 task, 13
27. Gurrin C, Smeaton AF, Byrne D, O'Hare N, Jones GJF, O'Connor N (2008) An examination of a large visual lifelog. In: Li H, Liu T, Ma W-Y, Sakai T, Wong K-F, Zhou G (eds) Information retrieval technology. Springer, Berlin, pp 537–542
28. Gurrin C, Smeaton AF, Doherty AR (2014) LifeLogging: personal big data. Foundations and Trends®in Information Retrieval 8(1):1–125. https://doi.org/10.1561/1500000033. Accessed 2021-04-09
29. Harvey M, Langheinrich M, Ward G Remembering through lifelogging: a survey of human memory augmentation | Elsevier Enhanced Reader. https://doi.org/10.1016/j.pmcj.2015.12.002
30. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Las Vegas, pp 770–778, https://doi.org/10.1109/CVPR.2016.90. http://ieeexplore.ieee.org/document/7780459/. Accessed 2022-03-20
31. Heller S, Amiri Parian M, Gasser R, Sauter L, Schuldt H (2020) Interactive Lifelog Retrieval with vitrivr. In: Proceedings of the third annual workshop on lifelog search challenge. ACM, Dublin, pp 1–6. https://doi.org/10.1145/3379172.3391715. Accessed 2021-03-31
32. Hürst W, Ouwehand K, Mengerink M, Duane A, Gurrin C (2018) Geospatial access to lifelogging photos in virtual reality. In: Proceedings of the 2018 ACM workshop on the lifelog search challenge. ACM, Yokohama, pp 33–37. https://doi.org/10.1145/3210539.3210547. Accessed 2021-04-13
33. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T (2014) Caffe: convolutional architecture for fast feature embedding. arXiv:1408.5093 [cs]
34. Karako K, Chen Y, Song P, Tang W Super-aged society: constructing an integrated information platform of self-recording lifelogs and medical records to support health care in Japan. BioScience Trends,
35. Khan OS, Larsen MD, Poulsen LAS, Jónsson BT, Zahálka J, Rudinac S, Koelma D, Worring M (2020) Exquisitor at the lifelog search challenge 2020. In: Proceedings of the third annual workshop on lifelog search challenge. ACM, Dublin, pp 19–22. https://doi.org/10.1145/3379172.3391718. Accessed 2021-03-31
36. Kim S, Yeom S, Kwon O-J, Shin D, Shin D (2018) Ubiquitous healthcare system for analysis of chronic patients' biological and lifelog data. IEEE Access 6:8909–8915. Conference Name: IEEE Access
37. Kovalčík G, Škrhak V, Souček T, Lokoč J (2020) VIRET tool with advanced visual browsing and feedback. In: Proceedings of the third annual workshop on lifelog search challenge. ACM, Dublin, pp 63–66. https://doi.org/10.1145/3379172.3391725
38. Le T-K, Ninh V-T, Tran M-T, Nguyen T-A, Nguyen H-D, Zhou L, Healy G, Gurrin C (2020) LifeSeeker 2.0: interactive lifelog search engine at LSC 2020. In: Proceedings of the third annual workshop on lifelog search challenge. ACM, Dublin, pp 57–62. https://doi.org/10.1145/3379172.3391724
39. Leibetseder A, Schoeffmann K (2020) lifeXplore at the lifelog search challenge 2020. In: Proceedings of the third annual workshop on lifelog search challenge. ACM, Dublin, pp 37–42. https://doi.org/10.1145/3379172.3391721. Accessed 2021-03-31

40. Li J, Zhang M, Ma W, Liu Y, Ma S (2020) A multi-level interactive lifelog search engine with user feedback. In: Proceedings of the third annual workshop on lifelog search challenge. ACM, Dublin, pp 29–35. https://doi.org/10.1145/3379172.3391720. Accessed 2021-03-31
41. Lin W-H, Hauptmann A (2006) Structuring continuous video recordings of everyday life using time-constrained clustering. https://doi.org/10.1184/R1/6609992.v1
42. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vision 60(2):91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94. Accessed 2022-05-12
43. Mejzlík F, Veselý P, Kratochvíl M, Souček T, Lokoč J (2020) SOMHunter for lifelog search. In: Proceedings of the third annual workshop on lifelog search challenge. ACM, Dublin, pp 73–75. https://doi.org/10.1145/3379172.3391727
44. Nguyen T-N, Le T-K, Ninh V-T, Tran M-T, Thanh Binh N, Healy G, Caputo A, Gurrin C (2021) Lifeseeker 3.0: an interactive lifelog search engine for lsc'21. In: Proceedings of the 4th annual on lifelog search challenge. Association for Computing Machinery, New York, pp 41–46. https://doi.org/10.1145/3463948.3469065
45. Ni J, Chen B, Allison NM, Ye X A hybrid model for predicting human physical activity status from lifelogging data | Elsevier Enhanced Reader. https://doi.org/10.1016/j.ejor.2019.05.035
46. Ninh V-T, Le T-K, Zhou L, Piras L, Riegler M Overview of ImageCLEFlifelog 2020: lifelog moment retrieval and sport performance lifelog, 17
47. Pech-Pacheco JL, Cristobal G, Chamorro-Martinez J, Fernandez-Valdivia J (2000) Diatom autofocusing in brightfield microscopy: a comparative study. In: Proceedings 15th international conference on pattern recognition. ICPR-2000, vol 3, pp 314–317
48. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G, Sutskever I (2021) Learning transferable visual models from natural language supervision. arXiv:2103.00020s [cs]. Accessed 2021-04-08
49. Rossetto L, Baumgartner M, Ashena N, Ruosch F, Pernischová R, Bernstein A (2020) LifeGraph: a knowledge graph for lifelogs. In: Proceedings of the third annual workshop on lifelog search challenge. ACM, Dublin, pp 13–17. https://doi.org/10.1145/3379172.3391717
50. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 [cs]. Accessed 2022-05-12
51. Sugawara J, Ochi D, Yamashita R, Yamauchi T, Saigusa D, Wagata M, Obara T, Ishikuro M, Tsunemoto Y, Harada Y, Shibata T, Mimori T, Kawashima J, Katsuoka F, Igarashi-Takai T, Ogishima S, Metoki H, Hashizume H, Fuse N, Minegishi N, Koshiba S, Tanabe O, Kuriyama S, Kinoshita K, Kure S, Yaegashi N, Yamamoto M, Hiyama S, Nagasaki M (2019) Maternity Log study: a longitudinal lifelog monitoring and multiomics analysis for the early prediction of complicated pregnancy. BMJ Open 9(2):025939. https://doi.org/10.1136/bmjopen-2018-025939. Accessed 2022-05-11
52. Tran L-D, Nguyen M-D, Binh NT, Lee H, Gurrin C (2020) Myscéal: an experimental interactive lifelog retrieval system for LSC'20. In: Proceedings of the third annual workshop on lifelog search challenge. ACM, Dublin, pp 23–28. https://doi.org/10.1145/3379172.3391719
53. Tran L-D, Nguyen M-D, Thanh Binh N, Lee H, Gurrin C (2021) Myscéal 2.0: a revised experimental interactive lifelog retrieval system for lsc'21. In: Proceedings of the 4th annual on lifelog search challenge. Association for Computing Machinery, New York, pp 11–16. https://doi.org/10.1145/3463948.3469064
54. Tran M-T, Nguyen T-A, Tran Q-C, Tran M-K, Nguyen K, Ninh V-T, Le T-K, Trang-Trung H-P, Le H-A, Nguyen H-D, Do T-L, Vo-Ho V-K, Gurrin C (2020) FIRST - flexible interactive retrieval system for visual lifelog exploration at LSC 2020. In: Proceedings of the third annual workshop on lifelog search challenge. ACM, Dublin Ireland, pp 67–72. https://doi.org/10.1145/3379172.3391726
55. Zhou L, Gurrin C, Healy G, Joho H, Nguyen T-B, Albatal R, Hopfgartner F (2022) Overview of the ntcir-16 lifelog-4 task. In: Proceedings of the 16th NTCIR conference on evaluation of information access technologies, Tokyo