SAMM Long Videos: A Spontaneous Facial Micro- and Macro-Expressions Dataset

Chuin Hong Yap, Connah Kendrick and Moi Hoon Yap
Department of Computing and Mathematics, Manchester Metropolitan University, Manchester, M1 5GD, UK

***Abstract*— With the growth of popularity of facial micro-expressions in recent years, the demand for long videos with micro- and macro-expressions remains high. Extended from SAMM, a micro-expressions dataset released in 2016, this paper presents SAMM Long Videos dataset for spontaneous micro- and macro-expressions recognition and spotting. SAMM Long Videos dataset consists of 147 long videos with 343 macro-expressions and 159 micro-expressions. The dataset is FACS-coded with detailed Action Units (AUs). We compare our dataset with Chinese Academy of Sciences Macro-Expressions and Micro-Expressions (CAS(ME)^2) dataset, which is the only available fully annotated dataset with micro- and macro-expressions. Furthermore, we preprocess the long videos using OpenFace, which includes face alignment and detection of facial AUs. We conduct facial expression spotting using this dataset and compare it with the baseline of MEGC III. Our spotting method outperformed the baseline result with F1-score of 0.3299.**

I. INTRODUCTION

Facial expression research is multidisciplinary with various applications, such as emotional study, behavioural psychology, security [1], well-being [2] and communication. In general, macro-expression refers to normal facial expression and micro-expression refers to a brief facial expression with duration of less than 500ms [3]. Due to its involuntary nature, it is an important cue in non-verbal communication.

In recent years, an international collaboration initiated by researchers [4], [5] has conducted workshops and challenges on datasets and methods for facial micro-expressions recognition and spotting [6]. However, in real-world scenario, the occurrences of micro- and macro-expressions could co-exist or occur in isolation. Therefore, spotting micro- and macro-expressions is a challenging task. To date, there is limited Facial Action Coding System (FACS) coded long videos dataset. This paper presents SAMM Long Videos dataset, which consists of FACS coded micro- and macro-expressions. We release the dataset and its annotation for research uses.

The rest of the paper is organised as follows: Section II describes the related work. Section III presents new analysis of the dataset, dataset preprocessing stage and performance metrics. Section IV discusses about the results of AU classification and spotting using this dataset. Section V concludes the paper.

II. RELATED WORK

CAS(ME)^2 dataset [7] is the only fully annotated dataset with both macro- and micro-expressions. There are 22 subjects and 87 long videos (in part A). The average duration is 148s. The facial movements are classified as macro- and micro-expressions. The video samples may contain multiple macro- or micro-expressions. The onset, apex and offset of these expressions were annotated and presented in a spreadsheet. Also, the eye blinks are labelled with the onset and offset frame. For MEGC II competition [5] in 2019, the authors of SAMM [8] released 79 long videos for the micro-expressions grand challenge, but with only micro-expressions annotated and was only made available for MEGC II.

In MEGC II, Li et al. [6] proposed temporal pattern extracted from local region for micro-expression spotting on two recently published datasets, i.e. SAMM [8] (79 long videos) and CAS(ME)^2 [7]. Even though CAS(ME)^2 is labelled with macro- and micro-expressions, the focus of the challenge is on micro-expressions spotting. Li et al. [6] demonstrated their Local Temporal Pattern (LTP) method outperformed Local Binary Pattern (LBP) approaches, LBP-χ^2-distance, by Moilanen et al. [9].

III. DATASET PROFILE AND EVALUATION

This section describes the specification of SAMM Long Videos, data preprocessing steps, facial AUs detection using OpenFace, facial movements detection method, and performance metrics for evaluation.

A. Experiment

The original version of SAMM dataset [8] is intended for facial micro-movements detection, which consists of a total of 32 subjects and has 7 videos each. The average length of videos is 35.3s. The original release of SAMM consisted of micro-movement clips with AUs annotated. Recently, the authors [10] introduced objective and emotion classes for the dataset. In 2018, MEGC I used the objective classes for the recognition challenge. In 2019, MEGC II's recognition challenge used the emotional classes from the database as ground truth. In addition to recognition challenge, the spotting challenge was introduced. However, due to the size of SAMM Long Videos dataset, the organisers only used a subset of 79 long videos, each contains one or more micro-movements, with a total of 159 micro-expressions. The index of onset, apex and offset frames of micro-movements were provided as the ground truth. Although the long videos were released for the grand challenge, the macro-expressions labels were not provided.

