Memento 2.0: An Improved Lifelog Search Engine for LSC'22

Naushad Alam
Insight Centre for Data Analytics, Dublin City University
Dublin, Ireland
naushad.alam2@mail.dcu.ie

Yvette Graham
School of Computer Science and Statistics, Trinity College
Dublin, Ireland
ygraham@tcd.ie

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

ABSTRACT

In this paper, we present Memento 2.0, an improved version of our system which first participated in the Lifelog Search Challenge 2021. Memento 2.0 employs image-text embeddings derived from two CLIP models (ViT-L/14 and ResNet-50x64) and adopts a weighted ensemble approach to derive a combined final ranking. Our approach significantly improves the performance over the baseline LSC'21 system. We additionally make important updates to the system's user interface after analysing the shortcomings to make it more efficient and better suited to the needs of the Lifelog Search Challenge.

CCS CONCEPTS

• **Information systems** → **Retrieval models and ranking**; **Search interfaces**.

KEYWORDS

lifelog, information retrieval, semantic image representation

ACM Reference Format:
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 Lifelog Search Challenge (LSC '22), June 27–30, 2022, Newark, NJ, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3512729.3533006

1 INTRODUCTION

Lifelogging is a process to actively capture and record the daily experiences of an individual, namely, known as the *lifelogger*. Given the low cost of wearable cameras, sensors like Fitbit, and data storage facilities, capturing one's life through images has become feasible. However, passive lifelogging, which is logging oneself automatically in a continuous periodic manner, can become quite a memory-intensive task over longer periods of time as data gathered can be in excess of 1 TB per individual per year [10].

An even larger challenge additionally lies in creating value from the large amount of data gathered. Research in the domain of lifelogging and at a broad level egocentric vision has seen huge interest from the research community in the recent years due to the endless use cases and application areas.

One major application area is employing lifelogs to augment memory, especially in people suffering from neurodegenerative diseases such as Alzheimer's and Parkinson's disease. Lifelogs can also be used to aide reminiscence therapy for people suffering from dementia [4].

The Lifelog Search Challenge (LSC) [12] has attracted significant participation over the past 4 years since its inception in 2018. The LSC uses a multimodal dataset comprising egocentric images captured using a wearable camera apart from data coming through wearable sensors such as location, biometrics, sleep unlike other publicly available egocentric datasets which are generally unimodal and are mostly domain specific like EPIC-Kitchen [5] and EGO-CH [20]. Recently released dataset Ego-4D [9] is a large-scale egocentric real-world video dataset similar to lifelogs in a sense that both these datasets diversely capture the real world from the first-person perspective. However, Ego-4D diverges from the original lifelog dataset in two ways; first it is a video dataset unlike lifelogs which capture images at regular time intervals, and is unimodal in nature recording only videos from the subjects contrary to lifelogs which are multimodal recording metadata as well in the form of GPS location, user activity, and biometric data. Additionally, the new dataset does not capture a single subject at length but instead pools shorter length video data from multiple subjects which makes it less suitable for use-cases like long term memory retrieval. The multimodal nature of the dataset capturing all nuances of daily life, the format and the time-sensitive nature of the competition is what makes LSC a very unique information retrieval challenge.

In this paper, we present Memento 2.0, an improved version of our earlier system to participate in the 2022 edition of Lifelog Search challenge [12]. Our proposed system aims to further improve the semantic gap that exists between queries and images by leveraging embeddings from 2 larger CLIP [19] models to create an ensemble ranking mechanism. Memento 2.0 largely borrows the user interface from its predecessor making a few important tweaks to give it an advantage over its earlier version in the competition. Moreover, our newer system further enhances the temporal search algorithm making it more scalable and flexible for the end-user. Similar to the earlier system, Memento 2.0 also supports temporal navigation but with the added flexibility of time duration, deciding how far to navigate into the past or future.

2 RELATED WORK

The Lifelog Search Challenge has seen growing participation from a large number of researchers from across the globe over the last 4 years. A total of 16 systems competed in the 2021 edition of LSC [11] proposing novel and interesting ideas to solve the problem of lifelog retrieval.

