Lifelog Moment Retrieval with Visual Concept Fusion and Text-based Query Expansion

Minh-Triet Tran, Tung Dinh-Duy, Thanh-Dat Truong,
Viet-Khoa Vo-Ho, Quoc-An Luong, and Vinh-Tiep Nguyen

University of Science, VNU-HCM
University of Information Technology, VNU-HCM
tmtriet@fit.hcmus.edu.vn
{ttdat,ddtung,vhvkhoa,lqan}@selab.hcmus.edu.vn
tiepnv@uit.edu.vn

Abstract. Lifelog data provide potential insight analysis and understanding about people in their daily activities. However, it is still a challenging problem to index lifelog data efficiently and to provide a user-friendly interface that supports users to retrieve moments of interest. This motivates our proposed system to retrieve lifelog moment based on visual concept fusion and text-based query expansion. We first extract visual concepts, including entities, actions, and places from images. Besides NeuralTalk, we also proposed a novel method using concept-encoded feature augmentation to generate text descriptions to exploit further semantics of images.

Our proposed lifelog retrieval system allows a user to search for lifelog moment with four different types of queries on place, time, entity, and extra biometric data. Furthermore, the key feature of our proposed system is to automatically suggest concepts related to input query concepts to efficiently assist a user to expand a query. Experimental results on Lifelog moment retrieval dataset of ImageCLEF 2018 demonstrate the potential usage of our method and system to retrieve lifelog moments.

Keywords: Lifelog Retrieval · Visual Concept · Visual Captioning · Text-based Query Expansion.

1 Introduction

With the increasing number of wearable cameras and smart devices, people can easily capture their daily emotional and memorial events Besides visual data, such as images or video clips, lifelog data [10] can be of different categories, such as audio clips, GPS or biometric data (heart rate, galvanic skin response, calorie burn, etc).

As a new promising trend for research and applications [6], lifelogging analysis [10] to retrieve moments of interest from lifelog data helps people to revive memories [18], verify events, find entities, or analyse people's social traits[7].

There are two main problems for lifelogging analysis and moment retrieval: (i) to analyze and efficiently index lifelog data, and (ii) to enhance usability and ease-of-use for users to input queries and retrieve moments of interest. To solve the first problem, we propose to extract concepts (entities and places) [24] and generate text descriptions from images. Currently, we gather lifelog images into visual shots but we still process each image independently. For the second problem, we develop a lifelog retrieval system that supports four types of queries (place, time, entity, and metadata)[24] and automatically suggests hints to users the related concepts from an input query.

Comparing to our initial system [24], our current system for lifelog data processing and retrieval has two main improvements to further exploit the semantic from text descriptions for images. First, we propose a new method for text description generation based on the spatial attention model by Kelvin Xu et. al[25] and the semantic attention model by Quanzeng You et. al[26]. Second, we develop the related concept recommendation module into our retrieval system for query expansion. From an initial concept, our system can automatically suggests potential related concepts for users to expand the query with the expectation to cover a wider range of retrieved results. Using our proposed method and system, we achieve the score of 0.479 for Lifelog moment retrieval (LMRT) with Lifelog data[6] in ImageCLEF 2018[15] , ranked second in the challenge of LMRT.

In Section 2, we briefly review recent achievements in Lifelog, place retrieval, and visual instance search. Then we propose our method to offline process lifelogging data in Section 3 and go deeply on image captioning in Section 4. Our system to assist users to find a moment of interest based on an arbitrary query is presented in Section 5. The conclusion and open questions for future work are discussed in Section 6.

2 Related Work

The comparison on evaluating the effectiveness of information access and retrieval systems operating over personal lifelog data has been considered for a long time. In 2012, the tasks in NTCIR-12 which are the first ones focus on known-item search and activity understanding applied over lifelog data [8]. The lifelog data is collected form 3 different volunteers wearing a camera to record visual daily life data for a month. In addition, they also provide a visual concept information using Caffe CNN-based visual concept detector. Because of the large lifelog data, many different analytic approaches and applications have been discussed in the workshop. The area of interest is widen to other aspect than the origin information retrieval purpose [1, 9]. While some team focus on improve the friendly UX/UI for an end-user, the others considered the crucial in privacy and data security. In addition, the way for preservation and maintenance of lifelogs is also discussed in the workshop.