The micro-movements interval is defined from onset to offset frame. In this dataset, all the micro- and macro-movements are labelled. Thus, the spotted frames can indi-

Fig. 1. Two examples of facial expressions from SAMM Long Videos dataset: (Top) Micro-expression; and (Bottom) Macro-expression. Both expressions with AU12. On apex frame (middle), the macro-expression shows higher intensity and visibility when compared to the micro-expression.

cate not only facial expressions but also other facial movements, such as eye blinks. The details of the experimental settings, eliciting process and coding process are described in Davison et al. [10].

B. Comparison of Facial Expressions and Dataset

Figure 1 shows two examples of facial expressions in SAMM Long Videos. The top row illustrates a micro-expression of brief AU12 with low intensity and the bottom row shows a macro-expression of AU12 with high intensity.

When compared to CAS(ME)^2, SAMM has more long videos. The resolution and frame rate of SAMM is higher than CAS(ME)^2. When compared to the number of facial expressions, SAMM has 159 micro-expressions (which is similar to the first release by Davison et al. [8]) and 343 macro-expressions (newly FACS-coded macro-expressions for this release). Table I compares the differences between SAMM and CAS(ME)^2.

TABLE I
A COMPARISON BETWEEN SAMM LONG VIDEOS AND CAS(ME)^2

Dataset	SAMM Long Videos	CAS(ME)^2
Number of Long Videos	147	87
Number of Videos with micro	79	32
Resolution	2040×1088	640×480
Frame rate	200	30
Number of Macro-expressions	343	300
Number of Micro-expressions	159	57

C. Dataset Preprocessing

We preprocess SAMM Long Videos using OpenFace which includes facial landmark detection, facial expression recognition, head pose and eye gaze estimation. For our case, we focus on face alignment and detection of AUs only.

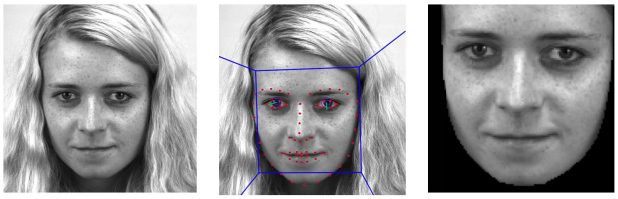

Fig. 2. An illustration of preprocessing steps using OpenFace: (Left) Original SAMM image; (Middle) Facial Landmark Detection; and (Right) Cropped face ROI.

1) Face alignment: OpenFace [11], [12] is a general-purpose toolbox for face recognition, which consists of face alignment algorithm using affine transformation. The facial landmarks are detected by Dlib's face landmark detector [13]. Figure 2 illustrates the original SAMM image, the facial landmarks and the aligned face image. OpenFace uses a similarity transform on current detected landmarks to a representation of landmarks from a neutral expression. This process maps the face texture to a common reference frame and removes changes due to scaling and in plane rotation. The output image has a dimension of 112×112 pixel with interpupilary distance of 45 pixels.

OpenFace utilises Convolutional Experts Constrained Local Model (CE-CLM) which uses deep networks to detect and track facial landmark features. The deep network was simplified from the original version which contained 180,000 parameters to around 90,000. This reduces the model size and speeds up the model by 1.5 times with minimal accuracy loss. Moreover, CE-CLM uses 11 initialisation hypotheses at different orientations which lead to 4 times performance improvement. It also utilises sparse response maps which further improved the model speed by 1.5 times.

