Detection of Semantic Risk Situations in Lifelog Data for Improving Life of Frail People

Thinhinane Yebda
thinhinane.yebda@u-bordeaux.fr
LaBRI, Université de Bordeaux
Talence, France

Jenny Benois-Pineau
jenny.benois-pineau@u-bordeaux.fr
LaBRI, Université de Bordeaux
Talence, France

Marion Pech
marion.pech@u-bordeaux.fr
BPH, Inserm, Université de Bordeaux
Bordeaux, France

Hélène Amièva
helene.amieva@u-bordeaux.fr
BPH, Inserm, Université de Bordeaux
Bordeaux, France

Cathal Gurrin
Cathal.Gurrin@dcu.ie
Dublin City University
Dublin, Ireland

ABSTRACT

The automatic recognition of risk situations for frail people is an urgent research topic for the interdisciplinary artificial intelligence and multimedia community. Risky situations can be recognized from lifelog data recorded with wearable devices. In this paper, we present a new approach for the detection of semantic risk situations for frail people in lifelog data. Concept matching between general lifelog and risk taxonomies was realized and tuned AlexNet was deployed for detection of two semantic risks situations such as risk of domestic accident and risk of fraud with promising results.

CCS CONCEPTS

• **Applied computing** → *Health care information systems*; • **Computing methodologies** → **Neural networks**.

KEYWORDS

Risk situations detection, Neural network, Classification, CNN networks

1 INTRODUCTION

Europe combines demographic ageing and gerontogrowth. This second phenomenon, defined by the increase in the number of elderly people, is manifested in France for instance by the 35.2% increase in the number of people over 65 between 1999 and 2014, while the number of people over 80 has increased by 78.9% [1].

Frailty is a geriatric syndrome which is characterized by decreased strength, low physical activity and weight loss associated with health complications. Frailty is manifested by biological (age-related) vulnerability to stressors and decreased physiological reserves. The five-point criteria for frailty that are recognized and commonly used for identification [2] are: self-reported exhaustion, slowed performance (by walking speed), weakness (by grip strength), involuntary weight loss (10 pounds in the past year), and low physical activity. They are composite results of several organ systems.

Numerous studies have been realized to determine the most frequent risk situations faced by frail people such as falls[3]. Today we are seeking for detection of much more complex and "semantic" risk situations in everyday life of frail people. In this paper, we present ongoing research into the detection of risks in lifelog data for risk prevention for frail people.

Demonstration description. We will present a video demonstration of project scenarios, the annotation of lifelog data with a task-tailored interface and also the results of detection of semantic risk situations which can be found in general purpose lifelog data. A part of demo will be real-time detection of semantic risk situation from predefined taxonomy. In the following we present our methodology and contributions.

The remainder of the paper is organized as follows: in section 2 we present related work, Section 3 presents our method for detection of semantic risk situations on the visual component of lifelog data, in section 4 we describe our experiment and present results of the detection of two semantic risk situations. Section 5 concludes this work and outlines its perspectives.

2 RELATED WORK

2.1 IT for support of independent living

In recent years, wearable technologies and the Internet of Things (IoT)-based applications aiming at supporting independent living of the elderly have considerably progressed. A frequent outcome targeted by such technologies is fall detection, which also constitutes one of the major causes of institutionalization [4]. Therefore, wearable sensors (WS) are being developed with the aim of becoming effective tools for prevention, early detection of falls, and monitoring general activities of daily living in older adults. A progress has been made possible thanks to the development of remote data

collection methods and technologies [5]. Regarding the applications for monitoring activities of daily living and independent living in older adults, several studies have shown the efficiency of WS and IoT systems from a technological point of view [6]-[7]-[8]. For remote monitoring, smartphone and smartwatch-based applications are frequently used in older adults [8]-[9]-[10]. The main advantage of these devices is that they are non-invasive and comfortable for users.

