MEGC2022: ACM Multimedia 2022 Micro-Expression Grand Challenge

Jingting Li
CAS Key Laboratory of Behavioral Science, Institute of Psychology
lijt@psych.ac.cn

Moi Hoon Yap
Centre for Advanced Computational Science, Manchester Metropolitan University
m.yap@mmu.ac.uk

Wen-Huang Cheng
National Yang Ming Chiao Tung University
whcheng@nycu.edu.tw

John See
Heriot-Watt University Malaysia
J.See@hw.ac.uk

Xiaopeng Hong
Harbin Institute of Technology
hongxiaopeng@ieee.org

Xiaobai Li
University of Oulu
xiaobai.li@oulu.fi

Su-Jing Wang*
CAS Key Laboratory of Behavioral Science, Institute of Psychology & Department of Psychology, University of the Chinese Academy of Sciences
wangsujing@psych.ac.cn

Adrian K. Davison
University of Manchester
adrian.davison@manchester.ac.uk

Yante Li
University of Oulu
yante.li@oulu.fi

Zizhao Dong
CAS Key Laboratory of Behavioral Science, Institute of Psychology
dongzz@psych.ac.cn

ABSTRACT

Facial micro-expressions (MEs) are involuntary movements of the face that occur spontaneously when a person experiences an emotion but attempts to suppress or repress the facial expression, typically found in a high-stakes environment. Unfortunately, the small sample problem severely limits the automation of ME analysis. Furthermore, due to the brief and subtle nature of ME, ME spotting is a challenging task, and the performance is still not satisfactory yet. This challenge focuses on two tasks, i.e., the micro- and macro-expression spotting task, and the ME Generation task.

CCS CONCEPTS

• **Computing methodologies** → **Computer vision**; • **Applied computing** → *Psychology*.

KEYWORDS

Micro-Expression, Spotting, Generation

ACM Reference Format:

Jingting Li, Moi Hoon Yap, Wen-Huang Cheng, John See, Xiaopeng Hong, Xiaobai Li, Su-Jing Wang*, Adrian K. Davison, Yante Li, and Zizhao Dong. 2022. MEGC2022: ACM Multimedia 2022 Micro-Expression Grand Challenge. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22), Oct. 10–14, 2022, Lisboa, Portugal.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3503161.3551601

MM '22, October 10–14, 2022, Lisboa, Portugal

ACM ISBN 978-1-4503-9203-7/22/10.
https://doi.org/10.1145/3503161.3551601

1 INTRODUCTION

Since the advent of deep learning techniques, face recognition and face-based emotion understanding have become popular research topics. Facial micro-expressions (MEs) are involuntary facial movements that occur spontaneously when a person is experiencing a certain emotion but tries to suppress or repress the facial expression, often seen in high-risk environments. ME is considered one of the clues to understanding complex human emotions. It has a wide range of potential applications in the real world, such as police interrogation, clinical diagnosis, depression analysis, etc. Computational analysis and automation of ME tasks is an emerging area of face research that has seen strong interest as recently as 2014. The emergence of several spontaneous facial ME datasets has provided the impetus for further developments in computation.

However, ME research suffers from the following challenges. First, ME elicitation and manual annotation are challenging and laborious. Therefore, the number of labeled ME samples is limited. There are only 11 published spontaneous ME databases, including CASME series (CASME [20], CASME II [19], CAS(ME)^2 [14], CAS(ME)^3 [8]), SMIC series (SMIC [12], SMIC-E [13], SMIC-E-long [16], 4DME [11]), SAMM series (SAMM [2], SAMM-LV [22]), and MMEW [1]. Second, ME annotation is subjective to different annotators. Therefore, standardization of ME labels is almost impossible. In addition, MEs have the characteristics of short duration and low intensity, and extracting their features is challenging. To solve this problem, we expect that recent advances in pattern recognition can help improve the performance of ME discovery and recognition.