Moreover, they keep on enrich the lifelog data by adding more information in semantic locations like a cafe, restaurant and physical activity such as walking, cycling and running in ImageCLEFlifelog 2017 [5]. The tasks on this lifelog data are (1) retrieval task which is the evaluation on the correctness of returned image followed several specific queries and (2) summarization task that summarize all the images according to specific requirment.

Location plays an important role in Lifelog Moment Retrieval Task (LMRT) that increases the accuracy of the whole system. In the task, the information about a place is very important so it is better to retrieve as much as possible diverse images. Duc-Tien et al. have introduced the method to deal with this problem using the dataset collected on Flickr [4, 3]. The precision of the method has improved up to 75%.

Sivic has introduced first Bag-of-visual-word (BoW) model which is on of the most state-of-the-art approaches for video retrieval [23]. The model follows the *key assumption*, which is the two similar images sharing the significant number of local path matched against each other. In order to boost the performance of Instance Search system, many techniques, such as RootSIFT feature [2], large vocabulary [20], soft assignment [21], multiple detectors and feature combination at late fusion, query-adaptive asymmetrical dissimilarities [28], are applied. In this paper, we focus on baseline model of BOVW for easy of implementation and better performance.

Word Representations in Vector Space is a problem that learns high-quality distributed vector representations, capturing a large number of precise syntactic and semantic word presentation [17, 16]. The state of the art is Word2Vec which is based on Skip-gram model introduced by Mikolov et. al. They improve the Skip-gram in time consuming by simplifying the 'hierarchical softmax' with 'negative sampling' and learning regular word representation.

Generating an natural language description for one specific image is challenging problem in Computer Vision. Although a lot of work concentrates on labeling images with fixed set on visual categories, their drawback are relying on hard-coded visual concepts and sentence templates that reducing complex visual scene. To overcome this problem, Karpathy and Fei-Fei introduced NeuralTalk, state-of-the-art model, to generate the caption of an image.

3 Proposed Method

3.1 Overview

We inherit our framework in the Lifelog Search Challenge [24] to solve the LMRT challenge. However, the main difference between the two methods is that we further take advantage of the semantic that can be provided by text descriptions generated from an image. To exploit different aspects as well as styles of semantics in text description, we use NeuralTalk and our Concept Augmentation with Attention method to generate different descriptions corresponding to one image with the expectation to get more insight information about that image.

Figure 1 presents the overview of our proposed method to offline process lifelogging data. Our method has four main components:

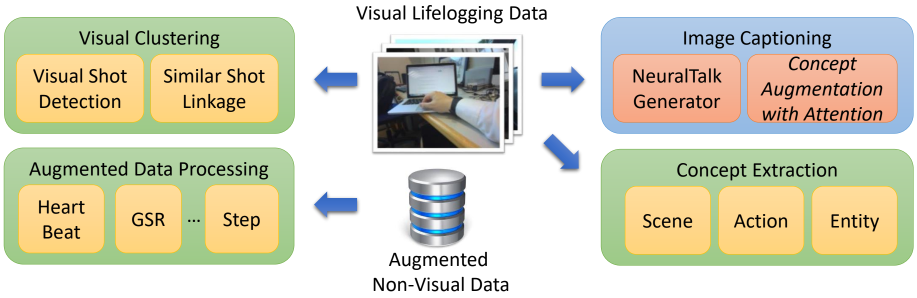

Fig. 1. Our Proposed Method in Lifelog Moment Retrieval Task

- Visual Clustering: we first group each sequence of similar contiguous images into a visual shot, then we link visually similar shots to a scene using visual retrieval with our Bag-of-Word framework.
- Concept Extraction: we extract three types of concepts from each image, including scene category and scene attributes, entities, and actions.
- Image captioning: besides NeuralTalk, we propose our new method using Concept Augmentation with Attention to generate text descriptions for an image.
- Augmented Data Processing: we process extra data for query refinement, including biometrics, blood pressure, blood sugar level, text data of computer activities, etc.

