2026 International Conference on Automatic Face and Gesture Recognition (FG)

Investigating the Relationship Between Micro-Expressions and Cognitive Load via a Novel Maze Paradigm

Lin Zhao[1,2], Jingting Li[1,2,*], Zizhao Dong[1], Yu Qian[1,3] and Su-Jing Wang[1,2,*]
[1] State Key Laboratory of Cognitive Science and Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, China
[2] Department of Psychology, University of Chinese Academy of Sciences, Beijing, 100049, China
[3] School of Computer Science, Jiangsu University of Science and Technology, Jiangsu, 212028, China

Abstract— **Facial micro-expressions (MEs) are involuntary facial movements that reveal genuine internal states. However, their relationship with cognitive load (CL) remains insufficiently understood, as most existing micro-expression studies rely on emotionally driven elicitation paradigms. In this paper, we introduce a novel maze-solving paradigm designed to induce MEs through controlled task complexity and time pressure, without relying on emotional stimuli. Based on this psychological experiment involving 50 participants, we collected a cognitive-load-oriented facial expression dataset using 4K high-resolution video recordings and five-channel facial electromyography (EMG). By utilizing EMG spikes as objective auxiliary timestamps, we rigorously annotated 1,421 expression segments. Behavioral analyses demonstrate that elevated CL significantly increases ME frequency, with MEs exhibiting higher sensitivity to cognitive depletion than macro-expressions under high-load conditions. Furthermore, we provide robust computational validation by employing a deep learning network to classify CL states directly from ME data. The model achieves 93.51% accuracy in binary classification (Easy vs. High Difficulty) and 65.75% in a granular five-level CL classification task. Our findings bridge psychological theory and affective computing, establishing MEs as a robust, non-intrusive visual marker for objective cognitive load assessment.**

I. INTRODUCTION

Facial expressions serve as a primary modality for social interaction, offering an intuitive window into individuals' emotions and intentions [11]. While humans naturally use facial expressions to convey internal states and infer those of others, evolutionary and social adaptations have equipped individuals with the ability to manipulate or suppress these expressions [13]. However, research indicates that suppressed emotions frequently leak through micro-expressions, rapid and involuntary facial movements characterized by low intensity and brief duration (typically 1/25 to 1/5 of a second), often localized to specific facial regions [8], [25], [19]. Neurologically, this phenomenon is rooted in the dual-pathway control of facial muscles: the cortical motor system governs voluntary facial movements (e.g., intentional masking), while subcortical circuits (such as the amygdala and basal ganglia) drive spontaneous, authentic expressions [28]. Micro-expressions occur as a brief conflict between these two pathways [14].

Corresponding authors: {lijt, wangsujing}@psych.ac.cn. This paper is supported in part by grants from the National Natural Science Foundation of China (62476269, 62276252), and in part, by a grant from the Youth Innovation Promotion Association CAS.

Cognitive load, a multidimensional construct, represents the demands placed on an individual's working memory during task performance [23]. Classical Cognitive Load Theory posits three interacting components: intrinsic load (inherent task complexity), extrinsic load (format of information presentation), and germane load (schema construction) [31]. According to cognitive resource allocation theory, human cognitive resources are strictly limited. As task difficulty escalates, individuals must allocate a disproportionate share of cognitive resources to the primary task, leaving fewer resources for secondary executive functions, such as emotional regulation and facial expression [16], [21]. When executive resources are depleted by high cognitive demand, the top-down cortical suppression mechanism falters, allowing subcortically driven spontaneous expressions to "leak" through [27]. Consequently, high cognitive load significantly impairs top-down emotional control, resulting in increased leakage of genuine emotions via micro-expressions.

Numerous studies employing high-stakes, stress-inducing paradigms (e.g., mock crime scenarios, mock interrogations) have demonstrated that individuals striving to conceal emotions under high psychological pressure and cognitive load exhibit a significantly higher frequency of micro-expressions [32], [9]. Therefore, micro-expressions can theoretically serve as an indirect, objective behavioral indicator of cognitive load. Furthermore, the emergence of micro-expressions may correlate with the depth of cognitive processing [3], as deeper processing monopolizes resources required for expression management.