Systems using virtual reality interfaces for lifelog querying have been popular since the first LSC challenge in 2018. In 2021 vitrivr-VR [23] and ViRMA [7] leveraged VR interface to do lifelog retrieval.

LSC '22, June 27–30, 2022, Newark, NJ, USA

ACM ISBN 978-1-4503-9239-6/22/06.
https://doi.org/10.1145/3512729.3533006

vitrivr-VR is an extension of the system vitrivr [14] and uses the same backend comprised of Cineast which is a feature extraction and query processing engine along with CottontailDB [8] while ViRMA used the M^3 model which translates the data into multi-dimensional space and is explored via the VR interface by projecting it into 3-D space. Similar to ViRMA [7], PhotoCube [22] also used the M^3 model and a web interface for search and navigation by creating a 3-D exploration cube on the screen.

MySceal 2.0 [24], who have been the winners for the last 2 editions of LSC approached the task by considering images as documents and proposed a concept weighing mechanism that determines the importance of an object by the area it occupies within the image. They further enhanced their system by adding color and text detection functionalities. LifeSeeker 3.0 [18] used a weighted Bag-of-Words model for free-text search and filtering.

Several systems employed user feedback to iteratively improve the search results in order to reach their target image. Exquisitor [15] trained a classifier based on user feedback without using any explicit query mechanism. They further added support to train multiple classifiers, results of which could be combined later to support queries that look for events with temporal relations. XQC [16] used the same backend as Exquisitor and proposed a mobile friendly interface to query lifelogs. SOMHunter [17] also leveraged user feedback which re-scores the images using a Bayesian style update.

FIRST 2.0 [25] tried to bridge the semantic gap between search queries and images by projecting them into a joint embedding space using a self-attention based model. Memento [1] leveraged image-text embeddings from the CLIP model to reduce the semantic gap between query and images. Voxento [2] used the same backend as Memento but employed an interactive voice enabled user interface.

Systems such as LifeGraph [21] and LifeConcept [3] tried to use external knowledge graphs to enrich the lifelog data in order to execute semantically complex queries easily.

Our proposed system Memento 2.0 further reduces the semantic gap between query and images through a weighted summation of scores from two larger recently released CLIP [19] models from OpenAI to rank the images. The approach significantly improves performance as compared to our baseline system Memento 1.0. The ability of our system to take natural language instructions gives us a clear advantage in the challenge as the LSC queries are usually very descriptive and give out finer details which directly go as input to the system without much tweaking.

3 SYSTEM OVERVIEW

In this section, we present an overview of the new LSC'22 Dataset and discuss the improvements in our search and ranking functionality, as well as the modifications in the user interface. We further discuss the new temporal search and navigation functionality of Memento 2.0.

3.1 LSC'22 Data

The Lifelog Search Challenge 2022 will use a multimodal dataset collected from a single individual across 18 months. The dataset includes egocentric point-of-view images collected using a narrative clip device during 2019-2020. Like previous editions of LSC the data also includes two sets of metadata but with significant changes,

- **Visual Concepts:** For every image in the dataset visual concepts are provided, which consist of, detected objects in the image, image caption along with caption confidence score and text detected from images using OCR models.
- **Metadata:** Like metadata for previous competitions, this dataset also contains location, timezone, elevation, and biometric data such as calories burnt, heart rate, step count, etc., along with new additions for LSC'22 like sleep data, sleep efficiency, music data etc.

3.2 Enhanced Image Representation

Our participating system last year used embeddings generated from the zero-shot CLIP model [19] using a Vision Transformer [6] (ViT-B/32) architecture which was able to efficiently capture the visual semantics without any sort of data specific fine-tuning. The system was able to correctly submit responses to 16 out of 23 LSC'21 queries and managed the 6th position on the leaderboard. Figure 2 shows the number of correct and incorrect submissions made by the participating systems where they are ordered left to right as per the final leaderboard. Voxento 2.0, which shared the same backend as ours stood at 4th position by answering 18 queries correctly. Interestingly, Memento 1.0 and Voxento 2.0 collectively answered 21 queries correctly out of the total 23 queries in LSC'21, surpassing LifeSeeker 3.0 tally of 20 which was the highest for last year's competition.This shows how well the CLIP model transfers to lifelog images and it's robustness towards distribution shift in data. The result further asserts the crucial role user interface and human factors play in this competition.

