Automatic Approach For Lifelog Moments Retrieval using Transfer Learning and Rule-based Matching

Fatma Ben Abdallah[1,2], Ghada Feki[1,3], Ons Saadallah[3], Amine Jarraya[3], and Chokri Ben Amar[1]

[1] REGIM-Lab.: REsearch Groups in Intelligent Machines, University of Sfax, National Engineering School of Sfax (ENIS), BP 1173, Sfax, 3038, Tunisia
{ben.abdallah.fatma,ghada.feki,chokri.benamar}@ieee.org,
{saadallahons,jarrayaemine}@gmail.com
http://www.regim.org/

[2] Higher Institute of Technological Studies, ISET Kairouan, 3199 Raccada, Tunisia
http://www.isetkr.rnu.tn/

[3] International Private School of Engineers, EPI Sousse, 4021 Sousse, Tunisia
https://www.episup.com//

Abstract. Visual lifelogging and egocentric vision have become more and more attractive and have been the subject of many challenges in recent years. This interest mainly focuses on exploring search and retrieval from lifelog, identifying lifelog activities, summarizing specific lifelog moments, gaining insights into daily lifelog activities, and developing an annotation approach of the multimodal lifelog data. From the results obtained during these challenges in the previous edition, retrieving lifelog moments remains an unsolved problem. Therefore in this paper, we describe the participation of the REGIMLab Team in the subtask ImageCLEF Lifelog 2020 Moment Retrieval (LMRT). This year, we test two strategies which both are based on fine-tuning. The first strategy is based on fine-tuning with MobilenetV2 using Kibana Query Language for retrieval. The second one is based on fine-tuning with Densenet201 using cosine similarity and word embedding for retrieval. A total of seven runs was submitted. The best result was reached by the third run using the first strategy with F1@= 0.189.

Keywords: Lifelog Image Retrieval · Fine-Tuning · Natural Language Processing · Word-Embedding

1 Introduction

Using a smartphone, a wearable camera and a smart watch with sensors and trackers has become the daily life of thousands of people around the world. The

democratization of this kind of gadget made possible to build several lifelog datasets which contain images, music listening activities, biometrics data (steps, heart rate, calorie burn, blood glucose), semantic locations visited, physical activities enhanced by metadata from visual concept detectors [30, 35]. Some of these datasets were used in several challenges like NTCIR[4], ImageCLEF lifelog [5] or Lifelog Search Challenge (LSC) [6]. By looking more closely at the different editions of these challenges, we notice that the subtask of lifelog retrieval moments in these challenges is always present from year to year. Indeed, last year's winner [25] obtained the score F1@10=0.61 and could be improved. This year, in the fourth edition of the ImageCLEF lifelog moment retrieval challenge (LMRT) which is a part of the Conference and Labs of the Evaluation Forum (CLEF 2020) [23], we proposed five main improvements comparing to our previous approach [4, 6]. Firstly, we performed an object detection using YOLOV3 [29] trained on Open dataset images V4 [24]. Secondly, we use a topic text extraction pipeline to delete irrelevant concepts from topics. After that, we use rule-based matching (R-BM) to extract, if it exists in the topic, the location (semantic name, longitude, latitude), the time (hour, date, part of the day) and the context (category(indoor/outdoor), object(color, bbox), activity). According to the extracted concept, a Kibana query is automatically generated thanks to the R-BM and returned relevant images. Also, we use web scrapping to add new image classes to the fine-tuning. The images are sorted according to their score obtained from the classification using the fine-tuning with MobilenetV2 [34] or by using the average score given by the Place CNN trained on Place 365 dataset and by the Faster R-CNN [30] trained on the COCO dataset.
The remainder of this paper covers the topics below:

- In section 2, we present existing lifelog moment retrieval approach.
- In section 3, we detail our automatic approach.
- Section 4 presents the experimental results obtained during the LMRT 2020 challenge.
- Section 5 provides some concluding remarks and future works suggestions.