Despite these theoretical foundations, existing research faces two primary limitations. First, conventional methods for measuring cognitive load predominantly rely on subjective self-reports (e.g., NASA Task Load Index(NASA-TLX)), which are retrospective and susceptible to reporting bias [15], [7], or physiological monitoring (e.g., electroencephalography (EEG)), which is highly intrusive and impractical for real-world applications [1]. Second, mainstream micro-expression elicitation paradigms heavily utilize intense emotional stimuli (e.g., emotive videos or mock crimes) [26]. In such contexts, the overlapping effects of emotional arousal (e.g., fear of being caught) and cognitive load (e.g., fabricating a lie) make it nearly impossible to determine whether the observed micro-expressions are a byproduct of intense emotion or pure cognitive exhaustion.

979-8-3315-7231-0/26/$31.00

Isolating these factors is crucial for developing accurate affective computing models that can operate in non-extreme, everyday cognitive tasks, such as driving, online learning, or complex problem-solving [24], [6].

To systematically investigate this relationship and address these gaps, the present study utilizes a maze-solving paradigm, which allows for precise manipulation of cognitive load through varying task complexity and time constraints. The maze-solving task places heavy demands on visuospatial working memory and executive planning without inherently triggering strong primary emotions [2]. Navigating mazes requires problem-solving strategies, such as trial-and-error, which imposes high cognitive load, or backtracking, which requires fewer resources [22]. Thus, cognitive load in this context is modulated by maze complexity and time pressure [34], effectively isolating cognitive stressors from purely emotional ones and thereby enhancing the ecological validity of the findings.

In summary, this study proposes a mutually influential relationship between cognitive load and micro-expressions, hypothesizing that increased cognitive load will elicit a higher frequency of micro-expressions. To validate this, we conducted a psychological behavioral experiment focusing on multimodal data acquisition, followed by a computational validation using a deep learning classification model.

The main contributions of this work are three-fold:

- We introduce a maze-solving task specifically designed to elicit micro-expressions through pure cognitive load and time pressure, effectively isolating cognitive stressors from the confounding emotional variables present in traditional high-stakes paradigms.
- We propose a controlled paradigm that manipulates cognitive load across two difficulty levels and three task types to collect micro-expression data, with synchronized 4K facial video and facial electromyography (EMG) enabling high-precision multimodal acquisition under varying cognitive load.
- We adopted a cross-disciplinary approach to robustly validate our hypothesis. First, comprehensive statistical analyses confirmed the behavioral mechanism: elevated cognitive load significantly increases the frequency of micro-expression leakage. Second, we provided computational corroboration using a standard deep learning classification model, demonstrating that these cognitively induced expressions possess distinct, machine-readable spatiotemporal patterns, thus proving their viability for automated cognitive state assessment.

II. Related work

A. Cognitive Load and Its Measurement

Cognitive load represents the volume of working memory resources required for task performance. Rooted in Sweller's Cognitive Load Theory (1988), it emphasizes that an inverse relationship exists between excessive cognitive load and task performance [30]. Accurate measurement of cognitive load is crucial and typically falls into three categories: subjective, behavioral, and physiological. Subjective evaluations often utilize the NASA Task Load Index (NASA-TLX) [37], a widely adopted psychometric instrument assessing workload across six dimensions (Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration). Behavioral markers include reaction times and error rates. Physiological measurements traditionally rely on EEG, functional near-infrared spectroscopy (fNIRS), or pupillometry [33], [10]. However, while highly accurate, these conventional physiological methods are often intrusive and require cumbersome equipment, limiting their ecological validity and deployment in real-world applications. This highlights the growing need for non-intrusive, vision-based cognitive load assessment methods.