This is the inaugural academic activity in this area of research. Our ambition is to conduct ME challenge yearly with continuity. We have held four ME Grand Challenges (MEGC)[1 2 3] in conjunction with FG2018 [24], FG2019 [15], FG2020 [6], and ACMMM2021 [10] and a ME Recognition Challenge (MER2020)[4] [17] in conjunction with ICIP2020.

In MEGC2022[5], we focus on two tasks in intelligent ME analysis: ME spotting and ME generation. ME spotting, as an inevitable first step in real applications, requires the algorithm to distinguish MEs from other types of head movements, such as macro-expressions (MaEs) and head movements, and to localize them in time. However, there is limited research on this in the community. Therefore, we set the challenge of spotting MaEs and MEs in long videos to promote community interest in this direction and ME spotting performance. In addition, since it is complicated to collect and label real ME samples, we propose the ME generation task, i.e., generating ME videos with different expression types. The task not only contributes to the data augmentation of ME, but also promotes the enhancement of ME feature extraction.

This year, eight teams participated in the spotting task and three teams participated in the generation task, respectively. The top three articles for each task were accepted.

2 SPOTTING CHALLENGE

The goal of this challenge is to spot MaE and ME interval in long video sequences.

2.1 Recommended Training Datasets

CAS(ME)^2: It consists of 22 participants and 98 long videos at 30 fps, including 300 MaEs and 57 MEs. As an improvement compared with MEGC2020 [6], a cropped version with only face region is provided for fair comparison in MEGC2021. Meanwhile, the authors of the SAMM dataset [2] released their corresponding long videos, i.e. the SAMM Long Videos dataset [22], which consists of 147 long videos at 200fps, including 343 macro-movements and 159 micro-movements in the long videos. In addition, the authors of the SMIC-E dataset also released their corresponding long videos, i.e. the SMIC-E-long dataset [22], including 162 long videos at 100 fps with 132 MEs.

2.2 Testing Dataset

For the first time, we released the unseen testing set (MEGC2022-testSet), which is introduced by Yap et al. [23] and contains 10 long videos, including 5 long videos from SAMM (SAMM Challenge dataset) [2] and 5 clips cropped from different videos in CAS(ME)^3 [8].

[1]http://www2.docm.mmu.ac.uk/STAFF/m.yap/FG2018Workshop.htm
[2]https://facial-micro-expressiongc.github.io/MEGC2019/
[3]https://megc2020.github.io/
[4]https://2020.ieeeicip.org/challenge/micro-expression-recognition-challenge/
[5]https://megc2022.github.io/challenge.html

2.3 Metrics

The metric we use was first presented at MEGC2019 and then used for the evaluation of spotting task in subsequent challenges, which is an interval-based evaluation method [6, 9].

Specifically, the true positive (TP) per interval in one video is first defined based on the intersection between the spotted interval and the ground-truth interval. The spotted interval $W_{spotted}$ is considered as TP if it fits the following condition:

$$\frac{W_{spotted} \cap W_{groundTruth}}{W_{spotted} \cup W_{groundTruth}} \geq k \tag{1}$$

where k is set to 0.5, $W_{groundTruth}$ represents the ground truth of the MaE or ME interval (onset-offset). If the condition is not fulfilled, the spotted interval is regarded as false positive (FP). **We consider that each ground-truth interval corresponds to at most one single spotted interval.**

The final evaluation is performed on the entire dataset, based on the overall F1-score of MaE and ME spotting performance. The champion of the challenge will be the best score for overall results in spotting MEs and MaEs. The participants should evaluate their results on grand challenge system[6], with evaluation codes used by the baseline method [23] to ensure fair comparison.

2.4 Methods

The baseline method and the corresponding baseline results for MEGC2021 and MEGC2022 are provided by Yap et al. [23]. They compare the frame differential motion via a convolutional model, each frame with two temporally local reference frames. Reference frames are sampled according to calculated MaE and ME duration.