2) Detection of AUs: OpenFace [11], [12] is capable of detecting the presence and intensity of AUs. Figure 3

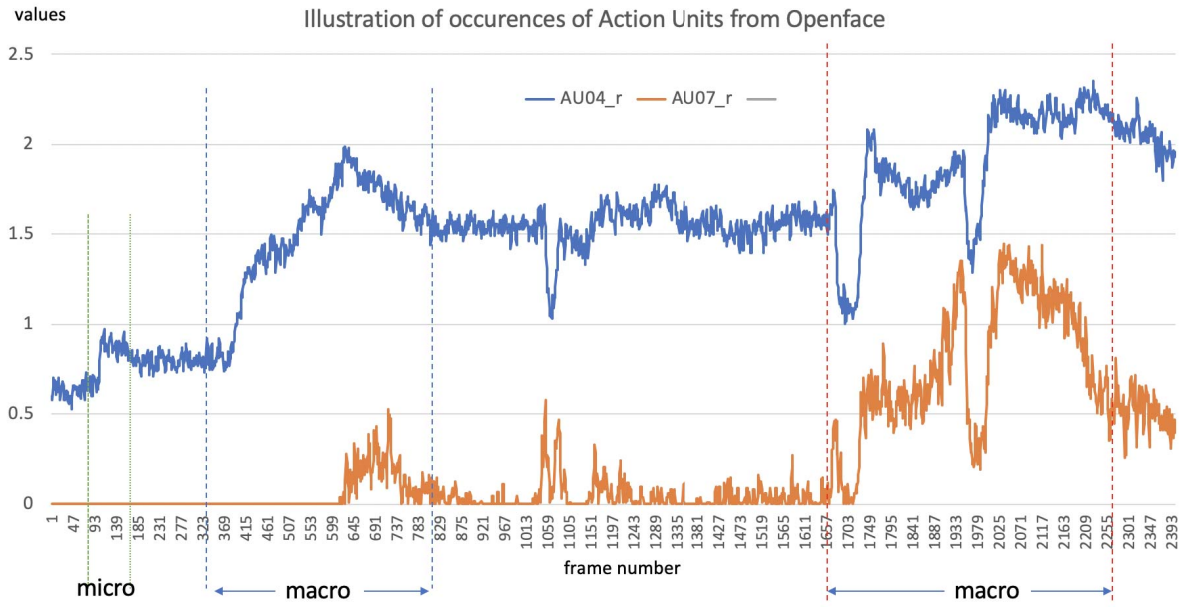

Fig. 3. Two AUs on long video clips, the blue line is AU4, and the orange line is AU7. There are one micro-expression and two macro-expressions found on this video clip.

illustrates the sample output of OpenFace on two AUs plotted on a graph, where there are one micro-expression and two macro-expressions in the video clip. OpenFace conducts facial expression recognition through detecting the intensity and presence of AUs. The intensity of AU is represented on a 5-point scale while the presence of AU is encoded 0 as absent and 1 as present. The AUs that are capable to be recognised by OpenFace are 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 28, and 45. The full name of the facial part represented by each AUs are shown in Table II. Linear kernel Support Vector Machines was used to detect AU occurrence while Support Vector Regression was use to detect the intensity of AU [14]. It also uses the concatenation of Histogram of Oriented Gradient (HOG) of the dimension reduced face image and facial shape features obtained from CE-CLM for each facial features. Note that AU45 involves detecting the presence of eye blink. It was removed from the output signal as it does not carry any significant information in facial expressions classification and spotting.

TABLE II
LIST OF AUS IN OPENFACE

AU	Full name
AU1	Inner Brow Raiser
AU2	Outer Brow Raiser
AU4	Brow Lowerer
AU5	Upper Lid Raiser
AU6	Cheek Raiser
AU7	Lid Tightener
AU9	Nose Wrinkler
AU10	Upper Lip Raiser
AU12	Lip Corner Puller
AU14	Dimpler
AU15	Lip Corner Depressor
AU17	Chin Raiser
AU20	Lip Stretcher
AU23	Lip Tightener
AU25	Lips Part
AU26	Jaw Drop
AU28	Lip Suck
AU45	Blink

D. Detection of Facial Movements

1) Signal smoothing: For spotting task, we combine the intensity of all the detected AUs, and normalised it to a scale of 0 to 1. Subsequently, Savitzky-Golay filter [15] with 21st order was used for noise reduction. Figure 4 illustrates a subject with the normalised sum of the intensity of AU and the smoothed signal plotted with respect to frame number.

