

Towards Objective Description of Eating, Socializing and Sedentary Lifestyle Patterns in Egocentric Images

Pedro Herruzo¹

pherruzo@ac.upc.edu

Laura Portell²

portell.laura@gmail.com

Alberto Soto²

alsoba13@gmail.com

Beatriz Remeseiro^{*3}

bremeseiro@uniovi.es

¹ DAMA-UPC

Universitat Politècnica de Catalunya
Barcelona, Spain

² Dept. of Mathematics and Computer
Science

Universitat de Barcelona
Barcelona, Spain

³ Department of Computer Science

Universidad de Oviedo
Gijón, Spain

Abstract

The objective description of lifestyle patterns from egocentric images captured by wearable cameras is considered the next step on health-tracking applications which goes alongside heartbeats, steps walked, or burned calories measured mainly by electronic wristbands. In this context, the classification of eating, socializing, and sedentary lifestyle patterns has been addressed in previous research, considering the following classes for each pattern, respectively: 1) no food, food but not eating, and eating; 2) not socializing, and socializing; and 3) not sedentary, and sedentary. This previous approach provides a solution that considers all the possible combinations of classes among these three patterns, thus solving a multi-class classification problem with 12 classes. Given the nature of the problem, we propose to address the classification of these three lifestyle patterns under a multi-task perspective, employing a general framework based on Inception-V3 convolutional neural network. The feature maps extracted from the employed network are used to perform a final lifestyle pattern categorization. It is worth to note that our proposed framework offers visual explanations of the results to accomplish the standards of the European Union in terms of explainability, thus increasing both transparency and confidence on the obtained results. Furthermore, in this work a web-based lifelogging tool for periodic visual summarization of lifestyle patterns is introduced. Our method has been tested on a dataset composed of more than 45,000 egocentric images, and the experimental results confirm the reliability of the selected method while outperforming state-of-the-art in terms of accuracy and F1-score.

* Corresponding author.

1 Introduction

With the advent of wearable cameras, egocentric lifelogging has become a very popular topic among computer vision researchers. The purpose of egocentric lifelogging is to describe daily experience of people who wear these cameras, by analyzing their continuous recordings [8]. Among the different applications of egocentric vision expressed in the literature, which include creating cognitive exercises for Alzheimer’s patients [13], or remembering personal experiences [15], tracking of daily habits for healthy purposes has received fewer attention. Note that other wearable devices have been used with the same target, e.g. activity trackers such as FitBit¹ —a wristband that measures calories and sleeping hours.

In this context, egocentric vision can be seen as a further step to provide useful information about a person on different health-related areas. For example, nutritionists could benefit from food recognition [9] to characterize eating patterns; whilst psychologists could profit from detecting social interactions [6] to describe behavior patterns, and from sentimental analysis [12] to depict people’s moods.

With the main aim of analyzing people’s lifestyle patterns, the so-called LAP dataset was presented in [14]. It contains more than 45,000 egocentric images annotated according to three different patterns: 1) eating pattern, with three classes: eating, food related non eating, or non food related; 2) socializing pattern, with two classes: socializing, not socializing; and 3) sedentary lifestyle pattern, with two classes: table, no table. Taking into account all the possible combinations between the different classes of the three patterns, the methodology proposed by Herruzo et al. [14] solves a multi-class classification problem with 12 classes, by means of traditional classifiers and convolutional neural networks (CNNs).

The main goal of our research is to propose a new approach to objectively describe the lifestyle patterns of the LAP aforementioned dataset, but using a multi-task perspective. In this manner, instead of solving a 12-class classification problem, our approach considers the different patterns as individual tasks. Our contributions are three-fold: 1) a method to automatically describe different people’s habits, such as when and for how long a person eats, interacts with others, and is seated; 2) a lifelogging tool for egocentric camera wearers that summarizes their daily activities; and 3) visual explanations to understand the pattern descriptions and provide confidence to the camera wearers.

The rest of this manuscript is structured as follows. Section 2 is devoted to providing a brief overview of related work. Section 3 describes the egocentric dataset used in this research, and presents the proposed method for the objective description of lifestyle patterns. Section 4 shows the experimental results and analyzes them in terms of visual explanations. Section 5 concludes the manuscript and includes future lines of research.