In MEGC2021, He et al. [28] utilize the traditional handcraft feature difference comparison method. Since the head shaking is an essential reason for the high false-positive rate of ME spotting, they propose a cutting box which is based on the optical flow of the nose region and is is adjusted several times to optimize the relative position between the face and the nose region. Thanks to this reliable face alignment, He et al. won first place in the spotting task of MEGC2021. Although both used optical flow, the second place team, i.e., Yu et al. [27] introduce a deep learning approach named location suppression spotting network (LSSNet) for the spotting task. In particular, fixed-length features in the sliding window are extracted by an I3D model. Location suppression modules are added to the pyramidal full convolutional neural network to reduce the proposals with longer and shorter intervals. Unlike the first two, who both use optical flow, the third place team, i.e., Yang et al. [21] focus on the individual components of facial muscle movement. Specifically, they propose a Concat-CNN model to learn the characteristic correlation between AUs of distinctive frames. The Concat-CNN uses three convolutional kernels with different sizes to observe features of different duration, and emphasizes both local and global mutation features by changing dimensionality (max-pooling size) of the output space.

In MEGC2022, the top two teams for the spotting task both used the traditional manual feature difference comparison method. Yu et al. [25] extract the optical flow from local facial regions and enhance the features of the expressions in the optical flow with low-pass

[6]https://megc2022.grand-challenge.org/

filters and EMD methods. Then, the sliding window method is used to locate the peaks of optical flow features and get intervals containing MEs or MaEs. Inspired by He et al. [28], Zhao et al. [31] remove global head movements using the optical flow feature in the nose region. Furthermore, Bayesian optimization is used to detect facial actions by feature curve changes. In contrast to the first two, the third place team, i.e., Leng et al. [7] use a deep learning approach and propose a spotting framework based on Apex and Boundary Perception Network (ABPN). ABPN mainly consists of three parts, i.e., a video encoding module that learns motion features, a probability evaluation module that predicts frame level auxiliary probabilities, and an expression proposal generation module for accurate localization.

2.5 Results & Analysis

The table lists the top three performance evaluations of the spotting tasks for the two-year challenge. Although MEGC2022 uses the unseen dataset, the CAS(ME)^3 acquisition environment is similar to CAS(ME)^2, and the SAMM and SAMM-LV environments are the same. Therefore, the results are comparable to some extent. Overall, the spotting performance this year is better than last year. In addition, most methods still chose optical flow to extract movement information of MaE and ME. Furthermore, the first places for two consecutive years were obtained by the feature difference comparison method instead of the deep learning-based method. In particular, reliable pre-processing, including face correction and elimination of global motion of faces, as well as more valid schemes such as filtering and Bayesian optimization during the detection of feature curve peaks, contribute to MaE and ME spotting. Also, effective interval fusion can reduce the number of FPs. On the other hand, the deep learning approach is limited by the small sample size problem, resulting in a weaker robust spotting performance.

Table 1: Overall F1-score of the spotting tasks for MEGC2021 and MEGC2022. [23] provides the baseline results.

		Both	SAMM-LV	CAS(ME)^2
MEGC2021	[28]	**0.3534**	**0.3638**	**0.3436**
	[27]	0.2717	0.2337	0.3257
	[21]	0.2452	0.2736	0.2019
	[23]	-	0.1193	0.0304
		Both	SAMM	CAS(ME)^3
MEGC2022	[25]	**0.3514**	**0.3846**	0.3333
	[31]	0.3467	0.3265	**0.3564**
(Unseen dataset)	[7]	0.2975	0.2264	0.3529
	[23]	0.1351	0.1176	0.1739

3 GENERATION CHALLENGE

The goal of this challenge is to generate specific ME (source) on the given template faces (target). By evaluating the authenticity and strength of the generated ME via psychologists' inspection, reliable ME generation can enable proper data augmentation of MEs, thereby promoting the further development of ME analysis.

