

Overview of ImageCLEF Lifelog 2020: Lifelog Moment Retrieval and Sport Performance Lifelog

Van-Tu Ninh⁴, Tu-Khiem Le⁴, Liting Zhou⁴, Luca Piras², Michael Riegler³,
Pål Halvorsen³, Mathias Lux⁵, Minh-Triet Tran⁶, Cathal Gurrin⁴, and
Duc-Tien Dang-Nguyen¹

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 the fourth edition of Lifelog challenges in ImageCLEF 2020. In this edition, the Lifelog challenges consist of two tasks which are Lifelog Moments Retrieval (LMRT) and Sport Performance Lifelog (SPLL). While the Lifelog Moments Retrieval challenge follows the same format of the previous edition, its data is a larger multimodal dataset based on the merger of three previous NTCIR Lifelog datasets, which contain approximately 191,439 images with corresponding visual concepts and other related metadata. The Sport Performance Lifelog, which is a brand new challenge, is composed of three subtasks that focus on predicting the expected performance of athletes who trained for a sport event. In summary, the ImageCLEF Lifelog 2020 receives 50 runs from six teams in total with competitive results.

1 Introduction

Due to the widely use of wearable devices, digital sensors, and smart phones which capture photos, biometric signals, and location information passively, a huge amount of daily-life data is recorded by many people everyday. As a result, there is an ever increasing reserach effort into developing methodologies for exploiting the potential of this data. Such lifelog data has been used for many retrieval and analytics challenges since the inaugoral NTCIR-12 - Lifelog task [6] in 2016.

There have been many reserch tasks addressed by these challenges, such as lifelog retrieval, data segmentation, data enhancement/annotation and interactive retrieval. Specifically in the ImageCLEF lifelog challenge, we note a number

Copyright © 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CLEF 2020, 22-25 September 2020, Thessalonik, Greece.

of different tasks, such as the Solve My Life Puzzle task in 2019 [5], Activity of Daily Living Understanding task in 2018 [4], or Lifelog Summarization task in 2017 [3]. Therefore, in the fourth edition of ImageCLEFlifelog tasks hosted in ImageCLEF 2020 [12], the organizers both propose a brand-new task which monitors the wellbeing and predicts the expected performance of the athletes training for a sporting event, as well as continuing to maintain the core Lifelog Retrieval Moments task with enriched dataset in terms of visual concepts, annotations, and scale.

Details of the two challenges and the data employed are provided in section 2. In section 3, submissions and results are presented and discussed. In final section 4, the paper is concluded with the discussion of final remarks and future work.

2 Overview of the Task

2.1 Motivation and Objectives

Personal lifelog data is continually increasing in volume due to the popularity of personal wearable/portable devices for health monitoring and life recording such as smartphones, smart watches, fitness bands, video cameras, biometric data devices and GPS or location devices. As a huge amount of data is created daily, there is a need for a systems that can analyse the data, index, categorize, summarize to gain deep insights from the data and support a user in some positive way.

Although many related workshops of lifelogging were held successfully for years such as three editions of NTCIR, annual Lifelog Search Challenge (LSC), and ImageCLEFlifelog 2019, we still aim to bring the attention of lifelogging to not only research groups but also to diverse audiences. Nevertheless, we continue to maintain the core task to encourage research groups to propose creative retrieval approaches to lifelog data, as well as nominating a new task to introduce a new challenge to the research community.

2.2 Challenge Description

Lifelog Moment Retrieval Task (LMRT)

In this task, the participants were required to retrieve a number of specific moments in a lifelogger’s life. Moments are defined as ”semantic events, or activities that happened throughout the day” [5]. For example, a participant would have been required to find and return relevant moments for the query “Find the moment(s) when the lifelogger was having an ice cream on the beach”. 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 and aimed towards compete relevance judgements. Figure 1 illustrates some examples of the moments when the lifelogger was shopping in a toy shop. In addition, listings 1 and 2 show all the

T.001 Having Beers in a Bar

Description: Find the moment(s) in 2015 or 2016 when u1 enjoyed beers in the bar.
Narrative: To be considered relevant, u1 must be clearly in a bar and drink beers.