Memento 2.0 aims to incorporate enhanced image embeddings to improve the search result quality as well as address the issues with the user interface of previous system. Our proposed system for LSC'22 leverages the embeddings generated using two larger recently released CLIP models, one of which is a ResNet-50 [13] model (ResNet50x64) while the other one is a Vision Transformer [6] model (ViT-L/14). We evaluated the performance of both these newer models separately on LSC'19 queries and found them to be superior than the older (ViT-B/32) model which Memento 1.0 used, where the performance of Vit-L/14 is better than ResNet50x64.

ViT-L/14 generates 768-d image embeddings while ResNet50x64 generates 1024-d embeddings. We observed the rankings of ViT-L/14 to be better for a majority of our evaluation queries, however ResNet's ranking was superior for the remaining ones which led us to experiment with a new ranking mechanism by taking weighted sum of the scores from these two models. We observed (discussed in Section 4) that the new weighted scoring mechanism taking in ViT-L/14 scores and ResNet50x64 scores in a 3:1 ratio to be outperforming individual models.

3.3 Improved Temporal Search and Navigation

Memento 2.0 supports an enhanced version of the temporal search functionality of its predecessor system. The scope of the older algorithm was limited as it could temporally search only across the top-2000 ranked images hence constraining the search space and making it less effective in certain circumstances.

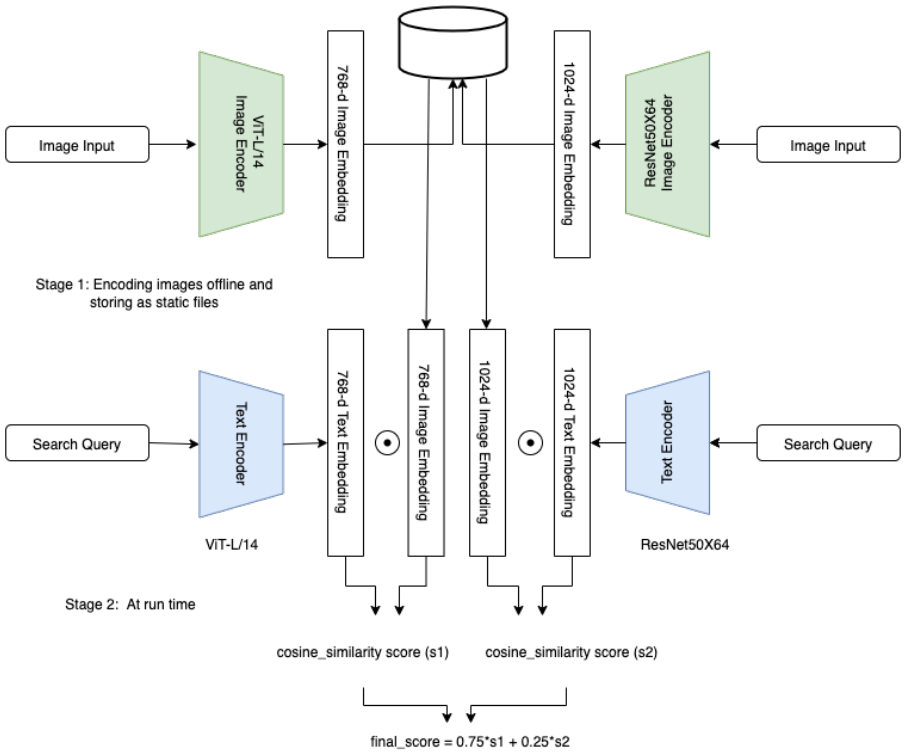

Figure 1: Memento 2.0: System Architecture. Initially (Stage 1), the raw lifelog images are encoded using the image encoders of respective models and stored as static files. Further at run time (Stage 2), the search query is passed on to the respective text encoders and cosine similarity is calculated by comparing respective image and text embeddings

