

## Application-Oriented Micro-Expression Analysis: From Data Acquisition to Intelligent Deployment

**Authors:** Li Jingting, Zhao Lin, Dong Zichao, Wang Sujing, Wang Sujing

**Date:** 2025-06-12T18:37:40+00:00

### Abstract

Micro-expressions refer to brief facial actions that individuals unconsciously display when attempting to suppress genuine emotions. Owing to their non-invasive nature, they possess significant application value in domains such as national security and public security. To address challenges including insufficient ecological validity, motion interference, and data privacy in practical scenarios, this research constructs a high-ecological-validity micro-expression elicitation paradigm grounded in physiological and behavioral psychological mechanisms, develops a perifacial electromyography-assisted coding system, and establishes a multi-scenario dynamic micro-expression database. The study further designs modules for eliminating interference from head movements and lip variations, integrates a self-supervised learning framework to tackle small-sample recognition issues, and leverages asynchronous federated learning to enable cross-scenario model deployment under privacy protection. Through the interdisciplinary convergence of psychology and computer science, this work proposes a micro-expression analysis framework that reconciles theoretical mechanisms with practical applications, thereby providing technical support for multi-domain applications.

### Full Text

## Micro-expression Analysis for Practical Applications: From Data Acquisition to Intelligent Deployment

**LI Jingting<sup>12</sup>, ZHAO Lin<sup>12</sup>, DONG Zizhao<sup>12</sup>, WANG Su-Jing<sup>12</sup>**

<sup>1</sup> State Key Laboratory of Cognitive Science and Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

<sup>2</sup> Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract:** Micro-expressions refer to brief facial movements that unconsciously leak when individuals attempt to conceal their true emotions. Their non-invasive characteristics hold significant application value in national security, public safety, and other domains. Addressing issues of insufficient ecological validity, motion interference, and data privacy in practical scenarios, this research constructs micro-expression elicitation paradigms with high ecological validity based on physiological and behavioral psychology mechanisms, develops a facial peripheral electromyography (EMG) signal-assisted coding system, and establishes a multi-scenario dynamic micro-expression database. By designing interference elimination modules for head movements and lip changes, combining self-supervised learning frameworks to solve small-sample recognition problems, and simultaneously utilizing asynchronous federated learning to achieve cross-scenario model deployment under privacy protection, this study proposes a micro-expression analysis framework that balances theoretical mechanisms and practical applications through interdisciplinary integration of psychology and computer science, providing technical support for multi-domain applications.

**Keywords:** micro-expression intelligent analysis, affective computing, high ecological validity, practical application

## 1 Problem Statement

Emotional fluctuations not only reveal an individual's inner world but also serve as intuitive manifestations of complex psychological states. Among various emotional cues, facial expressions are considered the most direct window into emotion. However, as the idioms "forcing a smile" and "hiding a dagger behind a smile" remind us, expressions are often deliberately controlled to conceal genuine emotional leakage. Micro-expressions, as brief and subtle facial movements, typically leak unconsciously or when control is impossible, becoming critical clues for revealing true emotions. Consequently, micro-expressions demonstrate tremendous application potential in public security, national security, and other fields.

Research and application of micro-expressions rely on solid theoretical foundations. As early as 1859, Charles Darwin noted in *The Expression of the Emotions in Man and Animals* that certain facial expressions are difficult to suppress even with great effort (Darwin, 1872). Neurophysiological research has found that voluntary and involuntary expressions are controlled by the pyramidal and extrapyramidal motor systems, respectively (Rinn, 1984). In 1969, Professor Paul Ekman proposed that conflict between voluntary and involuntary expressions may lead to micro-expressions, which can arise from leakage of self-inhibited preliminary expression impulses or be accidentally truncated during normal expression processes (Ekman & Friesen, 1969). Therefore, micro-expressions can reflect an individual's true emotions and are characterized by being fleeting, weakly visible, and locally occurring.

Manual detection and recognition of micro-expressions is extremely challenging;

even professionally trained experts achieve only slightly better than random guessing accuracy (Frank et al., 2009). This phenomenon reveals the limitations of manual micro-expression recognition, particularly the high demands on human resources and time, highlighting the urgency of improving intelligent micro-expression analysis technology. Current intelligent micro-expression analysis faces multiple challenges: First, the generation mechanisms of micro-expressions in different contexts remain incompletely understood, constraining the development of foundational theories. Second, the scarcity of high ecological validity micro-expression samples limits the training and application effectiveness of deep learning models. Third, the high difficulty of manual annotation makes it challenging to obtain sample sizes sufficient for algorithmic needs. Additionally, interference factors such as head pose variations and lip movements significantly impact analysis accuracy. Finally, data privacy issues hinder effective integration and sharing of cross-scenario data. Therefore, constructing an efficient data acquisition system and developing adapted intelligent analysis techniques constitute the key path to breaking through bottlenecks in micro-expression research and promoting practical applications.

## 2.1 Research Status of Micro-Expression Intelligent Analysis

As shown in [FIGURE:1], micro-expression intelligent analysis has received widespread attention since 2009, a progress driven by the release of a series of high-quality micro-expression databases. However, currently available micro-expression data comprises only approximately 10,000 samples, with insufficient data volume and ecological validity, limiting the development of deep learning technology in micro-expression analysis. Moreover, compared with micro-expression recognition (classifying known micro-expressions), research on micro-expression detection (discovering and locating the onset and offset moments of micro-expressions in videos) remains relatively scarce and faces severe challenges. This section will overview and analyze the research status of micro-expression intelligent analysis from three aspects: micro-expression databases, detection, and recognition.

Ecological validity is the key factor determining whether a micro-expression database is suitable for real-scenario micro-expression analysis and is directly related to micro-expression elicitation paradigms. This study categorizes published micro-expression databases into three generations based on elicitation methods, with ecological validity gradually increasing.

**First-generation micro-expression databases:** In the early stages of micro-expression analysis, sample collection involved recording performers attempting to pose fleeting facial expressions after observing standard expression samples. The USF-HD (Shreve et al., 2011) and Polikovsky (Polikovsky et al., 2009) databases are two posed micro-expression databases. However, micro-expressions are considered spontaneous and difficult to feign; imitated expressions include only external manifestations and lack authenticity. There-

fore, more natural micro-expression elicitation methods are needed to enhance database ecological validity.

**Second-generation micro-expression databases:** To address the aforementioned issues, psychology researchers attempted to elicit micro-expressions using emotional stimuli. For example, the “neutral face” paradigm requires subjects to maintain a neutral expression while viewing emotional stimulus materials, which can elicit spontaneous micro-expressions while avoiding unrelated facial movements. Databases such as the CASME series (Li et al., 2023; Qu et al., 2018; Yan et al., 2013; Yan et al., 2014), SMIC (X. Li et al., 2013), SAMM (A. K. Davison et al., 2016), MMEW (Ben et al., 2022), 4DME (X. Li et al., 2022), and DFME (Zhao et al., 2023) were collected through this method. Although these micro-expression samples possess strong spontaneity and were collected in well-controlled experimental environments, they still differ from real-world contexts. Therefore, to apply micro-expression detection and recognition to practical scenarios, further improvement of sample ecological validity is necessary.

**Third-generation micro-expression databases:** To address these limitations, collecting micro-expression detection samples with higher ecological validity represents an inevitable choice for advancing micro-expression analysis research. Husák et al. released the MEVIEW database (Husák et al., 2017, February) collected in natural environments, containing 40 annotated micro-expressions from poker games and television interview videos. However, these samples also suffer from uncontrollable factors such as lens focal length and head movement. Current micro-expression analysis research still requires collection in well-controlled laboratory environments. To this end, CAS(ME)3 released micro-expression samples collected through simulated crime scenarios, improving ecological validity while eliminating uncontrollable factors (J. Li, Dong, et al., 2023). However, the high ecological validity portion of CAS(ME)3 includes only 166 samples from 31 subjects, and the interactive context is limited to simulated interrogation Q&A, with limited sample size and application scenarios. Therefore, to promote practical application of micro-expression analysis, future work needs to construct more elicitation paradigms that align with real scenarios and collect larger-scale micro-expression data.