2 Related Work

The constitution's aim of the lifelog datasets is to characterize social pattern and behavior in egocentric photo-streams [1, 31], to localize and segment object [8, 15], to predict different food-related tasks [2], to localize and recognize food and ingredients [9, 10], to predict correct day and part of the day [13], to retrieve specific predefined moments in a lifelogger's life [11–13, 17–22, 27], to explore approach to event segmentation [12, 18], to visualize knowledge and insight about the lifelog data [17, 18, 21], to predict performances for an athlete who prepared for a sport event [27] and to detect and retrieve micro-activities [7].

[4] http://ntcir-lifelog.computing.dcu.ie/
[5] https://www.imageclef.org/2020/lifelog
[6] http://lsc.dcu.ie/
[7] http://ntcir-mart.computing.dcu.ie/

By analyzing the approach of the winning teams in the lifelog challenges, we have noticed several areas for improvement which we describe in the following. The user's interactivity and his involvement in the research process is a major asset which enabled a significant improvement of the results [14, 25, 28, 33, 36]. In addition, approaches [14, 25, 28, 33, 36] which enrich visual concepts using machine learning methods and vision-based algorithm performed better results than those who used only the metadata given by the challenge's organizers [16]. Furthermore, several teams employed natural language processing (NLP) [14, 28, 33] to convert the topic into interpretable concepts or to generate predefined synonyms. Also, investigation in object color detection is essential for some topics [25]. Finally, approaches using segmentation or clustering did not improve the results [32, 33].

3 Automatic Approach For Lifelog Moments Retrieval

In the following, we describe the improvement of our work since our first participation in ImageCLEF [3–7].

3.1 Strategy 1 : Fine-tuning with MobilenetvV2 using Kibana Query Language for retrieval

We proposed five main improvements comparing to our previous approach [4, 6]. Fig. 1 detailed our automatic approach for lifelog moments retrieval. The proposed approach is divided into two stages :

- an offline stage to perform text classification and mining, object detection and transfer learning
- and an online stage to achieve retrieval and web scrapping to add new classes to transfer learning.

In the offline process, we firstly performed an object detection using YOLOV3 [29] trained on Open dataset images V4 [24]. Secondly, we use a topic text extraction pipeline to delete irrelevant concepts from topics. After that, we use rule-based matching (R-BM) to extract, if it exists in the topic, the location (semantic name, longitude, latitude), the time (hour, date, part of the day) and the context (category(indoor/outdoor), object(color, bbox), activity). For the transfer learning process, we used the ground truth to automatically dispatch 14468 images into 104 classes.
In the online stage, if a new concept is extracted by the R-BM and does not belong to one of the labels lists (Places365, Sunattribute, Cocodataset or Opendataset) we operate a web scrapping to add new image classes to the fine-tuning. Then, we retrain the whole network with all classes. From a given user topic and thanks to the R-BM, a Kibana query is automatically generated and returns the relevant images. The images are sorted according to their score obtained from the classification using the fine-tuning with MobilenetV2 [34] or by using the average score given by the Places CNN trained on Places365 dataset and by the

Faster R-CNN [30] trained on the COCO dataset or by YOLOV3 trained on Open dataset images V4. The choice of the used CNN depends on the rules in the R-BM.
Similarly to our previous participation in the ImageCLEF Lifelog Moment Re-