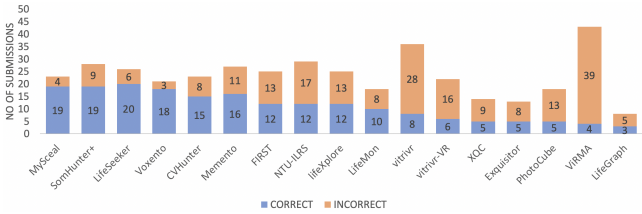

Figure 2: Number of Correct and Incorrect Submissions across participating systems in LSC'21. Systems are ordered left to right as per their position on the final leader board.

Our newly proposed algorithm performs a temporal search on a large chunk of the corpus which allows the user to search for a target event in the context of a temporally close past or future event. The users also gets to choose how far to look in the past or future by specifying a time duration before they initiate the search process.

The temporal search functionality of Memento 2.0 has the following execution steps:

(1) The user needs to inputs a main event and either a past event or a future event, or both to initiate a temporal search.
(2) Similar to the previous algorithm, this algorithm also tries to search and rank the main event first. We take a subset of ranked images (top-n) based on empirically derived cosine similarity threshold of 0.15 as searching temporally across the entire corpus in a brute-force manner is not feasible due to time constraints. Using the similarity score to subset the ranked list ensures that we don't leave behind any relevant image which was very likely with the hard threshold of top-2000 which the previous algorithm imposed.
(3) The algorithm iterates through the subset from (2), to search for past and future event in their respective search spaces. The user can choose how far it should look in the past or future temporal window by specifying the time duration before initiating the search process.
(4) The algorithm assigns temporal scores (past and future) to every image in the initial subset, 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 and initial cosine similarity score based on which the images are re-ranked and rendered on screen.

Memento 2.0, like its predecessor, supports sequential browsing of previous and next non-blurred images around a probable target image but now the user has the flexibility to choose how far to look by specifying a time duration. It was earlier limited by the number of images instead of time duration which in certain circumstances didn't help in the search process.

