Retrieving Events in Life Logging

Ergina Kavallieratou[1], Carlos R. del-Blanco[2] , Carlos Cuevas[2] and Narciso García[2]

[1] Department of Information & Communication Systems Engineering, University of the Aegean, Samos 83200, Greece

[2] Grupo de Tratamiento de Imágenes, ETSI Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain

kavallieratou@aegean.gr, {cda,ccr,narciso}@gti.ssr.upm.es

Abstract. This paper describes our contribution for the Lifelog Moment Retrieval (LMRT) challenge of ImageCLEF Lifelog2018. Lifelogging has a tremendous potential in many applications. However, the wide range of possible moment events along with the lack of fully annotated databases make this task very challenging. This work proposes an interactive and weakly supervised learning approach that can dramatically reduce the time to retrieve any kind of events in huge databases. Impressive results have been obtained in the referred challenge, reaching the first rank.

Keywords: Life Logging, Deep Learning, Supervised Learning.

1 Introduction

Lifelogging is the procedure of retrieving and tracking personal data during the daily life. The potential applications are endless, from memory retrieval [1] to surveillance [2]. Due to this fact, an increasing number of research works and events have been appearing in the last years, such as:

- the LifeLog of DARPA of the U.S. Department of Defense [3] and
- MyLifeBits by Gordon Bell of Microsoft [4-5].

Since then, many lifeloggers have tracked huge quantities of big data for various purposes [6][7]. The technology makes it easy to collect data automatically using sensors. However, there is not still a consolidate framework to analyze all this data to properly extract useful information for the target applications. On the other hand, a lot of discussion is taking place over the privacy and the ethics dimension of lifelogging [8]. Besides all, the huge potential of lifelogging has encouraged new efforts to advance in this line, such as the two competitions that have started the last two years:

- NTCIR Lifelog Task [9]
- ImageCLEFlifelog [10].

This year, the ImageCLEF Lifelog 2018 [11-12] has been divided into two subtasks (challenges):

- The LMRT challenge about Lifelog Moment Retrieval
- The ADLT challenge about Activities of Daily Living.

In this paper, three strategies are presented addressing the LMRT challenge. The participants had to retrieve a number of specific moments in a lifelogger's life. Moments were defined as semantic events or activities that happened throughout the day. The ground truth for this subtask was created using manual annotation. The dataset consisted of 50 days of data from a lifelogger, namely: images (1,500-2,500 per day from wearable cameras), visual concepts (automatically extracted visual concepts with varying rates of accuracy), semantic content (semantic locations, semantic activities) based on sensor readings (via the Moves App) on mobile devices, biometrics information (heart rate, galvanic skin response, calorie burn, steps, etc.), music listening history. The dataset is built based on the data available for the NTCIR-13 - Lifelog 2 Task, which contained a total of 80,440 images.

The rest of the paper is structured as follows. In section 2, the proposed three strategies are described, in section 3 the experimental results for some trials are presented, while the conclusions are included in the section 4.

2 Proposed Strategies

Three different strategies have been conceived for addressing the ImageCLEFlifelog 2018 challenge, with the purpose to accurately retrieve images that correspond to the ten proposed topics (Table 1, Fig.1). The first strategy, called *Two-class strategy*, a deep learning framework has been developed that considers every topic independently. This is, two classes are considered per topic, one representing the event o action described by the topic, and the other the absence of it. The second strategy, called *Ten-class strategy*, considers all the topics simultaneously. Thus, the developed deep neural network uses ten output classes, one per topic. And finally, the last strategy, called *Eleven-class strategy*, is an evolution of the second one that adds an additional output to consider events that do not belong to the 10 referred challenge topics.

In this section the proposed strategies, as well as the preprocessing and postprocessing stages, are described in detail.

Table 1. The topics of the SubTask 2: Lifelog moment retrieval (LMRT).

Topic ID	Topic Title
LST001	Preparing Salad
LST002	VR Experiments
LST003	My Presentations
LST004	Interviewed by a TV presenter
LST005	Dinner at Home
LST006	Assembling Furniture

LST007	Taking a coach/bus in foreign countries
LST008	Costa Coffee with friends
LST009	Using mobile phone or tablets in a vehicle
LST010	Graveyard

```
        <topic>
                <id>001</id>
                <uid>u1</uid>
                      <title>Preparing Salad</title>
                <description>Find the moments when I was preparing
salad.</description>
                <narrative>To be considered relevant, the moments
must show the lifelogger preparing a salad, in a kitchen or in an
office environment. Eating salad is not considered relevant.
Preparing other types of food is not considered
relevant.</narrative>
        </topic>
```

Fig. 1. Exemplary Topic Description.