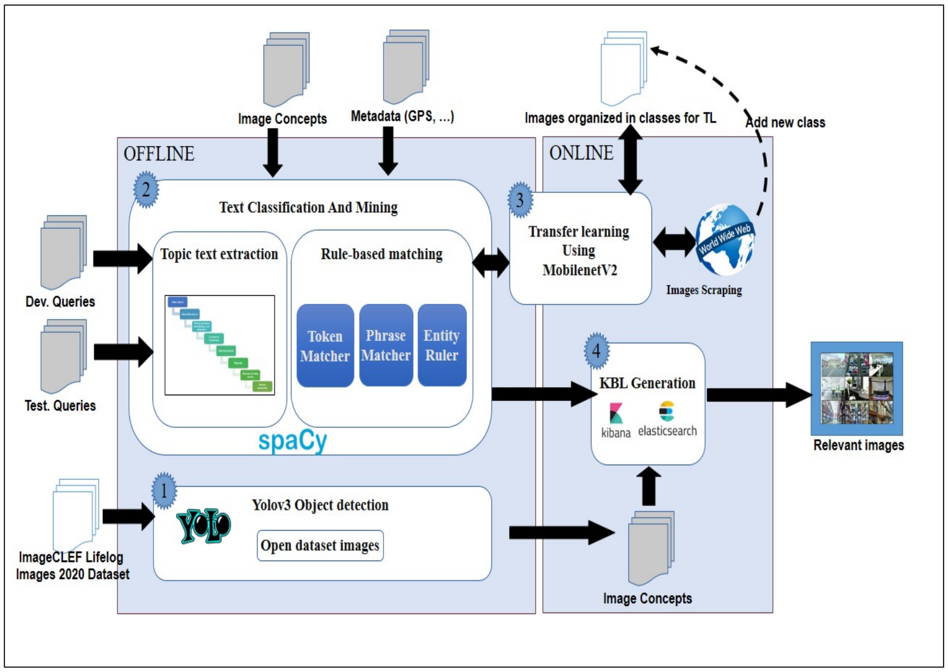

Fig. 1. Automatic approach for lifelog moments retrieval

trieval Task 2019 (LMRT), we used the ground truth of the development dataset (LMRT2019 and LMRT2020) to automatically dispatch images into categories for the fine-tuning. This year, we used Mobilenet v2 instead of Googlenet. We try several deep architectures. Better and faster accuracy during the training phase was reached by Mobilenet v2.
For fine-tuning, we replaced the last three layers of the network: a fully connected layer, a softmax layer, and a classification output layer. We froze the convolutional base before compiling and training the model to prevent the weights in a given layer from being updated during training. With 80% of the images for training and 20% for validation, we used data augmentation to prevent the network from overfitting. We use a learning rate equal to 10^{-4} for 10 epochs and batch size equal to 32.
Furthermore, instead of using Apache Cassandra and Cassandra Query Language (CQL), we used Elastic Search and Kibana Query Language (KBL) to perform retrieval on image concepts and metadata. Our approach can automatically extract from an initial query relevant concepts using text classification based on

rule-based matching. The rule-based matching is based on token matcher, phrase matcher and entity ruler. After that, the retrieval phase consists of searching the extracted query concepts in the file containing the image concepts using KBL automatically generated. The images are sorted according to their score obtained from the classification using the fine-tuning with MobilenetV2 [34] or by using the average score given by the Place CNN trained on Place 365 dataset and by the Faster R-CNN [30] trained on the COCO dataset.

3.2 Strategy 2 : Fine-tuning with Densenet201 using cosine similarity and word embedding for retrieval

To perform the fine-tuning, we start with freezing convolutional layers of Densenet201 in order to prevent the weight's update and only train the fully connected layers. The matching between queries and images is based on the Cosine Similarity method between queries and Images Description. We treat the textual queries with three-words embedding models that we build from scratch which are Word2vec, fastText, and Glove. The training results show that the Word2Vec model is better than fastText and Glove but in the evaluation results, the Glove showed a better performance than Word2Vec and fastText. The Glove model built from scratch gives a weak similarity score. As a result, we decide to support the matching process with the pertained Word2Vec model that has been trained on huge data like GoogleNews to give a good word representation for each query. The score of similarity has improved.