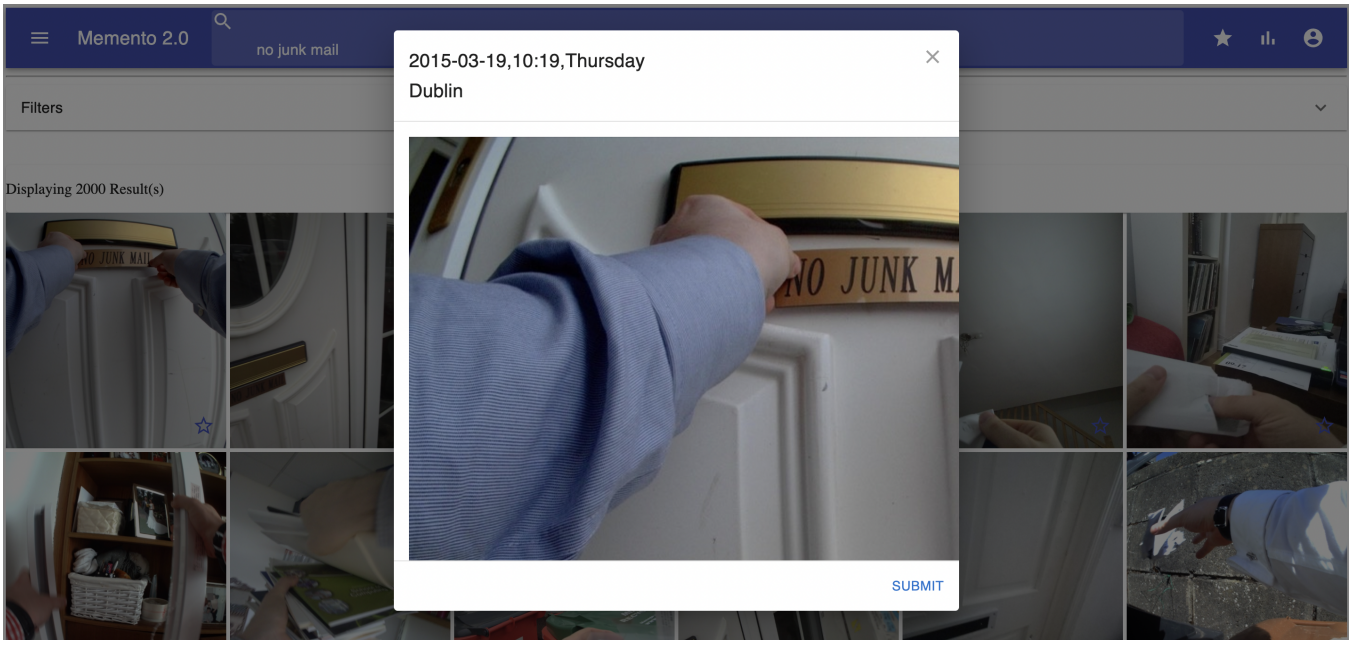

Figure 3: Memento 2.0: Primary Search Interface with zoom and submit popup window.

	t	@1	@3	@5	@10	@20	@50
ViT-B/32 (Baseline)	0 sec	8.33	25.00	29.17	29.17	37.50	50.00
	30 sec	8.33	25.00	25.00	33.33	33.33	54.17
	60 sec	12.50	29.17	29.17	41.67	54.17	75.00
ViT-L/14	0 sec	20.83	33.33	41.67	50.00	54.17	62.50
	30 sec	33.33	41.67	41.67	45.83	58.33	66.67
	60 sec	37.50	**45.83**	**45.83**	54.17	62.50	79.17
ResNet50X64	0 sec	25.00	29.17	29.17	29.17	45.83	58.33
	30 sec	25.00	33.33	41.67	50.00	58.33	62.50
	60 sec	25.00	37.50	41.67	54.17	58.33	75.00
Ensemble 1:1	0 sec	29.17	33.33	33.33	41.67	58.33	70.83
	30 sec	33.33	37.50	45.83	50.00	58.33	66.67
	60 sec	33.33	37.50	45.83	54.17	**70.83**	79.17
Ensemble 1:3	0 sec	29.17	33.33	33.33	41.67	50.00	66.67
	30 sec	25.00	33.33	41.67	50.00	54.17	62.50
	60 sec	33.33	37.50	41.67	50.00	58.33	79.17
Ensemble 3:1	0 sec	29.17	29.17	33.33	54.17	58.33	70.83
	30 sec	37.50	41.67	**45.83**	54.17	58.33	70.83
	60 sec	**41.67**	**45.83**	**45.83**	**62.50**	**70.83**	**83.33**

Table 1: Hit@K calculated for all 6 models at different amounts of elapsed times, t and K values across 24 evaluation topics for LSC'19. Highest value in each column is highlighted in bold

3.4 Modifications in the User Interface

Memento 2.0 borrows its user interface from the earlier system but with improvements and enhancements to make it even better suited to the time-sensitive nature of the Lifelog Search Challenge.

- **Modified Primary Search Interface:** Investigating the issues from LSC'21 we have made minor changes to the primary search interface. Firstly, we have added zoom and submit functionality which will help the user in two ways;
  (i) the user can quickly get a better look at the image as we missed out on a few queries from last year's competition which revealed information that was tough to spot. e.g. the query looking for an orange ride-on suitcase where 'ride-on suitcase' was written over it but was too small to read or the query revealing the name of the person written on the flight boarding pass;
  (ii) it will also help in scenarios where the user has identified the correct response and needs to quickly submit it for evaluation. The response submitting mechanism in Memento 1.0 was only through the starred images interface which consumed a few additional seconds hence the system losing its competitive edge despite being the fastest to locate the correct response.

