Overview of ImageCLEFlifelog 2019: Solve My Life Puzzle and Lifelog Moment Retrieval

Duc-Tien Dang-Nguyen[1], Luca Piras[2], Michael Riegler[3], Liting Zhou[4], Mathias Lux[5], Minh-Triet Tran[6], Tu-Khiem Le[4], Van-Tu Ninh[4], and Cathal Gurrin[4]

[1] University of Bergen, Bergen, Norway
[2] Pluribus One & University of Cagliari, Cagliari, Italy
[3] Simula Metropolitan Center for Digital Engineering, Oslo, Norway
[4] Dublin City University, Dublin, Ireland
[5] ITEC, Klagenfurt University, Klagenfurt, Austria
[6] University of Science, VNU-HCM, Ho Chi Minh City, Vietnam

Abstract. This paper describes ImageCLEFlifelog 2019, the third edition of the Lifelog task. In this edition, the task was composed of two subtasks (challenges): the Lifelog Moments Retrieval (LMRT) challenge that followed the same format as in the previous edition, and the Solve My Life Puzzle (Puzzle), a brand new task that focused on rearranging lifelog moments in temporal order. ImageCLEFlifelog 2019 received noticeably higher submissions than the previous editions, with ten teams participating resulting in a total number of 109 runs.

1 Introduction

Since 2016 with the first lifelog initiative, the NTCIR-12 - Lifelog task [7], research in lifelogging, *'a form of pervasive computing, consisting of a unified digital record of the totality of an individual's experiences, captured multimodally through digital sensors and stored permanently as a personal multimedia archive'* [5], is getting more attention, especially within the multimedia information retrieval community. However, given the huge volume of data that a lifelog would generate, along with the complex patterns of the data, we are just at the starting point of lifelog data organisation. We need to advance the next step of development and to unlock the potential of such data. There is a need new ways of organising, annotating, indexing and interacting with lifelog data, and that is the key motivation of the Lifelog task at ImageCLEF.

The ImageCLEFlifeLog2019 task at ImageCLEF 2019 [10], was the third edition of the task, with previous editions in 2017 [2] and 2018 [3], which were inspired by the fundamental image annotation and retrieval tasks of ImageCLEF since 2003. This year, the task continued to follow the general evolution of ImageCLEF, by applying the advanced deep learning methods and extending the focus to multi-modal approaches instead of only working just with image retrieval.

Comparing to the previous editions, in this third edition we merged the previous two sub-tasks (challenges): Activities of Daily Living understanding (ADLT) and Lifelog Moment Retrieval (LMRT) into a single challenge (LMRT) and proposed a brand new one: Solve My Life Puzzle (Puzzle), which focused on the new ways of organising lifelog data, in particular in rearranging lifelog moments.

The details of this year two challenges will be provided in section 2. This includes also the descriptions of data and resources. For the rest of the paper, in Section 3, submissions and results are presented and discussed and in the final section 4 the paper is concluded and final remarks and future work are discussed.

2 Overview of the Task

2.1 Motivation and Objectives

An increasingly wide range of personal devices, such as smartphones, video cameras as well as wearable devices that allow capturing pictures, videos, and audio clips for every moment of our lives are becoming available. Considering the huge volume of data created, there is a need for systems that can automatically analyse the data in order to categorize, summarize and also query to retrieve the information the user may need.

Despite the increasing number of successful related workshops and panels, lifelogging has seldom been the subject of a rigorous comparative benchmarking exercise as, for example, the new lifelog evaluation task at NTCIR-13 [8] or the last editions of the ImageCLEFlifelog task [2][3]. In this edition we aimed to bring the attention of lifelogging to a wide audience and to promote research into some of the key challenges of the coming years.

2.2 Challenge Description

Lifelog Moment Retrieval Task (LMRT)
In this task, the participants are required to retrieve a number of specific moments in a lifeloggers life. Moments are defined as semantic events, or activities that happened throughout the day. For example, a participant would have been required to find and return relevant moments for the query "Find the moment(s) when the user1 is cooking in the kitchen". In this edition, particular attention was to be paid to the diversification of the selected moments with respect to the target scenario. The ground truth for this subtask was created using a manual annotation process. Figure 1 illustrates some examples of the moments when the lifelogger was having coffee with friends". In addition, listings 1 and 2 list all the queries used in the challenge.

