Proceedings Paper:
Healy, G., Le, T.-K., Joho, H. et al. (2 more authors) (2020) Overview of NTCIR-15 MART. In: Proceedings of the 15th NTCIR Conference on Evaluation of Information Access Technologies. NTCIR-15 - Evaluation of Information Access Technologies, 08-11 Dec 2020, Online conference. National Institute of Informatics , pp. 299-303.

Overview of NTCIR-15 MART

Graham Healy
Dublin City University
Dublin, Ireland
graham.healy@dcu.ie

Tu-Khiem Le
The Insight Centre for Data Analytics,
Dublin City University
Dublin, Ireland
tukhiem.le4@mail.dcu.ie

Hideo Joho
Faculty of Library, Information and
Media Science, University of Tsukuba
Tsukuba, Japan
hideo@slis.tsukuba.ac.jp

Frank Hopfgartner
Information School, University of
Sheffield
Sheffield, United Kingdom
f.hopfgartner@sheffield.ac.uk

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

ABSTRACT

MART (Micro-activity Retrieval Task) was a NTCIR-15 collaborative benchmarking pilot task. The NTCIR-15 MART pilot aimed to motivate the development of first generation techniques for high-precision micro-activity detection and retrieval, to support the identification and retrieval of activities that occur over short time-scales such as minutes, rather than the long-duration event segmentation tasks of the past work. Participating researchers developed and benchmarked approaches to retrieve micro-activities from rich time-aligned multi-modal sensor data. Groups were ranked in decreasing order of micro-activity retrieval accuracy using mAP (mean Average Precision). The dataset used for the task consisted of a detailed lifelog of activities gathered using a controlled protocol of real-world activities (e.g. using a computer, eating, daydreaming, etc). The data included a lifelog camera data stream, biosignal activity (EOG, HR), and computer interactions (mouse movements, screenshots, etc). This task presented a novel set of challenging micro-activity based topics.

KEYWORDS

human activity detection, micro-activities, multi-modal sensing

1 INTRODUCTION

Extracting insightful or actionable information from personal sensor data (lifelogs) holds promise to improve an individual's productivity, health, and enable new application domains that use wearable sensor data [4, 5]. A core component of all these efforts is the detection of human activities, which is a pre-requisite for many lifelog applications. While activity segmentation approaches have been widely explored on lifelog data [6], the approaches taken have focused on identifying broad human activities of daily living (e.g. walking, eating, resting). Less attention has been given to micro-activities requiring high temporal precision, such as semantic workplace activities (e.g. pondering a problem, or having a short watercooler conversation), or information access/creation activities (e.g. writing an email, searching on the WWW). The MART pilot task aimed to motivate the development of a first generation of techniques for high-precision micro-activity detection and retrieval of micro-activities of daily living. This would support the identification and retrieval of activities that occur over short time-scales, such as minutes, rather than the long-duration event segmentation tasks of the past work.

Segmentation of personal sensor data (lifelogs) into indexable units is a key component of any functional lifelog retrieval system [7]. Any retrieval or activity-support tools for lifelog data need an accurate segmentation and retrieval model as a necessary underlying component for many use-cases. Research in this space heretofore has focused on temporal segmentation of macro rather than micro-activities (or events) as the related datasets are typically considered to be easier to collect and label. Existing experimental paradigms group activities together into large retrievable units in a process called event segmentation, which act as a blunt model for retrieval in that the detected real-world events are unlikely to be useful for many information retrieval challenges. It is our conjecture that Activities of Daily Living (at the macro-level) are more aligned with research into pervasive computing [18, 19], rather than information retrieval. Micro-activities, however, are more aligned with conventional information retrieval tasks and the proposed use-cases of lifelogs [7]. The challenge of identifying and retrieving micro-activities from multimodal data streams has heretofore lacked a rigorous investigative focus, particularly when the data is multi-modal involving bio-signals and other passively captured media. Notably however, a number of research efforts have been undertaken investigating a variety of sensor sources in isolation for activity detection [2, 12, 21]. It is our prediction that retrieval of micro-activities of daily life will be a key underlying mechanism supporting the use of lifelogs/worklogs for workplace/productivity enhancement, health-related applications and for personal productivity tools in general in the future [1, 14, 16].