Concerning detection and prevention of falls, different types of WS and IoT-based applications have been developed for older adults' care at home [10], [11], [12], [13], [14], [15], [16]. First, there is a smart and connected home health monitoring system [11]. This device is composed of hardware such as object sensors, a wearable human sensor with an alarm button, a gateway and there is a software for the data collection. This technology is based on deep learning and hidden Markov models with sensor orientation calibration [10][17]. Secondly, there is a waist-worn fall detector based on an Attitude and Heading Reference system and barometric sensor. This system showed 100% of sensibility for fall detection in diverse studies [12]-[13]. Another study reported the development of a novel passive Radio Frequency Identification using Doppler Frequency for fall detection. "Tagcare", the monitoring system's prototype has shown a high accuracy (98%) for movement detection and sudden falling [14]. Furthermore, wrist-worn wearables have been developed to improve fall detection [10]-[15]. Finally, a fall detection system has been developed in an indoor environment, based on IoT and a Big Data model "Ensemble-Random Forest" (RF). This device is based on a machine-learning algorithm, which showed a success rate of above 94% for accuracy, sensitivity and specificity in detecting three types of falls (forward, backward and lateral falls) and activities of daily living (walking, stairs climbing, and sitting) [16]. As may be seen, various technological solutions reveal to be potentially interesting tools to support independent living at home, and the recent results for the prevention of falls are promising. Nevertheless, challenges and barriers to the wider adoption of WS and IoT application still exist [18]. Today, few studies address the feasibility and acceptability of these devices outside a laboratory context, to detect the activities of an elderly person's daily routine, which is the next logical step to take [19]. A key component of the detection of risky situations for persons is the analysis of visual data recorded with wearable visual sensors. This kind of recordings has been called "lifelog". These data [20] have been used mainly for dairy generation in different studies, e.g. of patients suffering from Alzheimer disease [21] for recognition of familiar scenes [22]. The methodology nowadays consists of recognition of different situations accordingly to designed taxonomy by Deep Neural Networks classifiers. In the following subsection, we will briefly review available solutions.

2.2 Deep Learning in Lifelog visual content mining

Nowadays, practically all approaches we have been developing for any multimedia and multi-modal data analysis are "translated" into "Deep Learning" approaches. This is also the case of lifeLog data [23]. While for specific tasks in lifelog data new Deep Architectures are being designed [24], the standard backbones such as ResNet

[25] are suitable for recognition of egocentric scenes [26]. Nevertheless, data from risky situations are not very frequent in typical lifeLog data or in commonly used datasets [27]. Therefore a large training dataset of positive examples is not readily available for training a robust model, and it becomes more realistic to resort to lighter networks, such as AlexNet [28]. In this work, we do not propose a new architecture but adapt our architecture choice, data augmentation and hyperparameter tuning to our problem.

3 DETECTION OF SEMANTIC RISK SITUATIONS WITH DEEP CNNS

3.1 Taxonomy of risk situations

The wide range of risk situations encountered for fragile people have been identified in [29], [3]. These studies were used in [30] to identify a necessary set of sensors and data collection scenarios. According to these studies, the most common accidents among the elderly are falls. It is the third most frequent source of ill-health after heart attack and cancer. Nevertheless, the risks faced by older adults differ according to their medical history. Except for the cases of the elder's falls, there are other more specific situations such as Parkinson's falls, specific risks of people with Alzheimer's disease in addition to diabetics, risks for people with heart failure or respiratory failure and other chronic diseases. The situations of these patients differ according to the disease. For example, for Parkinson's patients, the most urgent risks are Parkinson's falls, which are very common, for Alzheimer's patients, there are stressful situations, loss of orientation and loss of direction. For people with diabetes, hyperglycemia and hypoglycemia are dangerous. Considering the many studies on risk situations for fragile people, in our work, we tackle risks which can be met in many situations of frailty. Without referring to a specific pathology, the risk situations we are interested in are among the most common faced by elderly in our society:

- Falls;
- Risk of falls;
- Risk of fraud;
- Risk of domestic accidents.