Comparing to our proposed method to offline process lifelogging data [24], our enhanced method further exploit the text descriptions of images. We use NeuralTalk to generate the baseline text descriptions, and we propose our new method with Concept Augmentation and Attention for better description generation. In this section, we briefly review the three components that we reuse from our framework [24], and we reserve Section 4 to present our new method for image captioning.

Visual Clustering: Instead of processing all images, we first group similar contiguous images into visual shots[24]. By this way, we can get better context of a scene in a visual shot. As images are captured with 45 second interval, the location and pose of an entity may change in two consecutive images. Therefore, we use FlowNet [14] to estimate the optical flow vectors at all corresponding pixels in two continuous frames, then determine if the two frames are similar.

To link shots corresponding to the same scene, we propose a solution to use our Bag-of-Visual-Word (BoVW) framework for visual retrieval[19]. For an image x in a visual shot S_i, we retrieve similar images in other shots. The distance between the two visual shots S_i and S_j is determined by the mean distance between their images.

$$distance(S_i, S_j) = mean\{distance(x, y), for\ x \in S_i\ and\ y \in S_j\} \tag{1}$$

To represent the similarity relationship between visual shots, we create an undirected graph with nodes as visual shots and edges. There is an undirected edge to link two nodes S_i and S_j if their distance is less than a threshold $max_{distance}$. Each connected component represents a cluster of similar visual shots, and is expected to represent the same or similar scene in real life.

Concept Extraction: We focus on three types of concepts that can be extracted from an image. We use MIT Place API [27] to determine the scene category and scene attributes of an image. To extract entities, we use Faster RCNN [22]. For possible action detection, we extract the main verb in each description generated from an image. Besides NeuralTalk, we propose our new method for image captioning, presented in Section4.

Augmented Data Processing: We feed the information about bodymetrics of the volunteers provided by the challenge to retrieve a better result. In this LMRT, for each bodymetric information, we cluster them into a range of value. Considering heart-rate for an example, we divide the range from 1 to 150 into fifteen 10-unit wide periods. Then the images is clustered into group based on those period.

4 Image Captioning with Concept Augmentation and Attention

Our proposed model is based on the spatial attention model in work of Kelvin Xu et. al[25] and the semantic attention model in work of Quanzeng You et. al[26]. Our model consists of two main modules: feature extraction and caption generation.

In the feature extraction module, with an input image I, we use a deep convolutional neural network to produce a $N \times N \times D$ feature map. We then use an object detection model to extract the labels of the objects in the image. The object detection model produces the probability that an object appears in the image or not. We choose k labels with the highest scores to avoid noise in the image. These labels are represented as one-hot vector and then multiplied with an embedding matrix $\mathbf{E}$ to produce L-dimension embedded vectors. Next, both the feature map and the embedded vectors are passed into the caption generation module.

In the caption generation module, the feature map and the embedded vectors are first processed through two different attention models. The first attention model uses a weight value α_i produced by the combination of information from previous hidden state and each feature from the feature map to show how much "attention" is on the feature vector a_i in the region i of the feature map. The image context vector at the current timestep t is then produced from the feature map and the weight value.

$$\alpha_{ti} = f_{softmax}(f_{attend1}(a_i, h_{t-1})) \tag{2}$$

$$z_{ta} = \sum_{i=1}^{N \times N} \alpha_{ti} \cdot a_i \tag{3}$$

The function $f_{softmax}$ helps the model to generate weights α_i for each region sumed up to 1 so that the context vector would be an expected context at timestep t.

We use the similar method in the second attention model for the embedded vectors of the labels. Each vector b_j in the k embedded vectors is multiplied with a weight value β_j and summed up to produce the label context vector.

$$\beta_{ti} = f_{softmax}(f_{attend2}(b_i, h_{t-1})) \tag{4}$$

$$z_{tb} = \sum_{j=1}^{k} \beta_j \cdot b_j \tag{5}$$

Finally, the image context vector z_a and the label context vector z_b are fed into an LSTM[13] to generate one word at each time step. The two context vectors are combined with the embedded vector of the word in the previous time step by a linear function which is also a fully connected layer in the model to get a context vector z. The LSTM takes in the vector z and produce a hidden state h_t at each time step t. The hidden state h_t is then passed through a fully connected layer to predict the next word in the caption. The predicted word is fed back into the attention models to produce a new set of weight values α, β and calculate a new context vector z for the next time step. Our entire model is showed in Figure 2 .