T.002 Building Personal Computer

Description: Find the moment(s) when u1 built my personal computer from scratch.
Narrative: To be considered relevant, u1 must be clearly in the office with the PC parts or uncompleted PCs on the table.

T.003 In A Toyshop

Description: Find the moment(s) when u1 was looking at items in a toyshop.
Narrative: To be considered relevant, u1 must be clearly in a toyshop where various toys are being examined.

T.004 Television Recording

Description: Find the moment(s) when u1 was being recorded for a television show.
Narrative: To be considered relevant, there must clearly be a television camera in front of u1. The moments the interviewer/cameramen is interviewing/recording u1 are also considered relevant.

T.005 Public Transport In Home Country

Description: Find the moment(s) in 2015 and 2018 when u1 was using public transports in my home country (Ireland).
Narrative: Taking any form of public transport in Ireland is considered relevant, such as bus, taxi, train and boat.

T.006 Seaside Moments

Description: Find moment(s) in which u1 was walking by the sea taking photos or eating ice-cream.
Narrative: To be considered relevant, u1 must be taking a walk by the sea or eating ice-cream with the sea is clearly visible.

T.007 Grocery Stores

Description: Find moment(s) in 2016 and 2018 when u1 was doing grocery shopping on the weekends.
Narrative: To be considered relevant, u1 must be clearly buys or visibly interacts with products in a grocery store on the weekends.

T.008 Photograph of The Bridge

Description: Find the moment(s) when u1 was taking a photo of a bridge.
Narrative: 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.009 Car Repair

Description: Find the moment(s) when u1 was repairing his car in the garden.
Narrative: Moments when u1 was repairing his car in the garden with the gloves on his hand. Sometimes, he also held the hammer and his phone and these moments are also considered relevant.

T.010 Monsters

Description: Find the moment(s) when u1 was looking at an old clock, with flowers visible, with a small monster watching u1.
Narrative: Moments when u1 was at home, looking at an old clock, with flowers visible, with a lamp and two small monsters watching u1 are considered relevant.

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

T.001 Praying Rite

Description: Find the moment when u1 was attending a praying rite with other people in the church.

Narrative: To be relevant, the moment must show u1 is currently inside the church, attending a praying rite with other people.

T.002 Lifelog data on touchscreen on the wall

Description: Find the moment when u1 was looking at lifelog data on a large touchscreen on the wall.

Narrative: To be relevant, the moment must show u1 was looking at his lifelog data on the touchscreen wall (not desktop monitor).

T.003 Bus to work - Bus to home

Description: Find the moment when u1 was getting a bus to his office at Dublin City University or was going home by bus.

Narrative: To be relevant, u1 was on the bus (not waiting for the bus) and the destination was his home or his workplace.

T.004 Bus at the airport

Description: Find the moment when u1 was getting on a bus in the aircraft landing deck in the airport.

Narrative: To be relevant, u1 was walking out from the airplane to the bus parking in the aircraft landing deck with many airplanes visible.

T.005 Medicine cabinet

Description: Find the moment when u1 was looking inside the medicine cabinet in the bathroom at home.

Narrative: To be considered relevant, u1 must be at home, looking inside the opening medicine cabinet beside a mirror in the bathroom.

T.006 Order Food in the Airport

Description: Find the moment when u1 was ordering fast food in the airport.

Narrative: To be relevant, u1 must be at the airport and ordering food. The moments that u1 was queuing to order food are also considered relevant.

T.007 Seafood at Restaurant

Description: Find moments when u1 was eating seafood in a restaurant in the evening time.

Narrative: The moments show u1 was eating seafood or seafood parts in any restaurant in the evening time are considered relevant.

T.008 Meeting with people

Description: Find the moments when u1 was a round-table meeting with many people and there was pink (not red) name-cards for each person.

Narrative: The moments show u1 was at a roundtable and having a meeting with many people, and with pink name-cards visible are considered relevant.

T.009 Eating Pizza

Description: Find the moments when u1 was eating a pizza while talking to a man.

Narrative: To be considered relevant, u1 must eat or hold a pizza with a man in the background.

T.010 Socialising

Description: Find the moments when u1 was talking to a lady in a red top, standing directly in front of a poster hanging on a wall.