T.001 Icecream by the Sea

Description: Find the moment when u1 was eating an icecream beside the sea
Narrative: To be relevant, the moment must show both the ice cream with cone in the hand of u1 as well as the sea clearly visible. Any moments by the sea, or eating an ice cream which do not occur together are not considered to be relevant.

T.002 Having Food in a Restaurant

Description: Find the moment when u1 was eating food or drinkingU1 was eating food in a restaurant while away from home. Any kinds of dishes are relevant. Only Drinking coffee and have dessert in a cafe won't be relevant.

T.003 Watching Videos

Description: Find the moment when u1 was watching video when using other digital devices.
Narrative: To be relevant, u1 must be watching videos in any location and any digital devices can be considered. For example: TV machine, tablet, mobile phone, laptop, desktop computer.

T.004 Photograph of a Bridge

Description: Find the moment when u1 was taking a photo of a bridge.
Narrative: U1 was walking on a pedestrian street and stopped to take a photo of a bridge. Moments when u1 was walking on a street without stopping to take a photo of a bridge are not relevant. Any other moment showing a bridge when a photo was not being taken are also not considered to be relevant.

T.005 Grocery Shopping

Description: Find the moment when u1 was shopping for food in a grocery shop.
Narrative: To be considered relevant, u1 must be clearly in a grocery shop and bought something from the it.

T.006 Playing a Guitar

Description: Find the moment when U1 or a man is playing guitar in view.
Narrative:Any use of guitars indoors could be considered relevant. Any type of Guitar could be considered as relevant.

T.007 Cooking

Description: Find moments when u1 was cooking food.
Narrative:The moments shows U1 was cooking food at any places are relevant.

T.008 Car Sales Showroom

Description: Find the moments when u1 was in a car sales showroom.
Narrative: u1 visited a car sales showroom a few times. Relevant moments show u1 indoors in a car sales showroom, either looking at cars or waiting for a salesman sitting at a table. Any moments looking at cars while outside of a showroom are not considered relevant.

T.009 Public Transportation

Description: Find the moments when U1 is taking the public transportation in any countries.
Narrative: To be considered relevant,the U1 must take a public transportation to other place. The moments that the U1 is driving a car is not relevant.

T.010 Paper or Book Reviewing

Description: Find all moments when u1 was reading a paper or book.
Narrative: To be relevant, the paper or book must be visible in front of U1 and sometimes U1 use a pen to mark on the paper or book.

Listing 1: Description of topics for the development set in LMRT.

T.001 In a Toyshop

Description: Find the moment when u1 was looking at items in a toyshop.
Narrative: To be considered relevant, u1 must be clearly in a toyshop. Various toys are being examined, such as electronic trains, model kits and board games. Being in an electronics store, or a supermarket, are not considered to be relevant.

T.002 Driving home

Description: Find any moment when u1 was driving home from the office.
Narrative: Moments which show u1 is driving home from the office is relevant. Driving from other place and to other place are not relevant.

T.003 Seeking Food in a Fridge

Description: Find the moments when u1 was looking inside a refrigerator at home.
Narrative: Moments when u1 is at home and looking inside a refrigerator are considered relevant. Moments when eating food or cooking in the kitchen are not considered relevant.

T.004 Watching Football

Description: Find the moments when either u1 or u2 was watching football on the TV.
Narrative: To be considered relevant, either u1 or u2 must be indoors and watching football on a television. Watching any other TV content is not considered relevant.

T.005 Coffee Time

Description: Find the moment when u1 was having coffee in a cafe.
Narrative:To be considered relevant, u1 must be in a cafe and having coffee alone or with another individual.

T.006 Having Breakfast at Home

Description:U1 was having breakfast at home and the breakfast time must be from 5:00 am until 9:00 am
Narrative: To be considered relevant, the moments must show some parts of the furniture being assembled.

T.007 Having Coffee with Two Persons

Description: Find the moment when u1 was having coffee with two person.
Narrative: Find the moment when u1 was having coffee with two person. One was wearing blue shirt and the other one was wearing white cloth. Gender is not relevant.

T.008 Using a Smartphone Outside

Description: Find the moment when u1 was using smartphone when he was walking or standing outside.
Narrative: To be considered relevant, u1 must be clearly using a smartphone and the location is outside.