3.1 Databases

CASME II [19]: CASME II contains 26 subjects and 255 ME sequences. All videos are at 200 fps to retain more facial information and the resolution is 640 × 480. The onset, apex, offset index for these expressions are given in the excel file. In addition, the eye blinks are labeled with onset and offset time.
SAMM [2] The original SAMM dataset [2] with 159 MEs. The index of onset, apex and offset frames of micro-movements are outlined in the ground truth excel file. The micro-movements interval is from onset frame to offset frame. In this database, all the micro-movements are labeled. Thus, the spotted frames can indicate not only ME but also other facial movements, such as eye blinks.
SMIC [12]: includes three subsets: SMIC-HS, SMIC-VIS and SMIC-NIR. SMIC-VIS and SMIC-NIR contrains 71 samples recorded by normal speed cameras with 25 fps of visual (VIS) and near inferred (NIR) light range, respectively. SMIC-HS recorded by 100 fps high-speed cameras contains 164 spontaneous ME clips from 16 subjects.

3.2 Metrics

3.2.1 Submission Video Format. Each database specifies three kinds of emotion samples (positive, negative, and surprised), i.e. source samples, as listed in Table 2.

Table 2: Sequence names of the assigned emotion samples for the generation task (Source).

Databases	CASME II	SAMM	SMIC-HS
Positive	EP01_01f	022_3_3	s3_po_05
Negative	EP19_06f	018_3_1	s11_ne_02
Surprise	EP01_13	007_7_1	s20_sur_01

The participants should generate the specified expressions on the provided template faces. Target faces in MEGC2021 are Asian female (from CASME I [20]) and western male (from SMIC-VIS) respectively (as shown in Fig. 1), i.e. target samples. This year, MEGC2022 increased diversity of target faces with four provided template faces, which are Asian female with glasses, Asian male without glasses (these two samples are selected from CAS(ME)^3 [8]), western male with glasses and western female without glasses respectively, as illustrated in Fig. 1. The expected output video should generated the facial micro expression from source samples videos on the target face.

The expected output video should generated the ME movement from source samples videos on the target face. The total number of videos for submission is 36, i.e. 4 templates × 3 emotions for each database. All submitted videos should be unified at 100fps, with a resolution of 256 × 256. The length of the generated video is based on the specified emotion sample and does not need to be normalized.

3.2.2 Evaluation Protocol. Each generated image will be evaluated based on the quality and action units. Specifically, the facial region will be divided into upper and lower parts and evaluated separately. By separating the face into two parts, evaluations can take into account partial facial movements that may occur.

Figure 1: Template faces for generation task (Target). The upper row shows the normalized template faces in MEGC2021. The lower row shows the template faces in MEGC2022. The eye area is obscured for privacy.

The quality and action unit of each block will be scored 0-3 by experts who have Facial Action Coding System (FACS) certification [5]. In addition, there will be a 'noise' category, which judges the overall generation's image quality and is also score 0-3. The following details the score categories: Score 0: Completely incorrect; Score 1: Poor; Score 2: Good; Score 3: Excellent. The maximum available score will be 9. The three experts who are FACS AU coders will evaluate the results without interfering with each other. The final score will be the average of scores provided by three experts.

3.3 Methods

In MEGC2021, the first place team, i.e., Zhang et al. [29] propose a novel method using the first-order motion model based on facial prior obtained by Region-Focusing Module. Specifically, Motion Prediction Module estimates facial motions using key points and local affine transformations. Meantime, the generative adversarial network (GAN) is applied, driving the target face to generate ME videos. The second place team, i.e., Fan et al. [4] use a deep motion re-targeting network that can learn landmarks in a self-supervised manner and generate dense motion between reference and desired images. Moreover, they apply deep transfer learning by borrowing knowledge from the MaE generation to alleviate the lack of training data. The third place team, i.e., Xu et al. [18] utilize GAN based on fine-grained AUs modulation to generate MEs sequence (FAMGAN) with different magnitudes of AUs.