Table 2. Corresponding Images per topic.

Topic ID	Corresponding Directories	Corresponding Split	#of images
LST001	home+work	Location	27,880
LST002	no activity	Activity	66,506
LST003	no activity	Activity	66,506
LST004	home+work	Location	27,880
LST005	home	Location	8,986
LST006	no activity	Activity	66,506
LST007	Transport	Activity	8,800
LST008	costa coffee	Location	601
LST009	transport+airplane	Activity	10,754
LST010	no place	Location	26,393

2.1 Preprocessing

In order to limit the big volume of images, considering the given metadata and the topics, we decided to split the images in the subdirectories, automatically, by using the Location and Activity tag of the metadata. Thus, two sets of directories were created, named after the names of the specific tag:

1. The Activity set was including just 3 directories: transport, airplane and walking, plus a fourth one called No-activity, including all the images with no information over activity.

2. The Location set was including 96 directories, plus a directory called No-place, including all the images where no named place was mentioned.

This automatic classification helped us to consider less images for a first retrieval to train our systems. Thus, for the presented topics (Table 1) corresponding directories were chosen, according to the description and the restrictions (Fig.1), as they are presented in Table 2.

2.2 Two-class strategy

This strategy (Fig.2) had to be repeated for each topic separately. For each image the question is: Does it satisfy the topic? Thus, for each topic we have two classes, namely: True, where the correct images are included; and False, all the others. After a first retrieval, applied to the corresponding directory, the system is retrained and tested over all data.

Considering the directory sets from preprocessing, the required steps include:

1. Manual choice of true images: In most cases about 10 images were selected as True, most of the times by the same event. Important exceptions were the topics 006 and 010 that there were few examples and, especially in 006, difficult to be found.
2. Training by using pretrained CNN: The pretrained Convolution Neural Networks AlexNet [13] or GoogleNet [14] were used.
3. Testing on the corresponding data (Table 1): The appropriate directories were chosen in accordance to the description and details given of the topic (Fig.1). The four co-authors discussed a lot over the various topics. However, many times we had to ask the organizers for explanations due to cultural differences and definitions.
4. Manually splitting the results to the two classes: here is where the maximum of five minutes of search time allowed per topic, was used. In fact, a simple application was created that was showing the True images and asking for a YES or NO entered by the user. The procedure was very fast. In most topics, 1-2 minutes were enough. The topic 008 required just few seconds. Exception was the topic 006. The negative results were so many that we just kept the True and False that were reached in 5 mins, so not all the images of the corresponding directory were used for the final training.
5. Training using the same pretrained CNN: the AlexNet or GoogleNet that were used in step 2 was also used here.
6. Testing on all data: the retrained CNN was applied to all 80,439 images.
7. Postprocessing, in order to adapt the results to the required format.

Three trials have been submitted by this strategy: one using AlexNet (subm#1), one using GoogleNet (subm#2) and one using the average of the two CNNs (subm#3).