B. Micro-expression Elicitation and Recognition

Micro-expressions are transient, low-intensity facial movements that reveal genuine, suppressed emotions, typically lasting between 1/25 and 1/5 of a second. In recent years, micro-expression analysis has garnered significant attention in the field of affective computing. Traditional psychological research heavily relies on high-stakes, stress-inducing paradigms (e.g., mock crime scenarios or intensely emotive video stimuli) to elicit these subtle movements [18]. On the computational front, the field has transitioned from hand-crafted spatiotemporal features (e.g., LBP-TOP [20]) to advanced deep learning architectures, such as dual-stream networks and attention mechanisms, to capture subtle facial dynamics [35]. Despite these methodological and algorithmic advances, a key limitation remains in the elicitation paradigms used to collect micro-expression data. Existing datasets predominantly rely on strong emotional stimuli to trigger facial reactions, such as high-stakes scenarios or emotionally evocative videos [4]. However, in many real-world human–computer interaction scenarios, facial expression may reflect non-emotional signals, such as fluctuations in cognitive load. This lacking between current data collection paradigms and practical application scenarios limits our understanding of how micro-expressions emerge under non-emotional cognitive stressors. Consequently, the mechanisms by which cognitive load influences the occurrence of micro-expressions remain largely unexplored.

C. Intersection with Micro-expressions

According to cognitive resource allocation theory, high mental engagement diminishes the capacity for secondary executive functions, such as conscious facial expression regulation [16], [17]. While affective computing utilizes facial features to infer cognitive states [38], noting behaviors like the "concentration face" or brow lowering (AU4) under high load [12], [5], these macro-expressions remain susceptible to conscious masking. Conversely, subcortically governed micro-expressions are fundamentally involuntary. To bridge this gap, this study objectively quantifies how pure cognitive load triggers micro-expressions. Through a dual-validation approach integrating behavioral analysis and deep learning classification, we position micro-expressions as a robust, non-intrusive visual marker for cognitive load measurement.

III. Method

A. Participant

Fifty participants (27 females; $M_{age} = 22.98$, $SD = 2.88$) were included in the final analysis and randomly assigned to three between-subjects conditions: Standard Maze ($n = 16$), Countdown Maze ($n = 17$), and Path Selection Maze ($n = 17$). Slight group inequalities resulted from excluding initially recruited participants with severe EMG artifacts or excessive head movements. All participants had normal/corrected vision and no history of psychiatric or neurological disorders. Written informed consent was obtained prior to the experiment, and participants received monetary compensation upon completion. The study protocol was approved by the Institutional Review Board (IRB) of the corresponding author's institution, adhering to the APA Ethical Principles of Psychologists and Code of Conduct.

B. Experimental Design and Materials

The study employed a 2 (Difficulty: Easy Difficulty vs. High Difficulty, within-subjects) $\times$ 3 (Maze Type: Standard, Countdown, Path Selection, between-subjects) mixed experimental design. Participants were randomly assigned to one of the three Maze Type conditions, and all participants completed both Easy Difficulty and High Difficulty levels within their assigned condition.

1) Task Implementation and Stimuli: The maze-solving task was programmed and presented using PsychoPy software. For each maze type condition, the stimulus pool consisted of 64 unique maze images. To ensure stimulus consistency, all maze images were procedurally generated using a dedicated maze-creation platform[1]. The topological complexity and solving difficulty of all mazes were strictly controlled through the configuration settings of the maze-generation platform, ensuring that these properties remained equivalent across all mazes within the same difficulty tier (Easy Difficulty or High Difficulty). To eliminate potential visual confounding factors, all maze stimuli were standardized: line thickness, black-and-white contrast, and overall dimensions remained constant across all conditions. The manipulation of difficulty was strictly restricted to topological complexity (i.e., the dead-ends), ensuring that any physiological changes observed were attributable to cognitive load rather than low-level visual features. To mitigate potential order effects and visual fatigue, the experiment utilized a counterbalanced block design. Rather than a fixed sequence for all participants, half of the participants completed the Easy Difficulty block first followed by the High Difficulty block, while the other half completed the blocks in the reverse order. Within each respective difficulty block, the presentation sequence of the specific maze stimuli was fully randomized.

