

Overview of ImageCLEFlifelog 2017: Lifelog Retrieval and Summarization

Duc-Tien Dang-Nguyen¹, Luca Piras², Michael Riegler³, Giulia Boato⁴,
Liting Zhou¹, and Cathal Gurrin¹

¹ Insight Centre for Data Analytics, Dublin City University
zhou.liting2@mail.dcu.ie, {duc-tien.dang-nguyen, cathal.gurrin}@dcu.ie

² DIEE, University of Cagliari
luca.piras@diee.unica.it

³ Simula Research Laboratory
michael@simula.no

⁴ DISI, University of Trento
boato@disi.unitn.it

Abstract. Despite the increasing number of successful related workshops and panels, lifelogging has rarely been the subject of a rigorous comparative benchmarking exercise. Following the success of the new lifelog evaluation task at NTCIR-12,¹ the first ImageCLEF 2017 LifeLog task aims to bring the attention of lifelogging to a wide audience and to promote research into some of the key challenges of the coming years. The ImageCLEF 2017 LifeLog task aims to be a comparative evaluation framework for information access and retrieval systems operating over personal lifelog data. Two subtasks were available to participants; all tasks use a single mixed modality data source from three lifeloggers for a period of about one month each. The data contains a large collection of wearable camera images, an XML description of the semantic locations, as well as the physical activities of the lifeloggers. Additional visual concept information was also provided by exploiting the Caffe CNN-based visual concept detector. For the two sub-tasks, 51 topics were chosen based on the real interests of the lifeloggers. In this first year three groups participated in the task, submitting 19 runs across all subtasks, and all participants also provided working notes papers. In general, the groups performance is very good across the tasks, and there are interesting insights into these very relevant challenges.

1 Introduction

The availability of a large variety of personal devices, such as smartphones, video cameras as well as wearable devices that allow capturing pictures, videos and audio clips of every moment of our life is creating vast archives of personal data where the totality of an individual's experiences, captured multi-modally through digital sensors are stored permanently as a personal multimedia archive.

¹ <http://ntcir-lifelog.computing.dcu.ie/NTCIR12/>

These unified digital records, commonly referred to as *lifelogs*, have gathered increasing attention in recent years within the research community. This happened due to the need for, and challenge of, building systems that can automatically analyse these huge amounts of data in order to categorize, summarize and query them to retrieve information that the user may need. For example, lifeloggers may want to recall some events that they do not remember clearly or to know some insights of their activities at work to improve the performance. Figure 1 shows an example of what a lifelogger wants to retrieve.



Fig. 1. An example lifelogging information need: 'Show me all eating moments'.

The ImageCLEF 2017 LifeLog task is inspired by the general image annotation and retrieval tasks that have been part of ImageCLEF since 2003. In the early years the focus was on retrieving relevant images from a web collection given (multilingual) queries, from 2006 onwards annotation tasks were also held, initially aimed at object detection, but more recently also covering semantic concepts [9,10,12,11]. In the last two editions [4,5], the image annotation task was expanded to concept localization and also natural language sentential description of images. As there is an increased interest in recent years in research combining text and vision, this year task, changing a little the focus of the retrieval object, aim at further stimulating and encouraging multi-modal research that uses text and visual data, and natural language processing for image retrieval and summarization.

This paper presents the overview of the first edition of the ImageCLEF 2017 LifeLog task, one of the four benchmark campaigns organized by ImageCLEF [6] in 2017 under the CLEF initiative². Section 2 describes the task in detail, including the participation rules and the provided data and resources. Section 3 presents and discusses the results of the submissions received for the task. Finally, Section 4 concludes the paper with final remarks and future outlooks.

² <http://www.clef-initiative.eu>

2 Overview of the Task

2.1 Motivation and Objectives

Based on the successful of NTCIR-12 lifelog task, we present here new tasks which aim to advance the state-of-the-art research in lifelogging as an application of information retrieval. By proposing these tasks at ImageCLEF, we intent to enlarge the association, by linking lifelogging researchers to the image retrieval community. We also hope that novel approaches based on multi-modal retrieval will be able to provide new insights from the personal lifelogs.

2.2 Challenge Description

The ImageCLEF 2017 LifeLog task³ aims to be a comparative evaluation of information access and retrieval systems operating over personal lifelog data. The task consisted of two sub-tasks, both allow participation independently. These sub-tasks are:

- Lifelog Retrieval Task (LRT);
- Lifelog Summarization Task (LST).

Lifelog retrieval task

The participants had to analyse the lifelog data and for several specific queries, return the correct answers. For example: *Shopping for Wine: Find the moment(s) when I was shopping for wine in the supermarket* or *The Metro: Find the moment(s) when I was riding a metro*. The ground truth for this sub-task was created by extending the queries from the NTCIR-12 dataset, which already provides a sufficient ground truth.

Lifelog summarization task

In this sub-task the participants had to analyse all the images and summarize them according to specific requirements. For instance: *Public Transport: Summarize the use of public transport by a user. Taking any form of public transport is considered relevant, such as bus, taxi, train, airplane and boat. The summary should contain all different day-times, means of transport and locations, etc.*

Particular attention had to be paid to the diversification of the selected images with respect to the target scenario. The ground truth for this sub-task was created utilizing crowdsourcing and manual annotations.