In MEGC2022, the first place team, Yu et al. [26] propose an end-to-end unsupervised motion transfer network. In particular, they adopt the thin-plate spline method to estimate the optical flow of the face motion. Furthermore, the face parsing method is employed to pay specific attention to the eyeglasses regions and thus ensure the reasonability of the deformation. Similar to the first place, the second place team, i.e., Zhao et al.[30] estimate non-linear ME motion using thin-plate spline transformation with a dense motion network. Then, the estimated ME motion transformations are sent to the generation network to synthesize the target ME. Meantime, the relative action units of the source ME is used as a constraint to encourage the network to ignore expression-irrelevant movements. The third place team, i.e., Fan et al. [3] propose a deep learning-based adaptive dual motion model. First, robust motions are extracted from two modalities: original color images and edge-based grayscale images with dual streams. Then they are fed into an adaptive motion fusion module for combing the motions adaptively to generate the dense motion.

3.4 Results & Analysis

We were pleasantly surprised by the videos submitted by the participants in both generation tasks. Some of the samples were evaluated by experts as being "fake-like". Compared with MEGC2021, the overall quality of MEGC2022 has been improved, even with an increasing diversity of faces. In particular, using the thin-plate spline method to estimate the face deformation makes the movements more natural and realistic. However, there are still problems with some of the generated views. For instance, there are blurred areas of the face; some videos do not show the facial expressions that should be generated; some videos have unnatural expression transitions, perhaps showing the maximum movement of the expression at the beginning of the video.

Table 3: Video evaluation of the MEGC2021 and MEGC2022 generation tasks. The total score of 18 videos for MEGC2021 is 162 and the total score of 36 videos for MEGC2022 is 324. The values in the table are normalized by dividing the coder's score for each participant's submitted videos by the total score.

MEGC2021				MEGC2022			
Method	[29]	[4]	[18]		[26]	[30]	[3]
Coder1	0.86	0.86	0.64	Coder1	0.74	0.73	0.61
Coder2	0.62	0.66	0.41	Coder2	0.58	0.55	0.56
Coder3	0.47	0.35	0.41	Coder3	0.66	0.58	0.44
Total	**1.95**	1.87	1.46	Total	**1.98**	1.86	1.61
mean	1.76			mean	**1.82**		

4 CONCLUSION & FUTURE CHALLENGES

Due to the small sample size problem of MEs and the subtle, brief, and localized nature of MEs themselves, intelligent ME analysis is extremely challenging. With the continuous development of spotting methods, we will propose closer to "in the wild" test videos in the future to promote the application of ME spotting in real-world scenarios. In addition, the evaluation of ME generation tasks is currently performed by the expert naked eye. How to establish an objective and reliable evaluation system remains a future task. Moreover, we will also try to add ME analysis tasks combining scene depth information, individual physiological signals, and other data in future challenges to explore multi-modal ME intelligent analysis methods.

ACKNOWLEDGMENTS

This work is supported by grants from the National Natural Science Foundation of China (62106256, U19B2032), the Academy of Finland (Grant 323287), The Manchester Metropolitan University VC PhD Studentship, and Ministry of Science and Technology of Taiwan (MOST-109-2223-E-009-002-MY3, MOST-110-2634-F-007-015).

REFERENCES

[1] Xianye Ben, Yi Ren, Junping Zhang, Su-Jing Wang, Kidiyo Kpalma, Weixiao Meng, and Yong-Jin Liu. 2021. Video-based Facial Micro-Expression Analysis: A Survey of Datasets, Features and Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).

[2] Adrian K Davison, Cliff Lansley, Nicholas Costen, Kevin Tan, and Moi Hoon Yap. 2018. Samm: A spontaneous micro-facial movement dataset. *IEEE Transactions on Affective Computing* 9, 1 (2018), 116–129.

[3] Xinqi Fan, Ali Shahid, and Hong Yan. 2022. Adaptive Dual Motion Model for Facial Micro-Expression Generation. In *Proceedings of the 30th ACM International Conference on Multimedia*. Association for Computing Machinery.

[4] Xinqi Fan, Ali Raza Shahid, and Hong Yan. 2021. Facial Micro-Expression Generation based on Deep Motion Retargeting and Transfer Learning. In *Proceedings of the 29th ACM International Conference on Multimedia*. Association for Computing Machinery, 4735–4739.

[5] W. V Friesen and P. Ekman. 1978. Facial action coding system: a technique for the measurement of facial movement. *Palo Alto* 3 (1978).