Analyzing the proposed taxonomy, it is clear that the different risk situations are not equivalent in semantic charge. Hence falls and fall onsets, i.e. risk-of-fall can be detected from analysis of accelerometer signals via thresholding of the magnitude. The second category of risks is much more "semantic". Domestic accidents can happen in any place of the person's home, but are more probable in kitchen, the risk of fraud or abuse supposes quite particular structure of a scene when a fragile subject faces a person he/she let enter in his/her home... Furthermore, the collection of such data on real-world cohort is a challenge for privacy issues. Therefore it is interesting to make a proof of concept on a general-purpose lifelog data. This is our case. In the followup we will shortly describe the LSC lifelog dataset [27] we work with and specify the matching of general-purpose lifelog concepts and potential risk situations.

3.2 Annotation of risk situations on general-purpose Lifelog data

The available LSC lifelog dataset comprises a large array of multi-modal data on an individual's life experience collected by wearable sensors [27]. The dataset was collected over a 27-day period by a single individual and includes images recorded with a wearable camera and multi-sensory data.

3.2.1 Image Data. The frequency of image acquisition is such that we have approximately $1,500$ images per day taken by a wearable camera. With 27 days of recording. The overall volume of the image lifelog database is of $41,664$ images, accompanied by the output of a concept detector that identifies the dominant visual concepts in each image from a total of 548 visual concepts such as sky, indoor, kitchen, computer, person, and so on. Accompanying this visual data was a music listening history, although we did not use this music data in thei research.

3.2.2 Multi-sensory data. These data are composed of biometric data, contextual data and also are supplied with annotations of a higher semantic level.
Biometric Data: the lifelog sensor kit records biometric data such as heart rate, galvanic skin response, calorie burns, and steps stored with a frequency of 1 minute. We also have data on blood pressure and blood sugar levels every day.
Contextual data: it consists of information on semantic places visited, physical activities, the mood of the day and a food diary.
Computer use: Represents a vector of data stored every minute, filtered using a blacklist, made anonymously and then filtered. We did not use this data in this work either.

Each piece of data described above is time-stamped, and this was used to align the data over time for indexing. We also used this temporal information to create a semantic annotation for each moment of collection, such as the day of the week or time of day (morning, afternoon, evening).

3.2.3 Annotation of risk situations. When considering risk situations for a person and without a ground truth in the data, we have used image data to identify potential risks. For this, concept matching to risky situations has been elaborated. The correspondence between the concepts in the LSC dataset and potential risk situation is given in table 1. The concepts of LSC were completed by psychologists, and are depicted in italic in the right column of table 1. For example, in the "risk-of-fall" situation was completed by "staircase".
The first step in annotation was performed *automatically*. Hereby parsing the metadata of LSC corpus, i.e. the file of concepts related to images, the images were labelled with the corresponding risk situation. For example, images with the "kitchen" concept were assigned to the "risk of domestic accident " class. Images without the LSC concepts from the correspondence table were assigned to the rejection class "other". At the second step, all annotations were verified *manually*. The images were re-assigned to the corresponding risk class (left column in the table 1) by an experienced psychologist.
In the present work, we use only visual data for risk situation recognition. Considering the matching of risk concepts (see left column of the table 1 to the LSC concepts (right column of the table 1) defined

Table 1: Concepts for risk situations from lifelog

Risk situations	Concepts
Fall	sky, grassy, ceiling, floor, night sky, clouds, roof, landing.
Risk of fall	shower, bathroom, bath, *staircase*.
Risk of fraud	person + door, woman + door, man + door.
Risk of domestic accident	kitchen, oven, microwave.

by psychologists one can notice that such risk situations as "fall" or "risk of fall" has a too heterogeneous set of visual concepts in LSC. It is illusory to try to recognize them only on images. Biometric data expressing the homeostasis of the person have to be considered, such as galvanic skin response, heart-rate... Nevertheless, for the LSC corpus, this is questionable as it was recorded by a healthy and very active person. A real-world corpus on fragile subjects is needed. Hence to make the proof of the concept we consider in this work only two risk situations: i)"risk of fraud" and ii)"risk of domestic accident".