2) Onset and offset frame acquisition: By using a custom analysis algorithm modified from [16], the onset and offset frame of the processed data were obtained and compared with the ground truth. This algorithm uses Daubechies wavelet [17] with scaling function of 2 and level 3 signal decomposition for signal smoothing. For peak detection, peak was defined as a point where it is higher than the 7 points that come immediate before and after it. A threshold value of 75th percentile of normalised AU intensity were also implemented. This filters out low local peaks originated from noise. The spotted interval is obtained by subtracting the onset from the offset. Since the frame rate of SAMM Long Videos is 200fps, a threshold of 100 frames (0.5s) was selected to classify the spotted intervals into macro-expressions ($>$0.5s) or micro-expressions ($\leq$0.5s).

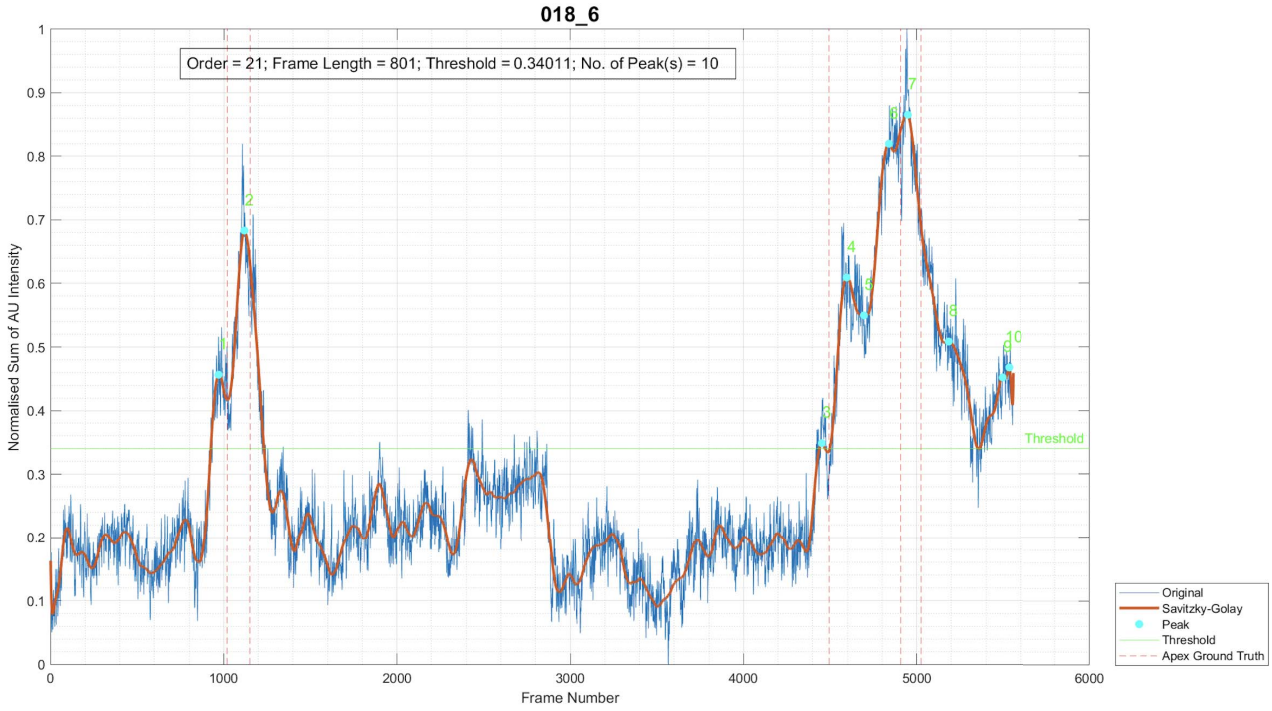

Fig. 4. Plot of Subject 018_6 for the normalised sum of AU intensity (blue line) and the filtered signal (red line) with peaks annotated. A threshold value of 75th percentile of normalised AU intensity was used to filter out the peaks originated from noise.

E. Performance Metrics

Since micro- and macro-expressions occur over a series of frames, the measurement (accuracy) based on the individual frame is not widely accepted in this domain. The preferable performance metrics used are based on the overlap of the frames (based on the concept of Intersect over Union (IoU) in computer vision). Following the spotting challenge in [6], a true positive (TP) is defined as

$$\frac{Predicted \cap GT}{Predicted \cup GT} \geq k \tag{1}$$

where k is set to 0.5, GT represents the ground truth expression interval (onset-offset), and *Predicted* represents the detected expression interval. Otherwise, the spotted interval is regarded as *false positive (FP)*. *False negative (FN)* occurs when the ground truth expression interval exists in the absence of predicted interval.