In this paper, we describe a new pilot task, that released a novel multi-modal micro-activity test collection for use by the IR community. Without running such a task, it is unlikely that many research groups would focus on this activity because such rich multimodal data is challenging to gather, understand and work with. Hence, we proposed this pilot task to motivate the exploration of new approaches supporting information access to micro-activities. Participating teams developed and benchmarked approaches to retrieve micro-activities from the rich time-aligned multi-modal sensor dataset employed for the task. This dataset was collected from individuals that followed a pre-defined protocol of real-world activities (e.g. using a computer, solving a problem, drinking, cleaning, etc) in a controlled environment. This allowed for a wide range

of consistent behaviours across the volunteers to be induced (e.g. reading, talking, etc). This was particularly important, as previous efforts to use multi-modal sensor data for lifelog data have relied upon the retroactive labelling of the collected data, which introduces inherent errors and problems with assigning labels. By design, this task did not require retrospective labelling, and in turn made each volunteer consistent in the data that was generated. Similarly, this enabled a large number of consistent induced micro-activities to be captured.

2 DATASET COLLECTION

The datasets used in NTCIR-15 MART were captured by instrumenting volunteers with a suite of multi-modal sensors alongside capturing computer interactions (via Loggerman software[1]) as they completed 20 pre-defined activities. The details of the protocol, the sensor signals captured during each experiment and the released data are detailed in this section.

2.1 Sensors Used

Each experimental volunteer (N=7) was equipped with a variety of sensors for data recording that included: (A) a lifelog camera capturing first-person perspective images at a rate of 2-3 images per minute using an Autographer wearable digital camera (worn on a lanyard), (B) EOG (Electrooculogram) capturing electrical signals associated with vertical (V-EOG) and horizontal (H-EOG) eye movements (direction and time) via a NeuroElectrics bluetooth amplifier, (C) heart rate via a pulse oximeter placed on the ear lobule, (D) tri-axial accelerometer readings from 3 locations (left forearm, right forearm and head) via a LSM9DS1 9DoF intertial measurement unit[2], and (E) detailed computer interaction using the Loggerman software. All sensor data was captured such that it could be accurately co-registered across time by using a common data recording computer. The forearm accelerometer sensors were attached to volunteers using customised Velcro straps. The Raspberry Pi and battery pack was either kept in the volunteer's pocket or attached to their belt. Loose cables were secured using Velcro. Volunteers reported that this setup did not restrict their movement and they were in fact able to move freely as they would to complete the activities.

2.2 Activity Structure

Each data collection session with a volunteer lasted approximately three hours, where the volunteers completed a range of predefined micro-activities (three repetitions of 20 different micro-activities) in a controlled environment, as per a pre-defined data gathering protocol, with variation in the sequencing of the different micro-activities. This allowed for a wide range of consistent behaviours across volunteers to be induced (e.g. reading, talking, etc). Each activity was performed continuously for ninety seconds. The experimenter in each case guided volunteers on when to start and when to stop an activity. In total data 420 activities (across the seven volunteers) were recorded.

[1]http://loggerman.org/

[2]the heart rate and forearm accelerometer sensors were connected to a Raspberry Pi 4 powered by a portable battery pack where readings were transmitted in real-time over a wireless network to the data recording computer to ensure proper time synchronisation.

The 20 activities completed by each volunteer in the experiment were:

- Act01: Writing/replying to an email.
- Act02: Reading text on screen (news websites and articles were not used).
- Act03: Editing a presentation on the computer.
- Act04: Zoning out while staring at a point in the room.
- Act05: Finance management (specifically using a calculator to total numbers present on paper or screen).
- Act06: A physical precision task that required both hands e.g. manipulating a circuit board.
- Act07: Document organisation where the subject needed to organise A4 sheets into a particular order e.g. by page number.
- Act08: Reading text on paper (written or printed).
- Act09: Counting/arranging physical currency (money).
- Act10: Writing with pen on paper e.g. on a blank sheet of paper or writing notes with a pen on printed text.
- Act11: Watching a YouTube video.
- Act12: Browsing (any) news website.
- Act13: Having a conversation with another person in the room (they could be directly facing this person or they might be out of view in the room).
- Act14: Making a telephone call (holding a cellular phone with either hand to their ear).
- Act15: Drinking/eating (eating or drinking anything).
- Act16: While seated, the subject closed their eyes and refrained from any movement for 90 seconds.
- Act17: Cleaning e.g. with a broom/hoover/cloth.
- Act18: Physical exercise. In this activity the subject was instructed to repeatedly sit up-and-down from their chair.
- Act19: Hand-eye coordination activity. In this task the subject was instructed to use both hands to 'play' with a tennis ball e.g. passing/throwing it between their hands.
- Act20: Walking/pacing around. In this task, the subject was instructed to pace the room continuously.

2.3 Released Data and Resources

The released dataset consisted of two components:

(1) Training set: 66% (280 activities) of the dataset with a set of training topic/activity descriptions.
(2) Test set: containing the remaining 33% (140 activities) and test topic/activity descriptions. Ground truth labels were withheld and used only by the organisers when evaluating the submissions. One sample of each activity (20) for each volunteer (7) was used for the test set.

Datasets were made available in a number of formats and configurations to promote accessibility and use i.e. both aggregated pre-processed data over the 90 second period for each activity and raw sensor data was provided. A baseline system was also developed and freely shared with participating researchers in order to provide a starting point from which they can build their own system. These are both detailed in the Appendix.

Table 1: mAP (mean Average Precision) evaluation results for the highest scoring submitted run per participating team. "mAP (Best)" shows the highest mAP achieved in the formal submission period by each team. *indicates a unique RunID was not provided by the team for the submission.

Team Name	mAP (Best)	RunID	Total runs
THUIR[11]	0.950	1	7
DCU[10]	0.901	9	10
NLP301[3]	0.851	*	10
UHAIK[20]	0.717	1,5	8
TMU19[13]	0.465	1	9

3 THE MART EVALUATION TASK

Groups participating in MART used the training data set to develop their automatic or interactive system approaches. A withheld test set was used in order to evaluate submissions. In total 7 groups signed up to MART but only 5 groups submitted valid runs: THUIR (Information Retrieval Group, Department of Computer Science and Technology, Tsinghua University, China), DCU (Dublin City University, Ireland), NLP301 (Department of Computer Science and Information Engineering, National Taiwan University, Taiwan), UHAIK (KDDI Research, University of Hyogo, Japan) and TMU19 (TMU-NLP, Taipei Medical University, Taiwan).

3.1 Evaluation Methodology

There were 140 activities in the test set (across all experimental volunteers), thus each query result for submission provided a ranked list of 140 activity id codes for that query. Since there were 20 queries in total (one for each activity), the submission provided for each run by a participating team had 2,801 rows (first row was for team id and the submission password). Submissions were made using a HTTP GET/POST to an online evaluation system. AP (average precision) was computed on the ranked list submitted for each query (in the order they were ranked in the submission), and the mean of the APs was calculated across all queries (e.g. act01, act02, act03). If a group made a submission with a ranked list for a query that was less than 140 (say only for a 'top 7' for each activity), the ranked list was extended to pad the difference (using the remaining activity-prediction labels in a randomized order).

4 EVALUATION RESULTS

A wide variety of approaches were investigated by the 5 active participating teams in MART. Notably, none the best ranked approaches used an interactive system, and instead relied on automated (machine-learning) methods to complete the task.