T.009 Wearing a Red Plaid Shirt

Description: Find the moment when U1 was wearing a red plaid shirt.
Narrative: To be relevant, the user1 was wearing a red plaid shirt in a day life.

T.010 Having a Meeting in China

Description: Find all moments when u1 was attending a meeting in China.
Narrative: To be relevant, the user1 must be in China and was having a meeting with others.

Listing 2: Description of topics for the test set in LMRT.

Fig. 1. Examples from the results of the query: *'Show all moment I was having coffee with friends.'*

Solve my Life Puzzle Task (Puzzle)
Given a set of lifelog images with associated metadata (e.g., biometrics, location, etc.), but no timestamps, the participants needed to analyse these images and rearrange them in chronological order and predict the correct day (e.g. Monday or Sunday) and part of the day (e.g. morning, afternoon, or evening). Figure 2 illustrates an example of this challenge.

2.3 Dataset

The data was a medium-sized collection of multimodal lifelog data over 42 days by two lifeloggers. The contribution of this dataset over previously released datasets was the inclusion of additional biometric data, a manual diet log and the inclusion of conventional photos. In most cases the activities of the lifeloggers were separate and they did not meet. However on a small number of occasions the lifeloggers appeared in data of each other. The data consists of:

- **Multimedia Content**. Wearable camera images captured at a rate of about two images per minute and worn from breakfast to sleep. Accompanying this image data was a time-stamped record of music listening activities sourced from `Last.FM`[1] and an archive of all conventional (active-capture) digital photos taken by the lifelogger.
- **Biometrics Data**. Using the FitBit fitness trackers[2], the lifeloggers gathered 24×7 heart rate, calorie burn and steps. In addition, continuous blood glucose monitoring captured readings every 15 minutes using the Freestyle Libre wearable sensor[3].

[1] Last.FM Music Tracker and Recommender - https://www.last.fm/
[2] Fitbit Fitness Tracker (FitBit Versa) - https://www.fitbit.com
[3] Freestyle Libre wearable glucose monitor - https://www.freestylelibre.ie/

(a) Evening-Monday (b) Morning-Tuesday (c) Morning-Wednesday

(d) Afternoon-Wednesday (e) Morning-Thursday (f) Afternoon-Friday

(g) Afternoon-Saturday (h) Morning-Sunday (i) Afternoon-Sunday

Fig. 2. Sample images from the Puzzle task and the predicted results.

- **Human Activity Data**. The daily activities of the lifeloggers were captured in terms of the semantic locations visited, physical activities (e.g. walking, running, standing) from the Moves app[4], along with a time-stamped diet-log of all food and drink consumed.
- **Enhancements to the Data**. The wearable camera images were annotated with the outputs of a visual concept detector, which provided three types of outputs (Attributes, Categories and Concepts). Two visual concepts which include attributes and categories of the place in the image are extracted using PlacesCNN [18]. The remaining one is detected object category and its bounding box extracted by using Faster R-CNN [14] trained on MSCOCO dataset [12].

Format of the metadata. The metadata was stored in a .csv files, which was called the minute-based table. The precise structured of it is described in Table 2. Additionally, extra metadata was included, such as visual categories and concepts descriptors. The format of the extra metadata could be found in Table 3.

[4] Moves App for Android and iOS - http://www.moves-app.com/

Table 1. Statistics of ImageCLEFlifelog 2019 Data

Characters	**Size**
Number of Lifeloggers	2
Number of Days	43 days
Size of the Collection	14 GB
Number of Images	81,474 images
Number of Locations	61 semantic locations
Number of Puzzle Queries	20 queries
Number of LMRT Queries	20 queries

2.4 Performance Measures

LMRT For assessing performance, classic metrics were deployed. These metrics were:

- 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 were the F1-measure@10, which gives equal importance to diversity (via CR@10) and relevance (via P@10).

Participants were allowed 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, 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) were encouraged.

Puzzle For the Puzzle task, we used Kendall's Tau score to measure the similarity of the temporal order between the participant's temporal arrangement and ground-truth for each query. The formula of Kendall's Tau is as follows:

$$\tau = \max\left(0, \frac{C - D}{C + D}\right) \tag{1}$$

where C and D are the number of concordant pairs and discordant pairs correspondingly between the participant's submission order and the one of ground-truth. In the original Kendall's Tau formula, the range of the formula is from $[-1, 1]$, however, we choose to narrow the range of the score to $[0, 1]$. It means that if the number of opposite ranking-direction pairs are greater than the quantity of the same ranking-direction pairs, participants would get nothing in Kendall's Tau score. The accuracy of part-of-day prediction was computed simply by dividing the number of correct predictions by the total number of predictions.