  Figure 3 shows a snapshot of the modified primary search interface with quick inspect and submit functionality.
- **Temporal Search Interface:** We have made modifications to the temporal search interface to accommodate the newer proposed algorithm. The interface now allows the user to choose a time duration specifying how far in the past or in the future to look for the event besides specifying what exactly to look for. Figure 4 shows a snapshot of the temporal search interface rendered over the primary search interface.

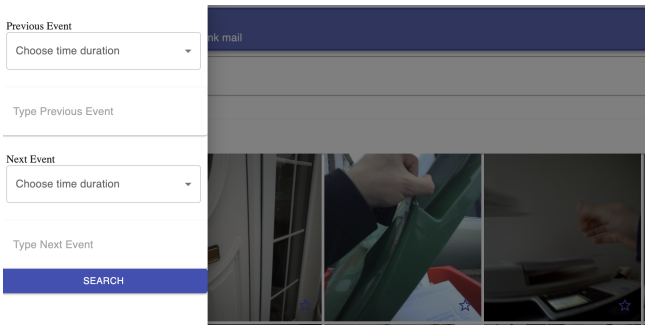

Figure 4: Temporal search interface rendered in an overlay window. It takes in past/future event along with time duration to constraint the search space on both sides

- **Quick Filtering Interface:** Memento 2.0 borrows the visual data filtering interface from last year's system but also includes an additional interface which is located on the main screen. The purpose here is quick accessibility of the most commonly used search filters without moving away from the search results.

4 SYSTEM EVALUATION

To validate the efficacy of the newer models as well as ensemble models over our baseline system Memento 1.0 we evaluated all approaches on 24 evaluation topics from LSC 2019.

Our evaluation approach is similar to last year's approach where for each query we only consider the information revealed to us by t=60 seconds. We manually create evaluation queries at t=0, 30 and 60 seconds for this exercise as the performance of the model depends a lot on the input query. We therefore handcraft the queries to keep it concise, mimicking the writing style of image captions. The reason behind considering only partial information for evaluating the system is the way how LSC queries are typically structured, where they reveal visually descriptive information early on and more specific pieces of information like time, date, place, etc., at later stages.The backend models however only understand visual concepts and don't do well when given specific information like date, time etc.

We evaluate the following 6 models on 24 evaluation queries from LSC 2019:

(1) ViT-B/32: Baseline model which powered the backend of Memento 1.0 in last year's challenge
(2) ViT-L/14: A larger Vision Transformer model released as successor to (1). It generates 768 dimensional image-text embeddings.
(3) ResNet50x64: A ResNet-50 model using 64x the compute of a ResNet-50. It generates 1024 dimensional image-text embeddings.
(4) Ensemble ViT-L/14 and ResNet50x64 (1:1): Weighted sum of cosine scores from both the models in 1:1 ratio.
(5) Ensemble ViT-L/14 and ResNet50x64 (1:3): Weighted sum of cosine scores from ViT-L/14 and ResNet50x64 in a 1:3 ratio.
(6) Ensemble ViT-L/14 and ResNet50x64 (3:1): Weighted sum of cosine scores from ViT-L/14 and ResNet50x64 in a 3:1 ratio.

We evaluate the models on Hit@K metric which can be defined as finding at least one target image among top-K images in the result set. Table 1 shows the hit percentages calculated from all 6 models for 24 LSC 2019 evaluation topics at different values of K and t. The Ensemble 3:1 model outperforms the larger ViT-L/14 and ResNet50x64 models at each K value by considerable margin. Moreover, the hit-ratio observed for Ensemble 3:1 at lower K values e.g 1,3,5 is significantly higher than all other models.

Overall at t=60, we observe that for 80.33% (20 out of 24 topics) of the topics, we have at least one target image in top-50 results.

5 CONCLUSION AND FUTURE WORK

In this work, we present Memento 2.0, an improved version of our last year's participating system in LSC'21. We discuss the robustness of the CLIP model to distribution shift in data and how well it generalizes to lifelogs in Section 3.2. We further made improvements to our search and ranking model using an ensembling approach which improves the performance by a significant margin. Additionally, we did a few small but important tweaks to the system's user interface to make it even more efficient. We plan to explore the feasibility of an end-to-end conversational system to partake in the future Lifelog Search Challenge.