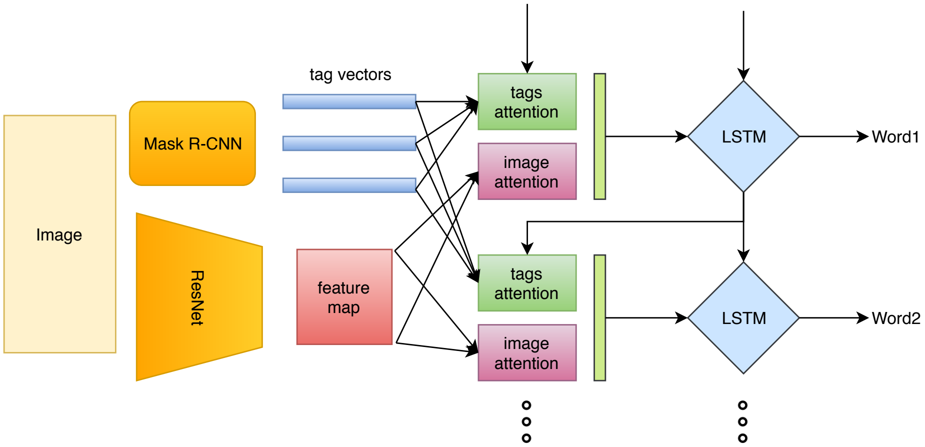

Fig. 2. Our image captioning model.

Unlike the work of Kelvin Xu et. al[25] which only uses the attention on image features and the work Quanzeng You et. al[26] which only looks at the image once and then uses attention on attributes, our model combines information from both features and labels. At each time step, the model will pay attention on certain region of image and on some certain tags to generate image caption.

Our model is trained on MS COCO dataset. In our implementation, we use the ResNet101[12] model as our convolutional neural network and the Mask R-CNN[11] as our object detector. The feature map is extracted from the last convolutional layer with size 14x14x1024. For the labels, we choose k=15 highest-score labels. Each label is embedded into a 512-dimension vector. The size of the final context vector is 2048.

5 Experiment

5.1 Strategy

In this Section, we present our system overview that can assist users to retrieve a memory in the past corresponding to a given query described in natural language text. Currently, we do not aim to make our system smart enough to automatically analyze and fully understand the query, but the system can help a user to step-by-step retrieve then filter with multiple criteria to get the appropriate results. Our system provides multiple strategies to query for past memory in lifelogging data, but it is the user who actively decides which sequence of strategies to search for a specific memory.

5.2 Query

Query 1: Find the moments when I was preparing salad. 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.

First of all, we consider on locations of the scenario. In this scenario, we focus on kitchen, office. Then we focus on the main context of the scenario. We mainly mining the main keyword "preparing salad". Through Word2Vec [17] model, we explore the similar and relevant words such as vegetable, fruit to extend the retrieval results. Finally, we get the candidate results and we manually choose the best results to submit. To create the variance of the result, we sort result based on time and choose the results at the different time. Figure 3 illustrates some image results of the query.

Query 2: Find the moments when I was with friends in Costa coffee. To be considered relevant, the moment must show at least a person together with the lifelogger in any Costa Coffee shop. Moments that show the user alone are not considered relevant.