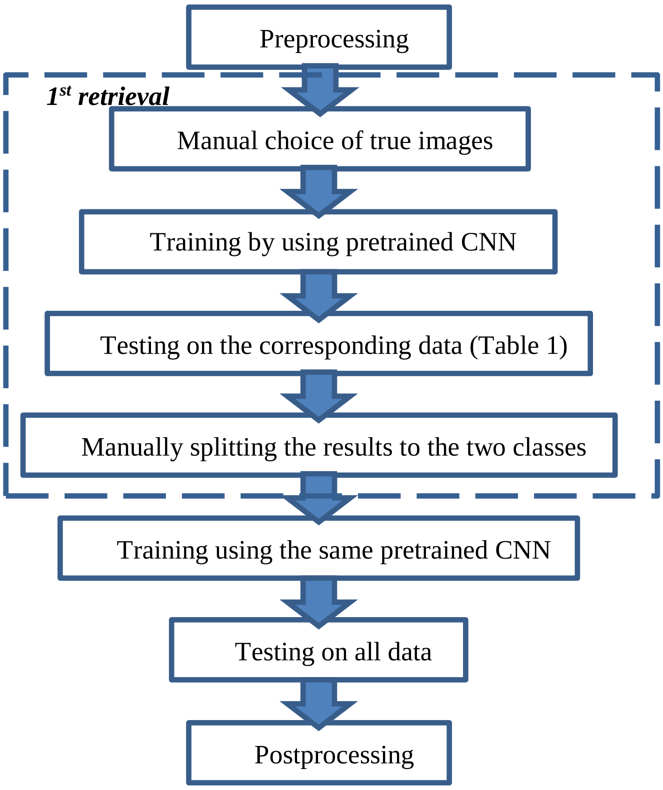

Fig. 2. The Two-class strategy.

2.3 Ten-class strategy

This strategy (Fig.3) is applied just once for the ten topics. However, it is required to have the result of the first retrieval of the Two-class strategy (§2.2) that includes the steps 1-4. Then the True classes of each topic is created by merging the results of the previous strategy for both AlexNet and GoogleNet. These will be the ten classes of this strategy. Thus, the strategy includes the steps:

1. Merging of the True classes of AlexNet and GoogleNet after the 1st retrieval (Fig.2) for each topic i.e. 10 classes.
2. Training a pretrained CNN: the AlexNet or GoogleNet, using the ten classes.
3. Testing on all data: the retrained CNN was applied to all 80,439 images.
4. Postprocessing to adapt the required format.

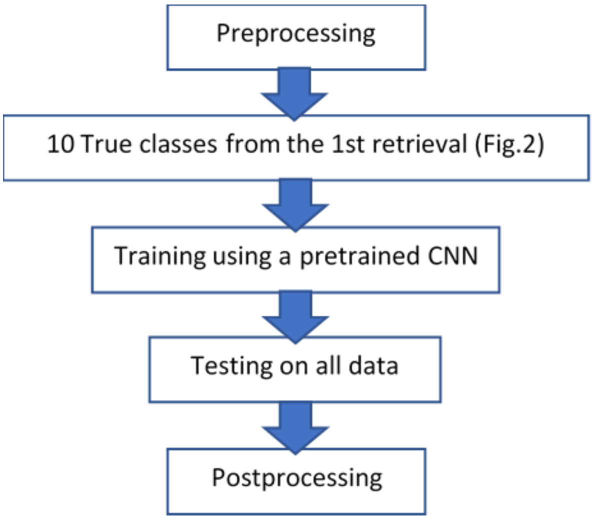

Fig. 3. The Ten-class strategy.

Two trials have been submitted by this strategy: one using AlexNet (subm#4) and one using GoogleNet (subm#5). The AlexNet trial proved to be our best submission.

2.4 Eleven-class strategy

This strategy (Fig.4) is very similar to the previous one, including one more class: the class that an image is included if doesn't belong to any other. For the training, this class was the merging all the False classes of the Two-class strategy, excluding the images that have already included to the classes of the Ten-class strategy. Thus, this strategy includes the steps:

1. Merging of the True classes of AlexNet and GoogleNet after the 1st retrieval (Fig.2) for each topic i.e. 10 classes
2. Merging the False classes of the Two-class strategy, excluding the images included at the 10 classes.
3. Training a pretrained CNN: the AlexNet or GoogleNet, using the eleven classes.
4. Testing on all data: the retrained CNN was applied to all 80,439 images.
5. Postprocessing to adapt the required format.

One trial has been submitted by this strategy using AlexNet (subm#6). It was not possible to submit in-time using GoogleNet (subm#0), since it required to much time for train due to the large number of images in the eleventh class, that is 37,063 images.