2.3 Dataset

The Lifelog dataset⁴ consists of data from three lifeloggers for a period of about one month each. The data contains 88,124 wearable camera images (approximately two images per minute), an XML description of 130 associated semantic

³ Challenge website at <http://www.imageclef.org/2017/lifelog>

⁴ Dataset available at <http://imageclef-lifelog.computing.dcu.ie/2017/>

Table 1. Statistics of Lifelog Dataset.

Number of Lifeloggers	3
Size of the Collection (Images)	88,124 images
Size of the Collection (Locations)	130 locations
Number of LRT Topics	36 (16 for devset, 20 for testset)
Number of LST Topics	15 (5 for devset, 10 for testset)

locations (e.g. Starbucks cafe, McDonalds restaurant, home, work) and the four physical activities: walking, cycling, running and transport of the lifeloggers at a granularity of one minute. A summary of the data collection is shown in Table 1.

In order to reduce the barriers-to-participation, the output of the Caffe CNN-based visual concept detector [7] was included in the test collection as additional metadata. This classifier provided labels and probabilities for 1,000 objects in every image. The accuracy of the Caffe visual concept detector is variable and is representative of the current generation of off-the-shelf visual analytics tools.

2.4 Topic and Ground-truth

Aside from the data, the test collection included a set of topics (queries) that were representative of the real-world information needs of lifeloggers. There were 36 ad-hoc search topics, 16 for the development set and 20 for the test set, representing the challenge of retrieval for the LRT task (see Tables 2 and 3) and 15 search topics, 5 for the development set and 10 for the test set, for the Summarization sub-task (see Tables 4 and 5). For full descriptions of the topics, please see the Appendix A.

The ground-truth of retrieval topics were created by the task organizers, with the verification of the lifeloggers. For summarization topics, task organizers manually classified the images into the clusters which are provided by the lifeloggers. All the results are then verified by the lifeloggers once more time before publishing.

Table 2: Search topics for the development set in Lifelog Retrieval Task

T1. Presenting/Lecturing
Query: Find the moment(s) when user u1 was lecturing a group of people in a classroom environment.
T2. On the Bus or Train
Query: Find the moment(s) when user u1 was taking a bus or a train in his home country.
T3. Having a Drink
Query: Find the moment(s) when user u1 was having a drink in a bar with someone.

T4. Riding a Red (and Blue) Train
Query: Find the moment(s) when user u1 was riding a red (and blue) coloured train.
T5. The Rugby Match
Query: Find the moment(s) when user u1 was watching rugby football on a television when not at home.
T6. Costa Coffee
Query: Find the moment(s) when user u1 was in Costa Coffee.
T7. Antiques Store
Query: Find the moment(s) when user u1 was browsing in an antiques store.
T8. ATM
Query: Find the moment(s) when user u1 was using an ATM machine.
T9. Shopping for Fish
Query: Find the moment(s) when user u1 was shopping for fish in the supermarket.
T10. Cycling home
Query: Find the moment(s) when user u2 was cycling home from work.
T11. Shopping
Query: Find the moment(s) in which user u2 was grocery shopping in the supermarket.
T12. In a Meeting
Query: Find the moment(s) in which user u2 was in a meeting at work with 2 or more people.
T13. Checking the Menu
Query: Find the moment(s) when user u2 was standing outside a restaurant checking the menu.
T14. Watching TV
Query: Find the moment(s) when user u3 was watching TV.
T15. Writing
Query: Find the moment(s) when user u3 was writing on a paper using a pen or pencil.
T16. Drinking in a Pub
Query: Find the moment(s) when user u3 was drinking in a pub with friends or alone.

Table 3: Search topics for the test set in Lifelog Retrieval Task

T1. Using laptop out of office
Query: Find the moment(s) in which user u1 was using his laptop outside the working places.

T2. On stage
Query: Find the moment(s) in which user u1 was giving a talk as a presenter/speaker.
T3. Shopping in the electronic market.
Query: Find the moment(s) in which user u1 was shopping in the electronic market.
T4. Jogging in the park
Query: Find the moments(s) in which user u1 was jogging in a park.
T5. In a Meeting 2
Query: Find the moment(s) that user u1 was in a meeting at work.
T6. Watching TV 2
Query: Summarize the moment(s) when user u1 was watching TV.
T7. Pizza and Friends
Query: Find the moment(s) in which user u1 was eating pizza in the restaurant with friends.
T8. Working in the air.
Query: Find the moment(s) in which user u1 was using computer in the airplane.
T9. Playing Guitar
Query: Find the moment(s) in which user u2 was playing guitar.
T10. Exercise in the gym
Query: Find the moment(s) in which user u2 was doing exercise in the gym.
T11. Eating fruits
Query: Find the moment(s) in which user u2 was having fruits.
T12. Brushing or washing face
Query: Find the moment(s) in which user u2 was brushing or washing face.
T13. Eating 2
Query: Find the moment(s) when user u2 was eating or drinking.
T14. At McDonald
Query: Find the moment(s) in which user u2 was at McDonald for eating or just for relaxing.
T15. Viewing a statue
Query: Find the moment(s) in which user u2 was viewing a statue.
T16. ATM
Query: Find the moment(s) when user u2 was using an ATM machine.
T17. Have party with friends at friends home.
Query: Find the moment(s) in which user u3 was attending a party with many friends at a friends home.
T18. Shopping in the butchers shop.
Query: Find the moment(s) in which user u3 was consuming in the butchers shop.