In Table 1, a ranked list is shown of the mAP (mean Average Precision) evaluation results for the highest scoring submitted runs per participating team. The first-ranked team (THUIR)[11] achieved a mAP of .95 on the withheld test set where the approach in their best submitted run used a combination of correlation-based feature selection and a rule-based GBDT (Gradient Boosting Decision Tree) classifier. The second-ranked team (DCU)[10] achieved a mAP of .901 on the withheld test set where the approach for their best submitted run used an Image-Tabular Pair-wise Similarity Model (IT-PS). The third-ranked team (NLP301)[3] achieved a mAP of .851 on the withheld test set, where the approach in their best submitted run used a supervised-based model that incorporated visual and biosignal features, along with a GRU network to capture slight variations in user's movements in the time-series data, and RoI (Region of Interest) features to detect computer activities. The fourth-ranked team (UHAIK)[20] achieved a mAP of .717, where the approach in their best submitted run used Super-LCC for feature selection and a SVM (Support Vector Machine). The fifth-ranked team (TMU19)[13] achieved a mAP of .465, where the approach in their best submitted run used a combination of image feature extraction and a BiLSTM (Bidirectional Long Short-Term Memory). Further details on the approaches explored by each team are available in the task participant papers [3, 10, 11, 13, 20].

Table 2: Average Precision evaluation results for the highest scoring submitted runs per participating team (per activity/topic). Note: *THU=THUIR, TMU=TMU19, UHA=UHAIK, NLPX=NLP301 and DCU=DCU.*

Act ID	THU	DCU	NLP	UHA	TMU	Mean
Act01	1.000	0.825	1.000	0.909	0.494	0.846
Act02	0.845	0.581	0.802	0.705	0.130	0.613
Act03	1.000	1.000	0.921	0.770	0.757	0.890
Act04	0.816	0.606	0.632	0.675	0.601	0.666
Act05	1.000	1.000	0.920	0.706	0.225	0.770
Act06	1.000	1.000	1.000	0.765	0.456	0.844
Act07	0.982	1.000	0.897	0.426	0.848	0.831
Act08	0.810	1.000	0.913	0.623	0.367	0.743
Act09	0.933	1.000	1.000	0.706	0.475	0.823
Act10	1.000	1.000	1.000	0.820	0.912	0.946
Act11	0.982	1.000	0.871	0.620	0.098	0.714
Act12	0.884	0.877	1.000	0.403	0.036	0.640
Act13	1.000	1.000	1.000	0.732	0.445	0.835
Act14	0.962	0.660	0.573	0.413	0.501	0.622
Act15	1.000	0.837	0.675	0.957	0.934	0.881
Act16	0.812	0.638	0.336	0.522	0.211	0.504
Act17	0.982	1.000	0.848	0.824	0.913	0.913
Act18	1.000	1.000	0.868	0.847	0.323	0.808
Act19	1.000	1.000	0.908	0.968	0.460	0.867
Act20	1.000	1.000	0.846	0.948	0.106	0.780
Mean	0.950	0.901	0.851	0.717	0.465	0.777

5 DISCUSSION

In Table 2, we show the average precision scores per activity/topic for the best submission for each team. Taking the mean of the average precision scores per activity (for each team's top performing submission), it can be seen that some activities were more difficult than others to correctly rank. For example, Act16 (closed eyes while seated) had the worst performance, as this was a difficult activity to correctly rank and distinguish from other activities. In Figure 1, we can see Act16 ('closed eyes and sitting still') was often confused with Act02 ('reading text on screen'), Act04 ('zoning out') and Act08