3.3 Architecture choice and tunning of CNN for risk situations recognition

Our preliminary tests on ResNet 26 have have not produced satisfactory results, therefore we applied to AlexNet [31] as our experience shows that less deep, but well-tunned networks allow for reasonable accuracies. When training the network we used a batch size of 64 images for the detection of the risk of domestic accident and a batch size of 10 images for the detection of the risk of fraud, proportionally to the number of original images in the training dataset. As the optimization method, we used the stochastic gradient descent (SGD) with Nesterov momentum [32]. In order to update the parameters W to minimize the objective function, $\mathcal{L}(W)$ the method assumes the iterative scheme,

$$V_{i+1} = \mu V_i - \eta \nabla \mathcal{L}(W_i + \mu V_i)$$
$$W_{i+1} = W_i + V_{i+1} \tag{1}$$

where V_i denotes velocity, μ the momentum coefficient ($\mu = 0.9$), η the learning rate, i the iteration and W_i the filters coefficients.

The exponential learning rate decay policy proved to be efficient in works with also not a large number of training data available [33]. We thus use it : $\eta = \eta_0.\gamma^i$ where η_0 is the initial learning rate, $\gamma \in [0, 1]$ gives the shape of learning rate decay curve and i the number of current training iteration. With a lower value of γ, the learning rate decreases faster at the beginning of the training, and a value close to 1, keeps a small decay for every iteration. We established $\gamma = 0.97$ for the risk of domestic accidents and higher values between 0.98 and 0.999 for the case of fraud risk after several experiments, due to the range of γ which gives an appropriate learning rate decay is $[0.95, 0.99]$. We defined the value of the initial learning rate η_0 after several experiments using the bisection method at $\eta_0 = 0.00125$ for the two risk situations on a search bar between $[0.0001, 0.1]$. The maximal number of epochs was 90. The training was done using the open-source MXNet Framework [34].

Table 2: Datasets used for training for each risk situation

Risk situation	Positive samples	Negative samples
Risk of domestic accident	406	600
Risk of fraud	120	180

4 EXPERIMENTS AND RESULTS

4.1 Visual Dataset

Thus we conducted experiments to recognize two risk situations on the visual part of LSC dataset [27]. It comprises 41664 images. We have selected four semantic risk classes, but have conducted experiments only on more semantic classes "risk of fraud" and "risk of domestic accident" 3. To avoid too similar images, we have limited our selection of images from the LSC dataset to one image per minute, which reduces the total number of images to 21,152. The classification problem we solved was binary. The positive example images in original LSC dataset for these two semantic risk situations are not numerous. The negative examples were selected proportionally. The figures are presented in Table 2. The constitution of these two datasets was done by taking the positive samples of a class and by randomly selecting ×1.5 images of the number of positives from the rest of the dataset.

For data augmentation, rotations, vertical and horizontal flips, left and right image translations and Gaussian noise addition were performed.

Finally, for the risk of a domestic accident, the training dataset consisted of 4224 images, 201 images for validation and 101 for testing. For the risk of fraud, the training dataset consists of 1254 images, 60 images for validation and 31 for testing.
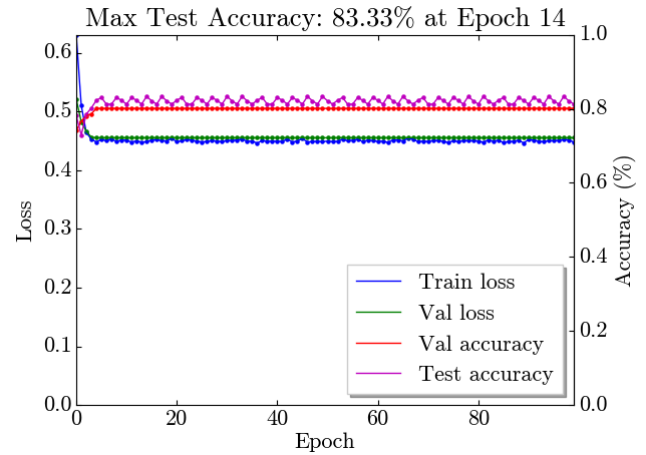
4.2 Results of detection of two semantic risk situations