4 Obtained results

Our code is written in Python and uses several third party libraries like Spacy, Elasticsearch, Pandas, Gensim, Glove-python, Beautiful Soup, Nltk. We used an Intel Core i7 with 16GB RAM in combination with NVIDIA GeForce GTX 1050 TI. We use also Google Colab[8] for Densenet201 and MobilenetV2 fine-tuning and detecting object using YOLOV3. Fig. 2 detailed official results of all teams that participated to the ImageCLEF LMRT 2020. The best team HCMUS obtained F1@10=0.811 with an interactive approach [26]. Our automatic approach ranked sixth in the challenge with F1@10=0.189.
The table 1 summarizes the submitted runs by our REGIMLAB team.

For the first strategy, we submitted the second, third and seventh run.

- The second run operated retrieval on the concepts given by the organizers.
- In the third run, we operated object detection for each image from the ImageCLEF 2020 lifelog dataset using YOLOV3 trained on Opendataset images which can detect 600 classes. The organizers used Faster R-CNN trained on COCO dataset (version 2014-2017) which can only detect 80 classes. The use of the YOLOV3 object detector improves the results.

[8] https://colab.research.google.com/

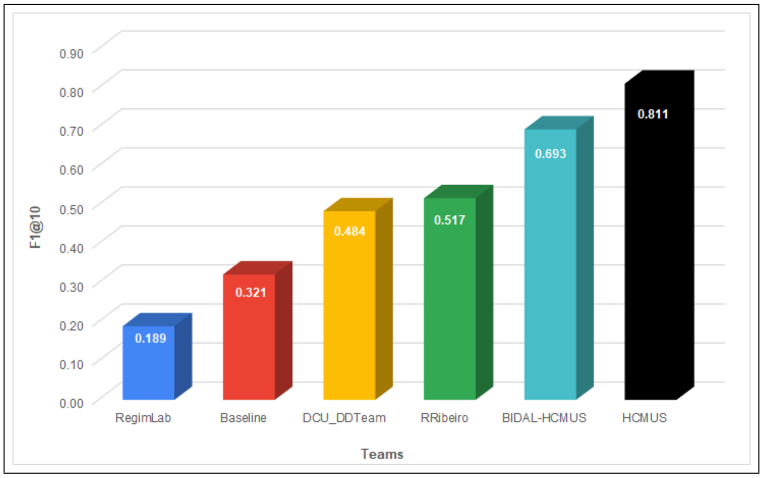

Fig. 2. ImageCLEF LMRT 2020 official results

- In the last run, we used transfer learning only if the concepts were not belonging to the labels' list (label sun attribute, Places attribute, COCO dataset attribute, YOLOV3 attribute). We used web scrapping to add new classes for the fine-tuning. This technique did not improve the results due to the quality of the downloaded images. This is confirmed by the values obtained in the table 2,3 and 4 for cut off points greater than 30. Some images containing text overlays, and some were irrelevant and further investigations needed to automatically detect and remove these uninformative images.

Best results for this first strategy were reached by the third run with F1@10= 0.189.

For the second strategy, we submitted the first, fourth, fifth and sixth run.

- For the first run, after extracting the keyword from the title, the description and the narrative of the topic using nltk preprocessing techniques, we trained the Word2Vec model using skip-gram algorithm from scratch using 150 epochs, learning rate=0.025, word embedding vector dimension=100 and min-count=1. For the transfer learning with Densenet201, we used a learning rate equal to 10^{-4} during 5 epochs and a batch size = 128 to proceed fast training. We obtained a network that is 80% accurate to predict the true labels.
- In the fourth run, we used the Glove model instead of Word2Vec trained with a learning rate = 0.05 during 100 epochs. From our empirical study, Glove was able to create semantic relationship better than fastText and Word2Vec. We used the Mix-Max Scaling technique to normalize the scores to a range between 0 to 1.

– In the fifth run, we used the same model as the first run but trained for 7 epochs. Further training helped improve the accuracy model from 80% to 90.81%. We used the Word2Vec model on Google News with the Glove model.
– In the sixth run, we used the same parameters as the first run and replace Word2Vec by Glove.