[1]https://www.mazegenerator.net/

2) Maze Types: As illustrated in Fig. 1,in all conditions, participants were required to actively perform the task rather than passively viewing the stimuli. They used the computer mouse to trace the path from the starting point to the exit. This active engagement ensured that participants were consistently processing spatial information and navigating the routes as instructed.

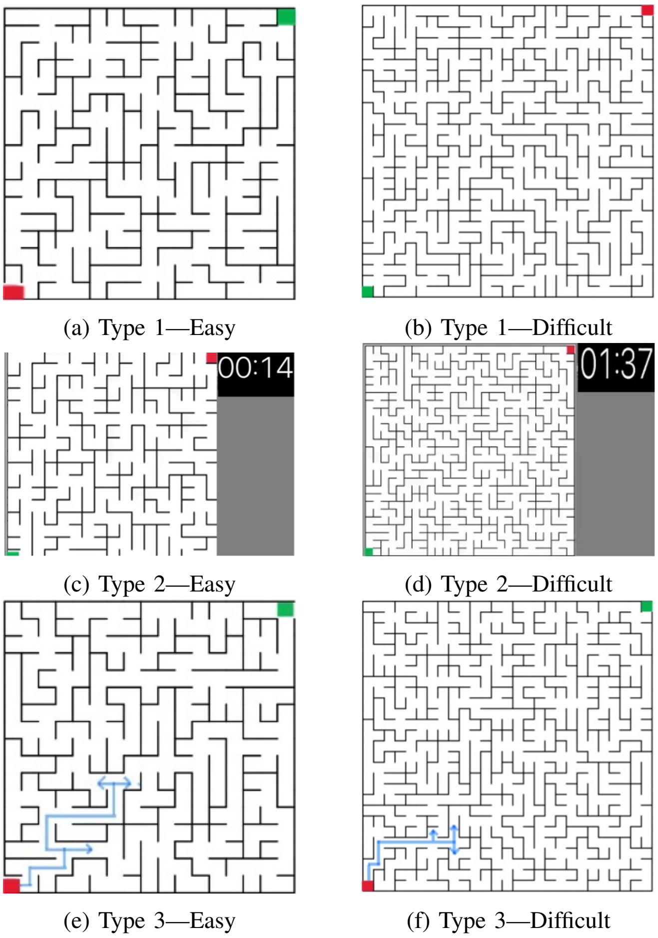
- Standard Mazes: Standard Mazes require participants to process a relatively large amount of visual-spatial information in order to determine the correct path from the starting point to the exit. Participants navigate the maze by continuously selecting the appropriate path based on visual cues. A trial is considered successfully completed once the participant reaches the exit, and no time limit is imposed in this condition.
- Countdown Mazes: Countdown Mazes replicate the structure of the Standard Maze task but introduce an additional countdown timer. The time constraint is designed to increase psychological pressure and elevate cognitive load during the task. Participants must reach the exit before the countdown expires for the trial to be considered successful. If a participant completes the maze before the timer ends, they may press the spacebar to proceed to the next maze trial.
- Path Selection Mazes: Path Selection Mazes maintain the same general maze structure as the Standard condition but provide three explicit candidate paths from which participants can choose. This design reduces the amount of visual cue processing required and therefore lowers cognitive demand to some extent. Participants complete the trial by selecting the correct path and successfully navigating to the exit, and no time limit is imposed in this condition.

3) Difficulty Levels: Within each task, maze stimuli were evenly divided into Group A (Simple, $n = 32$: demanding fewer cognitive resources and shorter reaction times) and Group B (Complex, $n = 32$: demanding extensive working memory resources and longer response times).