**Micro-expression detection** aims to accurately locate tiny and brief micro-expression segments in long videos, following two main approaches: first, comparing inter-frame feature differences, and second, classifying micro-expression frames versus non-micro-expression frames through machine learning.

First, the frame difference method calculates feature differences within temporal windows, filtering out facial movements with motion amplitudes approaching micro-expressions through thresholding. Commonly used features include Local Binary Patterns (LBP) (Moilanen et al., 2014, August), optical flow features MDMO (L.-W. Zhang et al., 2020), 3D-HOG (Davison et al., 2018, May), local spatiotemporal patterns (J. Li, Soladié, et al., 2023), among others. Second, to enhance the ability of detection methods to distinguish micro-expressions from

other facial movements, deep learning-based micro-expression detection methods have gradually become the focus, such as convolutional neural networks for multi-scale micro-expression segment detection in long videos (MESNet) (Wang et al., 2021), local bilinear convolutional neural networks (Pan et al., 2022), LGSNet (Yu et al., 2023), detection methods based on 3DCNN utilizing spatiotemporally oriented reference frames (Yap et al., 2022, October), action unit-aware graph convolutional networks (Yin et al., 2023), and detection models integrating VideoMAE (Xu et al., 2023, October).

However, current deep learning methods remain constrained by small sample sizes and cannot be applied to practical scenarios. Meanwhile, the Micro-Expression Grand Challenge (MEGC) has set micro-expression detection tasks in long videos (A. K. Davison et al., 2023, October; J. Li et al., 2021, October; J. Li et al., 2022, October; LI et al., 2020; See et al., 2019), greatly promoting academic development in micro-expression detection tasks. Comparing the top-three methods over the years reveals that traditional frame difference methods can remove slight head movements and external interference through effective preprocessing, filtering, and appropriate threshold settings, achieving good detection performance. However, in real-world scenarios, interference factors such as head pose and lighting changes cannot be eliminated through simple alignment or filtering, urgently necessitating the construction of effective interference elimination modules to achieve micro-expression detection in complex long videos.

**Micro-expression recognition** refers to classifying micro-expression clips according to specific emotion types, primarily employing two categories of methods: handcrafted feature methods and deep learning methods. Common handcrafted features include methods based on LBP-TOP, HOG, and optical flow. For instance, Li et al. employed a variant of HOG (X. Li et al., 2017), followed by Huang et al. who proposed LBP-TOP based on integral projection to enhance inter-class differences in micro-expression classification (Huang et al., 2017); Wang et al. proposed extracting LBP-TOP features separately in color spaces (Wang et al., 2015), and Liu et al. proposed a micro-expression recognition method based on main directional mean optical flow (Liu et al., 2015). However, due to the short duration and low intensity characteristics of micro-expressions, the representational capacity of these features for micro-expressions needs improvement.

In recent years, micro-expression recognition combined with deep learning has become the mainstream trend, with continuously improving recognition rates. Xie et al. proposed a micro-expression recognition framework that decouples facial motion information and identity information (Xie et al., 2022); Zhang et al. proposed a micro-expression recognition method based on Transformer spatiotemporal feature extraction (L. Zhang et al., 2022); Zong et al. utilized a sparse Transformer network to extract micro-expression-related features (Zhu et al., 2022); Mao et al. achieved objective class-based micro-expression recognition under partial occlusion through region-inspired relation reasoning networks

(Mao et al., 2022); Xia et al. applied networks trained on macro-expressions to micro-expression classification tasks to improve recognition performance (Xia et al., 2020, October); and Zhang et al. proposed a micro-expression recognition method based on graph neural networks (Y. Zhang et al., 2023, July).

Deep learning-based micro-expression analysis methods face small-sample problems for three main reasons: first, small sample sizes cause model overfitting; second, transfer learning-related methods have limited effectiveness in improving micro-expression analysis performance; and third, both sample numbers and network parameters in micro-expression analysis are constrained by sample size. Therefore, constructing a large-scale micro-expression database is crucial for advancing deep learning in intelligent micro-expression analysis.

## 2.2 Research Progress of Related Technologies

**Research Status of EMG-Based Facial Expression Recognition.** Over the past decades, accurate automatic facial expression annotation has been a challenging problem in affective computing, with computer vision-based annotation being the most common approach. In recent years, wearable EMG devices have gradually emerged as an alternative solution. For example, Ang et al. used 3-channel facial EMG signals to recognize angry, happy, and sad expressions (Ang et al., 2004, November). Chen et al. designed a head-mounted device to identify five expressions related to eyebrow movements (Chen et al., 2015). Sato et al. also designed a wearable facial EMG acquisition device to measure emotional valence (Sato et al., 2021). Compared with traditional facial EMG measurements using 8 electrode sets, Schultz et al. used only 4 sets (frontalis, corrugator supercilii, zygomaticus major), while the expression recognition rate decreased by less than 5% (Schultz & Pruzinec, 2010). Hamed et al. achieved 87% accuracy in distinguishing ten facial actions using 3 electrodes placed on the frontalis and temporalis muscles (Hamed et al., 2013). Lu et al. first characterized micro-expression motion intensity through facial action unit-based facial EMG measurements (Lu et al., 2022, October).

EMG signals have been widely applied in facial action recognition with high recognition rates. However, traditional experiments place electrodes directly on facial muscles, which, while ensuring signal acquisition quality, creates unnatural sensations, and electrode weight may affect facial movements. To address this issue, Gruebler et al. adopted distal electrode placement for facial muscle signal acquisition, showing good recognition results for smiling and frowning (Gruebler & Suzuki, 2014). Monica et al. also successfully achieved smile action unit detection through distal electrodes and computer vision methods (Perusquía-Hernández et al., 2021, December).

Due to the weak and transient characteristics of micro-expressions, visual-level annotation is extremely difficult. Therefore, designing a facial peripheral EMG signal capture system can effectively improve the efficiency of micro-expression coding.

**Research Status of Federated Learning.** Among numerous resource-constrained edge devices, large datasets exist. Federated learning technology can utilize these distributed data to train global models, thereby improving data representation accuracy and model applicability (Konečný et al., 2016). Simultaneously, it meets privacy protection requirements while ensuring data transmission compliance (Bonawitz et al., 2017). Federated learning systems consist of a central server and sub-nodes (edge devices), where sub-nodes train models locally on private data and the server aggregates models from each sub-node to update the global model. Federated learning has been applied in multiple domains. For instance, Google implemented federated learning in keyboard applications for Android users, advancing recommendation system development (Mansour et al., 2020); Cohen et al. noted that federated learning can solve problems of small data volume and insufficient labels while ensuring patient privacy (Price & Cohen, 2019).

Synchronous federated learning methods often suffer from decreased synchronization efficiency due to lagging nodes. To address this issue, asynchronous training is widely applied in traditional distributed stochastic gradient descent to cope with lagging nodes and heterogeneous delays (Lian et al., 2018, July). Examples include the FedMDS framework (Y. Zhang, Liu, et al., 2023) and FedAAM (R. Lu et al., 2024).

With the maturation of asynchronous federated technology, data privacy and transmission difficulties in micro-expression application scenarios have been effectively resolved. Through asynchronous federated learning frameworks, micro-expression intelligent analysis model deployment for multi-application scenarios can be achieved while considering node sample differences and transmission delays.