2 Related work

Despite the fact that observing nutritional and lifestyle habits has not attracted much attention yet in the literature, there are some relevant works focused on related topics. For example, De Barbaro [1] explained the importance of information recorded from sensors, like egocentric wearable cameras, to quantify how daily activities shape human developmental outcomes. In particular, some types of activities are highlighted for this purpose, including physical activity, quantity of interactions, and eating episodes, among others.

¹<https://www.fitbit.com/home>

Regarding social interactions, a model to estimate head pose and 3D location in egocentric videos is presented in [4]; whilst both distance and orientation of appearing individuals are exploited in [2] using egocentric images. Other daily activities have been considered in [9], including eating and socializing activities observed also in egocentric images. Different wearable sensors were analyzed in [6], using a multimodal dataset that includes video, audio, GPS and inertial sensing data. The experimentation carried out demonstrated that both video and audio are useful to predict people interactions.

To the best of our knowledge, there is only one previous research [24] (see Section 1) focused on eating patterns observed in egocentric images, to determine information such as how, where, when people eat. However, there are some works that present a baseline to deal with food recognition, using deep learning and combining features like time and multi-tasking. In this context, Yu et al. [24] presented a personalized food classifier with time and user dependency, whilst Khan et al. [16] use a multi-task approach to perform both food segmentation and volume estimation. With respect to alternative devices to wearable cameras, Bedri et al. [7] recognize eating episodes using a chewing sensor called IMU that could be mounted on eyeglasses, as suggested by the authors.

Finally, it is also worth to mention some relevant deep learning techniques for conventional images, i.e. non-egocentric ones, which solve the problem of discovering the relations between objects and activities. YOLO is one of the most popular object detectors [18], which builds the whole detection pipeline into a single network optimized end-to-end with very promising results. When the problem is to predict multiple labels for the same picture, each one corresponding to a different task, two main approaches should be highlighted: Caruana [10] proposed to use a different network for each task, with intermediate parameters shared allowing to backpropagate the errors of different tasks all together; whilst Abdulnabi et al. [1] proposed a multi-task CNN model that generates attribute-specific feature representations, and decomposes the overall model’s parameters into a latent task matrix to predict image attributes.

3 Materials and methods

This section describes the LAP dataset, composed of more than 45,000 egocentric images with three labels representing people’s habits. Next, we present a general framework to objectively describe these habits by analyzing nutritional activities, social interactions and environments; and a lifelogging tool for egocentric camera wearers, allowing them to visualize a summary of their daily activities.

3.1 LAP dataset

The LAP dataset was introduced in [24] aiming at representing different people’s lifestyle patterns. It contains 45,927 egocentric images ($1944 \times 592 \times 3$), acquired with a Narrative Clip² camera. The pictures were taken by four different people, in consecutive days (frame rate: 2fpm) and different environments. Table 1 shows the data distribution.

All the images were manually annotated with three labels, corresponding to three different patterns considered:

²<http://getnarrative.com/>

- Eating pattern: it allows to distinguish between non-food and food situations and, in the second case, it also takes into account whether the subject is eating or not. Three labels: non food related (NFR), food related non eating (FRNE), and food related eating (FRE).
- Socializing pattern: it allows to determine if a person is alone or not. Two labels: not socializing (NS), and socializing (S).
- Sedentary lifestyle pattern: it allows to determine if a person is far from a table or not, since the sedentary pattern is strongly related to being sat in front of a table. Two labels are used: no table (NT), and table (T).

	Environment	#Days	#Images
User 1	Working	18	16,453
User 2	Studying	18	14,326
User 3	Vacations	10	6,227
User 4	Working	9	8,291
Total		55	45,297

Table 1: LAP dataset specifications for the four users' images.

Table 2 illustrates the distribution of the egocentric images for each pattern and label. As can be observed, the dataset is noticeably imbalanced, especially with respect to the eating pattern with over 92% of the images labeled as NFR. In this sense, it should be noted that imbalanced datasets still remain a challenge for researchers.