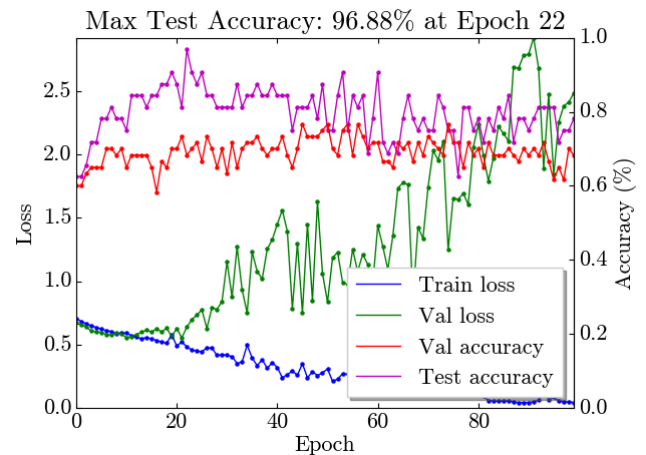The Alexnet model [31] presented in the section was used as a classifier 3.

We proceed with the analysis of the Alexnet model training curves for the two risk situations and illustrate them in figure 1 and figure 2, noting that the hyperparameters of the AlexNet model have been optimized for each of the risk situations separately, see section 3. As illustrated in figure 1, for the risk of domestic accidents we obtain rather good results and the model does not overfit on the whole training dataset, we get an accuracy of 81, 33% on the test set rather quickly at epoch 14. For the risk of fraud, good accuracy is obtained of 96, 88% on the test set at epoch 22, see figure 2 after this time, the overfitting is observed and validation loss diverges.

5 CONCLUSION AND PERSPECTIVES

This research is the first attempt to use deep CNN framework for semantic risk situations detection with visual content on a lifeLog data. We have tackled a complex problem of recognition of semantic risks situation in the daily lives of fragile people. Thus the definition of matching of general concepts to risk concepts was an important part of the work. We have used a well-known model Alexnet with the stochastic gradient descent (SGD) with Nesterov momentum,

Figure 1: Alexnet curves for risk of domestic accident. The maximum accuracy $(83, 33\%)$ **is attained at epoch 14**

Figure 2: Alexnet curves for risk of fraud. The maximum accuracy $(96, 88\%)$ **is attained at epoch 22**

but have thoroughly tunned parameters. The results are promising, and in perspective, we will combine the developed strategy with multi-sensory data on fragile subjects.

6 ACKNOWLEDGEMENT

REFERENCES

[1] A. Parant, "Blanchet mickaël, 2017, atlas des séniors et du grand âge en france. 100 cartes et graphiques pour analyser et comprendre, rennes, presses de l'ehesp, 120 p." *Population*, vol. 73, no. 4, pp. 852–853, 2018.
[2] N. Fedarko, "The biology of aging and frailty," *Clinics in Geriatric Medicine*, vol. 27, no. 1, pp. 27–37, 2 2011.
[3] S. S. Vemulapalli, M. B. Ulak, E. E. Ozguven, T. Sando, M. W. Horner, Y. Abdel-razig, and R. Moses, "Gis-based spatial and temporal analysis of aging-involved

accidents: a case study of three counties in florida," *Applied Spatial Analysis and Policy*, vol. 10, no. 4, pp. 537–563, Dec 2017.