[6] LI Jingting, Su-Jing Wang, Moi Hoon Yap, John See, Xiaopeng Hong, and Xiaobai Li. 2020. MEGC2020-the third facial micro-expression grand challenge. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*. IEEE, 777–780.

[7] Wenhao Leng, Sirui Zhao, Yiming Zhang, Shifeng Liu, Xinglong Mao, Hao Wang, Tong Xu, and Enhong Chen. 2022. ABPN: Apex and Boundary Perception Network for Micro- and Macro-Expression Spotting. In *Proceedings of the 30th ACM International Conference on Multimedia*. Association for Computing Machinery.

[8] Jingting Li, Zizhao Dong, Shaoyuan Lu, Su-Jing Wang, Wen-Jing Yan, Yinhuan Ma, Ye Liu, Changbing Huang, and Xiaolan Fu. 2022. CAS(ME)^3: A Third Generation Facial Spontaneous Micro-Expression Database with Depth Information and High Ecological Validity. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).

[9] Jingting Li, Catherine Soladie, Renaud Seguier, Su-Jing Wang, and Moi Hoon Yap. 2019. Spotting micro-expressions on long videos sequences. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE, 1–5.

[10] Jingting Li, Moi Hoon Yap, Wen-Huang Cheng, John See, Xiaopeng Hong, Xiaobai Li, and Su-Jing Wang. 2021. FME'21: 1st Workshop on Facial Micro-Expression: Advanced Techniques for Facial Expressions Generation and Spotting. In *Proceedings of the 29th ACM International Conference on Multimedia*. 5700–5701.

[11] Xiaobai Li, Shiyang Cheng, Yante Li, Muzammil Behzad, Jie Shen, Stefanos Zafeiriou, Maja Pantic, and Guoying Zhao. 2022. 4DME: A Spontaneous 4D Micro-Expression Dataset With Multimodalities. *IEEE Transactions on Affective Computing* (2022), 1–18. https://doi.org/10.1109/TAFFC.2022.3182342

[12] Xiaobai Li, Tomas Pfister, Xiaohua Huang, Guoying Zhao, and Matti Pietikäinen. 2013. A spontaneous micro-expression database: Inducement, collection and baseline. In *2013 10th IEEE International Conference and Workshops on Automatic face and gesture recognition (FG)*. IEEE, 1–6.

[13] Xiaobai Li, HONG Xiaopeng, Antti Moilanen, Xiaohua Huang, Tomas Pfister, Guoying Zhao, and Matti Pietikainen. 2017. Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods. *IEEE Transactions on Affective Computing* (2017).

[14] Fangbing Qu, Su-Jing Wang, Wen-Jing Yan, He Li, Shuhang Wu, and Xiaolan Fu. 2017. CAS(ME)^2: a database for spontaneous macro-expression and micro-expression spotting and recognition. *IEEE Transactions on Affective Computing* (2017).

[15] John See, Moi Hoon Yap, Jingting Li, Xiaopeng Hong, and Su-Jing Wang. 2019. Megc 2019–the second facial micro-expressions grand challenge. In *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE, 1–5.

[16] Thuong-Khanh Tran, Quang-Nhat Vo, Xiaopeng Hong, Xiaobai Li, and Guoying Zhao. 2021. Micro-expression spotting: A new benchmark. *Neurocomputing* 443 (2021), 356–368.

[17] Hong-Xia Xie, Ling Lo, Hong-Han Shuai, and Wen-Huang Cheng. 2020. Au-assisted graph attention convolutional network for micro-expression recognition. In *Proceedings of the 28th ACM International Conference on Multimedia*. 2871–2880.

[18] Yifan Xu, Sirui Zhao, Huaying Tang, Xinglong Mao, Tong Xu, and Enhong Chen. 2021. FAMGAN: Fine-grained AUs Modulation based Generative Adversarial Network for Micro-Expression Generation. In *Proceedings of the 29th ACM International Conference on Multimedia*. Association for Computing Machinery, 4813–4817.