	Id	Labels	%	#Images
Eating	0	NFR	92.47	41,883
	1	FRNE	3.32	1,503
	2	FRE	4.22	1,910
Total				45,297

	Id	Labels	%	#Images
Socializing	0	NS	61.75	27,971
	1	S	38.25	17,326
Total				45,297

	Id	Labels	%	#Images
Sedentary	0	NT	70.10	31,755
	1	T	29.90	13,542
Total				45,297

Table 2: Distribution of patterns and labels in the LAP dataset.

3.2 General framework

The problem to be solved is, given an input image, to categorize it into three different patterns: $Eating := \{0 = NFR, 1 = FRNE, 2 = FRE\}$, $Socializing := \{0 = NS, 1 = S\}$, and

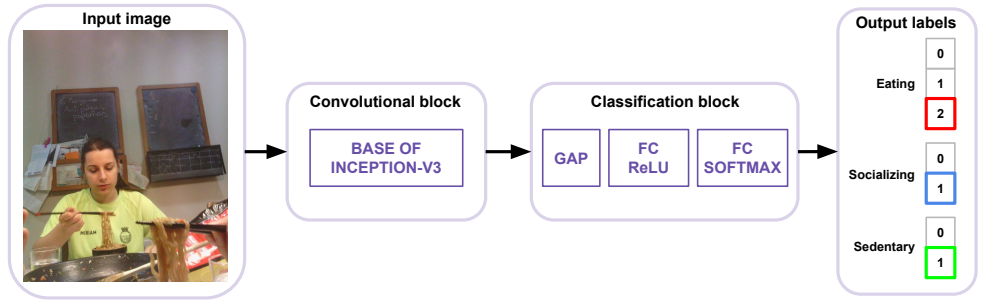


Figure 1: General framework made up the convolutional base of the Inception-V3 model, and a classification block composed of a Global Average Pooling (GAP), a Fully Connected (FC) layer with a Rectified Linear Unit (ReLU), and a FC layer with a Softmax.

$Sedentary := \{0 = NT, 1 = T\}$. Based on that, we propose the general framework illustrated in Figure 1, which is trained end-to-end for the classification of the three lifestyle patterns.

The first step to determine the habits of camera wearers is to extract a set of features from the egocentric input images. For this purpose, the input images are codified using a convolutional neural network; in particular, the convolutional base of the Inception-V3 model [20] with weights pre-trained on ImageNet [19], which are fine-tuned during the training phase. Note that the information obtained in this block is represented by 2048 feature maps (spatial resolution: 8×8), which are next used in the classification step.

Next stage is focused on classifying the feature maps previously extracted into the one target categories for each pattern. Based on the idea proposed by Caruana [10], which uses a different network for each task, we propose to use three different classifiers (one per pattern) that share the architecture. In this manner, the first classifier C^E predicts

$$\hat{y}^E \in Eating := \{0 = NFR, 1 = FRNE, 2 = FRE\},$$

the second classifier C^{SO} predicts

$$\hat{y}^{SO} \in Socializing := \{0 = NS, 1 = S\},$$

and the third classifier C^{SE} predicts

$$\hat{y}^{SE} \in Sedentary := \{0 = NT, 1 = T\}.$$

The architecture that defines these three classifiers is as follows: a Global Average Pooling (GAP) that transforms the 2048 feature maps in a vector composed of 2048 deep features, a Fully Connected (FC) layer with a Rectified Linear Unit (ReLU) [17] as activation function and 1024 units, and a FC layer with a Softmax to perform the final classification.

Bearing in mind the problem with imbalanced classes, described in Section 3.1, the cross entropy loss used in the classification step is combined with weights

$$w_i = \frac{M_j}{N_i} \text{ for } i \in j,$$

where N_i is the number of images in class $i \in \{NFR, FRNE, FRE, NS, S, NT, T\}$, and M_j is the maximum number of images in pattern $j \in \{Eating, Socializing, Sedentary\}$. For example, let $i = S$ with $N_i = 17326$, then $i \in j = Socializing$ and $M_j = \max\{27971(NS), 17326(S)\}$, thus obtaining the weight $w_i = \frac{27971}{17326} = 1.61$.