## 2.3 Contributions of This Paper

As facial cues that leak when individuals attempt to hide their true emotions, micro-expressions are widely applied in medical, public security, and other domains. However, the application of intelligent micro-expression analysis faces numerous challenges, such as small sample problems, insufficient model performance in complex scenarios, and data privacy and transmission limitations. To promote practical application of micro-expression analysis in specific scenarios (as shown in

), this research constructs systematic solutions around core challenges: addressing the scarcity of high ecological validity samples by designing multimodal real-interaction context micro-expression elicitation paradigms based on emotional-behavioral psychology principles; tackling manual annotation efficiency bottlenecks by developing a facial peripheral EMG signal-driven automated coding assistance system to improve annotation accuracy; overcoming interference from head movements and lip changes on detection by developing interference elimination modules based on motion compensation mechanisms; addressing insuffi-



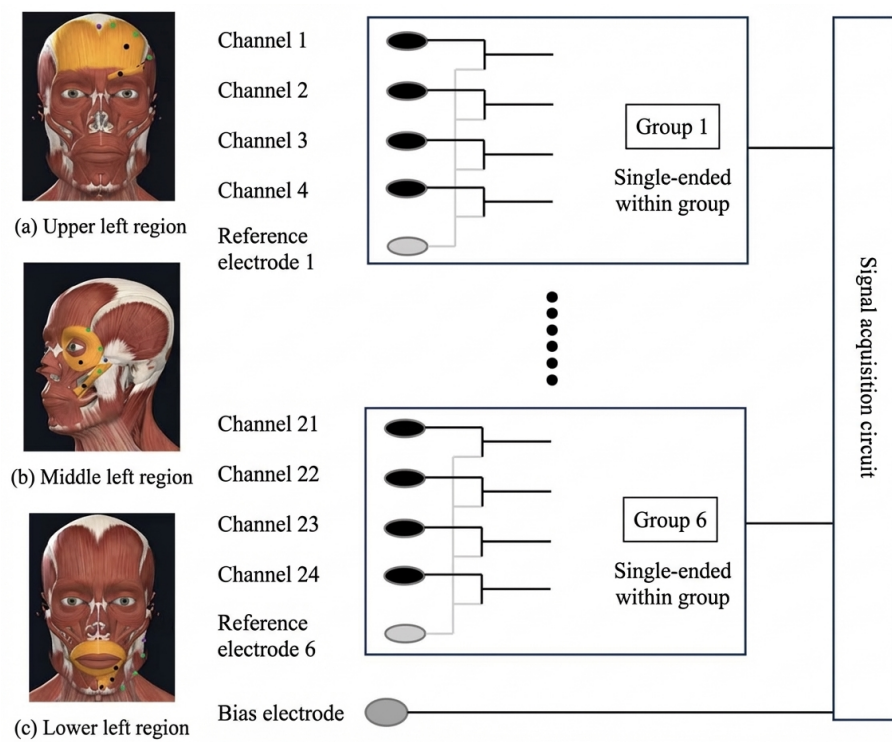


Figure 1: Figure 2



cient representation capability of deep learning models by constructing vertical domain-adapted self-supervised learning architectures to enhance the robustness of micro-expression feature extraction; and confronting data privacy and transmission limitations by innovatively integrating asynchronous federated learning frameworks with offline media transmission protocols to achieve secure and compliant model iterative updates. Based on interdisciplinary integration of psychology, computer science, and other disciplines, this study aims to establish a micro-expression intelligent analysis framework with solid theoretical foundations, promote performance improvements in micro-expression technology, and facilitate practical application of non-contact psychological state monitoring.

### 3 Research Framework

As shown in

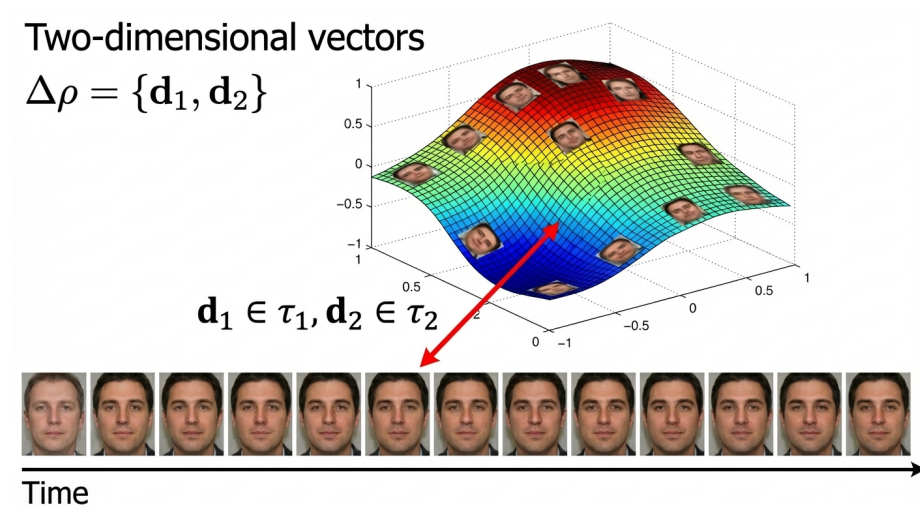


Figure 2: Figure 3

, facing core challenges encountered by micro-expression intelligent analysis in specific application scenarios, this study first designs facial peripheral EMG-based micro-expression auxiliary coding methods grounded in physiological and behavioral psychology research to construct multimodal, high ecological validity databases across different scenarios. Based on this data foundation, this research focuses on solving technical challenges in micro-expression intelligent analysis for specific application scenarios and achieving micro-expression intelligent analysis deployment across different scenarios based on federated learning. By integrating knowledge and technologies from multiple disciplines including psychology and computer science, this study aims to establish a micro-expression intelligent analysis framework that possesses both theoretical foundations and meets practical application requirements.

### 3.1 Basic Theory and Models

**Research on Micro-Expression Mechanisms in Specific Application Scenarios.** Addressing challenges of scarce high ecological validity micro-expression data and unclear micro-expression behavioral and physiological mechanisms, this study relies on psychological research foundations to investigate micro-expression mechanisms in interactive contexts. First, effective experimental paradigms for eliciting high ecological validity micro-expressions are designed to collect micro-expression data in heterogeneous scenarios. Subsequently, an auxiliary coding system based on facial peripheral EMG is designed to overcome time and labor consumption issues in manual coding, constructing a large-scale high ecological validity micro-expression database. Based on collected data, the behavioral and physiological variation patterns of micro-expressions in different interactive contexts are studied.

**Basic Model for Micro-Expression Intelligent Analysis in Specific Scenarios.** Head pose and lip movement changes interfere with detection in video data obtained from real, complex scenarios, causing distortion in the temporal and spatial features of micro-expressions and thereby affecting detection accuracy and recall rates. This study will develop efficient pluggable processing algorithms to accurately identify and eliminate interference caused by head movements and speaking, ensuring accurate extraction of micro-expression features. Through large-scale data acquisition, interference elimination modules, and mature micro-expression detection algorithms, efficient and low-cost solutions are provided for specific scenario applications.

**Self-Supervised Learning Model for Micro-Expressions Based on Large Vertical Domain Models.** To address small-sample problems in micro-expression analysis, this study proposes a self-supervised learning model for micro-expressions based on large vertical domain models. Leveraging developments in visual large models, this study constructs a two-level downstream task framework based on vertical domains. First, based on large amounts of unlabeled expression data, the model learns intensity variation patterns of facial action units; then, using small amounts of labeled data for micro-expression recognition training, it overcomes small-sample limitations. Through the image feature mining capabilities of large models, facial action variation patterns and micro-expression features are learned hierarchically, improving micro-expression recognition performance.

**Asynchronous Federated Intelligent Analysis System for Micro-Expressions.** Given data security concerns in micro-expression intelligent analysis application scenarios, this study plans to adopt asynchronous federated learning deployment methods. Without directly sharing raw data, data from different sources is utilized for model training, improving model generalization capability and accuracy. Federated learning ensures data privacy and security while enabling models to learn from multi-source data and improve performance. To enhance model adaptability to small samples and specific

scenarios, self-supervised learning and reinforcement learning will be combined at the sub-node level. Furthermore, this architecture not only achieves efficient model training and updating but can also adapt to dynamically changing application environments, providing security guarantees for the development of micro-expression intelligent analysis technology and its deployment in sensitive domains.