Table 2. Structure of Minute-based table.

Field name	Meaning	Example
minute_ID	Identity field for every minute, unique for every volunteer	u1_20180503_0000
utc_time	UTC Time with format: YYYYMMDD_HHMM_UTC	20180503_0000_UTC
local_time	Local time in volunteers timezone (from volunteers smart phone): YYYYMMDD_HHMM	20180503_0100
time_zone	The name of volunteers timezone	Europe/Dublin
lat	Latitude of volunteers position	53.386881
lon	Longitude of volunteers position	-6.15843
name	The name of the place corresponding to volunteers position	Home
song	The name of the song was playing at that time activity The activity that volunteer was doing at that time	walking, transport
steps	The number of volunteers steps collected by wearable devices	14
calories	The number of calories collected by wearable devices	1.17349994
historic_glucose (mmol/L)	The historic glucose index collected by wearable devices, measured in mmol/L	4.3
scan_glucose (mmol/L)	The scan glucose index collected by wearable devices, measured in mmol/L	4.8
heart_rate	The heart rate of volunteer at that time collected by wearable devices	73
distance	The distance collected by wearable devices	
img00_id to img_19_id	The image ID captured by the wearable camera at that time	u1_20180503_1627_i01
cam00_id to cam14_id	The image ID captured by volunteers smart phone at that time	u1_20180503_1625_cam_i00

Table 3. Structure of the Visual Concepts table.

Field name	**Meaning**	**Example**
image_id	Identity field for every image, including images from wearable camera and smart phone camera	u1_20180503_0617_i00
image_path	Image path to the corresponding image	
attribute_top1 to attribute_top10 (top 10 attributes)	The top 10 attributes predicted by using the PlaceCNN, trained on SUNattribute dataset	no horizon, man-made, metal, indoor lighting
category_topXX, category_topXX_score (top 05 categories)	The top 05 categories and their scores predicted by using the PlaceCNN, trained on Place 365 dataset.	chemistry_lab, 0.082
concept_class_topXX, concept_score_topXX, concept_bbox_topXX, (top 25 concepts)	Class name, bounding box and score of the top 25 objects with the highest score in each image. They are predicted by using Faster R-CNN, trained on the COCO dataset	person, 0.987673 508.568878 171.124496 513.541748 395.073303

Finally, the primary score was computed as the average of Kendall's Tau score and accuracy of part-of-day prediction for all queries.

2.5 Ground Truth Format

LMRT Task. The ground truth for the LMRT task was provided in two individual txt files: one file for the cluster ground truth and one file for the relevant image ground truth.

In the cluster ground-truth file, each line corresponded to a cluster where the first value was the topic id, followed by cluster id number, followed by the cluster user tag separated by comma. Lines were separated by an end-of-line character (carriage return). An example is presented below:

- 1, 1, Icecream by the Sea
- 2, 1, DCU_canteen
- ...
- 2, 8, Restaurant_5
- 2, 9, Restaurant_6
- ...

In the relevant ground-truth file, the first value on each line was the topic id, followed by a unique photo id, and then followed by the cluster id number (that corresponded to the values in the cluster ground-truth file) separated by comma. Each line corresponded to the ground truth of one image and lines were separated by an end-of-line character (carriage return). An example is presented below:

- 1, u1_20180528_1816_i00, 1
- 1, u1_20180528_1816_i02, 1
- 1, u1_20180528_1816_i01, 1
- 1, u1_20180528_1817_i01, 1
- 1, u1_20180528_1817_i00, 1
- 1, u1_20180528_1818_i02, 1
- ...
- 2, u1_20180508_1110_i00, 1
- 2, u1_20180508_1110_i01, 1
- 2, u1_20180508_1111_i00, 1
- 2, u1_20180508_1111_i01, 1
- 2, u1_20180508_1112_i01, 1
- ...