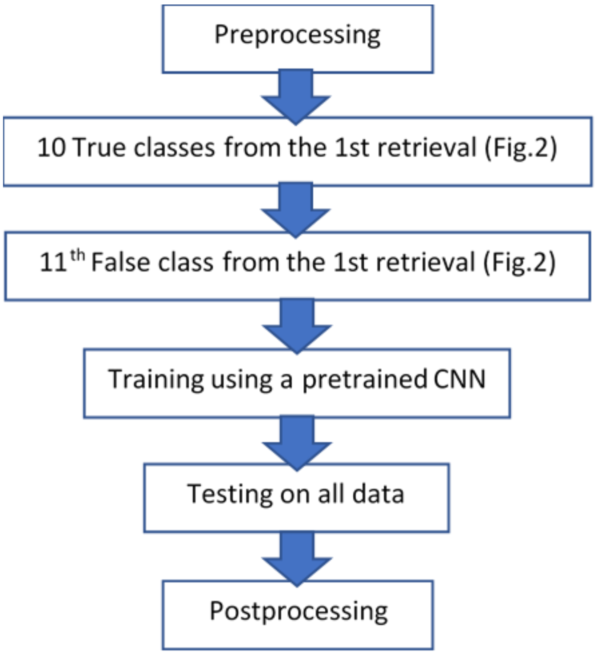

Fig. 4. The Eleven-class strategy.

2.5 Postprocessing

The Subtask 2 of ImageCLEFlifelog 2018 requires for the submissions a CSV file in the following format:

$$[\text{topic id, image id, confidence score}] \tag{1}$$

Where: - topic id: Number of the queried topic, e.g., 1 to 10 - image id: ID of a relevant image - confidence score: from 0 to 1. The CSV file should contain a diversified summarization in 50 images for each query.
The postprocessing procedure is creating the CSV file automatically and it is the same for the three strategies, using the probabilities of the classify level of the CNN. Thus, the images are ranked by the probabilities from high to low, for each result class (True of the Two-class strategy and the ten classes of Ten-class and Eleven-class strategies. As correct are chosen the first 50 images that:

- Have corresponding ID in metadata: the organizers were accepting as possible correct images only the images that were labeled in metadata with an ID number.
- Satisfy all the rules e.g. in topic 005, since dinner is required the time is required to be greater than 15.00.

3 Experimental results

For assessing performance, The organizers proposed the classic metrics for retrieval, specifically:

- 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.

All the presented results have been performed using Matlab in a computer with processor Intel(R) Core™ i7-7700HQ CPU@2.80 GHz x8 and GPU NVIDIA GeForce GTX 1060. Exception was the trial that was not submitted, due to extreme requirements in training. This was finally performed in a computer Intel(R) Core™ i9-7900X CPU@ 3.30 GHz x10 and GPU NVIDIA corporation device 1b02 x2.

Table 3. Indicative results of F1@10 for the proposed techniques.

Submission ID	**Strategy**	**CNN**	**F1@10**
subm#1	Two-class	AlexNet	0.504
subm#2	Two-class	GoogleNet	0.545
subm#3	Two-class	Average	0.477
subm#4	Ten-class	AlexNet	0.536
subm#5	Ten-class	GoogleNet	0.477
subm#6	Eleven-class	AlexNet	0.480
subm#0	Eleven-class	GoogleNet	0.542

Table 4. Results for all the trials of F1@X for X=5, 10, 20, 30, 40, 50.

Submission ID	F1@5	F1@10	F1@20	F1@30	F1@40	F1@50
subm#1	0.395	0.504	0.571	0.604	0.606	0.594
subm#2	0.520	0.545	0.562	0.547	0.523	0.522
subm#3	0.452	0.477	0.445	0.438	0.465	0.473
subm#4	0.543	0.536	0.543	0.552	0.562	0.556
subm#5	0.452	0.477	0.459	0.438	0.465	0.473
subm#6	0.480	0.480	0.495	0.521	0.528	0.549
subm#0	0.507	0.542	0.525	0.534	0.508	0.532

Official ranking metrics this year are the F1-measure@10, which gives equal importance to diversity (via CR@10) and relevance (via P@10). In table 3, indicative results of F1@10 are given for all the mentioned submissions (subm#1-6), plus the not submitted trial of the third strategy (subm#0).

In Table 4, F1@X for various cut off points are considered, with X=5, 10, 20, 30, 40, 50, for all the proposed techniques. Finally, in Tables 5-11, are given all the detailed results for the submission 1-6, plus the no-submitted trial.

Table 5. Detailed results for subm#1.