In the query, the information of time is not mentioned so we could not apply a day periods to retrieve an image. In addition, the number of images have a context in the coffee shop is numerous and the volunteer drink coffee in different shops, the user has to decide which shop is Costa. Moreover, because the volunteer visit a coffee shop in different day, for each day, we select only one image

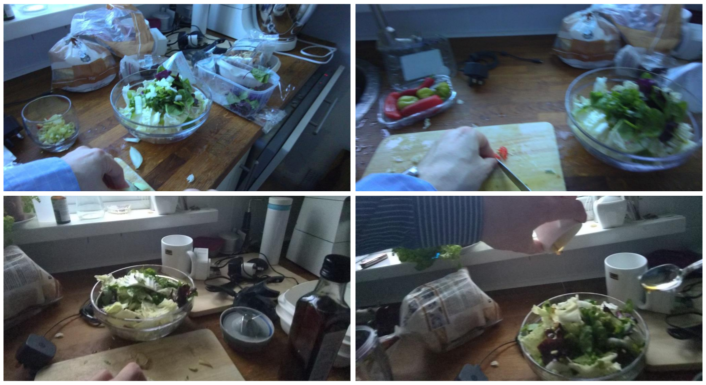

Fig. 3. The results of query 1

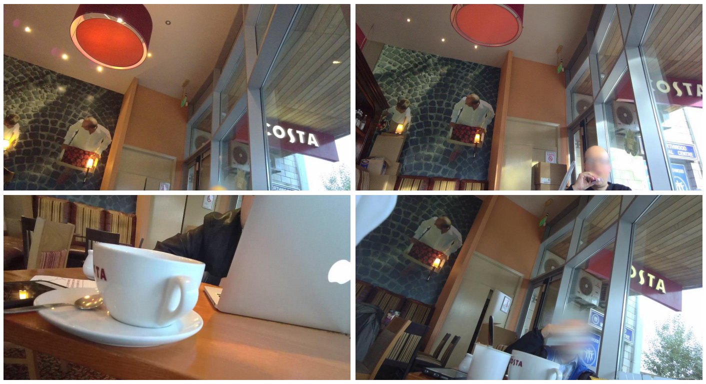

Fig. 4. The results of query 2

that meets the information in the query. Figure 4 show the retrieved images corresponding to the query.

Query 3: Find the moments when I was having dinner at home. Moments in which the user was having dinner at home are relevant. Dinner in any other location is not relevant. Dinner usually occurs in the evening time.

For this scenario, we consider time period and location. We mainly explore the context that is dinner indoor in the evening. Besides, we expand the results through keywords which are similar and relevant to dinner such as eat, drink, feed. With the retrieval list, we manually select the best results as a submission. Because the dinner is a daily activity, it almost exists every day so we select images at the different time and different days to create a variance of the submission. Figure 5 illustrates retrieval images corresponding to the query.

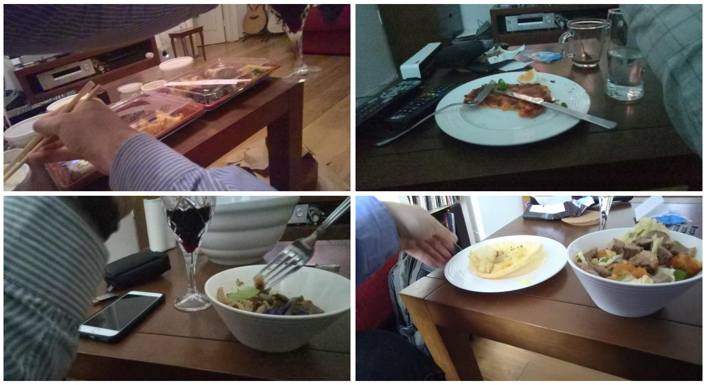

Fig. 5. The results of query 3

Query 4: Find the moments when I was giving a presentation to a large group of people. To be considered relevant, the moments must show more than 15 people in the audience. Such moments may be giving a public lecture or a lecture in the university.

The system could recognize that the scene of this query description is in the lecture hall where lot of students attend. The system suggests the "lecture' as a keyword for this scenario. Furthermore, "chair" and "table" are the object could be considered. From these keywords, the system return a list of images represent presentation. The user has to count the number of student appear in the image to decide which one is satisfied the query. The result images are shown in Figure 6.