[4] M. E. Tinetti and C. Kumar, "The patient who falls:"it's always a trade-off"," *Jama*, vol. 303, no. 3, pp. 258–266, 2010.

[5] S. Sabesan and R. Sankar, "Improving long-term management of epilepsy using a wearable multimodal seizure detection system," *Epilepsy & Behavior*, vol. 46, pp. 56–57, 2015.

[6] N. Hegde, M. Bries, T. Swibas, E. Melanson, and E. Sazonov, "Automatic recognition of activities of daily living utilizing insole-based and wrist-worn wearable sensors," *IEEE journal of biomedical and health informatics*, vol. 22, no. 4, pp. 979–988, 2017.

[7] N. Hegde and E. Sazonov, "Smartstep: A fully integrated, low-power insole monitor," *Electronics*, vol. 3, no. 2, pp. 381–397, 2014.

[8] P. Bellagente, C. Crema, A. Depari, P. Ferrari, A. Flammini, G. Lanfranchi, G. Lenzi, M. Maddiona, S. Rinaldi, E. Sisinni *et al.*, "Remote and non-invasive monitoring of elderly in a smart city context," in *2018 IEEE Sensors Applications Symposium (SAS)*. IEEE, 2018, pp. 1–6.

[9] M. Kheirkhahan, S. Nair, A. Davoudi, P. Rashidi, A. A. Wanigatunga, D. B. Corbett, T. Mendoza, T. M. Manini, and S. Ranka, "A smartwatch-based framework for real-time and online assessment and mobility monitoring," *Journal of biomedical informatics*, vol. 89, pp. 29–40, 2019.

[10] S. B. Khojasteh, J. R. Villar, C. Chira, V. M. González, and E. De la Cal, "Improving fall detection using an on-wrist wearable accelerometer," *Sensors*, vol. 18, no. 5, p. 1350, 2018.

[11] L. Maimoon, J. Chuang, H. Zhu, S. Yu, K.-S. Peng, R. Prayakarao, J. Bai, D. Zeng, S.-H. Li, H. Lu *et al.*, "Silverlink: developing an international smart and connected home monitoring system for senior care," in *International Conference on Smart Health*. Springer, 2016, pp. 65–77.

[12] P. Pierleoni, A. Belli, L. Palma, M. Pellegrini, L. Pernini, and S. Valenti, "A high reliability wearable device for elderly fall detection," *IEEE Sensors Journal*, vol. 15, no. 8, pp. 4544–4553, 2015.

[13] Y. Lin, C. Wang, J. Wang, and Z. Dou, "A novel dynamic spectrum access framework based on reinforcement learning for cognitive radio sensor networks," *Sensors*, vol. 16, no. 10, p. 1675, 2016.

[14] L. Zhu, R. Wang, Z. Wang, and H. Yang, "Tagcare: Using rfids to monitor the status of the elderly living alone," *IEEE Access*, vol. 5, pp. 11 364–11 373, 2017.

[15] J. Yuan, K. K. Tan, T. H. Lee, and G. C. H. Koh, "Power-efficient interrupt-driven algorithms for fall detection and classification of activities of daily living," *IEEE Sensors Journal*, vol. 15, no. 3, pp. 1377–1387, 2014.

[16] D. Yacchirema, J. S. de Puga, C. Palau, and M. Esteve, "Fall detection system for elderly people using iot and ensemble machine learning algorithm," *Personal and Ubiquitous Computing*, vol. 23, no. 5-6, pp. 801–817, 2019.

[17] D. Seo, B. Yoo, and H. Ko, "Data-driven smart home system for elderly people based on web technologies," in *International Conference on Distributed, Ambient, and Pervasive Interactions*. Springer, 2016, pp. 122–131.