Puzzle Task The ground truth was provided in only one individual .csv file. For each line in this file, first value was the id of the query, followed by the id of image provided by the organisers, followed by the order that the image should be arranged in this query in temporal manner, followed by the part-of-day prediction. Values in each line were separated by comma. Lines were separated by and end-of-line character (carriage return). The values in ground-truth file were sorted by query id, and then image id column in ascending order. An example is shown below:

- 1, 001.JPG, 17, 1
- 1, 002.JPG, 8, 1
- 1, 003.JPG, 9, 1
- 1, 004.JPG, 1, 1
- 1, 005.JPG, 2, 1
- ...
- 8, 001.JPG, 12, 3
- 8, 002.JPG, 16, 3
- 8, 003.JPG, 13, 3
- 8, 004.JPG, 15, 3
- 8, 005.JPG, 10, 1
- ...

Table 4. Official Results of the ImageCLEFlifelog 2019 LMRT Task.

Team	Run	P@10	CR@10	F1@10	Team	Run	P@10	CR@10	F1@10
Organiser [13]	RUN1*	0.41	0.31	0.29	**UATP** [15]	RUN1	0.03	0.01	0.02
	RUN2*	0.33	0.26	0.24		RUN2	0.08	0.02	0.03
ATS [17]	RUN1	0.10	0.08	0.08		RUN3	0.09	0.02	0.03
	RUN2	0.03	0.06	0.04		RUN4	0.1	0.02	0.03
	RUN3	0.03	0.04	0.04		RUN5	0.1	0.02	0.04
	RUN4	0.06	0.13	0.08		RUN6	0.06	0.06	0.06
	RUN5	0.07	0.06	0.05	**UPB** [6]	RUN1	0.17	0.22	0.13
	RUN6	0.07	0.13	0.08	**ZJUTCVR**	RUN1	**0.71**	0.38	0.44
	RUN7	0.08	0.19	0.1	[20]	RUN2†	0.74	0.34	0.43
	RUN8	0.05	0.11	0.07		RUN3†	0.41	0.31	0.33
	RUN9	0.10	0.14	0.10		RUN4†	0.48	0.35	0.36
	RUN11	0.14	0.16	0.12		RUN5†	0.59	0.5	0.48
	RUN12	0.35	0.36	0.25	**TUC_MI**	RUN1	0.02	0.10	0.03
BIDAL [4]	RUN1	0.69	0.29	0.37	[16]	RUN2	0.04	0.08	0.04
	RUN2	0.69	0.29	0.37		RUN3	0.03	0.06	0.03
	RUN3	0.53	0.29	0.35		RUN4	0.10	0.11	0.09
HCMUS [11]	RUN1	0.70	0.56	0.60		RUN5	0.08	0.13	0.09
	RUN2	0.70	**0.57**	**0.61**		RUN6	0.00	0.00	0.00
REGIM [1]	RUN1	0.28	0.16	0.19		RUN7	0.04	0.06	0.05
	RUN2	0.25	0.14	0.17		RUN8	0.04	0.01	0.02
	RUN3	0.25	0.10	0.14		RUN9	0.02	0.01	0.01
	RUN4	0.09	0.05	0.06		RUN10	0.15	0.15	0.12
	RUN5	0.07	0.09	0.06		RUN11	0.03	0.07	0.04
	RUN6	0.07	0.08	0.06		RUN12	0.06	0.11	0.06
						RUN13	0.01	0.01	0.01
						RUN14	0.06	0.21	0.09

Notes: * submissions from the organizer teams are just for reference.
† submissions submitted after the official competition.

3 Evaluation Results

3.1 Participating Groups and Runs Submitted

This year the number of participants as well as the number of submissions was considerably higher with respect to 2018: we received in total 50 valid submissions (46 official and 4 additional) for LMRT, and 21 (all are official) for Puzzle, from 10 teams representing over 10 countries. The submitted runs and their results are summarised in Tables 4 and 5.

3.2 Results

In this section we provide a short description of all submitted approaches followed by the official result of the task.

The Organiser team [13] provided a baseline approach to the LMRT task with a web-based interactive search engine called LIFER 2.0, which was based on a previous system [19] used at the LSC 2018 Lifelog Search Challenge. The authors submitted two runs which were obtained by letting two novice users perform interactive moment retrieval on the search engine for all ten queries. For the Puzzle task, the team proposed an activity mining approach which utilised Bag-of-Visual-Words (BOVW) methods. For each image in a query, the authors find the most relevant images in the training data based on the L2 distance of BOVW vectors to predict the part-of-day and chronological index of the images.