T19. Buying a ticket via ticket machine.
Query: Find the moment(s) in which user u3 was buying a ticket via ticket machine.
T20. Shopping 2
Query: Find the moment(s) in which user u3 doing shopping.

Table 4: Search topics for the development set in Lifelog Summarization Task

T1. Eating
Query: Summarize the moment(s) when user u1 was eating or drinking.
T2. Social Drinking
Query: Summarize the the social drinking habits of user u1.
T3. Shopping
Query: Summarize the moment(s) in which user u1 doing shopping.
T4. In a Meeting
Query: Summarize the activities of user u2 in a meeting at work.
T5. Watching TV
Query: Summarize the moment(s) when user u3 was watching TV.

Table 5: Search topics for the test set in Lifelog Summarization Task

T1. In a Meeting 2
Query: Summarize the activities of user u1 in a meeting at work.
T2. Watching TV 2
Query: Summarize the moment(s) when user u1 was watching TV.
T3. Using laptop outside the office
Query: Summarise the moment(s) in which user u1 was using his laptop outside the working places.
T4. Working at home
Query: Find the moment(s) in which user u1 was working at home.
T5. Eating 2
Query: Summarize the moment(s) when user u2 was eating or drinking.
T6. Social Drinking 2
Query: Summarize the the social drinking habits of user u2.
T7: Sightseeing
Query: Summarize the moments when the user u2 seeing street, people, landscape, etc. when he was traveling to other cities or countries.

T8. Transporting
Query: Summarize the moments when user u2 using public transportation.
T9. Preparing meals
Query: Find the moment(s) in which user u3 was preparing meals at home.
T10. Shopping 2
Query: Summarize the moment(s) in which user u3 was doing shopping.

2.5 Performance Measures

For the *Lifelog Retrieval Task* evaluation metrics based on NDCG (Normalized Discounted Cumulative Gain) at different depths were used, i.e., $NDCG@N$, where N varies based on the type of the topics, for the recall oriented topics N was larger (> 20), and for the precision oriented topics N was smaller N (5, 10 or 20).

In the *Lifelog Summarization Task* classic metrics were deployed:

- Cluster Recall at X ($CR@X$) a metric that assesses how many different clusters from the ground truth are represented among the top X results;
- Precision at X ($P@X$) measures the number of relevant photos among the top X results;
- F1-measure at X ($F1@X$) the harmonic mean of the previous two.

Various cut off points were considered, e.g., $X = 5, 10, 20, 30, 40, 50$. Official ranking metrics this year was the **F1-measure@10** or images, which gives equal importance to diversity (via $CR@10$) and relevance (via $P@10$).

Participants were also encouraged to undertake the sub-tasks in an interactive or automatic manner. For interactive submissions, a maximum of five minutes of search time was allowed per topic. In particular, the organizers would like to emphasize methods that allowed interaction with real users (via Relevance Feedback (RF), for example), i.e., beside of the best performance, the way of interaction (like number of iterations using RF), or innovation level of the method (for example, new way to interact with real users) has been evaluated.

3 Evaluation Results

3.1 Participation

This year, being the first edition of this challenging task, the participation was not so high but, taking into account the number of teams that downloaded the dataset (11 registered teams signed the copyright form), there are grounds for this number to increase considerably over coming iterations of the task. In total the three groups that took part in the task and submitted overall 19 runs. All three participating groups submitted a working paper describing their system, thus for these there were specific details available:

- **I2R:** [8] The team from Institute for Infocomm Research, A*STAR, Agency for Science Technology and Research (A*STAR), Singapore, represented by Ana Garcia del Molino, Bappaditya Mandal, Jie Lin, Joo Hwee Lim, Vigneshwaran Subbaraju and Vijay Chandrasekhar.
- **UPB:** [3] The team from University Politehnica of Bucharest, Romania, represented by Mihai Dogariu and Bogdan Ionescu.
- **Organizers:** [13] The team from Insight Centre for Data Analytics (Dublin City University), University of Cagliari, Simula Research Laboratory, University of Trento, was represented by Liting Zhou, Luca Piras, Michael Rieger, Giulia Boato, Duc-Tien Dang-Nguyen, and Cathal Gurrin.

Table 6 provide the main key details for the submitted runs of each group describing their system for each subtask. This table serves as a summary of the systems, and are also quite illustrative for quick comparisons. For a more in-depth look at the systems of each team, please refer to the corresponding papers.

3.2 Results for Subtask 1: Retrieval

Unfortunately only the Organizers team submitted runs for the Retrieval Subtask [13]. They proposed an approach composed by several step. First of all they grouped similar moments together based on time and concepts and, applying this chronological-based segmentation, they turned the problem of images retrieval into image segments retrieval. Then, starting from a topic query, they transformed it into small inquiries, where each of them is asking for a single piece of information of concepts, location, activity, and time. The moments that matched all of those requirements are returned as the retrieval results. In the end, in order to remove non-relevant images, a filtering step is applied on the retrieved images, by removing blurred and images that covered mainly by huge object or by the arms of the user.

On the Retrieval Subtask the Organizers team submitted 3 runs summarized in Table 7 The first run (*baseline*) exploited only time and the concepts information. Every single image has been considered as the basic unit and the retrieval just returns all images that contains the concepts extracted from the topics. They submitted this run as reference with the purpose that any other approach should obtain better performance than this. In the second run (*Segmentation*), the Organizers team introduced also the segmentation so as basic unit of retrieval has been used the segment, not image. In the last run (*Fine-Tuning*), the “translation” of the query into small inquiries has been applied.