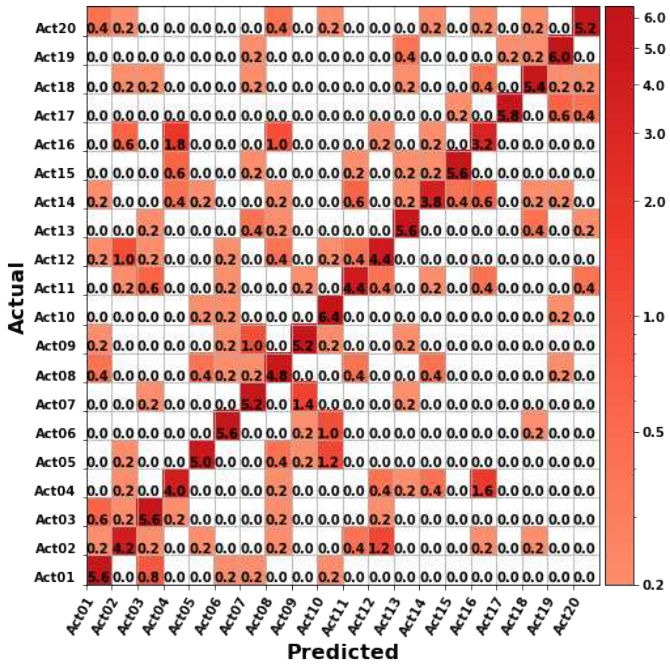

Figure 1: Averaged confusion matrix for best approach from each of the 5 participating teams. The first 7 predictions in the ranked list submitted for each activity were used to generate the plot. A perfect accuracy for all submissions for activities across all teams would appear as a 7 along the diagonal.

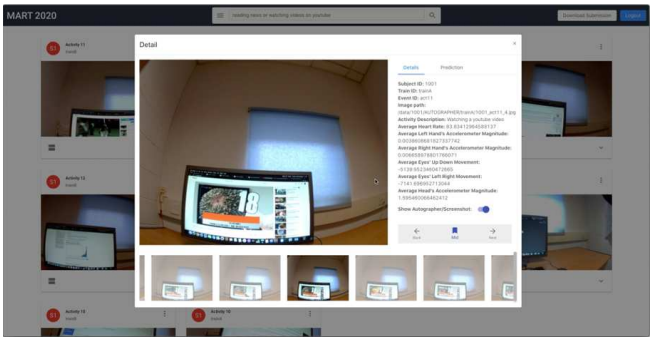

Figure 2: The baseline system which offers free-text search, displays activity's autographer/screenshot images with corresponding metadata and allows interactive ordering to arrange activity type prediction.

('reading text on paper'). Insights like these are important when designing future tasks in order to guide the focus of participating researcher's efforts by posing difficult to classify activity/topics that will encourage the development of techniques and models that can successfully leverage multi-modal signal sources in tandem for activity classification.

Another important observation from Table 2, is that while the best submitted run per participating team may have performed differently in terms of average precision across the activity/topics, for sixteen of the twenty of these, at least one submitted run scored perfect performance in terms of AP (i.e. AP=1). Only four activities/topics had a max performance (from the best submitted runs per team) with an average precision less than 1, namely: Act02 ('reading text on screen'), Act04 ('zoning out'), Act14 ('telephone call') and Act16 ('closed eyes and sitting still'). This indicates there is potential to combine the various approaches taken by different teams into a unified solution that could achieve a greater overall mAP for MART. Recalculating the mAP using the best average precision from the best submitted runs from teams for activities/topics, this new overall mAP score would be .972.

6 CONCLUSION

Five teams submitted runs along with a paper to NTCIR-15 MART although more teams signed up to participate. It is noteworthy that none of the best submitted runs from a team used an interactive system, and instead relied on automated techniques to automatically perform the task using labelled example training data. Of the five participating teams, the first-ranked team was THUIR [11] who achieved a mAP of .95 on the withheld test set.

Given the success of MART, future versions of this task will focus on developing a new larger dataset that incorporates more complex and diverse human activities. In particular, further sensor sources will be incorporated including EEG (Electroencephalography) as has been used in the prior NTCIR-13 task NAILS [9], along with data from other camera streams that will capture the environment [15] and the participants' facial expressions as they interact with the computer [17]. Future versions of MART will also incorporate an event segmentation task, where the time index and length of activities are not already pre-identified, in combination with retrieving activities of different time spans. Importantly, future versions of MART will incorporate a subject-independent retrieval task where unlike this pilot task, task participants will be required to build systems that can retrieve activities where training data may not be available for particular subjects.

In this paper we have described the creation of the MART dataset, including a description of the motivation and reasoning behind its construction. This was an initial pilot task for NTCIR-15, with a focus on topic/activity detection from rich multi-modal data.