Table 5. Official Results of the ImageCLEFlifelog 2019 Puzzle Task.

Team	Run	Kendall's Tau	Part of Day	Final Score
Organiser [13]	RUN1*	0.06	0.31	0.18
	RUN2*	0.03	0.35	0.19
	RUN3*	0.03	0.34	0.18
	RUN4*	0.05	0.49	0.27
BIDAL [4]	RUN1	0.12	0.30	0.21
	RUN2	0.08	0.31	0.20
	RUN3	0.06	0.28	0.17
	RUN4	0.12	0.38	0.25
	RUN5	0.10	0.30	0.20
	RUN6	0.09	0.29	0.19
	RUN7	0.15	0.26	0.21
	RUN8	0.07	0.30	0.19
	RUN9	0.19	0.55	0.37
	RUN10	0.17	0.50	0.33
	RUN11	0.10	0.49	0.29
DAMILAB	RUN6	0.02	0.40	0.21
	RUN7	0.02	0.47	0.25
HCMUS [9]	RUN03ME	**0.40**	**0.70**	**0.55**
	RUN3	**0.40**	0.66	0.53
	RUN04ME	**0.40**	**0.70**	**0.55**
	RUN4	**0.40**	0.66	0.53

Notes: * submissions from the organizer teams are just for reference.

The REGIM-Lab [1] focused on LMRT task by improving the system from their last year participation with NoSQL which offers distributed database and framework to handle huge data. They employed the ground-truth of development set to improve the fine-tuning phase for concept extraction. In addition, CQL Query was used to exploit complicated metadata. For the query analysis, the authors trained a LSTM classifier to enhance the query with relevant concepts.

The UPB team [6] proposed the algorithm to eliminate blurry images which contain less information using a blur detection system. Following that, a metadata restriction filter, which was created manually by users, was applied to the dataset to further remove uninformative images. The remaining images was then computed a relevance score based on given metadata description for query answering.

The UAPTBioinformatics (UAPT) team [15] proposed an automatic approach for LMRT task. The images are pre-processed through an automatic selection step to eliminate images with irrelevant information to the topics (feature extraction, machine learning algorithm, k-nearest neighbors, etc.) and more visual concepts were generated using various state-of-the-art models (Google Cloud Vision API, YOLOv3). Then they extracted relevant words from topics' titles and narratives, dividing them into five categories, and finally matching them with the annotation concepts of lifelog images using a word embedding model trained on Google News dataset. Moreover, an extra step to reuse unselected images from the pre-processing steps for image similarity matching was proposed to increase the performance of their system.

The TUC_MI team [16] proposed an automatic approach for LMRT task. They firstly extracted twelve types of concept from different pre-trained models to increase the annotation information for lifelog data (1191 labels in total). For image processing, two methods were introduced to transform images into vectors: image-based vectors and segment-based vectors. For query processing, they processed the query with Natural Language Processing techniques and introduced a token vector which has the same dimension as image/segment vector. Finally, they defined a formula to compare the similarity between image/segment and token vector and conducted an ablation study to find the best model that achieved the highest score in this task.

The HCMUS team [11] proposed to extract semantic concepts from images to adapt to lifelogger's habits and behaviours in daily life. They firstly identified a list of concepts manually, then trained object detectors to extract extra visual concepts automatically. Moreover, they also utilised object' region of interest to infer its color by K-Means clustering. To further understand the temporal relationship between events, they also integrated the visualization function for an event sequence in their retrieval system. For the Puzzle task, the HCMUS team [9] utilised a BOVW approach to retrieve the visually similar moments with reference to lifelog data to infer the probability of the time and order of the images. Before applying the proposed remedy, they tried to cluster the images into groups based on the provided concepts extracted from PlacesCNN, GPS location, and user activities.

The BIDAL team [4] participated in both Puzzle and LMRT tasks. For the LMRT task, they introduced their interactive system with two main stages. For stage 1, they generated many atomic clusters from the dataset based on rough concepts and utilising text annotation to create Bag-of-Words (BOW) vectors for each image. In stage 2, they generated the BOW vector for query texts and found similar images that suited the context and content of the query. They then used the output for result expansion by adding more images which were in the same cluster. Finally, an end-user chooses appropriate images for the query. For the Puzzle task, they proposed to use visual feature matching via two similarity functions between each image in both train data and test data as the initial step to filter out dissimilarity associative pairs (train-test image pairs). Finally, the remaining images are grouped based on temporal order of train data so that it forms the full set as test data to rearrange the images.