Narrative: To be relevant, u1 must be talking with a woman in red, who was standing right in front of a scientific research poster.

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 looking at items in a toyshop where various toys are being examined.’

queries used in the challenge.

Sport Performance Lifelog (SPLL)

Given a dataset of 16 people who train for a 5km run for the sport event (e.g., daily sleeping patterns, daily heart rate, sport activities, and image logs of all food consumed during the training period), participants are required to predict the expected performance (e.g., estimated finishing time, average heart rate and calories consumption) of the trained athlete.

2.3 Datasets

LMRT Task: The data is a combination of three previously released datasets of NTCIR-Lifelog Tasks: NTCIR-12 [6], NTCIR-13 [7], and NTCIR-14 [8]. It is a large multimodal lifelog data of 114 days from one lifelogger whose dates range from 2015 to 2018. It contains five main data types which are multimedia content, biometrics data, location and GPS, human activity data, and visual concepts and annotations of non-text multimedia content. Details of each type of data are as follows:

- **Multimedia Content:** Most of this data are non-annotated egocentric photos captured passively by two wearable digital cameras: OMG Autographer and Narrative Clip¹. The lifelogger wore the device for 16-18 hours per day to capture a complete visual trace of daily life with about 2-3 photos captured per minute during waking hours. The photo data was manually redacted to remove identifiable content and faces [9].

¹ Narrative Clip and Narrative Clip 2 - <http://getnarrative.com>

Table 1. Statistics of ImageCLEFlifelog 2020 LMRT Data

Characters	Size
Number of Lifeloggers	1
Number of Days	114 days
Size of the Collection	37.1 GB
Number of Images	191,439 images
Number of Locations	166 semantic locations
Number of LMRT dev queries	10 queries
Number of LMRT test queries	10 queries

- **Biometrics Data:** This data contains heart rate, calories, and movement speed using a Fitbit fitness tracker². The lifelogger wore the Fitbit wearable device everyday for 24 hours so as to record continuous biometrics data.
- **Location and GPS:** 166 semantic locations as well as GPS data (with and without location name) are recorded using both Moves app and smartphones. The GPS plays an important role to infer the time zone of lifelogger’s current location to convert the time of different wearable devices into one standard timestamp.
- **Human Activity Data:** This data is recorded by the Moves app which also provide some annotated semantic locations. It consists of four types of activities which are walking, running, transport, and airplane.
- **Visual Concepts and Annotations:** The passively auto-captured images are passed through two deep neural networks to extract visual concepts about scenes and visual objects. For scene identification, we still employ the PlacesCNN [25] as in the latest edition. For visual object detection, we employed Mask R-CNN [10] pre-trained on 80 items of MSCOCO dataset [15] which is used to provide the category of visual objects in the image as well as its bounding boxes.

Format of the metadata. The metadata was stored in a csv files, which was called the metadata table. The structure and meaning of each field in the table are described in Table 3. Additionally, visual categories and concepts descriptors are also provided in the visual concepts table. The format of it could be found in Table 4.

SPLL Task: The data was gathered using three different approaches: wearable devices (Fitbit Fitness Tracker (Fitbit Versa), Google Forms, and PMSYS. Biometric data of an individual (training athlete) is recorded using Fitbit Fitness Tracker, including 13 different fields of information such as daily heart rate, calories, daily sleeping patterns, sport activities, etc. Google Forms were used to collect information of meals, drinks, medications, etc. At the same time, information of subjective wellness, injuries, and training load was recorded by PMSYS system. In addition, image-logs of food consumed during the training period from at least 2 participants and self reported data like mood, stress, fatigue, readiness

² Fitbit Fitness Tracker (FitBit Versa) - <https://www.fitbit.com>

to train and other measurements also used for professional soccer teams [23]. The data was approved by the Norwegian Center for Research Data with proper copyright and ethical approval to release. Statistics of the ImageCLEFlifelog 2020 SPL data is shown in table 2.