Topic	P@5	CR@5	F1@5	P@10	CR@10	F1@10	P@20	CR@20	F1@20	P@30	CR@30	F1@30	P@40	CR@40	F1@40	P@50	CR@50	F1@50
LST001	0.8	0.333	0.471	0.9	0.333	0.486	0.75	0.667	0.706	0.733	0.667	0.698	0.65	1	0.788	0.66	1	0.795
LST002	0.4	1	0.571	0.3	1	0.462	0.2	1	0.333	0.233	1	0.378	0.25	1	0.4	0.22	1	0.361
LST003	1	0.333	0.5	1	0.333	0.5	1	0.667	0.8	1	0.667	0.8	1	0.667	0.8	1	0.667	0.8
LST004	1	1	1	1	1	1	1	1	1	0.967	1	0.983	0.975	1	0.987	0.98	1	0.99
LST005	1	0.042	0.08	1	0.042	0.08	0.5	0.042	0.077	0.533	0.083	0.144	0.65	0.083	0.148	0.72	0.083	0.149
LST006	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LST007	0.6	0.333	0.429	0.7	0.5	0.583	0.6	0.833	0.698	0.667	0.833	0.741	0.7	0.833	0.761	0.7	0.833	0.761
LST008	1	0.25	0.4	1	0.5	0.667	1	0.75	0.857	1	1	1	1	1	1	1	1	1
LST009	0	0	0	0.4	0.2	0.267	0.4	0.2	0.267	0.567	0.4	0.469	0.45	0.4	0.424	0.36	0.4	0.379
LST010	1	0.333	0.5	1	1	1	0.95	1	0.974	0.7	1	0.824	0.6	1	0.75	0.54	1	0.701
Mean	0.68	0.362	0.395	0.73	0.491	0.504	0.64	0.616	0.571	0.64	0.665	0.604	0.627	0.698	0.606	0.618	0.698	0.594

Table 6. Detailed results for subm#2.

Topic	P@5	CR@5	F1@5	P@10	CR@10	F1@10	P@20	CR@20	F1@20	P@30	CR@30	F1@30	P@40	CR@40	F1@40	P@50	CR@50	F1@50
LST001	0.8	0.667	0.727	0.8	0.667	0.727	0.85	0.667	0.747	0.833	0.667	0.741	0.8	0.667	0.727	0.8	1	0.889
LST002	0.2	1	0.333	0.2	1	0.333	0.2	1	0.333	0.2	1	0.333	0.15	1	0.261	0.12	1	0.214
LST003	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5	0.98	0.333	0.497
LST004	1	1	1	1	1	1	1	1	1	1	1	1	0.975	1	0.987	0.96	1	0.98
LST005	1	0.042	0.08	1	0.042	0.08	1	0.083	0.154	0.867	0.125	0.218	0.825	0.125	0.217	0.78	0.125	0.215
LST006	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LST007	1	0.667	0.8	1	0.667	0.8	1	0.667	0.8	0.967	0.667	0.789	0.975	0.667	0.792	0.96	0.667	0.787
LST008	1	0.5	0.667	1	0.75	0.857	1	1	1	1	1	1	1	1	1	1	1	1
LST009	0.2	0.2	0.2	0.2	0.2	0.2	0.15	0.2	0.171	0.133	0.2	0.16	0.125	0.2	0.154	0.1	0.2	0.133
LST010	0.8	1	0.889	0.9	1	0.947	0.85	1	0.919	0.567	1	0.723	0.425	1	0.596	0.34	1	0.507
Mean	0.7	0.541	0.52	0.71	0.566	0.545	0.705	0.595	0.562	0.657	0.599	0.547	0.627	0.599	0.523	0.604	0.633	0.522

Table 7. Detailed results for subm#3.