3.3 Results for Subtask 2: Summarization

For Subtask 2, participants were asked to analyse all the lifelog images and summarize them according to specific requirements (see the topics in Tables 4 and 5). All the three teams, I2R [8], UPB [3] and Organizers* [13], participated in this subtask. Table 8 shows the **F1-measure@10**, for all submitted runs by participants.

Table 6. Submitted runs for ImageCLEFlifelog 2017 task.

Lifelog Retrieval Subtask.		
Team	Run	Description
Organizers [13]	Baseline	Baseline method, fully automatic.
	Segmen- tation	Apply segmentation and automatic retrieval based on concepts.
	Fine- tunning	Apply segmentation and fine-tunning. Using all information.
Lifelog Summarization Subtask.		
I2R [8]	Run 1	Parameters learned for maximum F1 score. Using only visual information.
	Run 2	Parameters learned for maximum F1 score. Using visual and metadata information.
	Run 3	Parameters learned for maximum F1 score. Using metadata.
	Run 4	Re-clustering in each iteration; 20% extra clusters. Using visual, metadata and interactive.
	Run 5	No re-clustering. 100% extra clusters. Using visual, metadata and interactive.
	Run 6	Parameters learned for maximum F1 score. Using visual, meta- data, and object detection.
	Run 7	Parameters learned for maximum F1 score, w/ and w/o object detection. Using visual, metadata, and object detection.
	Run 8	Parameters learned for maximum F1 score. Using visual infor- mation and object detection.
	Run 9	Parameters learned for maximum precision. Using visual and metadata information.
	Run 10	No re-clustering. 20 % extra clusters. Using visual, metadata and interactive.
UPB [3]	Run 1	Textual filtering and word similarity using WordNet and Retina.
Organizers [13]	Baseline	Baseline method, fully automatic.
	Segmen- tation	Apply segmentation and automatic retrieval and diversification based on concepts.
	Filtering	Apply segmentation, filtering, and automatic diversification. Us- ing all information.
	Fine- tunning	Apply segmentation, fine-tunning, filtering, and automatic di- versification. Using all information.
	RF	Relevance feedback. Using all information.

Table 7. Lifelog Retrieval Subtask results.

Retrieval Subtask.		
Team	Description	NDCG
	Baseline	0.09
Organizers [13]	Segmentation	0.14
	Fine-Tuning	0.39

I2R achieved the best F1@10 measure (excluding the organizers’ runs) of 0.497 by building a multi-step approach. As first step they filtered out unin-

Table 8. Lifelog Summarization Subtask results.

Team	Summarization Subtask.	
	Best Run	F1@10
I2R [8]	Run 1	0.394
	Run 2	0.497
	Run 3	0.311
	Run 4	0.440
	Run 5	0.456
	Run 6	0.397
	Run 7	0.487
	Run 8	0.341
	Run 9	0.356
	Run 10	0.461
UPB [3]	Run 1	0.132
Organizers* [13]	Baseline	0.103
	Segmentation	0.172
	Filtering	0.194
	Fine Tuning	0.319
	Relevance Feedback	0.772

*Note: Results from the organizers team are just for reference.

formative images, i.e., the ones with very homogeneous colors and with a high blurriness. Then the system ranked the remaining images and clustered the top ranked images into a series of events using either k-means or a hierarchical tree. As final step they selected, in an iterative manner, as many images per cluster as to fill a size budget. They submitted two different sets of runs: automatic (Run 1–3,6–9) and interactive (Run 4, 5, and 10). In the first ones in order to select the key-frames, all frames in each cluster are ranked according to distance to the cluster center (for k-means clustering) or relevance score (for hierarchical trees), then, the selection is sorted according to each frames relevance score. In the interactive process, they give the user the opportunity of removing, replacing and adding frames refining the automatically generated summary. They obtained the best result in the Run 2 where used visual and metadata information and automatic frame selection. It is worth to note that, on the contrary, the organizers team considerably improved the results of their automatic approach with the *Fine Tuning* introducing the human-in-the-loop, i.e., thanks to relevance feedback.

UPB team proposed an approach that combines textual and visual information in the process of selecting the best candidates for the tasks requirements. The run that they submitted relied solely on the information provided by the organizers and no additional annotations or external data, nor feedback from the users had been employed. Additionally, a multi-stage approach has been used. The algorithm starts by analyzing the concept detectors output provided by the organizers and selecting for each image only the most probable concepts. From the list of the topics, each of them has been then parsed such that only

relevant words have been kept and information regarding location, activity and the targeted user are extracted as well. The images that did not fit the topic requirements have been removed and this shortlist of images is then subject to a clustering step. Finally, the results are pruned with the help of a similarity scores computed using WordNets builtin similarity distance functions.

The **Organizers** team submitted 5 runs for the Summarization Subtask applying the same strategy as in the retrieval subtask, in which the first three runs were to test the automatic approach with the increasing level of the ‘criteria’ as proposed in [1], while the last two runs are used to test the fine tuning and the relevance feedback approaches [2]. For the relevance feedback approach, they ran a simulation by exploiting the ground-truth annotated data. The results confirm what is highlighted in Section 3.2; applying segmentation improved both retrieval and summarization performance. From Table 8 it is quite clear that applying fine-tuning significantly improved performance but what is worth to note is the big gaps in results between the automatic approach with the fine-tuning and the fine-tuning with the human-in-the-loop (relevance feedback) approaches. This shows that a better natural language processing is needed as well as machine learning studies in this field.