### 3.2 Key Technologies

**Design of High Ecological Validity Psychology Paradigms.** As described in Section 2, compared with the unidirectional passive elicitation paradigms commonly adopted by existing micro-expression databases (such as eliciting micro-expressions through viewing video stimuli), this study innovatively constructs a multimodal interactive micro-expression elicitation framework: on one hand establishing situational pressure through classic lie-detection paradigms, and on the other hand introducing interpersonal interaction paradigms through role-playing to construct natural social scenarios. This combined application of dual paradigms can help achieve the following advances: 1) In addition to passive reception, experimental stimuli incorporate dynamic interaction, making the micro-expression elicitation mechanism more aligned with authentic human psychology during interpersonal communication; 2) By placing subjects in task environments with real-world mapping relationships, the study effectively enhances the psychological authenticity of emotion elicitation.

In summary, through four experimental paradigms with different ecological validities, the ecological validity of elicited micro-expressions is enhanced, exploring the diversity and complex mechanisms of micro-expressions from multiple angles and contexts, further tapping the application potential of micro-expressions in real scenarios, and improving automated analysis accuracy. Additionally, the multidimensional framework based on multiple experimental paradigms helps deepen understanding of the complex relationships between micro-expressions, emotions, and cognitive processes, providing theoretical support for their application in specific scenarios.

- a) **Video Elicitation Paradigm.** This paradigm serves as the traditional elicitation method for baseline reference in micro-expression databases. Specifically, video clips with high emotional valence are used as stimuli to elicit expressions, including seven emotions (happiness, anger, sadness, fear, anger, surprise, and neutral). Each emotion includes 2-3 videos, with each video lasting 1-3 minutes. These videos are the elicitation materials used in the CASME database series.
- b) **Intentional Deception Paradigm in Interactive Contexts.** The experiment employs face-to-face questioning with yes/no answers. Every two subjects form a group, playing the roles of “liar” and “lie detector” respectively, with roles swapped after each round of questioning. Before the experiment begins, subjects must complete a questionnaire containing

baseline questions, factual questions, and preference questions based on actual circumstances.

- c) **Active Lying Paradigm.** In this paradigm, the experimenter sets a speech topic such as “My Childhood,” requiring subjects to sit in a chair and deliver a free speech based on the topic, with the entire content being lies. Subjects are informed that after the experiment, others will watch the speech video, and they will receive monetary rewards if they can convince viewers that the speech content is true.
- d) **Simulated Crime Paradigm.** This study sets up a simulated crime room and a simulated interrogation room. In the simulated crime room, subjects can open boxes to find target items. Based on different behaviors, subjects are categorized as perpetrators, informants, or innocents. In the simulated interrogation room, subjects are informed that a theft has occurred in the laboratory and they have become key suspects. They must confront a newly developed computer lie-detection algorithm, with high rewards for successfully proving their innocence or partial deduction of participant fees otherwise.

**Multimodal Data Acquisition.** In scientific research, particularly in affective computing and micro-expression analysis domains, the precision of data acquisition is crucial, as its quality directly affects research accuracy and reliability. To capture subtle facial expression changes and related physiological signals as much as possible, as shown in [FIGURE:5], this study employs a high-performance equipment combination for data collection and analysis.

This study uses 4K, 120fps high-definition cameras for video recording. 4K resolution ensures clear presentation of facial details, providing rich facial spatial information for subsequent intelligent analysis. The 120fps high frame rate guarantees continuity of facial movements, supporting subsequent dynamic feature extraction. Simultaneously, depth cameras are used to obtain scene depth information, effectively eliminating visual interference caused by head or body movements and improving micro-expression detection and recognition accuracy, particularly under complex backgrounds or lighting conditions.

Additionally, multi-channel physiological instruments are used to record participants' physiological changes, including respiration, pulse, heart rate, and skin conductance. The emergence of micro-expressions is often accompanied by physiological reactions such as accelerated heartbeat and respiratory changes, providing objective indicators for micro-expression research. Through comprehensive analysis of facial micro-expressions and physiological data, emotional states can be fully understood, and the occurrence mechanisms and physiological foundations of micro-expressions can be deeply explored.

This study will also incorporate facial thermal cameras. Changes in facial temperature can reflect blood flow variations, indirectly indicating physiological changes in the body when emotions are elicited. The non-contact acquisition method of facial temperature and micro-expressions can provide new applica-

tion methods for lie detection, further expanding the research boundaries of micro-expression analysis in psychology and physiology.

**Facial Peripheral EMG-Based Auxiliary Coding System.** For both data-driven deep learning methods and statistical analysis-based behavioral mechanism analysis, facial micro-expression database sample sizes are significantly insufficient. In database construction, micro-expression annotation is a task requiring precision, often necessitating trained annotators to observe videos frame-by-frame to determine onset, apex, and offset frames of facial actions—a method extremely time-consuming and labor-intensive. Therefore, this study proposes using facial EMG signals to assist annotators in achieving semi-automated annotation of expression databases.

To achieve facial peripheral EMG signal acquisition, this study designs a multi-channel flexibly configurable EMG signal acquisition system. Directly attaching electrodes and related wires on facial muscle activity areas may impose psychological or physical burdens on subjects, thereby affecting natural emotional leakage and interfering with micro-expression authenticity. Moreover, electrodes and their wires would obstruct key facial regions, severely impacting model extraction and recognition of relevant facial detail features for subsequent computer vision-based facial expression (especially micro-expression) video analysis research. Therefore, we innovatively propose a facial peripheral EMG acquisition scheme. As shown in [FIGURE:6], by strategically placing electrodes in peripheral facial regions, we aim to completely avoid obstructing facial expressions, neither interfering with natural facial muscle movements nor leveraging the volume conduction effect of bioelectrical signals to objectively capture EMG signals generated by source muscle activity. Even when electrodes are not directly attached to source muscles, relevant muscle activity signals can still be detected by peripheral electrodes through tissue conduction. Our previous comparison with commercial EMG systems demonstrated the reliability of this self-developed EMG system (Zhang et al., 2023).

Regarding signal acquisition, as described in the previous paragraph, we do not directly measure target muscle activity but instead capture “crosstalk” signals from source muscle electrical activity transmitted to adjacent facial peripheral regions—the phenomenon of source muscle electrical activity propagating to and being detected in neighboring non-target muscle areas. To effectively process these mixed signals, we adopt a grouped single-level configuration scheme (as shown in Figure 3.3). Specifically, based on facial muscle anatomical structures and their correspondence with facial action units, we divide the face into six symmetric regions (three on each side). Since existing research results show that the basic proportional relationships of facial thirds and fifths and the positional structural relationships of the nose, mouth, and eyes do not change across individuals or expressions (Zhang, 2022).

Based on this foundation, combined with the symmetry of facial nerve innervation and the dynamic expression characteristics of emotions, we partition facial regions. During partitioning, we first locate facial regions according to verti-

cal order, initially dividing them into upper regions—the frontal region (covering frontalis and corrugator supercilii, primarily for eye-brow area movements), middle regions—the orbital-zygomatic region (covering orbicularis oculi and zygomaticus major muscle groups, mainly for expressing core emotions such as happiness), and lower regions—the oral-mandibular region (covering depressor anguli oris and platysma muscles, primarily for expressing social intentions). Subsequently, considering the asymmetric activation of left and right hemifaces during emotional expression (for instance, emotional intensity is stronger on the left face than the right during expression). Finally, the activation sequence of facial regions during emotion generation also differs, proceeding from front to back as: upper region, middle region, and lower region, which also aligns with the “cognition-emotion-behavior” neural transmission pathway. In summary, we divide the face into six regions: left upper (containing left frontalis, left corrugator supercilii), left middle (containing left orbicularis oculi, left zygomaticus major), left lower (containing left depressor anguli oris, left platysma), right upper (containing right frontalis, right corrugator supercilii), right middle (containing right orbicularis oculi, right zygomaticus major), and right lower (containing right depressor labii inferioris, right mentalis).