Table 2. Statistics of ImageCLEFlifelog 2020 SPL Data

Categories	Approach	Rate of entries	Number of entries
Calories	Fitbit	Per minute	3377529
Steps	Fitbit	Per minute	1534705
Distance	Fitbit	Per minute	1534705
Sleep	Fitbit	When it happens (usually daily)	2064
Lightly active minutes	Fitbit	Per day	2244
Moderately active minutes	Fitbit	Per day	2396
Very active minutes	Fitbit	Per day	2396
Sedentary minutes	Fitbit	Per day	2396
Heart rate	Fitbit	Per 5 seconds	20991392
Time in heart rate zones	Fitbit	Per day	2178
Resting heart rate	Fitbit	Per day	1803
Exercise	Fitbit	When it happens 100 entries per file	2440
Sleep score	Fitbit	When it happens (usually daily)	1836
Google Forms reporting	Google Form	Per day	1569
Wellness	PMSYS	Per day	1747
Injury	PMSYS	Per day	225
SRPE	PMSYS	Per day	783

2.4 Performance Measures

LMRT: Classic metrics are employed to assess the performance of this LMRT task. These metrics include:

- 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 are considered, e.g., X=5, 10, 20, 30, 40, 50. Official ranking metrics are the F1-measure@10, which gives equal importance to diversity (via CR@10) and relevance (via P@10). In particular, the final score to rank

Table 3. Structure of metadata table.

Field name	Meaning	Example
minute_ID	Identity field for every minute	20180503_0000
utc_time	UTC Time with format: UTC_YYYY-MM-DD_HH:MM.UTC	UTC_2018-05-03_00:00
local_time	Local time in lifelogger’s timezone (combined from lifelogger’s smart phone and other applications): YYYY-MM-DD_HH:MM	2018-05-03_01:00
timezone	The name of lifelogger’s timezone	Europe/Dublin
lat	Latitude of lifelogger’s position	53.386881
lon	Longitude of lifelogger’s position	-6.15843
semantic_name	The name of the place corresponding to lifelogger’s position	Home
elevation	Information of the floors that the lifelogger went up estimated by Fitbit	69
speed	The current speed when the lifelogging was moving	14.5
activity_type	The activity that lifelogger was doing at that time	walking, transport
calories	The number of calories collected by wearable devices	1.17349994
heart_rate	The heart rate of volunteer at that time collected by wearable devices	73
steps	The number of lifelogger’s steps collected by wearable devices	14

submissions of participants is the average F1-measure@10 of ten queries, which provides information of general performance of each interactive system for all 10 queries.

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

Table 4. Structure of the Visual Concepts table.

Field name	Meaning	Example
minute_id	Identity field for every image, including images from wearable camera and smart phone camera	20180503_0000
utc_time	UTC Time with format: UTC.YYYY-MM-DD.HH:MM	UTC_2018-05-03.00:00
image_path	Image path to the corresponding image	2016-08-08/20160808_084305_000.jpg
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 Mask R-CNN, trained on the COCO dataset	person, 0.987673 508.568878 171.124496 513.541748 395.073303

real users) were encouraged.

SPLL: For this task, we employ two evaluation metrics to rank the submissions of participants. The primary score is to check how accurately the participants can predict whether it was an improvement or a deterioration after the training process by comparing the sign of the actual change value to the predicted one. The secondary score is the absolute difference between the actual change and the predicted one. The primary score is ranked in descending order, and if there is a draw in the primary score, the secondary score is used to re-rank the teams.

2.5 Ground Truth Format

LMRT Task. The development 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 the cluster id number. Lines were separated by an end-of-line character (carriage return). An example is presented below:

```
- 1, 1
- 1, 2
- ...
- 2, 1
- 2, 2
- ...
```