3.4 Limitations of the challenge

The major limitation that we learned from the task is about the difficulty of the topics. Many topics require huge effort on natural language processing to make the system understand the topics, which limit major of the teams, which are mainly from the image retrieval community. We also learned that the scope of the subtasks should be better defined since the summarization subtask already covers the retrieval task. As the result, most of the teams only interested in doing the second subtask.

As the ultimate goal is to provide insights from lifelogs, the current two subtasks only provide basic information, which is far away meaningful information. Thus, a subtask that better focuses on the quantified self, i.e., knowledge mined from self-tracking data, should be considered.

4 Discussions and Conclusions

A large gap between signed-up teams and submitted runs from the teams was observed. This can have two reasons, (i) due to the amount of data that has to be processed some teams might not be able to do so. (ii) the task seemed to be very complex requiring participants not just only to process single types of data but different ones such as audio and video, etc. For future iterations of the task it will be important to support teams by providing pre-extracted features or maybe access to hardware for the computation. Nevertheless, the submitted runs show that multimodal analysis is not used often. A closer contact with the teams during the whole task could help to find out individual bottle necks of the teams that prevents them from using other modalities and support them to overcome

these bottlenecks. All in all the task was quite successful for the first year and tacking into account that lifelogging is a rather new and not common field. The task helped to raise more awareness for lifelogging in general but also to point at the potential research questions such as the previous mentioned multimodal analysis, system aspects for efficiency, etc. For a possible next iteration of the task the dataset should be enhanced with more data and pre-extracted visual and multi-modal features. Furthermore, a platform should be established that can help the organizers to communicate and support the participants during their participation period.

References

1. Dang-Nguyen, D.T., Piras, L., Giacinto, G., Boato, G., De Natale, F.G.: A hybrid approach for retrieving diverse social images of landmarks. In: 2015 IEEE International Conference on Multimedia and Expo (ICME). pp. 1–6 (2015)
2. Dang-Nguyen, D.T., Piras, L., Giacinto, G., Boato, G., De Natale, F.G.: Multimodal retrieval with diversification and relevance feedback for tourist attraction images. *ACM Transactions on Multimedia Computing, Communications, and Applications* (2017), accepted
3. Dogariu, M., Ionescu, B.: A Textual Filtering of HOG-based Hierarchical Clustering of Lifelog Data (September 11-14 2017)
4. Gilbert, A., Piras, L., Wang, J., Yan, F., Dellandréa, E., Gaizauskas, R.J., Villegas, M., Mikolajczyk, K.: Overview of the imageclef 2015 scalable image annotation, localization and sentence generation task. In: Working Notes of CLEF 2015 - Conference and Labs of the Evaluation forum, Toulouse, France, September 8-11, 2015. (2015)
5. Gilbert, A., Piras, L., Wang, J., Yan, F., Ramisa, A., Dellandrea, E., Gaizauskas, R., Villegas, M., Mikolajczyk, K.: Overview of the ImageCLEF 2016 Scalable Concept Image Annotation Task. In: CLEF2016 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org, Évora, Portugal (September 5-8 2016)
6. Ionescu, B., Müller, H., Villegas, M., Arenas, H., Boato, G., Dang-Nguyen, D.T., Dicente Cid, Y., Eickhoff, C., Garcia Seco de Herrera, A., Gurrin, C., Islam, B., Kovalev, V., Liauchuk, V., Mothe, J., Piras, L., Riegler, M., Schwall, I.: Overview of ImageCLEF 2017: Information extraction from images. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction 8th International Conference of the CLEF Association, CLEF 2017. Lecture Notes in Computer Science, vol. 10456. Springer, Dublin, Ireland (September 11-14 2017)
7. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22Nd ACM International Conference on Multimedia. pp. 675–678. MM '14, ACM, New York, NY, USA (2014), <http://doi.acm.org/10.1145/2647868.2654889>
8. Molino, A.G.D., Mandal, B., Lin, J., Lim, J.H., Subbaraju, V., Chandrasekhar, V.: VC-I2R@ImageCLEF2017: Ensemble of Deep Learned Features for Lifelog Video Summarization (September 11-14 2017)
9. Thomee, B., Popescu, A.: Overview of the ImageCLEF 2012 Flickr Photo Annotation and Retrieval Task. In: CLEF 2012 working notes. Rome, Italy (2012)

10. Villegas, M., Paredes, R.: Overview of the ImageCLEF 2012 Scalable Web Image Annotation Task. In: Forner, P., Karlgren, J., Womser-Hacker, C. (eds.) CLEF 2012 Evaluation Labs and Workshop, Online Working Notes. Rome, Italy (September 17-20 2012)
11. Villegas, M., Paredes, R.: Overview of the ImageCLEF 2014 Scalable Concept Image Annotation Task. In: CLEF2014 Working Notes. CEUR Workshop Proceedings, vol. 1180, pp. 308–328. CEUR-WS.org, Sheffield, UK (September 15-18 2014)
12. Villegas, M., Paredes, R., Thomee, B.: Overview of the ImageCLEF 2013 Scalable Concept Image Annotation Subtask. In: CLEF 2013 Evaluation Labs and Workshop, Online Working Notes. Valencia, Spain (September 23-26 2013)
13. Zhou, L., Piras, L., Riegler, M., Boato, G., Dang-Nguyen, D.T., Gurrin, C.: Organizer Team at ImageCLEFlifelog 2017: Baseline Approaches for Lifelog Retrieval and Summarization (September 11-14 2017)

A Topics List 2017

The following tables present the descriptions of 51 search topics for the Image-CLEF 2017 LifeLog Retrieval and Summarization Tasks.