Within each region, as shown in [FIGURE:7], a single-level electrode configuration is adopted (4 recording electrodes plus 1 reference electrode), aiming to maximize collection of local “crosstalk” signals within that region, which primarily originate from target AU-related muscles beneath the region (such as corrugator supercilii, zygomaticus major, etc.). Through this grouped design and consideration of inter-regional distances, we primarily confine mixed signal sources within target regions, thereby significantly reducing interference from muscle activities in other facial regions. Thus, each group obtains “crosstalk” from its region, making the acquired EMG signals better reflect the activity state of target regions and improving signal relative accuracy and regional specificity. However, distal EMG measurement means that signals collected by a single electrode channel during facial actions may be combinations of signals from multiple muscle sources to varying degrees. We employ Independent Component Analysis (ICA) signal processing algorithms to separate independent signal sources related to specific facial action units from acquired mixed EMG signals. Subsequently, EMG signal onset and offset moments are detected. Specifically, surface EMG (sEMG) signals are superpositions of numerous motor unit action potentials (MUAPs), and it is widely accepted that actual sEMG signal amplitude distributions are non-Gaussian, especially during muscle activity periods (Sato & Kochiyama, 2023), which satisfies a key ICA assumption. However, facial muscles are dense and exhibit synergistic movements; signals recorded by single electrodes may indeed contain activities from multiple muscle sources and crosstalk, making the assumption of completely independent source signals difficult to perfectly satisfy in practice. Nevertheless, this is precisely an important motivation for applying ICA to process facial EMG signals. As Sato et al. noted, facial EMG signals are susceptible to crosstalk and interference from non-expression actions (such as speaking, chewing), and ICA can separate these



mixed signals, thereby reducing crosstalk effects and extracting purer target muscle activity signals (Sato & Kochiyama, 2023). Our proposed multi-channel regional acquisition design combined with ICA aims to separate dominant, relatively independent signal sources as much as possible under signal mixing conditions to improve regional activity detection accuracy. While guaranteeing absolutely physiologically independent muscle unit signals cannot be ensured, ICA serves here as an effective signal processing tool for improving target signal signal-to-noise ratio and discriminability. Additionally, it must be emphasized that the primary goal of using ICA in this paper is to accurately locate facial muscle action occurrence regions and action onset/offset timing, rather than precisely quantifying the motion intensity of each independent muscle unit (or ICA-separated component).

In this study, the core value of the facial peripheral EMG system lies in providing coders with objective, real-time facial muscle activity information as an efficient auxiliary tool. Given that micro-expressions are characterized by short duration and weak amplitude, they are often difficult for the human eye to accurately capture and code, while EMG signals can precisely record these subtle muscle activities. Therefore, with synchronously recorded EMG data, coders can obtain objective physiological indicators regarding whether specific AUs (facial action units) are activated, activation intensity, and duration, significantly improving micro-expression annotation speed and accuracy. It must be clarified that although subjects wear facial peripheral EMG devices during video acquisition, since we employ a peripheral facial electrode layout that does not obstruct core facial regions directly related to expression production, these synchronous video samples with precise timestamps can themselves be directly used for training facial micro-expression recognition models. Thus, facial peripheral EMG technology offers dual advantages: on one hand, it greatly improves training dataset label quality and annotation efficiency; on the other hand, synchronously acquired video samples effectively expand micro-expression databases, helping alleviate the long-standing small-sample problem in this field.

In summary, our proposed facial peripheral EMG acquisition scheme aims to balance the need for objective physiological signal measurement with the need to avoid interfering with subjects' natural expressions and subsequent visual analysis, ensuring signal quality through specific electrode configurations and signal processing strategies, and positioning it as an effective auxiliary means for improving micro-expression manual coding efficiency and accuracy, thereby laying a foundation for subsequent high-quality dataset construction and intelligent recognition model research.

**Micro-Expression Mechanism Research Based on Statistical Analysis.** After data cleaning and organization of annotated micro-expression samples, descriptive statistical analysis is first conducted on the data, including micro-expression frequency and duration, as well as mean values and standard deviations of physiological data features such as heart rate and skin conductance activity. Subsequently, comparative analysis across paradigms is performed, ex-



amining differences in micro-expressions under three paradigms: passive lying, active expression (lying), and simulated crime. Data across different paradigms are compared using t-tests, with p-values calculated to determine result significance. Based on statistical test results, comparisons are interpreted, discussing possible reasons and significance of differences across paradigms. Additionally, Pearson correlation coefficients are used to explore associations between micro-expression frequency, intensity, and physiological responses (such as heart rate changes). For example, Pearson correlation coefficients between micro-expression frequency (X) and heart rate (Y) are calculated, and the correlation strength is evaluated based on obtained r-values.

**Video Segmentation Technology Based on Low-Dimensional Manifold Head Pose Sparse Features.** Head movements during interaction severely impact micro-expression detection. In preprocessing, traditional processing pipelines often remove head movements based on facial rigid structures, i.e., the nose region. However, this method can only handle cases with very small head movement amplitudes in videos. Considering that individuals rarely undergo dramatic head pose changes during interaction, particularly during micro-expression leakage, as shown in [FIGURE:8], this study proposes a video segmentation method based on low-dimensional manifold head pose sparse features, dividing complex long videos into short video segments with gentle head pose changes for micro-expression detection.

Specifically, head pose information of individuals is first extracted using computer vision tools, and head pose sparse features are extracted based on low-dimensional manifolds. Then, through manifold learning based on locally linear embedding, adjacency graphs between data points are constructed to preserve local or global geometric structures of data points, ultimately obtaining embedded representations of data in low-dimensional spaces that reveal intrinsic geometric structures in head movement data while preserving key characteristics of head motion. After obtaining low-dimensional representations, sparse coding techniques are used to select the most representative features, which can accurately reconstruct the translation and rotation dynamics of head movements with minimal elements.

**Multimodal Lip Movement Change Capture Algorithm Combining Video, Speech, and Text.** In addition to head movement effects on micro-expression detection, lip movement changes caused by speaking constitute another important interference factor. Relying solely on visual signal processing may lead to elimination of non-speaking lip movements, consequently losing corresponding emotional expressions. Therefore, as shown in [FIGURE:9], this study proposes a multi-channel algorithm based on text, speech, and visual signals that can not only capture lip movement changes but also extract emotional background information based on linguistic information.

First, the Wav2Vec network is used to extract text information from raw audio waveforms in the video's audio track. Subsequently, phoneme alignment techniques are employed to temporally align extracted text with dialogue in

the video. Based on alignment results, dialogue segments in the video are segmented. On one hand, since micro-expressions often occur during interaction, features from video and speech signals during these stages are extracted to learn emotional cues manifested in the video. On the other hand, through voiceprint feature registration and matching, the experimenter (questioner) and subject (micro-expression research target) are distinguished, selecting video segments when the subject is speaking. After determining the speaking moments of the speaker, facial preprocessing is performed to extract lower facial features and compare feature differences based on sliding windows, with lip movement changes extracted using specific strategies. In subsequent micro-expression detection processes, segments with lip movement changes receive special treatment, focusing only on their upper face. Additionally, based on text features, speech features, and video features, emotional background information extraction is achieved through feature fusion.

**Self-Supervised Learning Model for Micro-Expressions Based on Large Vertical Domain Models.** The foundation of micro-expression recognition is the model's ability to effectively learn facial action features. However, the inherently weak motion characteristics of micro-expressions and small sample sizes make robust micro-expression recognition algorithms difficult to achieve. Meanwhile, visual large models possess extremely strong image feature learning capabilities. As shown in [FIGURE:10], leveraging vertical domain large models, this study establishes a two-level downstream task framework based on sample scales. First, large-scale expression data is used to develop large models' ability to mine facial expression features through self-supervised learning tasks; then, further fine-tuning is performed on smaller-scale micro-expression datasets. Through hierarchical progression from large models to specific niche domains, micro-expression recognition efficacy is improved.

First, for vertical domain facial action learning, this study constructs sample pairs through facial action unit (AU) intensity. Specifically, videos in large expression video databases record rich facial expressions. Appropriate samples are selected from these videos as candidate sets to obtain samples more suitable for self-supervised networks to learn facial action features. The [OpenFace] toolkit is first used to detect AU activation state information. Subsequently, to exclude interference from individual biometric features during comparison, onset-apex sample pairs are constructed to enable the network to focus on facial actions. Therefore, this step comprises two stages: one part is the selection of apex frames with maximum motion amplitude, and the other is the inference of corresponding onset frames for the apex. Frames with AU intensity greater than a set threshold are designated as apex frames for the current sample pair. According to the conventional definition of micro-expression duration, i.e., less than 500ms, and considering that frames with maximum motion amplitude are mostly distributed in the middle region of micro-expression segments, after obtaining apex frames, based on video frame rate (30fps in this example), 8-9 frames are pushed forward as candidate onset frames. Similarly, for non-RGB

derived modalities such as depth modality, we preserve the same segments in depth modality sequences according to time.