Topic	P@5	CR@5	F1@5	P@10	CR@10	F1@10	P@20	CR@20	F1@20	P@30	CR@30	F1@30	P@40	CR@40	F1@40	P@50	CR@50	F1@50
LST001	0.6	0.333	0.429	0.6	0.667	0.632	0.6	0.667	0.632	0.533	0.667	0.593	0.525	0.667	0.587	0.48	0.667	0.558
LST002	0.6	1	0.75	0.4	1	0.571	0.2	1	0.333	0.133	1	0.235	0.1	1	0.182	0.08	1	0.148
LST003	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5
LST004	1	1	1	0.9	1	0.947	0.95	1	0.974	0.967	1	0.983	0.95	1	0.974	0.9	1	0.947
LST005	0.8	0.083	0.151	0.9	0.083	0.153	0.85	0.125	0.218	0.8	0.125	0.216	0.75	0.208	0.326	0.74	0.25	0.374
LST006	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LST007	0.8	0.333	0.471	0.5	0.333	0.4	0.4	0.333	0.364	0.5	0.5	0.5	0.525	0.667	0.587	0.54	0.667	0.597
LST008	1	0.25	0.4	1	0.5	0.667	1	0.5	0.667	1	0.5	0.667	1	0.75	0.857	1	1	1
LST009	0.8	0.2	0.32	0.9	0.2	0.327	0.65	0.2	0.306	0.667	0.2	0.308	0.65	0.2	0.306	0.62	0.2	0.302
LST010	0.4	0.667	0.5	0.4	1	0.571	0.3	1	0.462	0.233	1	0.378	0.2	1	0.333	0.18	1	0.305
Mean	0.7	0.42	0.452	0.66	0.512	0.477	0.595	0.516	0.445	0.583	0.532	0.438	0.57	0.583	0.465	0.554	0.612	0.473

Table 8. Detailed results for subm#4.

Topic	P@5	CR@5	F1@5	P@10	CR@10	F1@10	P@20	CR@20	F1@20	P@30	CR@30	F1@30	P@40	CR@40	F1@40	P@50	CR@50	F1@50
LST001	1	0.667	0.8	0.9	1	0.947	0.75	1	0.857	0.7	1	0.824	0.625	1	0.769	0.6	1	0.75
LST002	0.6	1	0.75	0.3	1	0.462	0.2	1	0.333	0.167	1	0.286	0.175	1	0.298	0.16	1	0.276
LST003	0.8	0.333	0.471	0.9	0.333	0.486	0.75	0.667	0.706	0.767	0.667	0.713	0.825	0.667	0.737	0.86	0.667	0.751
LST004	1	1	1	1	1	1	1	1	1	0.933	1	0.966	0.95	1	0.974	0.92	1	0.958
LST005	1	0.042	0.08	0.8	0.083	0.151	0.85	0.125	0.218	0.767	0.167	0.274	0.775	0.25	0.378	0.76	0.292	0.422
LST006	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LST007	1	0.5	0.667	0.7	0.667	0.683	0.65	0.833	0.73	0.633	0.833	0.72	0.65	0.833	0.73	0.64	0.833	0.724
LST008	1	0.25	0.4	1	0.25	0.4	1	0.5	0.667	1	0.75	0.857	1	0.75	0.857	1	0.75	0.857
LST009	0.8	0.4	0.533	0.6	0.4	0.48	0.3	0.4	0.343	0.433	0.4	0.416	0.575	0.4	0.472	0.54	0.4	0.46
LST010	0.8	0.667	0.727	0.6	1	0.75	0.4	1	0.571	0.3	1	0.462	0.25	1	0.4	0.22	1	0.361
Mean	0.8	0.486	0.543	0.68	0.573	0.536	0.59	0.653	0.543	0.57	0.682	0.552	0.583	0.69	0.562	0.57	0.694	0.556

Table 9. Detailed results for subm#5.

Topic	P@5	CR@5	F1@5	P@10	CR@10	F1@10	P@20	CR@20	F1@20	P@30	CR@30	F1@30	P@40	CR@40	F1@40	P@50	CR@50	F1@50
LST001	0.6	0.333	0.429	0.6	0.667	0.632	0.6	0.667	0.632	0.533	0.667	0.593	0.525	0.667	0.587	0.48	0.667	0.558
LST002	0.6	1	0.75	0.4	1	0.571	0.2	1	0.333	0.133	1	0.235	0.1	1	0.182	0.08	1	0.148
LST003	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5
LST004	1	1	1	0.9	1	0.947	0.95	1	0.974	0.967	1	0.983	0.95	1	0.974	0.9	1	0.947
LST005	0.8	0.083	0.151	0.9	0.083	0.153	0.85	0.125	0.218	0.8	0.125	0.216	0.75	0.208	0.326	0.74	0.25	0.374
LST006	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LST007	0.8	0.333	0.471	0.5	0.333	0.4	0.5	0.5	0.5	0.5	0.5	0.5	0.525	0.667	0.587	0.54	0.667	0.597
LST008	1	0.25	0.4	1	0.5	0.667	1	0.5	0.667	1	0.5	0.667	1	0.75	0.857	1	1	1
LST009	0.8	0.2	0.32	0.9	0.2	0.327	0.65	0.2	0.306	0.633	0.2	0.304	0.65	0.2	0.306	0.62	0.2	0.302
LST010	0.4	0.667	0.5	0.4	1	0.571	0.3	1	0.462	0.233	1	0.378	0.2	1	0.333	0.18	1	0.305
Mean	0.7	0.42	0.452	0.66	0.512	0.477	0.605	0.532	0.459	0.58	0.532	0.438	0.57	0.583	0.465	0.554	0.612	0.473

Table 10. Detailed results for subm#6.