Table 9: Description of search topics for the development set in LRT.

T1. Presenting/Lecturing
Query: Find the moment(s) when user u1 was lecturing a group of people in a classroom environment.
Description: A lecture can be in any classroom environment and must contain more than one person in the audience who are sitting down. The moments from entry to exit of the classroom are relevant. A classroom environment has desks and chairs with students. Discussion or lecture encounters in which the audience are standing up, or outside of a classroom environment are not considered relevant.
T2. On the Bus or Train
Query: Find the moment(s) when user u1 was taking a bus or a train in his home country.
Description: The user normally drives a car. On some occasions he takes public transport and leaves the car at home. Moments in which the user is on a train or a bus are relevant only when he is in his home country. Moments in which the user is on public transport in other countries are not relevant. Moments in taxis are also considered non-relevant.
T3. Having a Drink
Query: Find the moment(s) when user u1 was having a drink in a bar with someone.
Description: Any moment in which the user is clearly seen having a beer or other drink in a bar venue is considered relevant. Having a drink in any other location (e.g. a cafe), or without another person present is not considered relevant. The type of drink is not relevant once it is presumed alcoholic in nature and not tea/coffee.
T4. Riding a Red (and Blue) Train
Query: Find the moment(s) when user u1 was riding a red (and blue) coloured train.
Description: In order to be considered relevant, the moment must contain an external view of the red (and blue) train followed by a period of time spent riding the train. Moments that just show a red train in the field of view are not considered relevant if the user does not ride the train.
T5. The Rugby Match
Query: Find the moment(s) when user u1 was watching rugby football on a television when not at home.

Description: Moments that show rugby football on a television when the user is not at home are considered relevant. To be considered relevant the moment(s) must show the entirety or part of the TV screen and be of sufficient duration to indicate the act of observation. It does not matter which teams are playing. Any point from the start to the end of this sports event is considered relevant.

T6. Costa Coffee

Query: Find the moment(s) when user u1 was in Costa Coffee.

Description: Moments that show the user consuming coffee/food in a Costa Coffee outlet are considered relevant. Any other consumption of food / drink is not considered relevant. Costa Coffee is clearly identified by the red coloured logo on the cups or the logo in the environment. The moments from entry to exit of the Costa Coffee outlet are relevant.

T7. Antiques Store

Query: Find the moment(s) when user u1 was browsing in an antiques store.

Description: Moments which show the user browsing for antiques in antiques stores are relevant. If the user exits an antique store and enters another shortly afterwards, then this would be considered two moments. The antiques stores can be identified by the presence of a large number of old objects of art/furniture/decoration arranged on/in display units.

T8. ATM

Query: Find the moment(s) when user u1 was using an ATM machine.

Description: The ATM Machine can be from any bank and in any location. To be relevant, the user must be directly in front of the machine with no people between the user and machine. Moments that show an ATM machine without showing the user directly in front of the machine are not considered relevant.

T9. Shopping for Fish

Query: Find the moment(s) when user u1 was shopping for fish in the supermarket.

Description: To be considered relevant the moment must show the user inside the supermarket on a shopping activity. The user must be clearly shopping and interacting with objects, including fish, in the supermarket. If the user is in a supermarket and does not buy fish, the shopping event is not considered to be relevant.

T10. Cycling home

Query: Find the moment(s) when user u2 was cycling home from work.

Description: The relevant moments must show the user cycling a bicycle from his/her point of view. Cycling home from work is relevant. Cycling to work or cycling to/from other destinations are not considered to be relevant.

T11. Shopping

Query: Find the moment(s) in which user u2 was grocery shopping in the supermarket.

Description: To be relevant, the user must clearly be inside a supermarket and shopping. Passing by or otherwise seeing a supermarket are not considered relevant if the user does not enter the supermarket to go shopping.
T12. In a Meeting
Query: Find the moment(s) in which user u2 was in a meeting at work with 2 or more people.
Description: To be considered relevant, the moment must occur at meeting room and must contain at least two colleagues sitting around a table at the meeting. Meetings that occur outside of my place of work are not relevant.
T13. Checking the Menu
Query: Find the moment(s) when user u2 was standing outside a restaurant checking the menu.
Description: To be considered relevant, the user must be checking the menu of a restaurant while outside the restaurant. Reading the menu inside the restaurant is not considered relevant. Other views of restaurants are not considered relevant if the user is not reading the menu outside.
T14. Watching TV
Query: Find the moment(s) when user u3 was watching TV.
Description: To be relevant, TV set must be on and entirely or partially visible during the moments. The user must be watching TV for a period of time not less than 5 minutes. Moments which show the user was watching TV while having meals are considered relevant. Moments in which the user is using desktop computer or laptop to watch TV shows are not considered relevant.
T15. Writing
Query: Find the moment(s) when user u3 was writing on a paper using a pen or pencil.
Description: To be considered relevant the user must be writing some information on a paper using a pen or a pencil. The writing behaviour must be visible. It does not matter what type of pen is being used or the type of paper. It does not matter what the user is writing.
T16. Drinking in a Pub
Query: Find the moment(s) when user u3 was drinking in a pub with friends or alone.
Description: Relevant moments show the user drinking in a pub. Drinking at home or in any place other than a pub are not considered to be relevant. The user may be with a friend, or alone.