3.3 Lifelogging tool

With the main aim of giving value to the egocentric camera wearers, we have developed a web-based lifelogging tool to analyze their egocentric images based on the lifestyle patterns considered. This web application offers users the possibility to get visualizations that summarize their daily activity in an efficient way. The lifelogging tool is composed of three main functionalities, as illustrated in Figure 2, and provides: 1) statistics about the usage of the service, 2) a weekly visualization that informs the users about their lifestyle patterns, and 3) a daily visualization that allows the users traverse their photos taken in certain moments of the day. Note that the lifelogging tool can be used for other types of lifestyle patterns, and it is available for download from our Github³.

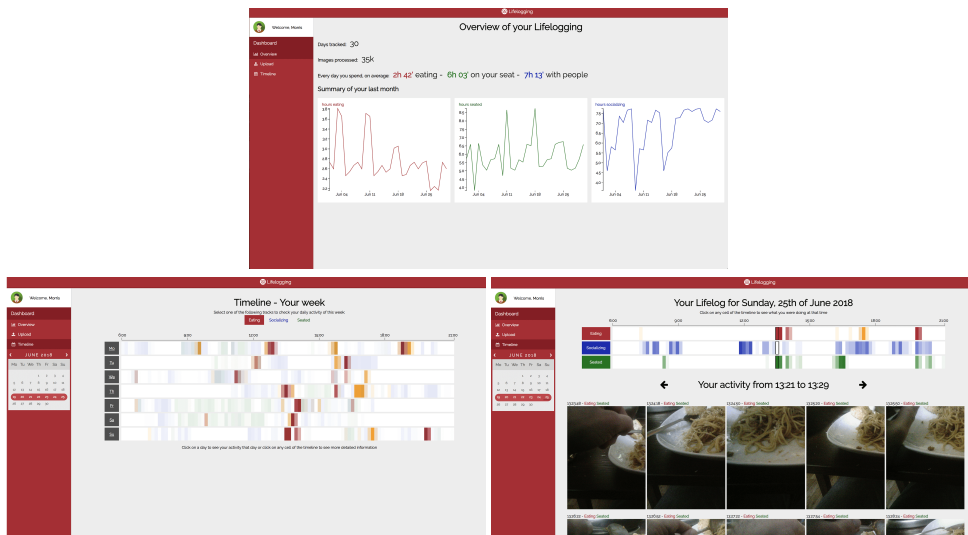


Figure 2: Different screenshots of the lifelogging tool that include, general information on the application usage (top), weekly summary with eating patterns highlighted (bottom, left), and daily summary (bottom, right). Best seen in electronic form.

4 Experimental results

In this section, we report the results obtained in some experiments carried out to test the performance of our approach. First, we describe the experimentation in detail, including other methods used to compare with our framework. Next, we present the results obtained and include some discussion. Finally, we analyze the results by means of visual explanations.

4.1 Experimentation

The performance of our approach was evaluated over the LAP dataset described in Section 3.1, using the original division into training, validation and test sets [14]. Additionally, the original images have been resized to $299 \times 299 \times 3$.

³<https://github.com/beareme/LAP-Lifelogging-Tool>

Our implementation of the general framework⁴ is on Keras [14], with TensorFlow as backend. Note that the proposed method uses the stochastic gradient descent optimizer, with a momentum of 0.9, and a learning rate of 0.001. With respect to the training process, we used a batch size of 16 and 100 epochs. Additionally, we applied some popular techniques of image data augmentation: flipping, adding Gaussian noise, and rotations with different degrees in the range $[-30, 30]$.

The method proposed in [14] was used as a baseline. Moreover, in order to analyze different approaches of supervised learning, we considered three machine learning algorithms: k-nearest neighbors (kNN), support vector machines (SVM), and gradient boosting machines (GBM). In order to apply them to the problem at hand, we considered two different options to be used as feature extractors: the Incremental-PCA (IPCA) [6], and the convolutional base of Inception-V3 [21] pre-trained on ImageNet.