Best results for this second strategy were reached by this fifth run with F1@10= 0.162.

Table 1. Submitted runs of the REGIMLab Team

Cut-off	F1@10	Strategy	Fine-Tuning	Word Embedding	Adding metadata
RUN_01	0.054	1^{st}	Densenet201	Word2Vec	-
RUN_02	0.174	2^{nd}	-	-	-
RUN_03	**0.189**	2^{nd}	-	-	from YOLOV3
RUN_04	0	1^{st}	Densenet201	Glove	-
RUN_05	0.162	1^{st}	Densenet201	Glove + Word2Vec	-
RUN_06	0.038	1^{st}	Densenet201	Glove	-
RUN_07	**0.189**	2^{nd}	MobilenetV2	-	from YOLOV3

The results of the runs submitted to the LMRT 2020 subtask are detailed in tables 2,3 and 4.

5 Conclusion and Perspectives

This paper presents our automatic approach for lifelog moment retrieval at the ImageCLEF LMRT 2020. This third version employs a rule-based matching with token matcher and entity ruler to extract relevant concepts from the topic and to automatically generate a Kibana query. The results demonstrate the feasibility of the process. The use of YOLOv3 object detector has improved the performance which confirms that the enhancement of the results is closely linked to the feature extraction and object detector. Like precedent LMRT edition, the best results were reached by an interactive approach, which confirms the necessity to use human in the loop method to obtain good achievement. So as future work, to improve efficiency, we will investigate in formal knowledge representation of the lifelog domain like ontologies or knowledge base. We will also include the user in the relevance feedback and try to use our own object detector trained on a lifelog dataset.

Table 2. Precision at X (P@X)

Cut-off	P@5	P@10	P@20	P@30	P@40	P@50
RUN_01	**0.080**	0.040	0.020	0.013	0.010	0.009
RUN_02	**0.200**	0.160	0.140	0.130	0.122	0.118
RUN_03	**0.220**	0.170	0.145	0.134	0.129	0.124
RUN_04	0	0	**0.010**	0.007	0.010	0.009
RUN_05	**0.200**	0.190	0.145	0.130	0.140	0.139
RUN_06	**0.060**	0.030	0.025	0.017	0.018	0.015
RUN_07	**0.220**	0.170	0.140	0.127	0.114	0.111

Table 3. Cluster Recall at X (CR@X)

Cut-off	CR@5	CR@10	CR@20	CR@30	CR@40	CR@50
RUN_01	**0.083**	0.083	0.083	0.083	0.083	0.083
RUN_02	**0.183**	0.217	0.229	0.229	0.240	0.240
RUN_03	**0.208**	0.242	0.254	0.254	0.365	0.365
RUN_04	0	0	**0.033**	0.033	0.083	0.083
RUN_05	**0.146**	0.158	0.158	0.158	0.158	0.158
RUN_06	**0.050**	0.050	0.083	0.083	0.133	0.133
RUN_07	**0.208**	0.242	0.242	0.242	0.354	0.354

Table 4. F1-measure at X (F1@X)

Cut-off	F1@5	F1@10	F1@20	F1@30	F1@40	F1@50
RUN_01	**0.081**	0.054	0.032	0.023	0.018	0.016
RUN_02	**0.181**	0.174	0.158	0.152	0.151	0.151
RUN_03	**0.203**	0.189	0.167	0.159	0.162	0.162
RUN_04	0	0	0.015	0.011	**0.018**	0.016
RUN_05	0.157	**0.162**	0.145	0.135	0.138	0.137
RUN_06	**0.055**	0.038	0.038	0.028	0.031	0.027
RUN_07	**0.203**	0.189	0.160	0.150	0.147	0.147

6 Acknowledgments

The research leading to these results has received funding from the Ministry of Higher Education and Scientific Research of Tunisia under the grant agreement number LR11ES48.