Then, the comparison of the spotting accuracy will be measured in *Precision*, *Recall*, and $F1-Score$ as in equation (2), equation (3) and equation (4), respectively.

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1-score = \frac{2TP}{2TP+FP+FN} \tag{4}$$

For AUs detection, we report the *Accuracy* of our results which is the number of AUs correctly detected by OpenFace divided by the total number of AUs in ground truth.

TABLE III
ACCURACY OF AU CLASSIFICATION OF OPENFACE

	Micro	Macro	Combined
Matched AU	80	245	278
Ground Truth	172	501	590
Accuracy	0.4651	**0.4890**	0.4712

TABLE IV
COMPARISON BETWEEN AU CLASSIFICATION OF VIDEOS CONTAINING ONLY MICRO- OR MACRO-EXPRESSION

	Micro only videos	Macro only videos
Matched AU	9	104
Ground Truth	31	231
Accuracy	0.2903	**0.4502**

IV. RESULT AND DISCUSSION

For AU classification, AU presence in each video were compared with ground truth. This measures the ability of OpenFace to detect and distinguish AUs. The *Accuracy* for AU classification is shown in Table III.

OpenFace does not classify AU presence by macro- or micro-expression and we observe that there are 84 overlapped AUs (between macro- and micro-expressions). As a result, the detected AU in these two classes contains

TABLE V

RESULTS OF MACRO- AND MICRO-SPOTTING IN SAMM LONG VIDEOS COMPARED TO THE BASELINE OF MEGC III.

	Our Results			**Baseline of MEGC III** [18]		
Expression	macro-expression	micro-expression	overall	macro-expression	micro-expression	overall
Total number	343	159	502	343	159	502
TP	**172**	6	**178**	22	**29**	51
FP	328	71	399	334	1407	1741
FN	171	153	324	321	130	451
Precision	**0.3440**	**0.0779**	**0.3085**	0.0618	0.0202	0.0285
Recall	**0.5015**	0.0377	**0.3546**	0.0641	**0.1824**	0.1016
F1-score	**0.4081**	**0.0508**	**0.3299**	0.0629	0.0364	0.0445

false positives and *false negatives* from the overlapped AUs. To compare the performance of OpenFace on micro- and macro-movements detection, we analyse the micro-only videos and macro-only videos. The comparison between the performance metric of videos containing only macro- or micro-expression is shown in Table IV. All videos used in this comparison contains either with macro- or micro-expressions which gives a fairer evaluation in the detection accuracy of these two classes. There are 18 micro-only and 68 macro-only videos in this dataset. As we observe, micro-expression detection has a lower performance compared to macro-expression detection. This indicates that OpenFace has lower accuracy in detecting AU of micro-expression.

For facial movements detection, the results and comparison with baseline results [18] are shown in Table V. A sorting algorithm were used to combine consecutive detection of TPs into a single measurement. As a result, it was found that there are misclassification of micro-expressions as a single macro-expression. A check on the TPs of micro- and macro-expression was also performed and found that there are 43 micro- misclassified as macro-expressions and 9 macro- misclassified as micro-expressions. All of them were labelled as FPs and FNs. Moreover, by following the criteria set by MEGC III spotting challenge, the overlapping between the prediction and ground truth interval of the TPs found was calculated using equation (1). It was found that 93 TPs did not fulfil the benchmark and they were labelled as FPs and FNs as well. Overall, the sorting algorithm works well as we had verified that the total number of ground truth facial expressions matched the sum of TPs and FNs.

In Table V, it is shown that our spotting method outperformed the baseline. In both results, macro-expression spotting yields the highest F1-score. This confirms that macro-expression spotting is easier compared to micro-expression spotting. It can also be explained as macro-expression has longer interval (> 0.5s) and hence easier to be spotted by the spotting algorithm. Moreover, OpenFace was trained on facial expressions datasets, which have no labelled micro-expressions in the training set, which explain the low F1-score in spotting micro-expressions.