Subsequently, frame difference estimation is used to highlight regions where facial movements significantly occur, providing representations of temporal changes in micro-expressions. Specifically, let  $I_{onset}^i$  and  $I_{apex}^i$  denote the onset and apex frames in RGB view for the  $i$ -th candidate segment, where  $N$  represents the number of labeled samples. By subtracting onset frames from apex frames, we obtain frame differences  $\Delta I_{RGB}^i = I_{apex}^i - I_{onset}^i$ . Additionally, we extend this method to other modalities of micro-expressions, such as depth and grayscale. For any onset and apex frames in each modality, represented as  $D_{onset}^i$  and  $D_{apex}^i$ , we compute  $\Delta D^i = D_{apex}^i - D_{onset}^i$ .

This study proposes Multi-view Contrastive Learning (MvCL) to distinguish samples from two different distributions. Specifically, for each candidate segment, a corresponding multimodal information segment (additional modality) is provided, such as grayscale or depth maps. “Mutual information” exists between two different modalities. Therefore, feature representations obtained from RGB images and additional modalities of the same facial expression after passing through feature extractors should be close to each other in high-dimensional feature space, while pushing away from feature representations of other expressions. [FIGURE:11] illustrates the self-supervised contrastive learning process using RGB and additional modalities.

Through self-supervised contrastive learning using multi-view contrast, effective learning of micro-expression feature representations from two different perspectives—RGB images and additional modalities—can be achieved. This multimodal self-supervised learning method can utilize mutual information between two modalities to improve model generalization and feature expression capabilities, thereby achieving better performance in micro-expression recognition tasks.

After pre-training, feature representations obtained from self-supervised contrastive learning pre-training are transferred to downstream micro-expression recognition tasks. Specifically, all layers except the linear classification layer in the pre-trained network are first frozen, and a classifier is trained using cross-entropy loss on this basis for final micro-expression classification.

**Asynchronous Federated Intelligent Analysis System.** Due to micro-expression application scenario constraints, this study designs a micro-expression intelligent analysis deployment architecture based on asynchronous federated learning. As shown in [FIGURE:12], the central server first constructs the basic model for intelligent analysis networks, then for different scenarios, parameter uploading and iteration based on delayed rounds and scenario sample distributions are performed in different communication rounds.

Specifically, the central server model, i.e., the central server hosting the basic intelligent analysis network, is responsible for coordinating the learning process and aggregating updates from various nodes. Meanwhile, different nodes send

updates to the central server at different time points, and the server does not need to wait for all nodes to synchronize.

The model parameters, and the central server performs parameter aggregation after receiving updates. In the parameter integration strategy, three main influencing factors are thoroughly considered to achieve more accurate and efficient model training processes.

First, addressing sample number differences across node scenarios, this study adopts a sample parameter weighting method, ensuring nodes with larger sample sizes receive relatively larger weights during parameter integration, a strategy that helps improve overall model learning efficiency and accuracy.

Second, considering inconsistencies in update rounds between nodes and the central server, this study introduces the concept of stale parameters, adjusting the influence of stale parameters by setting decreasing functions negatively correlated with communication rounds, thereby ensuring the central server can effectively utilize the latest global information for parameter updates.

Finally, to address potential communication delay issues in practical applications, this study also sets local node iteration parameters, encouraging nodes to perform self-iteration and optimization of local models while waiting for central server updates, thereby continuously enhancing model adaptability and robustness.

Through the organic combination of these three strategies, this study aims to explore a novel parameter integration mechanism that can both fully utilize the distributed characteristics of federated learning and ensure efficient and accurate model training.

**Incremental Learning Model Based on Specific Scenario Data Accumulation.** Addressing the problem of insufficient micro-expression samples during the initial stage of sub-node model construction, this study proposes a strategy for gradually accumulating and optimizing models. As shown in [FIGURE:13], first, a self-supervised facial action learning model is constructed using existing base models and large amounts of unlabeled data. Subsequently, based on a small batch of micro-expression samples obtained through manual annotation, the self-supervised model is fine-tuned and optimized for downstream tasks. Additionally, the project will employ incremental learning methods to gradually improve model accuracy and robustness through progressive feedback. This approach not only ensures the model possesses certain recognition capabilities in early stages but also allows continuous improvement over time and with data accumulation, ultimately achieving efficient and accurate micro-expression recognition.

Specifically, according to application scenario requirements, a learning process that continuously adapts to new data is designed. In this process, we maintain a dynamically growing dataset composed of already annotated samples.

Let  $S_{labeled}$  be the set of manually annotated samples, and  $S_{correct}$  be the set

of samples correctly recognized by the model. The cumulative dataset  $D = S_{labeled} \cup S_{correct}$  is the union of both. Over time and as the model runs,  $S_{correct}$  will gradually increase, continuously expanding from  $S_{labeled}$ . In each iteration step, the model will be trained using the current cumulative dataset  $D$  to improve its micro-expression recognition capability. After new samples are predicted by the model, if the prediction confidence is high, these samples can be automatically added to  $S_{correct}$ ; otherwise, they may need manual annotation before being added to  $S_{labeled}$ .

## 4 Theoretical Construction and Innovation

As a brief and tiny facial movement that is nearly invisible to the naked eye, micro-expression causes numerous challenges for the development and application of its intelligent analysis. Meanwhile, emotion recognition based on micro-expressions has broad application potential in public and national security domains. This study proposes a practical micro-expression intelligent analysis solution for specific application scenarios, with the following features and innovations:

First, at the psychological level, research on high ecological validity micro-expression elicitation paradigms and their behavioral and physiological mechanisms is conducted. Previous micro-expression databases were mostly collected through “neutral face” elicitation paradigms, with large differences in ecological validity from real scenarios. This study designs interactive context micro-expression elicitation paradigms based on behavioral and emotional psychology, collecting micro-expression data close to real situations. Simultaneously, through psychological statistical analysis of physiological and behavioral data, behavioral and physiological mechanisms of micro-expressions under different interactive scenarios are revealed, laying theoretical foundations for micro-expression-based emotional monitoring.

Second, at the technical level, this study constructs a sensitive and robust basic model for micro-expression intelligent analysis. The research designs a distributed facial peripheral EMG acquisition system based on facial muscle action topological structures, improving micro-expression coding efficiency. Meanwhile, head pose segmentation and multimodal lip movement change recognition technologies are employed to reduce interference factors in complex videos and improve micro-expression detection accuracy. Through self-supervised learning models, micro-expression recognition performance is enhanced despite limited micro-expression sample sizes.

Finally, in terms of application promotion, this study constructs a micro-expression intelligent analysis cloud+local deployment architecture matching different specific application scenarios. Addressing data privacy and transmission limitations, an asynchronous federated learning framework is adopted, combining incremental learning sub-node models and central server models to support deployment and verification in real application scenarios.