[19] Wen-Jing Yan, Xiaobai Li, Su-Jing Wang, Guoying Zhao, Yong-Jin Liu, Yu-Hsin Chen, and Xiaolan Fu. 2014. CASME II: An improved spontaneous micro-expression database and the baseline evaluation. *PloS one* 9, 1 (2014), e86041.

[20] Wen-Jing Yan, Qi Wu, Yong-Jin Liu, Su-Jing Wang, and Xiaolan Fu. 2013. *CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces*. IEEE, 1–7.

[21] Bo Yang, Jianming Wu, Zhiguang Zhou, Megumi Komiya, Koki Kishimoto, Jianfeng Xu, Keisuke Nonaka, Toshiharu Horiuchi, Satoshi Komorita, Gen Hattori, et al. 2021. Facial Action Unit-based Deep Learning Framework for Spotting Macro-and Micro-expressions in Long Video Sequences. In *Proceedings of the 29th ACM International Conference on Multimedia*. Association for Computing Machinery, 4794–4798.

[22] Chuin Hong Yap, Connah Kendrick, and Moi Hoon Yap. 2020. Samm long videos: A spontaneous facial micro-and macro-expressions dataset. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*. IEEE, 771–776.

[23] Chuin Hong Yap, Moi Hoon Yap, Adrian Keith Davison, Connah Kendrick, Jingting Li, Su-Jing Wang, and Ryan Cunningham. 2022. 3D-CNN for Facial Micro- and Macro-expression Spotting on Long Video Sequences using Temporal Oriented Reference Frame. In *Proceedings of the 30th ACM International Conference on Multimedia*. Association for Computing Machinery.

[24] Moi Hoon Yap, John See, Xiaopeng Hong, and Su-Jing Wang. 2018. Facial micro-expressions grand challenge 2018 summary. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 675–678.

[25] Jun Yu, Zhongpeng Cai, Zepeng Liu, Guochen Xie, and Peng He. 2022. Facial Expression Spotting Based on Optical Flow Features. In *Proceedings of the 30th ACM International Conference on Multimedia*. Association for Computing Machinery.

[26] Jun Yu, Guochen Xie, Zhongpeng Cai, Peng He, and Qiang Ling. 2022. Micro Expression Generation with Thin-plate Spline Motion Model and Face Parsing. In *Proceedings of the 30th ACM International Conference on Multimedia*. Association for Computing Machinery.

[27] Wang-Wang Yu, Jingwen Jiang, and Yong-Jie Li. 2021. LSSNet: A Two-stream Convolutional Neural Network for Spotting Macro-and Micro-expression in Long Videos. In *Proceedings of the 29th ACM International Conference on Multimedia*. Association for Computing Machinery, 4745–4749.

[28] He Yuhong. 2021. Research on Micro-Expression Spotting Method Based on Optical Flow Features. In *Proceedings of the 29th ACM International Conference on Multimedia*. Association for Computing Machinery, 4803–4807.

[29] Yi Zhang, Youjun Zhao, Yuhang Wen, Zixuan Tang, Xinhua Xu, and Mengyuan Liu. 2021. Facial Prior Based First Order Motion Model for Micro-expression Generation. In *Proceedings of the 29th ACM International Conference on Multimedia*. Association for Computing Machinery, 4755–4759.

[30] Sirui Zhao, Shukang Yin, Huaying Tang, Rijin Jin, Yifan Xu, Tong Xu, and Enhong Chen. 2022. Fine-grained Micro-Expression Generation based on Thin-Plate Spline and Relative AU Constraint. In *Proceedings of the 30th ACM International Conference on Multimedia*. Association for Computing Machinery.

[31] Yuan Zhao, Xin Tong, Zichong Zhu, Jianda Sheng, Lei Dai, Lingling Xu, Xuehai Xia, Jiang Yu, and Jiao Li. 2022. Rethinking Optical Flow Methods for Micro-Expression Spotting. In *Proceedings of the 30th ACM International Conference on Multimedia*. Association for Computing Machinery.