Table 10: Description of search topics for the test set in LRT.

T1. Using laptop out of office
Query: Find the moment(s) in which user u1 was using his laptop outside the working places.

Description: To be consider to relevant, the user should use his laptop, for work or for entertainment out of his working place.
T2. On stage
Query: Find the moment(s) in which user u1 was giving a talk as a presenter/speaker.
Description: The user may be sitting or standing on a stage, facing many audience. A microphone should appear occasionally at the front of the user.
T3. Shopping in the electronic market.
Query: Find the moment(s) in which user u1 was shopping in the electronic market.
Description: Find the moment(s) that user u1 was at the electronic market. Spending time at normal supermarket is not considered relevant.
T4. Jogging in the park
Query: Find the moments(s) in which user u1 was jogging in a park.
Description: Find the moment(s) that user u1 was was jogging in a park. Walking or jogging in other places are not considered relevant.
T5. In a Meeting 2
Query: Find the moment(s) that user u1 was in a meeting at work.
Description: To be considered relevant, the moment must occur at meeting room and must contain at least two colleagues sitting around a table at the meeting. Meetings that occur outside of the work place are not relevant.
T6. Watching TV 2
Query: Summarize the moment(s) when user u1 was watching TV.
Description: To be relevant, TV set must be on and entirely or partially visible during the moments. Moments which show the user was watching TV while having meals are considered relevant. Moments in which the user is using desktop computer or laptop to watch TV are not considered relevant.
T7. Pizza and Friends
Query: Find the moment(s) in which user u1 was eating pizza in the restaurant with friends.
Description: The location must be a restaurant. The user should be eating the pizza together with his friend(s) (the friends can eat other food).
T8. Working in the air.
Query: Find the moment(s) in which user u1 was using computer in the airplane.
Description: To be relevant, the user must be using computer in an airplane. Using computer for entertainment is not considered relevant.
T9. Playing Guitar
Query: Find the moment(s) in which user u2 was playing guitar.
Description: To be considered relevant, the moment must clearly show the user is playing his guitar.
T10. Exercise in the gym
Query: Find the moment(s) in which user u2 was doing exercise in the gym.

Description: To be considered relevant, the moment must clearly show the user is doing exercise in the gym. Chatting or not doing exercise are not considered relevant.
T11. Eating fruits
Query: Find the moment(s) in which user u2 was having fruits.
Description: To be considered relevant, the moment must clearly show the user is eating some fruit, no matter where and when he was.
T12. Brushing or washing face
Query: Find the moment(s) in which user u2 was brushing or washing face.
Description: To be considered relevant, the moment must clearly show the user is brushing or washing face
T13. Eating 2
Query: Find the moment(s) when user u2 was eating or drinking.
Description: To be relevant, the images must show entirely or partially visible food/drink.
T14. At McDonald
Query: Find the moment(s) in which user u2 was at McDonald for eating or just for relaxing.
Description: To be considered relevant, the moment must clearly show the user is in McDonald.
T15. Viewing a statue
Query: Find the moment(s) in which user u2 was viewing a statue.
Description: To be considered relevant, the moment must clearly show a statue, at any possible location while the user was standing or walking.
T16. ATM
Query: Find the moment(s) when user u2 was using an ATM machine.
Description: The ATM Machine can be from any bank and in any location. To be relevant, the user must be directly in front of the machine with no people between the user and machine. Moments that show an ATM machine without showing the user directly in front of the machine are not considered relevant.
T17. Have party with friends at friends home.
Query: Find the moment(s) in which user u3 was attending a party with many friends at a friends home.
Description: To be relevant, the user should be at a party his friend's home, whether indoor or outdoor. Some food and drink should be visualized.
T18. Shopping in the butchers shop.
Query: Find the moment(s) in which user u3 was consuming in the butchers shop.
Description: To be relevant, the user must be at the butcher's shop, no matter what the user bought. Buying meet in the supermarket is not relevant.
T19. Buying a ticket via ticket machine.

Query: Find the moment(s) in which user u3 was buying a ticket via ticket machine.
Description: The ticket may include movie ticket, food ticket, any transport ticket. Using automatic ticket machine must be relevant, no matter what kinds of ticket and whether the user bought any ticket. Using ATM , Vending machine are not relevant.
T20. Shopping 2
Query: Find the moment(s) in which user u3 doing shopping.
Description: To be relevant, the user must clearly be inside a supermarket or shopping stores (includes book store, convenience store, pharmacy, etc). Passing by or otherwise seeing a supermarket are not considered relevant if the user does not enter the shop to go shopping.

Table 11: Description of search topics for the development set in LST.