7 APPENDIX

7.1 Data Pre-processing

In order to lower the barrier for participation in MART, we provided additional pre-processed metadata that included features extracted from each sensor source used for an activity. These features were by no means exhaustive and instead were intended to facilitate participation. Since each sensor source captured discrete time-series values for the time period of the activity, a set of summary statistics were calculated for each activity's time-series including: (A) the minimum value, (B) the maximum value, (C) the median value, (D) the mean value and (E) the standard deviation. As participating teams were provided with the raw signal data (e.g. sensor and images) for each activity, they were able to generate additional

summary statistics and features as they needed. Many participating teams reported that they used these pre-processed features.

Summary statistics were calculated on the pre-processed values for each sensor source:
Visual - Autographer - The ResNet101[8] pre-trained deep Convolutional Neural Network was used to extract predictions for 1,000 classes (ImageNet) on the Autographer images captured during the activity. Pre-processed features (in the form of summary statistics) were generated on the softmax values for each class over the time period of the activity.
Heart Rate - Instantaneous heart rate values were extracted (1 Hz) from the pulse oximeter signal using an approach based on signal autocorrelation with a sliding window (4 seconds). The heart rate for the window was calculated via the peak autocorrelative lag.
Acceleromator - X,Y and Z acceleromator readings from each sensor were processed to extract the time-series magnitude values. A gravity constant (of 1) was subtracted from the magnitude values.
Electroocugraphy - Raw HEOG and VEOG signals were bandpassed filtered between 2 Hz and 20 Hz in order to remove slow-drift type artefacts in the signals.
LoggerMan - From the mouse movement data the pixel distance travelled (via Euclidean distance), the time lag between mouse movements and the instantaneous velocities were calculated, and summary statistics were generated for each of these. Preprocessing or feature extraction was not carried out on the LoggerMan screenshots.
Preprocessed features were provided in a csv file for each activity X volunteer combination, which could be easily loaded using a data manipulation tool such as python-pandas.

7.2 Baseline System

A baseline system was also provided to participating researchers, which was intended to support data exploration while providing the basis for a basic interactive search engine upon which task participants could build their approach. The system comprised a user interface and an API server which indexed the dataset based on the semantic metadata (preprocessed features) as described in section 7.1.

Task participants could use the baseline to easily navigate through the dataset by executing a text query in the baseline system. By doing this, one can investigate the dataset activity-by-activity and recognise the differences between activities, as a way to help to develop insights and design suitable approaches for the task. The system presents the activities in blocks, each showing images from the Autographer and LoggerMan screenshots, and corresponding metadata, as shown in Figure 2.

The source code for the baseline system was made available to the participants, and in particular to support the development of interactive retrieval system approaches. It was intended that a team would able to leverage the API server to return customised results to the interface, where the interface already had a built-in interactive ranking functionality allowing a user to easily adjust their prediction order. The baseline system also had functionality to generate a submission file that aggregated the prediction results of all activities in test dataset for run submissions for MART.

8 ACKNOWLEDGMENTS

This work is funded as part of Dublin City University's Research Committee and the Insight Centre for Data Analytics (which is supported by Science Foundation Ireland under Grant Number SFI/12/RC/2289_2).