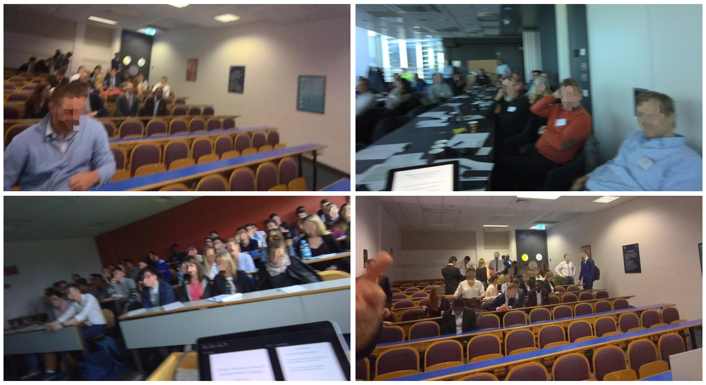

Fig. 6. The results of query 4

5.3 Result

Table 1 illustrates our result at LMRT challenge [6]. Our system is current the assistant retrieval system which helps user can retrieve the image by the textual query. Our long term goal aims to the automatically retrieval system with the prior textual query given by user.

Table 1. Result on ImageCLEF 2018 Lifelog - LMRT challenge.

Rank	Team	Score
1	AIlabGTi	0.545
2	**HCMUS**	**0.479**
3	Regim_Lab	0.424
4	NLP-lab	0.395
5	CAMPUS-UPB	0.216

The current result demonstrates the potential use of our system for moment retrieval in lifelog data. Our strategy is searching by keywords attention-based on an image description. The location and time period are mainly considered to filter the result corresponding to the context of the query.

6 Conclusion

Our proposal system assists user can retrieve lifelog moments based on the different types of queries (i.g. place, time, entity, extra biometric data). Leverage our system in Lifelog Search Challenge workshop to this challenge, we further explore the image caption to gain a better result. Besides, we proposed a novel method for generating image caption, it makes the description for every image can be diverse. For novice users, they could not know how to search with their existent keywords. To deal with this problem, we proposed a Keywords Recommendation by Word2Vec. We build-up a keywords dictionary which helps users can select the useful keywords for searching based on their existent keywords.

However, there are some weakness in our system. The retrieved results are not diverse. Diversity results play an important role in retrieval systems, especially in the Lifelog Moment Retrieval system. It helps to achieve a comprehensive and complete view on the query. Diversification of search results allows for better and faster search, gaining knowledge about different perspectives and viewpoints on retrieved information sources.

In following works, we will push new filters to remove noise and non-relative retrieved results. Additionally, we use new way to visualize images, specifically, we will cluster images into many group based on their features. It helps users can have a good visualization and easier to select images. Furthermore, our long-term goal is build-up a system which automatically searches with a raw textual query, is learning on strategies of users.

References

1. LTA '16: Proceedings of the First Workshop on Lifelogging Tools and Applications. ACM, New York, NY, USA (2016)
2. Arandjelović, R., Zisserman, A.: Three things everyone should know to improve object retrieval. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2911–2918. CVPR '12, Washington, DC, USA (2012)
3. Dang-Nguyen, D.T., Piras, L., Giacinto, G., Boato, G., Natale, F.G.B.D.: A hybrid approach for retrieving diverse social images of landmarks. In: 2015 IEEE International Conference on Multimedia and Expo (ICME). pp. 1–6 (June 2015)
4. Dang-Nguyen, D.T., Piras, L., Giacinto, G., Boato, G., Natale, F.G.B.D.: Multimodal retrieval with diversification and relevance feedback for tourist attraction images. ACM Trans. Multimedia Comput. Commun. Appl. **13**(4), 49:1–49:24 (Aug 2017)
5. 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)
6. Dang-Nguyen, D.T., Piras, L., Riegler, M., Zhou, L., Lux, M., Gurrin, C.: Overview of ImageCLEFlifelog 2018: Daily Living Understanding and Lifelog Moment Retrieval. In: CLEF2018 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Avignon, France (September 10-14 2018)
7. Dinh, T.D., Nguyen, D., Tran, M.: Social relation trait discovery from visual lifelog data with facial multi-attribute framework. In: Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods, ICPRAM 2018, Funchal, Madeira - Portugal, January 16-18, 2018. pp. 665–674 (2018)
8. Gurrin, C., Joho, H., Hopfgartner, F., Zhou, L., Albatal, R.: Overview of ntcir-12 lifelog task (2016), the authors acknowledge the financial support of Science Foundation Ireland (SFI) under grant number SFI/12/RC/2289 and the input of the DCU ethics committee and the risk &amp; compliance officer. We acknowledge financial support by the European Science Foundation via its Research Network Programme ?Evaluating Information Access Systems?.
9. Gurrin, C., Giro-i Nieto, X., Radeva, P., Dimiccoli, M., Dang-Nguyen, D.T., Joho, H.: Lta 2017: The second workshop on lifelogging tools and applications. In: Proceedings of the 2017 ACM on Multimedia Conference. pp. 1967–1968. MM '17, ACM, New York, NY, USA (2017)
10. Gurrin, C., Smeaton, A.F., Doherty, A.R.: Lifelogging: Personal big data. Found. Trends Inf. Retr. **8**(1), 1–125 (Jun 2014)
11. He, K., Gkioxari, G., Dollar, P., Girshick, R.: Mask r-cnn. In: The IEEE International Conference on Computer Vision (ICCV) (Oct 2017)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016)
13. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (Nov 1997)
14. Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T.: Flownet 2.0: Evolution of optical flow estimation with deep networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017. pp. 1647–1655 (2017)