6 ACKNOWLEDGEMENTS

This publication 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.

REFERENCES

[1] Naushad Alam, Yvette Graham, and Cathal Gurrin. 2021. Memento: A Prototype Lifelog Search Engine for LSC'21. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 53–58. https://doi.org/10.1145/3463948.3469069

[2] Ahmed Alateeq, Mark Roantree, and Cathal Gurrin. 2021. Voxento 2.0: A Prototype Voice-Controlled Interactive Search Engine for Lifelogs. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 65–70. https://doi.org/10.1145/3463948.3469071

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

[4] 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*. ACM, Dublin Ireland, 383–387. https://doi.org/10.1145/3372278.3391927

[5] Dima Damen, Hazel Doughty, Giovanni Maria Farinella, , Antonino Furnari, Jian Ma, Evangelos Kazakos, Davide Moltisanti, Jonathan Munro, Toby Perrett, Will Price, and Michael Wray. 2021. Rescaling Egocentric Vision: Collection, Pipeline and Challenges for EPIC-KITCHENS-100. *International Journal of Computer Vision (IJCV)* (2021). https://doi.org/10.1007/s11263-021-01531-2

[6] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv:2010.11929 [cs]* (June 2021). http://arxiv.org/abs/2010.11929 arXiv: 2010.11929.

[7] Aaron Duane and Bjorn Þór Jónsson. 2021. ViRMA: Virtual Reality Multimedia Analytics 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, 29–34. https://doi.org/10.1145/3463948.3469067

[8] Ralph Gasser, Luca Rossetto, Silvan Heller, and Heiko Schuldt. 2020. Cottontail DB: An Open Source Database System for Multimedia Retrieval and Analysis. In *Proceedings of the 28th ACM International Conference on Multimedia* (Seattle, WA, USA) *(MM '20)*. Association for Computing Machinery, New York, NY, USA, 4465–4468. https://doi.org/10.1145/3394171.3414538

[9] Kristen Grauman, Andrew Westbury, Eugene Byrne, Zachary Chavis, Antonino Furnari, Rohit Girdhar, Jackson Hamburger, Hao Jiang, Miao Liu, Xingyu Liu, Miguel Martin, Tushar Nagarajan, Ilija Radosavovic, Santhosh Kumar Ramakrishnan, Fiona Ryan, Jayant Sharma, Michael Wray, Mengmeng Xu, Eric Zhongcong Xu, Chen Zhao, Siddhant Bansal, Dhruv Batra, Vincent Cartillier, Sean Crane, Tien Do, Morrie Doulaty, Akshay Erapalli, Christoph Feichtenhofer, Adriano Fragomeni, Qichen Fu, Abrham Gebreselasie, Cristina Gonzalez, James Hillis, Xuhua Huang, Yifei Huang, Wenqi Jia, Weslie Khoo, Jachym Kolar, Satwik Kottur, Anurag Kumar, Federico Landini, Chao Li, Yanghao Li, Zhenqiang Li, Karttikeya Mangalam, Raghava Modhugu, Jonathan Munro, Tullie Murrell, Takumi Nishiyasu, Will Price, Paola Ruiz Puentes, Merey Ramazanova, Leda Sari, Kiran Somasundaram, Audrey Southerland, Yusuke Sugano, Ruijie Tao, Minh Vo, Yuchen Wang, Xindi Wu, Takuma Yagi, Ziwei Zhao, Yunyi Zhu, Pablo Arbelaez, David Crandall, Dima Damen, Giovanni Maria Farinella, Christian Fuegen, Bernard Ghanem, Vamsi Krishna Ithapu, C. V. Jawahar, Hanbyul Joo, Kris Kitani, Haizhou Li, Richard Newcombe, Aude Oliva, Hyun Soo Park, James M. Rehg, Yoichi Sato, Jianbo Shi, Mike Zheng Shou, Antonio Torralba, Lorenzo Torresani, Mingfei Yan, and Jitendra Malik. 2022. Ego4D: Around the World in 3,000 Hours of Egocentric Video. *arXiv:2110.07058 [cs]* (March 2022). http://arxiv.org/abs/2110.07058 arXiv: 2110.07058.