The ZJUTCVR team [20] pre-processed the images with blur/cover filters to eliminate the blurred and occluded images. Then, they proposed three approaches to handle the remaining lifelog images: the two-class approach, the eleven-class approach, and the clustering approach. For two-class approach, the authors divided query topics into directories and ran a test on each directory with a fine-tuned CNN. After that, the results are classified into two classes based on the relevance to topic description. The eleven-class approach shared the same process with the previous method, but the results are split into 11 classes, where 10 classes are corresponding to 10 query topics and the 11th class contains irrelevant images to all 10 topics. With the clustering approach, the team inherited the procedure of two-class approach with a modification after the first-round retrieval by clustering images with LVQ algorithm.

The ATS team [17] approached the LMRT task with 11 automatics runs and 1 interactive run. All automatic runs shared the same process with 4 components: Interpretation, Subset selection, Scoring and Refinement, but differed in the configuration of selecting the approach of each component. The interpretation state provided keywords and synonyms approaches which utilised WordNet and Word2Vec to diversify the results. The choice of subset is highly reliant on the configuration to use partial match or entire dataset to test. A scoring process was used to produce final ranking with three settings: label counting, topic similarity and SVM. The Refinement step offered multiple approaches: weighting, thresholding, visual clustering and temporal clustering. Finally, the team conducted ablation study to find the best configuration. The interactive run was done by letting user filter the subset of dataset and choose automatic approach of each component to complete the query.

The official results are summarised in Tables 4 and 5. For LMRT (Table 4), eight teams participated and the highest F1@10 was submitted by HCMUS [11] at their second run (RUN2) with a score of **0.61** which is considerably higher than the results obtained by novice human (the Organiser team results). For this task, the common approach was to enrich visual concepts from images through different CNNs, transform everything data to vector through BOW, Word2Vec, etc., cluster/segment sequential images, and apply different feature similarity search methods to find images with suitable context and concepts. The best score was achieved by building object color detectors to extract visual concepts which adapt to each user's daily life.

In the Puzzle task, four teams have participated and the highest score was obtained at the value of **0.55** by two runs (RUN03ME and RUN04ME), also from the HCMUS team. With the exception of the BIDAL team that utilised different functions and visual features to evaluate the similarity between each train-test pairs and proposed new grouping algorithm to decide query arrangement, most teams used BOVW to retrieve images from training data, in order to rearrange the test images based on the temporal order of the retrieved results. The highest score was achieved by applying BOVW with one million clusters, which would be a promising approach to conduct further research and improvement.

4 Discussions and Conclusions

The submitted approaches in this year confirmed the trend from last year: all approaches are exploiting multi-modal instead of using only visual information. We also confirmed the importance of deep neural networks in solving these challenges: all ten participants are using directly tailored-built deep networks or exploiting the semantic concepts extracted by using deep learning methods. Unlike previous editions, we received more semi-automatic approaches, which combine human knowledge with state-of-the-art multi-modal information retrieval. Regarding the number of the signed-up teams and the submitted runs, the task keeps growing, with the highest number of registrations and participated teams. It is also a great successful that team retention rate is high with two third non-organiser teams from last year continued to participate in this year. This again confirms how interesting and challenging lifelogging is. As next steps, we do not plan to enrich the dataset but rather provide richer and better concepts, improve the quality of the queries and narrow down the application of the challenges.

5 Acknowledgement

This publication has emanated from research supported in party by research grants from Irish Research Council (IRC) under Grant Number GOIPG/2016/741 and Science Foundation Ireland under grant numbers SFI/12/RC/2289 and SFI/13/RC/2106.

The authors thank to Thanh-An Nguyen, Trung-Hieu Hoang, and the annotation team of Software Engineering Laboratory (SELab), University of Science, VNU-HCM for supporting in building the data collection as well as giving valuable discussions for the Lifelog moment retrieval (LMRT) task.