Finally, the performance measures used to compare all the methods are the following ones: accuracy, the percentage of correctly classified images; and F1-score, the harmonic mean of precision and recall. We should point out the adequacy of the F1-score when working with imbalanced datasets.

4.2 Results

All the experimental results are shown in Table 3, which includes both accuracy and F1-score for all the methods stated in previous section. Note that the column Triad makes reference to the three patterns together; that is, a success in this case means that the three predicted labels are correct. This approach is the one considered in [14], which solves the problem as a 12-class classification. Additionally, we evaluated the performance of our framework by considering the three patterns individually.

	Triad		Eating		Socializing		Sedentary	
	Acc (%)	F1	Acc (%)	F1	Acc (%)	F1	Acc (%)	F1
Previous approach [14]	60.53	0.64	-	-	-	-	-	-
IPCA + kNN	52.36	0.41	92.60	0.89	65.45	0.56	73.48	0.65
IPCA + SVM	42.73	0.38	88.51	0.88	59.70	0.57	68.10	0.63
IPCA + GBM	48.69	0.37	92.61	0.89	63.73	0.55	71.99	0.64
Inception-V3 + kNN	54.83	0.53	94.07	0.93	70.63	0.70	79.58	0.79
Inception-V3 + SVM	58.42	0.59	94.01	0.94	74.06	0.83	81.95	0.83
Inception-V3 + GBM	54.24	0.51	94.23	0.93	70.05	0.69	79.45	0.78
Our approach	69.91	0.72	85.55	0.89	86.33	0.86	91.68	0.92

Table 3: Experimental results for the different methods considered, in terms of accuracy and F1-score. Triad column refers to the three patterns, whilst the other three columns provide the metrics for each individual pattern. Best results are in bold.

Regarding the triad column, it can be observed that our approach provides the most competitive results in terms of both accuracy and F1-score, with an improvement over 9% and 0.08, respectively, compared to the previous approach [14], ranked in the second position. When analyzing the individual lifestyle patterns, we can see that the performance metrics obtained are much higher, as expected due to the simplification of the problem. The approaches done through IPCA with traditional classifiers perform poorly than the rest of methods compared, mainly because IPCA provides a reduction of the feature space against the abstract representations computed by the deep learning approaches, including Inception-V3 with traditional classifiers and our framework.

⁴<https://github.com/lauraportell/LAP-LifestylePatterns-Classification>

Concerning the deep learning approaches, all of them achieve very competitive results for the three different patterns due to the high power of representation provided by CNNs. Worth of mention is that our approach outperforms the other methods in two out of the three lifestyle patterns, with a minimum improvement in terms of accuracy around 10%. With respect to the eating pattern, the results obtained by the Inception-V3 with the traditional classifiers is slightly better; in particular, the performance measures obtained are very similar regardless the algorithm. This is mainly motivated by the fact that the eating pattern is defined by three different classes, and the architecture considered in our approach for the classification stage has one single hidden layer.

Summarizing, our approach provides very competitive results when analyzing the three patterns individually and outperforms any other method in terms of the complete triad, thus demonstrating its adequacy to the problem at hand by providing a good trade-off between all the lifestyle patterns.

Finally, Figure 3 depicts the confusion matrices for each individual pattern, obtained when evaluating our approach. Note that these confusion matrices refer to the different classes using their identifiers (see Id column in Table 2). As can be observed, both socializing and sedentary lifestyle patterns perform satisfactorily, with a balance between both classes. Regarding the eating pattern, the main missclassifications are in the FRNE class (Id 1). Almost one fourth of the images of this class were classified as belonging to the FRE class (Id 2), probably due to the subtle visual difference between both classes.

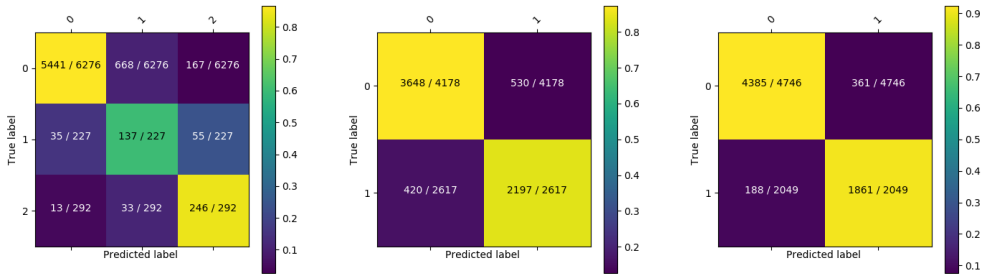


Figure 3: Confusion matrices obtained when evaluating the proposed method, from left to right: eating, socializing, and sedentary lifestyle patterns.