[18] M. M. Baig, S. Afifi, H. GholamHosseini, and F. Mirza, "A systematic review of wearable sensors and iot-based monitoring applications for older adults–a focus on ageing population and independent living," *Journal of medical systems*, vol. 43, no. 8, p. 233, 2019.

[19] B. Klaassen, B. J. van Beijnum, and H. J. Hermens, "Usability in telemedicine systems—a literature survey," *International journal of medical informatics*, vol. 93, pp. 57–69, 2016.

[20] A. R. Doherty, K. Pauly-Takacs, N. Caprani, C. Gurrin, C. J. A. Moulin, N. E. O'Connor, and A. F. Smeaton, "Experiences of aiding autobiographical memory using the sensecam," *Human-Computer Interaction*, vol. 27, no. 1-2, pp. 151–174, 2012.

[21] R. J. A. S. K. I. Louise Hopper, Eamonn Newman, "Dementia ambient care: Home-based monitoring and enablement of people with dementia," in *9th Panhellenic Conference on Alzheimer's Disease and 1st Mediterranean on Neurodegenerative Diseases Conference,PICAD 2015*, 2015, pp. 1–4.

[22] E. Talavera, N. Petkov, and P. Radeva, "Towards unsupervised familiar scene recognition in egocentric videos," *CoRR*, vol. abs/1905.04093, 2019.

[23] C. Gurrin, K. Schoeffmann, H. Joho, A. Leibetseder, L. Zhou, A. Duane, D.-T. Dang-Nguyen, M. Riegler, L. Piras, M.-T. Tran, J. Lokoc, and W. Huerst, "[invited papers] comparing approaches to interactive lifelog search at the lifelog search challenge (lsc2018)," *ITE Transactions on Media Technology and Applications*, vol. 7, no. 2, pp. 46–59, 2019.

[24] E. S. Aimar, P. Radeva, and M. Dimiccoli, "Social relation recognition in egocentric photostreams," in *ICIP*. IEEE, 2019, pp. 3227–3231.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, cite arxiv:1512.03385Comment: Tech report. [Online]. Available: http://arxiv.org/abs/1512.03385

[26] I. González-Díaz, J. Benois-Pineau, J. Domenger, D. Cattaert, and A. de Rugy, "Perceptually-guided deep neural networks for ego-action prediction: Object grasping," *Pattern Recognit.*, vol. 88, pp. 223–235, 2019.

[27] C. Gurrin, K. Schoe, H. Joho, and B. Munzer, "A Test Collection for Interactive Lifelog Retrieval," in *MMM 2019, the 25th International Conference on MultiMedia Modeling*, Thessaloniki, Greece, 2019.

[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[29] B. Tavernier, C. Helmer, F. Portet, H. Amieva, I. Carrière, J.-F. Dartigues, J. A. Ávila Funes, K. Ritchie, L. M. Gutiérrez-Robledo, M. L. Goff, and P. Barberger-Gateau, "Frailty Among Community-Dwelling Elderly People in France: The Three-City Study," *The Journals of Gerontology: Series A*, vol. 63, no. 10, pp. 1089–1096, 10 2008.

[30] T. Yebda, J. Benois-Pineau, H. Amieva, and B. Frolicher, "Multi-sensing of fragile persons for risk situation detection: devices, methods, challenges," in *2019 International Conference on Content-Based Multimedia Indexing (CBMI)*. IEEE, 2019, pp. 1–6.

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[32] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 675–678.

[33] A. M. Obeso, J. Benois-Pineau, M. S. García-Vázquez, and A. A. Ramírez-Acosta, "Saliency-based selection of visual content for deep convolutional neural networks - application to architectural style classification," *Multimedia Tools Appl.*, vol. 78, no. 8, pp. 9553–9576, 2019.

[34] T. Chen, M. Li, Y. Li, M. Lin, N. Wang, M. Wang, T. Xiao, B. Xu, C. Zhang, and Z. Zhang, "Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems," *arXiv preprint arXiv:1512.01274*, 2015.