V. CONCLUSION

This paper presents SAMM Long Videos dataset on facial micro- and macro-expressions. We evaluated the performance of OpenFace facial behaviour tools in AUs detection. As OpenFace was not designed to detect facial micro-expressions, it achieved a reasonable results with *Accuracy* of 0.4712 and $F1-Score$ of 0.3553. It performs poorly in micro-only videos with *Accuracy* of 0.2903 in contrast to 0.4502 in macro-only videos. For spotting, our overall results outperformed the baseline result of MEGC III in all three performance metrics of *Precision*, *Recall*, and $F1-Score$.

SAMM Long Videos will benefit researchers in many disciplines, such as affective computing, human behavioural, computer vision and machine learning community.

Future work will extend the dataset by improving class balance by including long videos that contain more micro-expressions. More effective method of elicitation of micro-expression should be investigated.

SAMM Long Videos dataset is publicly available at http://www2.docm.mmu.ac.uk/STAFF/M.Yap/dataset.php.

VI. ACKNOWLEDGEMENT

This project was supported by Manchester Metropolitan University Vice-Chancellor's PhD Studentship. The authors gratefully acknowledge the contribution of FACS Coders and collaborators in Davison et al. [8].

REFERENCES

[1] P. Ekman, "Lie catching and microexpressions," *The philosophy of deception*, p. 118–133, 2009.

[2] J. Endres and A. Laidlaw, "Micro-expression recognition training in medical students: a pilot study," *BMC medical education*, vol. 9, no. 1, p. 47, 2009.

[3] P. Ekman and W. V. Friesen, "Nonverbal leakage and clues to deception," *Psychiatry*, vol. 32, no. 1, p. 88–106, 1969.

[4] M. H. Yap, J. See, X. Hong, and S.-J. Wang, "Facial micro-expressions grand challenge 2018 summary," in *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on*. IEEE, 2018, pp. 675–678.

[5] J. See, M. H. Yap, J. Li, X. Hong, and S.-J. Wang, "Megc 2019–the second facial micro-expressions grand challenge," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE, 2019, pp. 1–5.

[6] J. Li, C. Soladié, R. Séguier, S.-J. Wang, and M. H. Yap, "Spotting micro-expressions on long videos sequences," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE, 2019, pp. 1–5.

[7] F. Qu, S.-J. Wang, W.-J. Yan, H. Li, S. Wu, and X. Fu, "Cas(me)^2: A database for spontaneous macro-expression and micro-expression spotting and recognition," *IEEE Transactions on Affective Computing*, vol. 9, no. 4, pp. 424–436, 2017.

[8] A. K. Davison, C. Lansley, N. Costen, K. Tan, and M. H. Yap, "Samm: A spontaneous micro-facial movement dataset," *IEEE Transactions on Affective Computing*, vol. 9, no. 1, pp. 116–129, 2018.

[9] A. Moilanen, G. Zhao, and M. Pietikäinen, "Spotting rapid facial movements from videos using appearance-based feature difference analysis," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*. IEEE, 2014, p. 1722–1727.

[10] A. Davison, W. Merghani, and M. Yap, "Objective classes for micro-facial expression recognition," *Journal of Imaging*, vol. 4, no. 10, p. 119, 2018.

[11] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "Openface: an open source facial behavior analysis toolkit," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2016, pp. 1–10.

[12] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 59–66.

[13] D. E. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, no. Jul, pp. 1755–1758, 2009.

[14] T. Baltrušaitis, M. Mahmoud, and P. Robinson, "Cross-dataset learning and person-specific normalisation for automatic action unit detection," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 6. IEEE, 2015, pp. 1–6.

[15] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures." *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[16] B. M. Hewitt, M. H. Yap, E. F. Hodson-Tole, A. J. Kennerley, P. S. Sharp, and R. A. Grant, "A novel automated rodent tracker (art), demonstrated in a mouse model of amyotrophic lateral sclerosis," *Journal of neuroscience methods*, vol. 300, pp. 147–156, 2018.

[17] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE transactions on information theory*, vol. 36, no. 5, pp. 961–1005, 1990.

[18] Y. He, S.-J. Wang, J. Li, and M. H. Yap, "Spotting macro-and micro-expression intervals in long video sequences," *arXiv preprint arXiv:1912.11985*, 2019.