In the relevant ground-truth file, the first value on each line was the topic id, followed by a unique photo id which is image name without the extension, 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, b00001216_21i6bq_20150306_174552e, 1
- 1, b00001217_21i6bq_20150306_174713e, 1
- 1, b00001218_21i6bq_20150306_174751e, 1
- 1, b00002953_21i6bq_20150316_203635e, 2
- 1, b00002954_21i6bq_20150316_203642e, 2
- ...
- 2, b00000183_21i6bq_20150313_072410e, 1
- 2, b00000184_21i6bq_20150313_072443e, 1
- 2, b00000906_21i6bq_20150312_171852e, 2
- 2, b00000908_21i6bq_20150312_172005e, 2
- 2, b00000909_21i6bq_20150312_172040e, 2
- ...
```

SPLL Task. The ground truth was provided in one txt file. For each line in this file, the first value was the id of the sub-task which is 1, 2 or 3 (since the SPLL task is split into three sub-tasks), followed by the id of individual (p01, p02, ..., p16), followed by the actual change in the status of the individual after the training period. Although the three sub-tasks has different requirements, their output format is the same, which is a number indicating the change before and after training with preceding '+' sign if the change is an increase, or '-' sign if the change is a decrease. If there is no change after the training process, a 0 value without a preceding sign is also allowed. Values in each line were separated by comma. Lines were separated by and end-of-line character (carriage return). An example is shown below:

```
- 1, p01, +8
- 1, p10, +86
- ...
```

Table 5. Official results of the ImageCLEFlifelog 2020 LMRT task.

Team	Run	P@10	CR@10	F1@10	Team	Run	P@10	CR@10	F1@10
Organiser	RUN1*	0.19	0.31	0.21	HCMUS	RUN1	0.79	0.73	0.72
	RUN2*	0.23	0.44	0.27		RUN2	0.78	0.73	0.72
	RUN3*	0.36	0.38	0.32		RUN3	0.79	0.69	0.71
REGIM	RUN1	0.04	0.08	0.05	RUN4	0.80	0.74	0.74	
	RUN2	0.16	0.22	0.17	RUN5	0.81	0.77	0.75	
	RUN3	0.17	0.24	0.19	RUN6	0.81	0.79	0.77	
	RUN4	0.00	0.00	0.00	RUN7	0.82	0.81	0.79	
	RUN5	0.19	0.16	0.16	RUN8	0.77	0.76	0.74	
	RUN6	0.03	0.05	0.04	RUN9	0.85	0.81	0.81	
	RUN7	0.17	0.24	0.19	RUN10	0.86	0.81	0.81	
UATP	RUN1	0.02	0.07	0.03	BIDAL	RUN1	0.69	0.68	0.65
	RUN2	0.02	0.07	0.03		RUN2	0.68	0.63	0.58
	RUN3	0.50	0.58	0.52		RUN3	0.68	0.69	0.65
DCU-DDTeam	RUN1	0.07	0.13	0.09	RUN4	0.70	0.69	0.66	
	RUN2	0.22	0.39	0.25	RUN5	0.72	0.69	0.66	
	RUN3	0.44	0.63	0.41	RUN6	0.73	0.69	0.67	
	RUN4	0.58	0.53	0.48	RUN7	0.75	0.65	0.64	
	RUN5	0.16	0.36	0.21	RUN8	0.73	0.69	0.67	
					RUN9	0.73	0.70	0.69	
					RUN10	0.74	0.70	0.69	

Notes: * submissions from the organizer teams are just for reference. The results in this paper are official version of ImageCLEFlifelog 2020 tasks.

- 2, p01, +0.2
- 2, p10, +0
- 2, p16, +2.3
- ...
- 3, p01, +0.2

3 Evaluation Results

3.1 Participating Groups and Runs Submitted

This year, we obtained 50 valid submissions in both two tasks of ImageCLEFlifelog tasks from 6 teams, which is not as high as in previous year. However, the results of these submissions show a significant improvement in the final scores compared to ImageCLEFlifelog 2019. In particular, there were 38 submissions in LMRT with 6 teams participating in the task, while only one non-organizer team submitted 10 runs in SPL task. The submitted runs and their results are summarised in Tables 5 and 6.

3.2 Results

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

The Organizer team continue to provided a baseline approach for the LMRT task with a web-based interactive search engine, which is an improved version

Table 6. Official Results of the ImageCLEFlifelog 2020 SPLL Task.

Team	Run	Primary Score	Secondary Score
Organiser	RUN1*	0.47	313.30
	RUN2*	0.41	203.10
BIDAL	RUN1	0.77	306.90
	RUN2	0.52	309.10
	RUN3	0.59	254.70
	RUN4	0.59	372.60
	RUN5	0.53	375.20
	RUN6	0.65	319.60
	RUN7	0.71	250.20
	RUN8	0.82	245.60
	RUN9	0.82	128.00
	RUN10	0.65	112.00

Notes: * submissions from the organizer teams are just for reference. The results in this paper are official version of ImageCLEFlifelog 2020 tasks.

of LIFER 2.0 system [18] which was used at ImageCLEFlifelog 2019. The interactive elements of this system comprise three features: free-text querying and filtering, visual similarity image search, and elastic sequencing to view nearby moments. The system, which focuses on experimenting the efficiency of free-text query features, is an early version of LifeSeeker 2.0 interactive search engine [14]. For the query processing procedure, the authors use natural language processing to parse the query into meaningful terms and employ Bag-of-Words to retrieve and rank relevant documents. The dictionary in Bag-of-Words is split into three dictionaries for filtering by using terms matching: time, location, and visual concepts. The authors extract more detailed visual concepts inferred from deep neural networks pre-trained on Visual Genome dataset [13] which were shown to be extremely useful for the retrieval process. For the SPLL task, the organizer team provided baseline approaches for all three sub-tasks, which used the exercise data from Fitbit Tracker, self-reporting, and food-images only. The authors propose a naive solution which computes the difference between consecutive rows of data from exercise activities and self-reporting including distance, exercise duration, calories, and weight; then categorises them into positive and negative groups based on sign of the value ('+' or '-') and calculate the average of the two groups. Finally, they sum the two average to obtain the results. In addition, they also try to build a Linear Regression Model to predict the pace change and a Convolutional Neural Network to detect the type of food for manual calories inference.

The REGIM-Lab approaches the LMRT Task with the same strategies as their work in ImageCLEFlifelog 2019 [1], which used the ground truth of the development dataset from both LMRT 2019 and LMRT 2020 to automatically categorise images into categories for deep neural network fine-tuning with MobileNet v2 and DenseNet and visual concepts clustering. However, the difference

is that they use Elastic Search and Kibana Query Language (KBL) to perform retrieval on image concepts and metadata instead Apache Cassandra and Cassandra Query Language (CQL). Moreover, they attempt to enrich more visual concepts using YOLO v3 [19] trained on OpenImage dataset. They also treat the textual queries with three-word embedding models built from scratch which are Word2vec, FastText, and Glove

HCMUS focused on the LMRT task only this year. Their retrieval system has three components which are query by caption, query by similar image, and query by place and time from the metadata. For query by caption, they encoded images using Faster R-CNN [20] to extract object-level features, then applied self-attention to learn interaction between them. For query sentence, they used RoBERTa model [16] to encode sentences. Finally, two feed forward networks were deployed to map image and text features to the common space correspondingly. Therefore, when a sentence is given, their model ranked all images based on cosine distance between the encoded images and the encoded query sentence to find the most relevant images to the description of the sentence. For query by similar image, the same strategy was applied with ResNet152 image encoder [11] instead of Faster R-CNN. For query by place and time from metadata, they simply find all moments based on the given semantic locations and view the images which are before and after a specific moments.