T1. Eating
Query: Summarize the moment(s) when user u1 was eating or drinking.
Description: User u1 wants to know insight of his eating/drinking habits. He would like to have a summary of what, when, where, and whom together he was eating or drinking. To be relevant, the images must show entirely or partially visible food/drink. Blurred or out of focus images are not relevant. Images that are covered (mostly by the lifelogger’s arm) are not relevant, even if they are recorded while the user was eating.
T2. Social Drinking
Query: Summarize the the social drinking habits of user u1.
Description: Drinking in a bar, away from home would be considered relevant. Moments drinking alcohol at home would not be considered social drinking. Drinking alone does not classify as social drinking. Blurred or out of focus images are not relevant. Images that are covered (mostly by the lifelogger’s arm) are not relevant.
T3. Shopping
Query: Summarize the moment(s) in which user u1 doing shopping.
Description: To be relevant, the user must clearly be inside a supermarket or shopping stores (includes book store, convenient store, pharmacy, etc). Passing by or otherwise seeing a supermarket are not considered relevant if the user does not enter the shop to go shopping. Blurred or out of focus images are not relevant. Images that are covered (mostly by the lifelogger’s arm) are not relevant
T4. In a Meeting
Query: Summarize the activities of user u2 in a meeting at work.

Description: This is an extension of topic 12 from the retrieval subtask. To be considered relevant, the moment must occur at meeting room and must contain at least two colleagues sitting around a table at the meeting. Meetings that occur outside of the work place are not relevant. Different meetings have to be summarized as different activities. Blurred or out of focus images are not relevant. Images that are covered (mostly by the lifelogger’s arm) are not relevant.
T5. Watching TV
Query: Summarize the moment(s) when user u3 was watching TV.
Description: This is an extension of topic 14 from the retrieval subtask. To be relevant, TV set must be on and entirely or partially visible during the moments. Moments which show the user was watching TV while having meals are considered relevant. Moments in which the user is using desktop computer or laptop to watch TV shows are not considered relevant. Blurred or out of focus images are not relevant. Images that are covered (mostly by the lifelogger’s arm) are not relevant.

Table 12: Description of search topics for the test set in LST.

T1. In a Meeting 2
Query: Summarize the activities of user u1 in a meeting at work.
Description: To be considered relevant, the moment must occur at meeting room and must contain at least two colleagues sitting around a table at the meeting. Meetings that occur outside of the work place are not relevant. Different meetings have to be summarized as different activities. Blurred or out of focus images are not relevant. Images that are covered (mostly by the lifelogger’s arm) are not relevant.
T2. Watching TV 2
Query: Summarize the moment(s) when user u1 was watching TV.
Description: To be relevant, TV set must be on and entirely or partially visible during the moments. Moments which show the user was watching TV while having meals are considered relevant. Moments in which the user is using desktop computer or laptop to watch TV shows are not considered relevant. Blurred or out of focus images are not relevant. Images that are covered (mostly by the lifelogger’s arm) are not relevant.
T3. Using laptop outside the office
Query: Summarise the moment(s) in which user u1 was using his laptop outside the working places.
Description: To be consider to relevant, the user should use his laptop, for working or for entertainment out of his working place. Blurred or out of focus images are not relevant. Images that are covered (mostly by the lifelogger’s arm) are not relevant.
T4. Working at home

Query: Find the moment(s) in which user u1 was working at home.
Description: To be consider to relevant, the user should be using computer for work, reviewing an article or taking some notes at home. Using computer for entertainment is not relevant. Blurred or out of focus images are not relevant. Images that are covered (mostly by the lifelogger's arm) are not relevant.
T5. Eating 2
Query: Summarize the moment(s) when user u2 was eating or drinking.
Description: User u2 wants to know insight of his eating/drinking habits. He would like to have a summary of what, when, where, and whom together he was eating or drinking. To be relevant, the images must show entirely or partially visible food/drink. Blurred or out of focus images are not relevant. Images that are covered (mostly by the lifelogger's arm) are not relevant, even if they are recorded while the user was eating.
T6. Social Drinking 2
Query: Summarize the the social drinking habits of user u2.
Description: Drinking in a bar, away from home would be considered relevant. Moments drinking alcohol at home would not be considered as social drinking. Drinking alone does not classify as social drinking. Blurred or out of focus images are not relevant. Images that are covered (mostly by the lifelogger's arm) are not relevant.
T7: Sightseeing
Query: Summarize the moments when the user u2 seeing street, people, landscape, etc. when he was traveling to other cities or countries.
Description: Photos taken inside public transport are not relevant. Sightseeing in his hometown is not relevant. Blurred or out of focus images are not relevant. Images that are covered (mostly by the lifelogger's arm) are not relevant.
T8. Transporting
Query: Summarize the moments when user u2 using public transportation.
Description: Photos taken inside a car or a taxi are not relevant. Blurred or out of focus images are not relevant. Images that are covered (mostly by the lifelogger's arm) are not relevant.
T9. Preparing meals
Query: Find the moment(s) in which user u3 was preparing meals at home.
Description: To be considered relevant, the moment must clearly show the user is preparing meals in the kitchen. Eating is not relevant. Blurred or out of focus images are not relevant. Images that are covered (mostly by the lifelogger's arm) are not relevant.
T10. Shopping 2
Query: Summarize the moment(s) in which user u3 was doing shopping.

Description: To be relevant, the user must clearly be inside a supermarket or shopping stores (includes book store, convenience store, pharmacy, etc). Passing by or otherwise seeing a supermarket is not considered relevant if the user does not enter the shop to go shopping. Blurred or out of focus images are not relevant. Images that are covered (mostly by the lifelogger's arm) are not relevant.