REFERENCES

[1] Muhammad Bilal Amin, Oresti Banos, Wajahat Ali Khan, Hafiz Syed Muhammad Bilal, Jinhyuk Gong, Dinh-Mao Bui, Soung Ho Cho, Shujaat Hussain, Taqdir Ali, Usman Akhtar, et al. 2016. On curating multimodal sensory data for health and wellness platforms. *Sensors* 16, 7 (2016), 980.
[2] Ferhat Attal, Samer Mohammed, Mariam Dedabrishvili, Faicel Chamroukhi, Latifa Oukhellou, and Yacine Amirat. 2015. Physical human activity recognition using wearable sensors. *Sensors* 15, 12 (2015), 31314–31338.
[3] Tai-Te Chu, Yi-Ting Liu, Chia-Chung Chang, An-Zi Yen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2020. NLP301 at the NTCIR-15 Micro-activity Retrieval Task: Incorporating Region of Interest Features into Supervised Encoder. In *Proceedings of the NTCIR-15 Conference.*
[4] Maria Cornacchia, Koray Ozcan, Yu Zheng, and Senem Velipasalar. 2016. A survey on activity detection and classification using wearable sensors. *IEEE Sensors Journal* 17, 2 (2016), 386–403.
[5] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, Rami Albatal, Graham Healy, and Duc-Tien Dang Nguyen. 2020. Experiments in Lifelog Organisation and Retrieval at NTCIR. In *Evaluating Information Retrieval and Access Tasks.* Springer, 187–203.
[6] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, Van-Tu Ninh, Tu-Khiem Le, Rami Albatal, Duc-Tien Dang-Nguyen, and Graham Healy. 2019. Advances in lifelog data organisation and retrieval at the ntcir-14 lifelog-3 task. In *NII Conference on Testbeds and Community for Information Access Research.* Springer, 16–28.
[7] Cathal Gurrin, Alan F Smeaton, and Aiden R Doherty. 2014. Lifelogging: Personal big data. *Foundations and trends in information retrieval* 8, 1 (2014), 1–125.
[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 770–778.
[9] Graham Healy, Tomás E Ward, Cathal Gurrin, and Alan F Smeaton. 2017. Overview of NTCIR-13 NAILS task. (2017).
[10] Tu-Khiem Le, Manh-Duy Nguyen, Ly-Duyen Tran, Van-Tu Ninh, Cathal Gurrin, and Graham Healy. 2020. DCU team at the NTCIR-15 Micro-activity Retrieval Task. In *Proceedings of the NTCIR-15 Conference.*
[11] Jiayu Li, Ziyi Ye, Weizhi Ma, Min Zhang, Yiqun Liu, and Shaoping Ma. 2020. THUIR at the NTCIR-15 Micro-activity Retrieval Task. In *Proceedings of the NTCIR-15 Conference.*
[12] Haojie Ma, Wenzhong Li, Xiao Zhang, Songcheng Gao, and Sanglu Lu. 2019. AttnSense: Multi-level Attention Mechanism For Multimodal Human Activity Recognition.. In *IJCAI.* 3109–3115.
[13] Duy-Duc Le Nguyen, Yu-Chi Lang, and Yung-Chun Chang. 2020. TMU19 at the NTCIR-15 Micro-activity Retrieval Task. In *Proceedings of the NTCIR-15 Conference.*
[14] Henry Friday Nweke, Ying Wah Teh, Ghulam Mujtaba, and Mohammed Ali Al-Garadi. 2019. Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions. *Information Fusion* 46 (2019), 147–170.
[15] Suneth Ranasinghe, Fadi Al Machot, and Heinrich C Mayr. 2016. A review on applications of activity recognition systems with regard to performance and evaluation. *International Journal of Distributed Sensor Networks* 12, 8 (2016).
[16] Mark C Schall Jr, Richard F Sesek, and Lora A Cavuoto. 2018. Barriers to the adoption of wearable sensors in the workplace: A survey of occupational safety and health professionals. *Human factors* 60, 3 (2018), 351–362.
[17] Sabrina Stöckli, Michael Schulte-Mecklenbeck, Stefan Borer, and Andrea C Samson. 2018. Facial expression analysis with AFFDEX and FACET: A validation study. *Behavior research methods* 50, 4 (2018), 1446–1460.
[18] Emmanuel Munguia Tapia, Stephen S Intille, and Kent Larson. 2004. Activity recognition in the home using simple and ubiquitous sensors. In *International conference on pervasive computing.* Springer, 158–175.
[19] Jiaxuan Wu, Yunfei Feng, and Peng Sun. 2018. Sensor fusion for recognition of activities of daily living. *Sensors* 18, 11 (2018), 4029.
[20] Takuma Yoshimura, Pham Huulong, Ryota Mibayashi, Rui Kimura, and Hiroaki Ohshima. 2020. Behavioral Classification Using Feature Selection in the Micro Activity Retrieval Task. In *Proceedings of the NTCIR-15 Conference.*
[21] Mi Zhang and Alexander A Sawchuk. 2011. A feature selection-based framework for human activity recognition using wearable multimodal sensors.. In *BodyNets.* 92–98.