Topic	P@5	CR@5	F1@5	P@10	CR@10	F1@10	P@20	CR@20	F1@20	P@30	CR@30	F1@30	P@40	CR@40	F1@40	P@50	CR@50	F1@50
LST001	1	0.333	0.5	1	0.667	0.8	0.85	0.667	0.747	0.8	1	0.889	0.75	1	0.857	0.68	1	0.81
LST002	0.8	1	0.889	0.5	1	0.667	0.35	1	0.519	0.233	1	0.378	0.175	1	0.298	0.14	1	0.246
LST003	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5	1	0.667	0.8
LST004	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.98	1	0.99
LST005	1	0.042	0.08	1	0.083	0.154	1	0.083	0.154	1	0.083	0.154	1	0.083	0.154	1	0.083	0.154
LST006	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LST007	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5	1	0.5	0.667	1	0.667	0.8	1	0.667	0.8
LST008	1	0.5	0.667	1	0.5	0.667	1	0.5	0.667	1	0.75	0.857	1	1	1	1	1	1
LST009	0.6	0.2	0.3	0.4	0.2	0.267	0.3	0.2	0.24	0.267	0.2	0.229	0.225	0.2	0.212	0.26	0.2	0.226
LST010	0.4	0.333	0.364	0.2	0.333	0.25	0.45	1	0.621	0.367	1	0.537	0.3	1	0.462	0.3	1	0.462
Mean	0.78	0.408	0.48	0.71	0.445	0.48	0.695	0.512	0.495	0.667	0.587	0.521	0.645	0.628	0.528	0.636	0.662	0.549

Table 11. Detailed results for the no-submitted subm#0.

Topic	P@5	CR@5	F1@5	P@10	CR@10	F1@10	P@20	CR@20	F1@20	P@30	CR@30	F1@30	P@40	CR@40	F1@40	P@50	CR@50	F1@50
LST001	1	0.667	0.8	1	0.667	0.8	0.95	0.667	0.784	0.867	0.667	0.754	0.825	0.667	0.737	0.78	1	0.876
LST002	0.6	1	0.75	0.5	1	0.667	0.3	1	0.462	0.3	1	0.462	0.225	1	0.367	0.18	1	0.305
LST003	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5	1	0.333	0.5
LST004	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
LST005	1	0.042	0.08	1	0.042	0.08	1	0.042	0.08	1	0.083	0.154	1	0.083	0.154	1	0.25	0.4
LST006	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LST007	0.8	0.167	0.276	0.8	0.5	0.615	0.85	0.5	0.63	0.9	0.5	0.643	0.9	0.5	0.643	0.92	0.5	0.648
LST008	1	0.5	0.667	1	0.75	0.857	1	0.75	0.857	1	1	1	1	1	1	1	1	1
LST009	0.2	0.2	0.2	0.1	0.2	0.133	0.05	0.2	0.08	0.133	0.2	0.16	0.1	0.2	0.133	0.1	0.2	0.133
LST010	1	0.667	0.8	0.9	0.667	0.766	0.75	1	0.857	0.5	1	0.667	0.375	1	0.545	0.3	1	0.462
Mean	0.76	0.458	0.507	0.73	0.516	0.542	0.69	0.549	0.525	0.67	0.578	0.534	0.642	0.578	0.508	0.628	0.628	0.532

4 Conclusions

This paper describes our proposal for the Lifelog Moment Retrieval (LMRT) challenge of ImageCLEF Lifelog2018. The competition was quite challenging as it required to handle a huge number of images for retrieving moments for ten specific topics. 3 different strategies were proposed in order to respond to the 10 topics. All of them used deep learning and specifically AlexNet and GoogleNet.