References

1. Aghaei, M.; Dimiccoli, M.; Ferrer, C. C. & Radeva: P. Towards social pattern characterization in egocentric photo-streams Computer Vision and Image Understanding, Elsevier (2018)

2. Aguilar, E.; Bolaños, M. & Radeva, P.: Regularized uncertainty-based multi-task learning model for food analysis, Journal of Visual Communication and Image Representation, volume 60, pp. 360-370 (2019)
3. Abdallah, F. B.; Feki, G.; Ammar, A. B.; & Amar, C. B.: A new model driven architecture for deep learning-based multimodal lifelog retrieval. In: poster Proceedings: International Conference WSCG, CSRN 2803, pp 8-17, Czech Republic(2018)
4. Abdallah, F.B.; Feki, G.; Ezzarka, M.; Ammar, A.B. & Amar, C.B.: Regim Lab Team at ImageCLEF Lifelog Moment Retrieval Task 2018. Working Notes of CLEF 2018, France(2018)
5. Abdallah, F. B.; Feki, G.; Ammar, A. B. & Amar, C. B.: Multilevel Deep Learning-Based Processing for Lifelog Image Retrieval Enhancement. In: Proceedings of the IEEE International Conference SMC, pp 1348-1354, Japan(2018)
6. Abdallah, F.B.; Feki, G.; Amar, C.B.: Big Data For Lifelog Moments Retrieval Improvement. In: CLEF2019 Working Notes. CEUR Workshop Proceedings, Switzerland (2019)
7. Abdallah, F. B.; Feki, G.; Ammar, A. B. & Amar, C. B.: Transfer Learning for Improving Lifelog Image Retrieval, Computer Analysis of Images and Patterns (CAIP), In: 18th International Conference Proceedings Springer, volume 11678, pp. 284-295, Italy (2019)
8. Bolaños, M. & Radeva P.: Ego-object discovery. preprint arXiv:1504.01639 (2015).
9. Bolaños, M. & Radeva, P.: Simultaneous food localization and recognition, In : the 23rd International Conference on Pattern Recognition (ICPR), pp. 3140-3145 (2016)
10. Bolaños, M.; Ferrà, A. & Radeva, P.: Food Ingredients Recognition through Multi-label Learning, In Proceedings of the 3rd International Workshop on Multimedia Assisted Dietary Management (ICIAP Workshops) (2017)
11. Dang-Nguyen, D.-T.; Piras, L.; Riegler, M.; Boato, G.; Zhou, L. & Gurrin, C.: Overview of ImageCLEFlifelog 2017: Lifelog Retrieval and Summarization, CLEF2017 Working Notes, CEUR-WS.org (2017)
12. 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, CLEF2018 Working Notes, CEUR-WS.org (2018)
13. Dang-Nguyen, D.-T.; Piras, L.; Riegler, M.; Tran, M.-T.; Zhou, L.; Lux, M.; Le, T.-K.; Ninh, V.-T. & Gurrin, C.: Overview of ImageCLEFlifelog 2019: Solve my life puzzle and Lifelog Moment Retrieval, CLEF2019 Working Notes, CEUR-WS.org (2019)
14. 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)
15. Dimiccoli, M.; Bolaños, M.; Talavera, E.; Aghaei, M.; Nikolov, S. G. & Radeva, P.: SR-clustering: Semantic regularized clustering for egocentric photo streams segmentation, Computer Vision and Image Understanding, volume 155, pp. 55-69 (2017)
16. Dogariu, M.; Ionescu, B.: Multimedia Lab @ ImageCLEF 2019 Lifelog Moment Retrieval Task. In: CLEF2019Working Notes. CEUR Workshop Proceedings, CEURWS. org <http://ceur-ws.org> (2019)
17. Gurrin, C.; Joho, H.; Hopfgartner, F.; Zhou, L.; & Albatal, R. (2016).: Overview of NTCIR-12 Lifelog Task. Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, Japan (2016)
18. Gurrin, C.; Joho, H.; Hopfgartner, F.; Zhou, L.; Gupta, R.; Albatal, R.; Nguyen, D.T.D.: Overview of NTCIR-13 Lifelog-2 Task. Proceedings of the Thirteenth NTCIR conference (NTCIR-13), Tokyo, Japan (2017)