4) Apparatus: To capture subtle facial micro-expressions, a 4K resolution, 60 fps high-speed camera continuously recorded participants' frontal faces. Simultaneously, facial EMG signals were recorded at a sampling rate of 1000 Hz using a specialized acquisition system developed by our institution. The EMG sensors precisely targeted five key muscle regions: the corrugator supercilii, orbicularis oculi, zygomaticus major, levator labii superioris, and depressor anguli oris.

C. Procedure

Participants were seated in a quiet, uniformly lit laboratory room in front of a monitor, and the experimenter provided a standardized overview of the task. Specifically, participants were instructed to navigate the presented mazes from the starting point to the exit as accurately and quickly as possible. Throughout the experiment, participants were instructed to maintain a neutral facial expression and avoid deliberate facial movements. For the Countdown Maze condition,



(a) Type 1—Easy (b) Type 1—Difficult
(c) Type 2—Easy (d) Type 2—Difficult
(e) Type 3—Easy (f) Type 3—Difficult

Fig. 1: Illustration of the maze type and difficulty. Type 1: Standard Maze; Type 2: Countdown Maze; Type 3: Path Selection Maze

participants were informed that if they could not complete the current maze image within the time limit, that specific maze would be considered a failure, and they would directly proceed to the next maze image.

Prior to the formal experiment, participants completed a practice block consisting of 2 trials (one Easy Difficulty and one High Difficulty maze) to familiarize themselves with the experimental paradigm and response mechanism. The formal experiment comprised a total of 64 trials per participant, with two groups evenly assigned to the Easy and High Difficulty conditions (32 trials each).

The specific experimental procedure is illustrated in Fig. 2. Each trial commenced with a central fixation cross ("+"). This fixation cross served to eliminate visual afterimages and allow facial muscles and cognitive states to return to a baseline level. Following the fixation cross, the maze stimulus was presented until the participant made a response or a pre-defined timeout occurred. Multimodal data (video and EMG) were continuously and synchronously recorded throughout the entire task execution.

Immediately following the completion of the experimental blocks, participants completed the NASA-TLX to evaluate their subjective cognitive load across six dimensions: mental demand, physical demand, temporal demand, perceived performance, effort, and frustration. Based on their cumulative scores, participants' cognitive load states were strictly categorized into five percentile-based classes: extremely low (0–20%), low (21–40%), moderate (41–60%), high (61–80%), and extremely high (81–100%).

IV. Two-Stage Micro-Expression Screening Pipeline

To ensure the reliability of the collected data, we employed a two-stage screening pipeline consisting of algorithm-based pre-screening followed by EMG-assisted human verification.

A. Algorithm-Based Micro-Expression Pre-Screening

For the initial screening of potential micro-expression segments, we employed an optical flow–based micro-expression spotting algorithm with boundary calibration, which achieved the top performance in the latest long-video micro-expression spotting challenge [29]. Facial regions were first detected, cropped, and aligned, after which regions of interest (ROIs) were defined to extract optical flow features. Candidate micro-expression segments were then identified based on the magnitude of the processed optical flow signals and further refined through boundary calibration[36].

B. EMG-Assisted Human Verification

Following the algorithm-based pre-screening stage, candidate segments were further validated using facial EMG signals and human verification. EMG recordings were examined to identify subtle muscle activations corresponding to the detected segments. Subsequently, FACS-certified human coders inspected the video clips to confirm the presence of micro-expressions and refine the temporal boundaries when necessary. This EMG-assisted verification step effectively removed false positives and ensured the reliability of the final micro-expression samples.

C. Cognitive Load Annotation

For computational validation, participants' cumulative NASA-TLX scores were mapped into five percentile-based classes to represent granular cognitive stress levels (Class 1 to 5, corresponding to 0-20% to 81-100% intervals). Additionally, for the binary classification task, trials from the Easy and High Difficulty conditions were categorically labeled as"Low Load" and "High Load."

V. Behavioral Results

From the 50 participants' video data, 1,421 facial expressions were successfully coded and cross-validated with EMG signals. The detailed distribution of these elicited expressions is depicted in Tab. I.