References

1. Abdallah, F.B., Feki, G., Ammar, A.B., , Amar, C.B.: Big Data For Lifelog Moments Retrieval Improvement. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org> (2019)
2. Dang-Nguyen, D.T., Piras, L., Riegler, M., Boato, G., Zhou, L., Gurrin, C.: Overview of ImageCLEFlifelog 2017: Lifelog Retrieval and Summarization. In: CLEF 2017 Labs Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org> (2017)
3. 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. In: CLEF2018 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org> (2018)
4. Dao, M.S., Vo, A.K., Phan, T.D., , Zettsu, K.: BIDAL@imageCLEFlifelog2019: The Role of Content and Context of Daily Activities in Insights from Lifelogs. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org> (2019)

5. Dodge, M., Kitchin, R.: 'Outlines of a world coming into existence': Pervasive computing and the ethics of forgetting. Environment and Planning B: Planning and Design 34(3), 431–445 (2007)
6. Dogariu, M., Ionescu, B.: Multimedia Lab @ ImageCLEF 2019 Lifelog Moment Retrieval Task. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org> (2019)
7. Gurrin, C., Joho, H., Hopfgartner, F., Zhou, L., Albatal, R.: Overview of NTCIR-12 Lifelog Task (2016)
8. Gurrin, C., Joho, H., Hopfgartner, F., Zhou, L., Gupta, R., Albatal, R., Dang-Nguyen, D.T.: Overview of NTCIR-13 Lifelog-2 Task. In: Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies (2017)
9. Hoang, T.H., Tran, M.K., Nguyen, V.T., Tran, M.T.: Solving Life Puzzle with Visual Context-based Clustering and Habit Reference. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org> (2019)
10. Ionescu, B., Müller, H., Péteri, R., Cid, Y.D., Liauchuk, V., Kovalev, V., Klimuk, D., Tarasau, A., Abacha, A.B., Hasan, S.A., Datla, V., Liu, J., Demner-Fushman, D., Dang-Nguyen, D.T., Piras, L., Riegler, M., Tran, M.T., Lux, M., Gurrin, C., Pelka, O., Friedrich, C.M., de Herrera, A.G.S., Garcia, N., Kavallieratou, E., del Blanco, C.R., Rodríguez, C.C., Vasillopoulos, N., Karampidis, K., Chamberlain, J., Clark, A., Campello, A.: ImageCLEF 2019: Multimedia retrieval in medicine, lifelogging, security and nature. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the 10th International Conference of the CLEF Association (CLEF 2019), LNCS Lecture Notes in Computer Science, Springer (2019)
11. Le, N.K., Nguyen, D.H., Nguyen, V.T., Tran, M.T.: Lifelog Moment Retrieval with Advanced Semantic Extraction and Flexible Moment Visualization for Exploration. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org> (2019)
12. Lin, T., Maire, M., Belongie, S.J., Bourdev, L.D., Girshick, R.B., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: common objects in context. CoRR abs/1405.0312 (2014), http://arxiv.org/abs/1405.0312
13. Ninh, V.T., Le, T.K., Zhou, L., Piras, L., Riegler, M., Lux, M., Tran, M.T., Gurrin, C., Dang-Nguyen, D.T.: LIFER 2.0: Discover Personal Lifelog Insight by Interactive Lifelog Retrieval System. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org> (2019)
14. Ren, S., He, K., Girshick, R.B., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. CoRR abs/1506.01497 (2015), http://arxiv.org/abs/1506.01497
15. Ribeiro, R., Neves, A.J.R., Oliveira, J.L.: UAPTBioinformatics working notes at ImageCLEF 2019 Lifelog Moment Retrieval (LMRT) task. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org> (2019)
16. Taubert, S., Kahl, S.: Automated Lifelog Moment Retrieval based on Image Segmentation and Similarity Scores. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org> (2019)
17. Tournadre, M., Dupont, G., Pauwels, V., Cheikh, B., Lmami, M., , Ginsca, A.L.: A Multimedia Modular Approach to Lifelog Moment Retrieval. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org> (2019)

18. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A.: Places: A 10 million image database for scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (2017)
19. Zhou, L., Hinbarji, Z., Dang-Nguyen, D.T., Gurrin, C.: Lifer: An interactive lifelog retrieval system. In: Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge. pp. 9–14. LSC '18, ACM, New York, NY, USA (2018), http://doi.acm.org/10.1145/3210539.3210542
20. Zhou, P., Bai, C., Xia, J.: ZJUTCVR Team at ImageCLEFlifelog2019 Lifelog Moment Retrieval Task. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org> (2019)