19. Gurrin, C.; Schöffmann, K.; Joho, H.; Dang-Nguyen, D. T.; Riegler, M. & Piras, L.: LSC '18: Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge, Association for Computing Machinery (2018)
20. Gurrin, C.; Schöffmann, K.; Joho, H.; Dang-Nguyen, D. T.; Riegler, M. & Piras, L.: LSC'19: Proceedings of the ACM Workshop on Lifelog Search Challenge, Association for Computing Machinery (2019)
21. 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, Proceedings of the 14th NTCIR conference (NTCIR-14), pp. 14-26 (2019)
22. Gurrin, C.; Le, T.-K.; Ninh, V.-T.; Dang-Nguyen, D.-T.; Jónsson, B. Þór.; Lokoš, J.; Hürst, W.; Tran, M.-T. & Schöffmann, K.: Introduction to the Third Annual Lifelog Search Challenge (LSC'20), Proceedings of the 2020 International Conference on Multimedia Retrieval, Association for Computing Machinery,pp. 584–585 (2020)
23. 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.; l 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.; Ştefan, L. D. & Constantin, M. G.: Overview of the ImageCLEF 2020: Multimedia Retrieval in Lifelogging, Medical, Nature, and Internet Applications, Experimental IR Meets Multilinguality, Multimodality, and Interaction, LNCS Lecture Notes in Computer Science, Springer (2020)
24. Kuznetsova, A.; Rom, H.; Alldrin, N.; Uijlings, J. R. R.; Krasin, I.; Pont-Tuset, J.; Kamali, S.; Popov, S.; Malloci, M.; Duerig, T. & Ferrari, V.: The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale, CoRR, abs/1811.00982 (2018)
25. 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, CEURWS.org <http://ceur-ws.org> (2019)
26. Mai-Nguyen, A.-V.; Phan, T.-D.; Vo, A.-K.; Tran, V.-L.; Dao, M.-S. & Zettsu, K.: BIDAL-HCMUS@LSC2020: An Interactive Multimodal Lifelog Retrieval with Query-to-Sample Attention-Based Search Engine. Proceedings of the Third Annual Workshop on Lifelog Search Challenge, ACM, pp. 43–49 (2020)
27. Ninh, V.-T.; Le, T.-K.; Zhou, L.; Piras, L.; Riegler, M.; l Halvorsen, P.; Tran, M.-T.; Lux, M.; Gurrin, C. & Dang-Nguyen, D.-T.: Overview of ImageCLEF Lifelog 2020:Lifelog Moment Retrieval and Sport Performance Lifelog, CLEF2020 Working Notes, CEUR-WS.org (2020)
28. 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)
29. Redmon, J. & Farhadi, A.: YOLOv3: An Incremental Improvement, CoRR, abs/1804.02767 (2018)
30. 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
31. Talavera, E.; Wuerich, C.; Petkov, N. & Radeva, P.: Topic Modelling for Routine Discovery from Egocentric Photo-streams, Pattern Recognition, pp. 107330 (2020)

32. 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)
33. 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)
34. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A. & Chen, L.-C.: MobileNetV2: Inverted Residuals and Linear Bottlenecks, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
35. 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)
36. Zhou, P.; Bai, C.; Xia, J.: ZJUTCVR Team at ImageCLEFlifelog2019 Lifelog Moment Retrieval Task. In: CLEF2019Working Notes. CEURWorkshop Proceedings, CEUR-WS.org <http://ceur-ws.org> (2019)