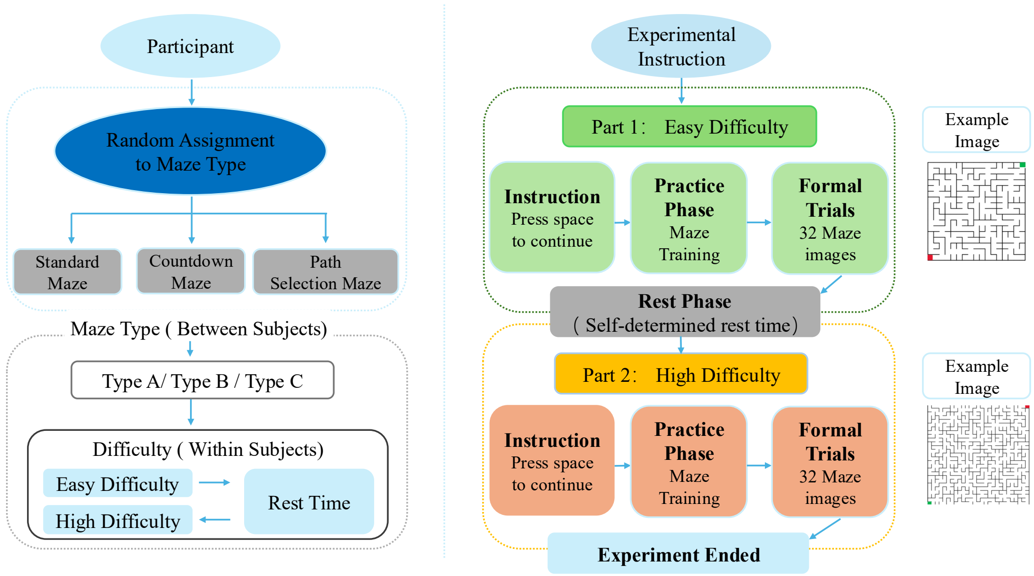


Fig. 2: Cognitive load experiment procedure

TABLE I: Micro- and macro-expression counts for different maze types

Type	Easy Difficulty		High Difficulty	
	Macro-expression	Micro-expression	Macro-expression	Micro-expression
Standard Maze	76	118	114	196
Countdown Maze	79	118	126	201
Path Selection Maze	64	97	96	134

A. Manipulation Check: Effect of Maze Difficulty on Cognitive Load

As a manipulation check, a paired-samples t-test was conducted on the subjective cognitive load (NASA-TLX) scores collapsed across all maze types. The results indicated that the high difficulty maze induced a significantly higher cognitive load ($M = 3.90$, $SD = 1.05$) compared to the easy difficulty maze ($M = 2.48$, $SD = 1.07$), $t(49) = -11.37$, $p < .001$, Cohen's $d = 1.61$. As illustrated in Fig. 3, this results confirms that the task complexity successfully imposed a greater cognitive burden on participants.

To further delineate the specific impact of varying task designs on participants' cognitive states, we analyzed the distribution of subjective cognitive load scores across the three distinct maze types (Path Selection, Standard, and Countdown) under both Easy difficulty and High Difficulty conditions. As illustrated in Fig. 4, the High Difficulty condition consistently elicited higher cognitive load scores across all maze paradigms.

Specifically, the descriptive statistics revealed distinct distribution patterns. For the Path Selection Maze, the cognitive load was $M = 2.19$ ($SD = 1.05$) in the Easy difficulty condition and $M = 3.94$ ($SD = 1.00$) in the High Difficulty condition. For the Standard Maze, scores were $M = 2.24$ ($SD = 0.97$) for the Easy difficulty condition and $M = 3.71$ ($SD = 1.05$) for the High Difficulty condition. Notably, the Countdown Maze induced the highest overall cognitive load, with scores reaching $M = 2.76$ ($SD = 0.97$) in the Easy difficulty condition and peaking at $M = 4.06$ ($SD = 1.03$) in the High Difficulty condition.