DCU-DDTeam interactive search engine is the improved version of their Mysceál system in LSC'20 [24] and follows the same pipeline. The visual concepts of each image are the combination of the given metadata and outputs from DeepLabv3+ [2] and the enriched metadata extracted from Microsoft Computer Vision API. These annotations, along with other information such as locations and time, were then indexed in the Elastic Search. The input query is analyzed to extract the main information and enlarged by their expansion mechanism. They are then combined with the indexed database to find matching images which are then ordered by their ranking function. In this version, they introduced three changes in the previous system including visual similarity, the user interface, and the summary panel. The visual similarities between images were measured by using cosine distance between visual features composed of SIFT [17] and VGG16 features. For the user interface, the authors remove the triad of searching bars as in the original version and reorganised the interface to explore cluster events more efficiently. The summary panel consists of the “Word List” panel which is the area on the screen showing the results of their query expansion with adjustable scores allowing the user to emphasize the concepts that they need to retrieve.

The BIDAL team is the only non-organizer team participating in both the LMRT and SPLL tasks. For the LMRT task, the authors generated clusters by employing a scene recognition model trained on the Google Fixmatch method [22]. They then used an attention mechanism to match the input query with the correct samples, which were then utilized to find other relevant moments. For the SPLL task, they summarized information from various interval attributes, removed several unnecessary attributes, and generated some new attributes. Then,

they trained several typical time-series neural network structure including Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) using the generated attributes or a set of attributes, or some pre-defined seeding attributes.

The UA.PT Bioinformatics team continued to employ the approaches from last year challenge [21] to test the performance of their automatic lifelog search engine in the attempt to enrich visual concepts and labels by utilising many different object detection networks including Faster R-CNN and Yolov3 [19] pre-trained on COCO dataset [15]. The image retrieval procedure was then done on a text-based vector space by computing similarity score between the text labels extracted from the images and the visual concepts. Finally, a threshold was set to choose the results for each topic. As the results prove that this automatic approach did not work, the authors developed a web-based interactive search engine with a timestamp-clustering visualization to select the moments instead of defining a threshold to choose the results automatically. The algorithms for searching relevant moments are mostly the same as automatic approach except for three new features which are: narrowing searching items by text matching between manually analysed query and the indexed database containing concepts of each image.

The official results are summarised in Tables 5 and 6. There are six teams participating in the LMRT task with the highest F1@10 score of 0.81 was achieved by HCMUS (Table 5). Most of the teams tried to enrich the visual concepts by deploying different CNNs for objects and places detection, then performing text analysis on the query and text matching. Some additional features were also added in most interactive systems such as searching for visually similar images, terms weighting for results re-ranking, context understanding before performing search, etc. The highest scoring approach by the HCMUS team, considered visual vector features extracted from CNNs when making the comparison between feature vectors to find relative moments.

In the SPLL task, only one non-organizer team participated and they managed to achieve good scores. For the prediction of performance change, their approach gained 0.82 and 128.0 in terms of prediction accuracy and L1 distance between the prediction value and actual change respectively.

4 Discussions and Conclusions

In the ImageCLEFlifelog 2020, most of the submitted results managed to gain high scores in both tasks. Although the set of topics for LMRT task is different from previous edition, participants managed to search for the relative moments on a large-scale dataset while still achieving higher scores than the results in previous edition. This proves that the proposed features and query mechanisms actually enhance the performance of their retrieval systems. Most of the teams enrich semantic visual concepts using many different CNNs pretrained on different datasets such as COCO, OpenImage, and Visual Genome before indexing and querying; retrieve relative images based on text matching and text retrieval