Except of the amount of images, other facts that we had to deal with was the cultural differences e.g. what time is dinner for the specific country, as well as the differences in definitions e.g. for some people, vehicle is what is moving on the road while for others can be any transport mean. Last but no least, the explanation of the topics by the participants could also be a problem e.g. what Assembling Furniture includes?

The detailed results, given by the organizers and presented in section 3, require much more experimentation and further study. For example, the topic LST004 Interviewed by a TV presenter, almost always gave a result very close to 1, while the LST006 Assembling Furniture gave always 0. The last one means that no correct image was among the ones we chose as True. Thus, the organizers could consider the possibility of giving 1-2 correct images per topic, at the beginning of the competition.

In any case, it is a challenge that can create many new research fields and worth to be considered.

Acknowledgements

This work has been partially supported by the Ministerio de Economía, Industria y Competitividad (AEI/FEDER) of the Spanish Government under project TEC2016-75981 (IVME).

References

1. Allen, Anita L.: Dredging up the past: Lifelogging, memory, and surveillance. The University of Chicago Law Review, vol. 75, no 1, p. 47-74 (2008).
2. O'Hara, K., Tuffield, M. M., & Shadbolt, N.: Lifelogging: Privacy and empowerment with memories for life. Identity in the Information Society, 1(1), 155-172 (2008).
3. Magazine, G.: LifeLog: DARPA looking to record lives of interested parties. https://www.geek.com/news/lifelog-darpa-looking-to-record-lives-of-interested-parties-552879/ (2013), retrieved on 28-5-2018.
4. Gemmell, J., Bell, G., Lueder, R., Drucker, S., & Wong, C.: MyLifeBits: fulfilling the Memex vision. In Proceedings of the tenth ACM international conference on Multimedia (pp. 235-238). ACM.(2002).
5. Gemmell, J., Bell, G., & Lueder, R.: MyLifeBits: a personal database for everything. Communications of the ACM, 49(1), 88-95, (2006).
6. Sueda, K., Miyaki, T., & Rekimoto, J.: Social geoscape: visualizing an image of the city for mobile UI using user generated geo-tagged objects. In International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services (pp. 1-12). Springer, Berlin, Heidelberg, (2011).
7. Heo, S., Kang, K., & Bae, C.: Lifelog collection using a smartphone for medical history form. In IT Convergence and Services (pp. 575-581). Springer, Dordrecht, (2011).
8. Jacquemard, T., Novitzky, P., O'Brolcháin, F., Smeaton, A. F., & Gordijn, B.: Challenges and opportunities of lifelog technologies: A literature review and critical analysis. Science and engineering ethics, 20(2), 379-409, (2014).
9. Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, Rami Albatal: Overview of NTCIR-12 Lifelog Task. Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, Tokyo, Japan, (2016).
10. Duc-Tien Dang-Nguyen, Luca Piras, Michael Riegler, Giulia Boato, Liting Zhou, Cathal Gurrin: Overview of ImageCLEFlifelog 2017: Lifelog Retrieval and Summarization. CLEF2017 Working Notes, Dublin, Ireland, vol 1866, (2017).
11. Duc-Tien Dang-Nguyen and Luca Piras and Michael Riegler and Liting Zhou and Mathias Lux and Cathal Gurrin: Overview of ImageCLEFlifelog 2018: Daily Living Understanding and Lifelog Moment Retrieval. CLEF2018 Working Notes. CEUR Workshop Proceedings. (2018).
12. Bogdan Ionescu and Henning Muller and Mauricio Villegas and Alba Garcia Seco de Herrera and Carsten Eickhoff and Vincent Andrearczyk and Yashin Dicente Cid and Vitali Liauchuk and Vassili Kovalev and Sadid A. Hasan and Yuan Ling and Oladimeji Farri and Joey Liu and Matthew Lungren and Duc-Tien Dang-Nguyen and Luca Piras and Michael Riegler and Liting Zhou and Mathias Lux and Cathal Gurrin: Overview of ImageCLEF 2018}: Challenges, Datasets and Evaluation. Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Ninth International Conference of the CLEF Association (CLEF 2018). LNCS Lecture Notes in Computer Science, Springer, (2018).
13. Krizhevsky, A., Sutskever, I., & Hinton, G. E.: Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems. pp. 1097-1105, (2012).
14. Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich: Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1-9. (2015).