## References

- Ang, L. B. P., Belen, E. F., Bernardo, R. A., Boongaling, E. R., Briones, G. H., & Coronel, J. B. (2004). Facial expression recognition through pattern analysis of facial muscle movements utilizing electromyogram sensors. *2004 IEEE Region 10 Conference TENCN 2004.*, C, 600-603 Vol. 3. <https://doi.org/10.1109/TENCN.2004.1414843>
- Ben, X., Ren, Y., Zhang, J., Wang, S.-J., Kpalma, K., Meng, W., & Liu, Y.-J. (2022). Video-based facial micro-expression analysis: A survey of datasets, features and algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9), 5826–5846. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2021.3067464>
- Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 1175–1191. <https://doi.org/10.1145/3133956.3133982>
- Chen, Y., Yang, Z., & Wang, J. (2015). Eyebrow emotional expression recognition using surface EMG signals. *Neurocomputing*, 168, 871–879. <https://doi.org/10.1016/j.neucom.2015.05.037>
- Darwin, C. (1872). *The descent of man, and selection in relation to sex* (Vol. 2). D. Appleton.
- Davison, A. K., Lansley, C., Costen, N., Tan, K., & Yap, M. H. (2016). Samm: A spontaneous micro-facial movement dataset. *IEEE Transactions on Affective Computing*, 9(1), 116–129.
- Davison, A. K., Li, J., Yap, M. H., See, J., Cheng, W.-H., Li, X., Hong, X., & Wang, S.-J. (2023). MEGC2023: ACM multimedia 2023 ME grand challenge. *Proceedings of the 31st ACM International Conference on Multimedia*, 9625–9629. <https://doi.org/10.1145/3581783.3612833>
- Davison, A., Merghani, W., Lansley, C., Ng, C.-C., & Yap, M. H. (2018). Objective micro-facial movement detection using FACS-based regions and baseline evaluation. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 642–649. <https://doi.org/10.1109/FG.2018.00101>
- Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception†. *Psychiatry*, 32(1), 88–106. <https://doi.org/10.1080/00332747.1969.11023575>
- Frank, M., Herbasz, M., Sinuk, K., Keller, A., & Nolan, C. (2009). I see how you feel: Training laypeople and professionals to recognize fleeting emotions. *The Annual Meeting of the International Communication Association*. Sheraton New York, New York City, 1–35.
- Gruebler, A., & Suzuki, K. (2014). Design of a wearable device for reading positive expressions from facial EMG signals. *IEEE Transactions on Affective*

*tive Computing*, 5(3), 227–237. IEEE Transactions on Affective Computing. <https://doi.org/10.1109/TAFFC.2014.2313557>

Hamed, M., Salleh, S.-H., Astaraki, M., & Noor, A. M. (2013). EMG-based facial gesture recognition through versatile elliptic basis function neural network. *BioMedical Engineering OnLine*, 12(1), 73. <https://doi.org/10.1186/1475-925X-12-73>

Huang, X., Wang, S.-J., Liu, X., Zhao, G., Feng, X., & Pietikäinen, M. (2019). Discriminative spatiotemporal local binary pattern with revisited integral projection for spontaneous facial micro-expression recognition. *IEEE Transactions on Affective Computing*, 10(1), 32–47. IEEE Transactions on Affective Computing. <https://doi.org/10.1109/TAFFC.2017.2713359>

Husák, P., Cech, J., & Matas, J. (2017). Spotting facial micro-expressions “in the wild”. *22nd Computer Vision Winter Workshop (Retz)*, 1–9. <http://cmp.felk.cvut.cz/~cechj/ME/>

Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2017). Federated learning: Strategies for improving communication efficiency (No. arXiv:1610.05492). arXiv. <https://doi.org/10.48550/arXiv.1610.05492>

Li, J., Dong, Z., Lu, S., Wang, S.-J., Yan, W.-J., Ma, Y., Liu, Y., Huang, C., & Fu, X. (2023). 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*, 45(3), 2782–2800. IEEE Transactions on Pattern Analysis and Machine Intelligence. <https://doi.org/10.1109/TPAMI.2022.3174895>

Li, J., Soladié, C., & Séguier, R. (2023). Local temporal pattern and data augmentation for spotting micro-expressions. *IEEE Transactions on Affective Computing*, 14(1), 811–822. IEEE Transactions on Affective Computing. <https://doi.org/10.1109/TAFFC.2020.3023821>

LI, J., Wang, S.-J., Yap, M. H., See, J., Hong, X., & Li, X. (2020). MEGC2020—The third facial micro-expression grand challenge. *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, 777–780. <https://doi.org/10.1109/FG47880.2020.00035>

Li, J., Yap, M. H., Cheng, W.-H., See, J., Hong, X., Li, X., & Wang, S.-J. (2021). FME’21: 1st workshop on facial micro-expression: advanced techniques for facial expressions generation and spotting. *Proceedings of the 29th ACM International Conference on Multimedia*, 5700–5701. <https://doi.org/10.1145/3474085.3478579>

Li, J., Yap, M. H., Cheng, W.-H., See, J., Hong, X., Li, X., Wang, S.-J., Davison, A. K., Li, Y., & Dong, Z. (2022). MEGC2022: ACM multimedia 2022 micro-expression grand challenge. *Proceedings of the 30th ACM International Conference on Multimedia*, 7170–7174. <https://doi.org/10.1145/3503161.3551601>



- Li, X., Cheng, S., Li, Y., Behzad, M., Shen, J., Zafeiriou, S., Pantic, M., & Zhao, G. (2022). 4DME: A spontaneous 4D micro-expression dataset with multimodalities. *IEEE Transactions on Affective Computing*, 14(4),
- Li, X., Hong, X., Moilanen, A., Huang, X., Pfister, T., Zhao, G., & Pietikäinen, M. (2018). Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods. *IEEE Transactions on Affective Computing*, 9(4), 563–577. *IEEE Transactions on Affective Computing*. <https://doi.org/10.1109/TAFFC.2017.2667642>
- Li, X., Pfister, T., Huang, X., Zhao, G., & Pietikäinen, M. (2013). A spontaneous micro-expression database: Inducement, collection and baseline. *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 1–6. <https://doi.org/10.1109/FG.2013.6553717>
- Lian, X., Zhang, W., Zhang, C., & Liu, J. (2018). Asynchronous decentralized parallel stochastic gradient descent. *Proceedings of the 35th International Conference on Machine Learning*, 3043–3052. <https://proceedings.mlr.press/v80/lian18a.html>
- Liu, Y.-J., Zhang, J.-K., Yan, W.-J., Wang, S.-J., Zhao, G., & Fu, X. (2016). A main directional mean optical flow feature for spontaneous micro-expression recognition. *IEEE Transactions on Affective Computing*, 7(4), 299–310. *IEEE Transactions on Affective Computing*. <https://doi.org/10.1109/TAFFC.2015.2485205>
- Lu, R., Zhang, W., Li, Q., He, H., Zhong, X., Yang, H., Wang, D., Xu, Z., & Alazab, M. (2024). Adaptive asynchronous federated learning. *Future Generation Computer Systems*, 152, 193–206. <https://doi.org/10.1016/j.future.2023.11.001>
- Lu, S., Li, J., Wang, Y., Dong, Z., Wang, S.-J., & Fu, X. (2022). A more objective quantification of micro-expression intensity through facial electromyography. *Proceedings of the 2nd Workshop on Facial Micro-Expression: Advanced Techniques for Multi-Modal Facial Expression Analysis*, 11–17. <https://doi.org/10.1145/3552465.3555038>
- Mansour, Y., Mohri, M., Ro, J., & Suresh, A. T. (2020). Three approaches for personalization with applications to federated learning (No. arXiv:2002.10619). *arXiv*. <https://doi.org/10.48550/arXiv.2002.10619>
- Mao, Q., Zhou, L., Zheng, W., Shao, X., & Huang, X. (2022). Objective class-based micro-expression recognition under partial occlusion via region-inspired relation reasoning network. *IEEE Transactions on Affective Computing*, 13(4), 1998–2016. *IEEE Transactions on Affective Computing*. <https://doi.org/10.1109/TAFFC.2022.3197785>
- Moilanen, A., Zhao, G., & Pietikäinen, M. (2014). Spotting rapid facial movements from videos using appearance-based feature difference analysis. *2014 22nd International Conference on Pattern Recognition*, 1722–1727. <https://doi.org/10.1109/ICPR.2014.303>

Pan, H., Xie, L., & Wang, Z. (2022). Spatio-temporal convolutional emotional attention network for spotting macro- and micro-expression intervals in long video sequences. *Pattern Recognition Letters*, 162, 89–96. <https://doi.org/10.1016/j.patrec.2022.09.008>

Perusquía-Hernández, M., Dollack, F., Tan, C. K., Namba, S., Ayabe-Kanamura, S., & Suzuki, K. (2021). Smile action unit detection from distal wearable electromyography and computer vision. *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, 1–8. <https://doi.org/10.1109/FG52635.2021.9667047>

Polikovskiy, S., Kameda, Y., & Ohta, Y. (n.d.). Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor. *3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009)*, P16/1-P16/6. <https://doi.org/10.1049/ic.2009.0244>

Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature Medicine*, 25(1), 37–43.