4.3 Visual explanations

The General Data Protection Regulation (GDPR) [23] of the European Union demands explanations when algorithms take decisions that affect people, aiming at transparent systems. As our approach is built for final users, it must provide confidence when predicting their lifestyle patterns from egocentric images. For this reason, our system includes visual explanations for model decisions based on Grad-CAM [20]. First, this method computes the importance α_k^c of a kernel k for a target class y^c by taking the derivative from the target class y^c with respect to the activation produced by this kernel A^k . Second, it sums the positive part of the weighted kernels activation to create the final heat-map. Third, it augments the size of this map up to the original size of the image and add them all.

Figure 4 shows three representative examples of how Grad-CAM provides visual explanations regarding the predictions of our model. As can be observed, users can see that

the predictions made by the system have rough visual arguments, thus increasing their confidence of the lifelogging tool.

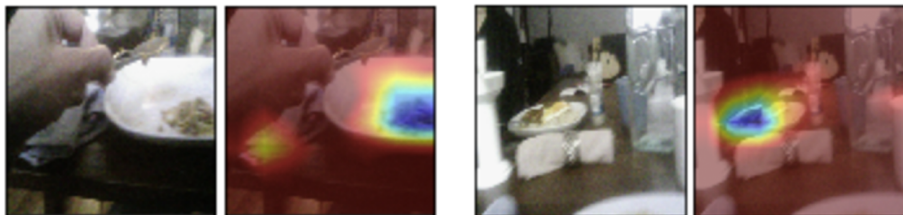


Figure 4: Grad-CAM highlights the image features selected by the model to make predictions (color scale: from blue to red, where blue represents the most relevant areas). The lifestyle patterns of these images are: food related eating (left), and food related non eating (right).

From a research perspective, Grad-CAM allows us to extract valuable information on how CNNs make predictions. Analyzing Figure 4 in detail, we can see that our model is able to learn that a close dish with food inside is highly correlated with the food related eating (FRE) class, whilst a similar situation, e.g. another person is eating but not the camera wearer, corresponds to the food related non eating (FRNE) class.

5 Conclusions

Egocentric lifelogging has been demonstrated to be a useful tool to track daily activities. The objective description of lifestyle patterns of wearable camera users represent useful information for health purposes, comparable to the data acquired with other wearable devices such as wristbands.

In this research, we proposed a general framework based on state-of-art convolutional neural networks that allow to characterize lifestyle patterns in three different groups: eating, socializing and sedentary lifestyle. The experimental results demonstrated the adequacy of our approach compared to other baseline methods, including previous research that solved the problem as multi-class classification. Furthermore, we present a lifelogging tool that allows wearable camera users to upload their pictures and get a detailed analysis of their daily activities, in terms of the three lifestyle patterns. Finally, the proposed method includes visual explanations to satisfy the European GDPR demands.

Our future lines of research plan to introduce time dependency into our model, using recurrent neural networks such as long short-term memory (LSTM). In addition to this, we would like to explore how visual explanations can be used as part of the training process to improve the performance of the model.

Acknowledgments

This research was supported by grants TIN2015-65069-C2-2-R and TIN2017-89244-R from the MINECO (Spanish Ministry of Economy and Competitiveness), 2017SGR-856 (MACDA) from AGAUR (Generalitat de Catalunya), and IDI-2018-000176 from the Principado de Asturias Regional Government, partially supported with ERDF funds. Pedro Herruzo acknowledges the support of the Secretariat of Universities and Research of the Department

of Economy and Knowledge of the Generalitat de Catalunya, under the Industrial Doctorate Grant 2017 DI 52.