15. Ionescu, B., Müller, H., Villegas, M., de Herrera, A.G.S., Eickhoff, C., Andrearczyk, V., Cid, Y.D., Liauchuk, V., Kovalev, V., Hasan, S.A., Ling, Y., Farri, O., Liu, J., Lungren, M., Dang-Nguyen, D.T., Piras, L., Riegler, M., Zhou, L., Lux, M., Gurrin, C.: Overview of ImageCLEF 2018: Challenges, datasets and evaluation. In: 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, Avignon, France (September 10-14 2018)
16. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. CoRR **abs/1301.3781** (2013)
17. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 26, pp. 3111–3119. Curran Associates, Inc. (2013)
18. Nguyen, V.T., Le, K.D., Tran, M.T., Fjeld, M.: Nowandthen: A social network-based photo recommendation tool supporting reminiscence. In: Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia. pp. 159–168. MUM '16, ACM, New York, NY, USA (2016)
19. Nguyen, V., Ngo, T.D., Tran, M., Le, D., Duong, D.A.: A combination of spatial pyramid and inverted index for large-scale image retrieval. IJMDEM **6**(2), 37–51 (2015)
20. Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Object retrieval with large vocabularies and fast spatial matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2007)
21. Philbin, J., Isard, M., Sivic, J., Zisserman, A.: Lost in quantization: Improving particular object retrieval in large scale image databases. In: In CVPR (2008)
22. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems (NIPS) (2015)
23. Sivic, J., Zisserman, A.: Video google: A text retrieval approach to object matching in videos. In: Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2. pp. 1470–. ICCV '03, IEEE Computer Society, Washington, DC, USA (2003)
24. Truong, T.D., Dinh-Duy, T., Nguyen, V.T., Tran, M.T.: Lifelogging retrieval based on semantic concepts fusion. In: Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge. pp. 24–29. LSC '18, ACM, New York, NY, USA (2018)
25. Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R., Bengio, Y.: Show, attend and tell: Neural image caption generation with visual attention. In: Bach, F., Blei, D. (eds.) Proceedings of the 32nd International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 37, pp. 2048–2057. PMLR, Lille, France (07–09 Jul 2015)
26. You, Q., Jin, H., Wang, Z., Fang, C., Luo, J.: Image captioning with semantic attention. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016)
27. Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A.: Learning deep features for scene recognition using places database. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 27, pp. 487–495. Curran Associates, Inc. (2014)
28. Zhu, C., Jegou, H., Satoh, S.: Query-adaptive asymmetrical dissimilarities for visual object retrieval. In: IEEE International Conference on Computer Vision, ICCV 2013, Sydney, Australia, December 1-8, 2013. pp. 1705–1712. IEEE (2013)