Qu, F., Wang, S.-J., Yan, W.-J., Li, H., Wu, S., & Fu, X. (2018). CAS(ME)<sup>2</sup>: A database for spontaneous macro-expression and micro-expression spotting and recognition. *IEEE Transactions on Affective Computing*, 9(4), 424–436. *IEEE Transactions on Affective Computing*. <https://doi.org/10.1109/TAFFC.2017.2654440>

Rinn, W. E. (1984). The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin*, 95(1), 52–77. <https://doi.org/10.1037/0033-2909.95.1.52>

Sato, W., & Kochiyama, T. (2023). Crosstalk in Facial EMG and Its Reduction Using ICA. *Sensors*, 23(5), 2720. <https://doi.org/10.3390/s23052720>

Sato, W., Murata, K., Uraoka, Y., Shibata, K., Yoshikawa, S., & Furuta, M. (2021). Emotional valence sensing using a wearable facial EMG device. *Scientific Reports*, 11(1), 5757.

Schultz, I., & Pruzinec, M. (2010). Facial expression recognition using surface electromyography. *Unpublished Doctoral Dissertation*. *Karlsruhe Institute of Technology*. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=7370f7109318a0ed91ae8a8737>

See, J., Yap, M. H., Li, J., Hong, X., & Wang, S.-J. (2019). MEGC 2019 – the second facial micro-expressions grand challenge. *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, 1–5. <https://doi.org/10.1109/FG.2019.8756611>

Shreve, M., Godavarthy, S., Goldgof, D., & Sarkar, S. (2011). Macro- and micro-expression spotting in long videos using spatio-temporal strain. *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 51–56. <https://doi.org/10.1109/FG.2011.5771451>

Wang, S.-J., He, Y., Li, J., & Fu, X. (2021). MESNet: A convolutional neural

network for spotting multi-scale micro-expression intervals in long videos. *IEEE Transactions on Image Processing*, 30, 3956–3969. IEEE Transactions on Image Processing. <https://doi.org/10.1109/TIP.2021.3064258>

Wang, S.-J., Yan, W.-J., Li, X., Zhao, G., Zhou, C.-G., Fu, X., Yang, M., & Tao, J. (2015). Micro-expression recognition using color spaces. *IEEE Transactions on Image Processing*, 24(12), 6034–6047. IEEE Transactions on Image Processing. <https://doi.org/10.1109/TIP.2015.2496314>

Xia, B., Wang, W., Wang, S., & Chen, E. (2020). Learning from macro-expression: A micro-expression recognition framework. *Proceedings of the 28th ACM International Conference on Multimedia*, 2936–2944. <https://doi.org/10.1145/3394171.3413774>

Xie, T., Sun, G., Sun, H., Lin, Q., & Ben, X. (2022). Decoupling facial motion features and identity features for micro-expression recognition. *PeerJ Computer Science*, 8, e1140.

Xu, K., Chen, K., Sun, L., Lian, Z., Liu, B., Chen, G., Sun, H., Xu, M., & Tao, J. (2023). Integrating VideoMAE based model and optical flow for micro- and macro-expression spotting. *Proceedings of the 31st ACM International Conference on Multimedia*, 9576–9580. <https://doi.org/10.1145/3581783.3612868>

Yan, W.-J., Li, X., Wang, S.-J., Zhao, G., Liu, Y.-J., Chen, Y.-H., & Fu, X. (2014). CASME II: An improved spontaneous micro-expression database and the baseline evaluation. *PLOS ONE*, 9(1), e86041. <https://doi.org/10.1371/journal.pone.0086041>

Yan, W.-J., Wu, Q., Liu, Y.-J., Wang, S.-J., & Fu, X. (2013). CASME database: A dataset of spontaneous micro-expressions collected from neutralized faces. *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 1–7. <https://doi.org/10.1109/FG.2013.6553799>

Yap, C. H., Yap, M. H., Davison, A., Kendrick, C., Li, J., Wang, S.-J., & Cunningham, R. (2022). 3D-CNN for facial micro- and macro-expression spotting on long video sequences using temporal oriented reference frame. *Proceedings of the 30th ACM International Conference on Multimedia*, 7016–7020. <https://doi.org/10.1145/3503161.3551570>

Yin, S., Wu, S., Xu, T., Liu, S., Zhao, S., & Chen, E. (2023). AU-aware graph convolutional network for macro and micro-expression spotting. *2023 IEEE International Conference on Multimedia and Expo (ICME)*, 228–233. <https://doi.org/10.1109/ICME55011.2023.00047>

Yu, W.-W., Jiang, J., Yang, K.-F., Yan, H.-M., & Li, Y.-J. (2024). LGSNet: A two-stream network for micro- and macro-expression spotting with background modeling. *IEEE Transactions on Affective Computing*, 15(1), 223–240. IEEE Transactions on Affective Computing. <https://doi.org/10.1109/TAFFC.2023.3266808>

Zhang, L. (2022). Animation expression control based on facial region division. *Scientific Programming*, 2022(1), 5800099. <https://doi.org/10.1155/2022/5800099>

Zhang, L., Hong, X., Arandjelović, O., & Zhao, G. (2022). Short and long range relation based spatio-temporal transformer for micro-expression recognition. *IEEE Transactions on Affective Computing*, 13(4), 1973–1985. IEEE Transactions on Affective Computing. <https://doi.org/10.1109/TAFFC.2022.3213509>

Zhang, L.-W., Li, J., Wang, S.-J., Duan, X.-H., Yan, W.-J., Xie, H.-Y., & Huang, S.-C. (2020). Spatio-temporal fusion for macro- and micro-expression spotting in long video sequences. *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, 734–741. <https://doi.org/10.1109/FG47880.2020.00037>

Zhang, J., Huang, S., Li, J., Wang, Y., Dong, Z., & Wang, S.-J. (2023). A perifacial EMG acquisition system for facial-muscle-movement recognition. *Sensors*, 23(21), Article 21. <https://doi.org/10.3390/s23218758>

Zhang, Y., Liu, D., Duan, M., Li, L., Chen, X., Ren, A., Tan, Y., & Wang, C. (2023). FedMDS: An efficient model discrepancy-aware semi-asynchronous clustered federated learning framework. *IEEE Transactions on Parallel and Distributed Systems*, 34(3), 1007–1019. IEEE Transactions on Parallel and Distributed Systems. <https://doi.org/10.1109/TPDS.2023.3237752>

Zhang, Y., Wang, H., Xu, Y., Mao, X., Xu, T., Zhao, S., & Chen, E. (2023). Adaptive graph attention network with temporal fusion for micro-expressions recognition. *2023 IEEE International Conference on Multimedia and Expo (ICME)*, 1391–1396. <https://doi.org/10.1109/ICME55011.2023.00241>

Zhao, S., Tang, H., Mao, X., Liu, S., Zhang, Y., Wang, H., Xu, T., & Chen, E. (2024). DFME: A new benchmark for dynamic facial micro-expression recognition. *IEEE Transactions on Affective Computing*, 15(3), 1371–1386. IEEE Transactions on Affective Computing. <https://doi.org/10.1109/TAFFC.2023.3341918>

Zhou, H., Huang, S., Li, J., & Wang, S.-J. (2023). Dual-ATME: Dual-branch attention network for micro-expression recognition. *Entropy*, 25(3), Article 3. <https://doi.org/10.3390/e25030460>

Zhu, J., Zong, Y., Chang, H., Xiao, Y., & Zhao, L. (2022). A sparse-based transformer network with associated spatiotemporal feature for micro-expression recognition. *IEEE Signal Processing Letters*, 29, 2073–2077. IEEE Signal Processing Letters. <https://doi.org/10.1109/LSP.2022.3211200>

Source: ChinaXiv — Machine translation. Verify with original.