[10] Cathal Gurrin. 2021. Personal Data Matters: New Opportunities from Lifelogs. In *2021 16th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*. 1–3. https://doi.org/10.1109/iSAI-NLP54397.2021.9678155

[11] Cathal Gurrin, Klaus Schoeffmann, Bjorn Thor Jonsson, Duc Tien Dang Nguyen, Jakub Lokoc, Luca Rossetto, Minh-Triet Tran, Wolfgang Hurst, and Graham Healy. 2021. An Introduction to the Fourth Annual Lifelog Search Challenge, LSC'21. In *ICMR '21, The 2021 International Conference on Multimedia Retrieval*. ACM, Taipei, Taiwan.

[12] Cathal Gurrin, Liting Zhou, Graham Healy, Bjorn Thor Jonsson, Duc Tien Dang Nguyen, Jakub Lokoc, Minh-Triet Tran, Wolfgang Hurst, Luca Rossetto, and Klaus Schoeffmann. 2022. An Introduction to the Fifth Annual Lifelog Search Challenge, LSC'22. In *ICMR '22, The 2022 International Conference on Multimedia Retrieval*. ACM, Newark, NJ, USA.

[13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Las Vegas, NV, USA, 770–778. https://doi.org/10.1109/CVPR.2016.90

[14] 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

[15] Omar Shahbaz Khan, Aaron Duane, Björn Þór Jónsson, Jan Zahálka, Stevan Rudinac, and Marcel Worring. 2021. Exquisitor at the Lifelog Search Challenge 2021: Relationships Between Semantic Classifiers. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 3–6. https://doi.org/10.1145/3463948.3469255

[16] Emil Knudsen, Thomas Holstein Qvortrup, Omar Shahbaz Khan, and Björn Þór Jónsson. 2021. XQC at the Lifelog Search Challenge 2021: Interactive Learning on a Mobile Device. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 89–93. https://doi.org/10.1145/3463948.3469063

[17] Jakub Lokoč, František Mejzlik, Patrik Veselý, and Tomáš Souček. 2021. Enhanced SOMHunter for Known-Item Search in Lifelog Data. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 71–73. https://doi.org/10.1145/3463948.3469074

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

[19] 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. *arXiv:2103.00020 [cs]* (Feb. 2021). http://arxiv.org/abs/2103.00020 arXiv: 2103.00020.

[20] Francesco Ragusa, Antonino Furnari, Sebastiano Battiato, Giovanni Signorello, and Giovanni Maria Farinella. 2020. EGO-CH: Dataset and fundamental tasks for visitors behavioral understanding using egocentric vision. *Pattern Recognition Letters* 131 (March 2020), 150–157. https://doi.org/10.1016/j.patrec.2019.12.016

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

[22] Jihye Shin, Alexandra Waldau, Aaron Duane, and Björn Þór Jónsson. 2021. PhotoCube 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, 59–63. https://doi.org/10.1145/3463948.3469073

[23] Florian Spiess, Ralph Gasser, Silvan Heller, Luca Rossetto, Loris Sauter, Milan van Zanten, and Heiko Schuldt. 2021. Exploring Intuitive Lifelog Retrieval and Interaction Modes in Virtual Reality with Vitrivr-VR. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 17–22. https://doi.org/10.1145/3463948.3469061

[24] 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

[25] Hoang-Phuc Trang-Trung, Thanh-Cong Le, Mai-Khiem Tran, Van-Tu Ninh, Tu-Khiem Le, Cathal Gurrin, and Minh-Triet Tran. 2021. Flexible Interactive Retrieval SysTem 2.0 for Visual Lifelog Exploration 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, 81–87. https://doi.org/10.1145/3463948.3469072