algorithms; perform visual similar image search. We also note many interesting approaches from teams to enhance the affordance and interaction of the retrieval systems, including integrating filter mechanism into free-text search, considering adding visual vector features into the final encoded vector, clustering images into events, etc.

Regarding the number of teams and submitted runs, only 6 teams participated in the LMRT task, including an organizer team, which produced 50 submissions in total. Each team was allowed to submit up to 10 runs. For the LMRT task, among five teams which participated in ImageCLEFlifelog 2019 (including the organizer team), four teams managed to obtain better results with the highest F1-score up to 0.81. The mean (SD) increase of final F1-score from these five teams is 0.25 (0.18). The new team from Dublin City University also managed to achieve the 4th rank with a 0.48 F1-score. For the SPLL task, as the task is new, only one team from The Big Data Analytics Laboratory submitted 10 runs. Their best submission achieves an accuracy of performance change and the absolute difference between the prediction and actual change are 0.82 and 128 respectively, which is a good result.

5 Acknowledgement

This publication has emanated from research supported in part 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.

References

1. Abdallah, F.B., Feki, G., Ammar, A.B., Amar, C.B.: Big data for lifelog moments retrieval improvement. In: CLEF (2019)
2. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: ECCV (2018)
3. 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)
4. 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: CLEF (2018)
5. Dang-Nguyen, D.T., Piras, L., Riegler, M., Zhou, L., Lux, M., Tran, M.T., Le, T.K., Ninh, V.T., Gurrin, C.: Overview of imagecleflifelog 2019: Solve my life puzzle and lifelog moment retrieval. In: CLEF (2019)
6. Gurrin, C., Joho, H., Hopfgartner, F., Zhou, L., Albatal, R.: Overview of ntcir-12 lifelog task. In: NTCIR (2016)
7. Gurrin, C., Joho, H., Hopfgartner, F., Zhou, L., Gupta, R., Albatal, R., Dang-Nguyen, D.T.: Overview of ntcir-13 lifelog-2 task (2017)
8. Gurrin, C., Joho, H., Hopfgartner, F., Zhou, L., Ninh, V.T., Le, T.K., Albatal, R., Dang-Nguyen, D.T., Healy, G.: Overview of the ntcir-14 lifelog-3 task (2019)

9. Gurrin, C., Smeaton, A.F., Doherty, A.R.: Lifelogging: Personal big data. *Found. Trends Inf. Retr.* 8, 1–125 (2014)
10. He, K., Gkioxari, G., Dollár, P., Girshick, R.B.: Mask r-cnn. 2017 IEEE International Conference on Computer Vision (ICCV) pp. 2980–2988 (2017)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 770–778 (2016)
12. Ionescu, B., Müller, H., Péteri, R., Abacha, A.B., Datla, V., Hasan, S.A., Demner-Fushman, D., Kozlovski, S., Liauchuk, V., Cid, Y.D., Kovalev, V., Pelka, O., Friedrich, C.M., de Herrera, A.G.S., Ninh, V.T., Le, T.K., Zhou, L., Piras, L., Riegler, M., Halvorsen, P., Tran, M.T., Lux, M., Gurrin, C., Dang-Nguyen, D.T., Chamberlain, J., Clark, A., Campello, A., Fichou, D., Berari, R., Brie, P., Dogariu, M., Ștefan, L.D., Constantin, M.G.: Overview of the ImageCLEF 2020: Multimedia retrieval in lifelogging, medical, nature, and internet applications. In: *Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the 11th International Conference of the CLEF Association (CLEF 2020)*, vol. 12260. LNCS Lecture Notes in Computer Science, Springer, Thessaloniki, Greece (September 22–25 2020)
13. Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., Chen, S., Kalantidis, Y., Li, L.J., Shamma, D.A., Bernstein, M.S., Fei-Fei, L.: Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International Journal of Computer Vision* 123, 32–73 (2016)
14. Le, T.K., Ninh, V.T., Tran, M.T., Nguyen, T.A., Nguyen, H.D., Zhou, L., Healy, G., Gurrin, C.: Lifeseeker 2.0: Interactive lifelog search engine at lsc 2020. *Proceedings of the Third Annual Workshop on Lifelog Search Challenge* (2020)
15. Lin, T.Y., Maire, M., Belongie, S.J., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. *ArXiv abs/1405.0312* (2014)
16. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach. *ArXiv abs/1907.11692* (2019)
17. Lowe, D.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60, 91– (11 2004)
18. 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: Discovering personal lifelog insights using an interactive lifelog retrieval system. In: *CLEF* (2019)
19. Redmon, J., Farhadi, A.: Yolov3: An incremental improvement. *ArXiv abs/1804.02767* (2018)
20. Ren, S., He, K., Girshick, R.B., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 1137–1149 (2015)
21. Ribeiro, R., Neves, A.J.R., Oliveira, J.L.: Ua.pt bioinformatics at imageclef 2019: Lifelog moment retrieval based on image annotation and natural language processing. In: *CLEF* (2019)
22. Sohn, K., Berthelot, D., Li, C.L., Zhang, Z., Carlini, N., Cubuk, E.D., Kurakin, A., Zhang, H., Raffel, C.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *ArXiv abs/2001.07685* (2020)
23. Thambawita, V., Hicks, S., Borgli, H., Pettersen, S., Johansen, D., Johansen, H., Kupka, T., Stensland, H., Jha, D., Grønli, T.M., Fredriksen, P.M., Eg, R., Hansen, K., Fagernes, S., Biørn-Hansen, A., Dang Nguyen, D.T., Hammer, H., Jain, R., Riegler, M., Halvorsen, P.: Pmdata: A sports logging dataset (02 2020)

24. Tran, L.D., Nguyen, M.D., Binh, N.T., Lee, H., Gurrin, C.: Myscéal: An experimental interactive lifelog retrieval system for lsc'20. Proceedings of the Third Annual Workshop on Lifelog Search Challenge (2020)
25. 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)