The authors gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used in this research.

References

- [1] Abrar H Abdulnabi, Gang Wang, Jiwen Lu, and Kui Jia. Multi-task CNN model for attribute prediction. *IEEE Transactions on Multimedia*, 17(11):1949–1959, 2015.
- [2] Maedeh Aghaei, Mariella Dimiccoli, and Petia Radeva. With whom do I interact? Detecting social interactions in egocentric photo-streams. In *23rd International Conference on Pattern Recognition*, pages 2959–2964, 2016.
- [3] Eduardo Aguilar, Marc Bolaños, and Petia Radeva. Food recognition using fusion of classifiers based on cnns. In *International Conference on Image Analysis and Processing*, pages 213–224, 2017.
- [4] Stefano Alletto, Giuseppe Serra, Simone Calderara, Francesco Solera, and Rita Cucchiara. From ego to nos-vision: Detecting social relationships in first-person views. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 580–585, 2014.
- [5] Akshay Balsubramani, Sanjoy Dasgupta, and Yoav Freund. The fast convergence of incremental PCA. In *Advances in Neural Information Processing Systems*, pages 3174–3182, 2013.
- [6] Sophia Bano, Tamas Suveges, Jianguo Zhang, and Stephen J Mckenna. Multimodal egocentric analysis of focused interactions. *IEEE Access*, 6:37493–37505, 2018.
- [7] Abdelkareem Bedri, Richard Li, Malcolm Haynes, Raj Prateek Kosaraju, Ishaan Grover, Temiloluwa Prioleau, Min Yan Beh, Mayank Goel, Thad Starner, and Gregory Abowd. EarBit: using wearable sensors to detect eating episodes in unconstrained environments. *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):37, 2017.
- [8] Marc Bolanos, Mariella Dimiccoli, and Petia Radeva. Toward storytelling from visual lifelogging: An overview. *IEEE Transactions on Human-Machine Systems*, 47(1):77–90, 2017.
- [9] Alejandro Cartas, Juan Marín, Petia Radeva, and Mariella Dimiccoli. Recognizing activities of daily living from egocentric images. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 87–95, 2017.
- [10] Rich Caruana. Multitask learning. *Machine learning*, 28(1):41–75, 1997.
- [11] François Chollet et al. Keras: The Python Deep Learning library, 2015. URL <https://keras.io/>.
- [12] Kaya de Barbaro. Automated sensing of daily activity: A new lens into development. *Developmental psychobiology*, 2019.

- [13] Aiden R Doherty, Chris JA Moulin, and Alan F Smeaton. Automatically assisting human memory: A SenseCam browser. *Memory*, 19(7):785–795, 2011.
- [14] Pedro Herruzo, Laura Portell, Alberto Soto, and Beatriz Remeseiro. Analyzing first-person stories based on socializing, eating and sedentary patterns. In *New Trends in Image Analysis and Processing - ICIAP 2017*, volume 10590, pages 109–119, 2017.
- [15] Yedid Hoshen and Shmuel Peleg. An egocentric look at video photographer identity. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4284–4292, 2016.
- [16] Sheema Khan, Kashif Ahmad, Tahir Ahmad, and Nasir Ahmad. Food items detection and recognition via multiple deep models. *Journal of Electronic Imaging*, 28(1): 013020, 2019.
- [17] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *27th International Conference on Machine Learning*, pages 807–814, 2010.
- [18] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 779–788, 2016.
- [19] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Fei-Fei Li. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.
- [20] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *IEEE International Conference on Computer Vision*, pages 618–626, 2017.
- [21] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016.
- [22] Estefania Talavera, Nicola Strisciuglio, Nicolai Petkov, and Petia Radeva. Sentiment Recognition in Egocentric Photostreams. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 471–479, 2017.
- [23] Paul Voigt and Axel von dem Bussche. *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Springer Publishing Company, Incorporated, 1st edition, 2017.
- [24] Qing Yu, Masashi Anzawa, Sosuke Amano, Makoto Ogawa, and Kiyoharu Aizawa. Food image recognition by personalized classifier. In *IEEE International Conference on Image Processing*, pages 171–175, 2018.