These distributional findings robustly corroborate our experimental design rationale. The Path Selection Maze, requiring minimal cue processing, elicited the lowest baseline cognitive load in the Easy difficulty condition. Conversely, the introduction of strict time limits in the Countdown Maze successfully imposed additional extrinsic cognitive load and psychological pressure, resulting in the highest reported cognitive stress among all conditions.

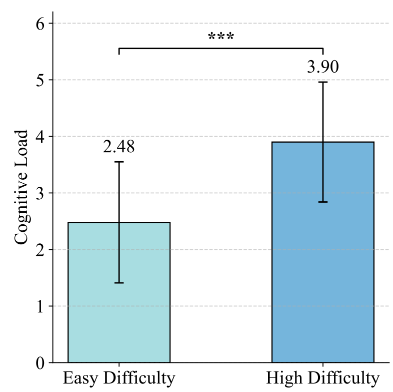


Fig. 3: The differences in cognitive load induced by different maze difficulties

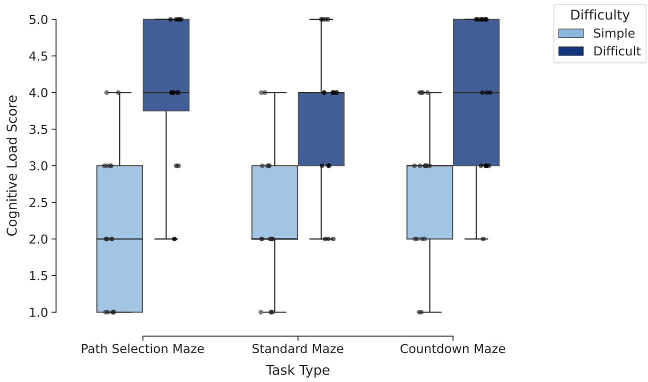


Fig. 4: Distribution of cognitive load across task types

B. Hypothesis Testing: Frequency of Micro-expressions vs. Macro-expressions

Paired-samples t-tests were conducted to examine differences in the frequency of elicited micro- and macro-expressions across conditions (see Fig. 5). In the Easy Difficulty condition, micro-expressions ($M = 7.06$, $SD = 2.56$) significantly outnumbered macro-expressions ($M = 4.00$, $SD = 2.21$), $t(49) = 7.66$, $p < .001$, $d = 1.08$ (a large effect size). Similarly, in the High Difficulty condition, micro-expressions ($M = 10.94$, $SD = 3.12$) were significantly more frequent than macro-expressions ($M = 6.38$, $SD = 2.70$), $t(49) = 8.70$, $p < .001$, $d = 1.23$. Crucially, the high difficulty maze elicited a greater total number of micro-expressions than the easy difficulty maze. This supports the hypothesis that as cognitive resources are heavily taxed, the top-down control over subtle facial musculature is weakened, leading to more frequent spontaneous micro-expression leakage.

VI. Computational Validation: Deep Learning Classification

To provide objective, computational evidence for our psychological hypothesis that distinct cognitive load states manifest as discriminable micro-expression patterns, we framed this as a machine learning classification task. If a deep learning model can accurately classify cognitive load levels purely from micro-expression data, it substantiates the behavioral findings.

A. Model Architecture

We adopted the Feature Refinement (FR) network as our validation model [39]. The FR architecture utilizes a dual-stream Inception network backbone to learn expression-shared features, coupled with specific modules to capture distinct characteristics before final fusion for classification. The FR method was selected due to its reproducibility and its demonstrated efficacy in capturing micro-expression features. Consequently, it has been widely adopted as a baseline approach for micro-expression recognition.

TABLE II: Ten-fold cross-validation accuracy for different classification tasks.

Task	Accuracy (%)
Binary classification	93.51
Five-class classification	65.75

B. Dataset and Implementation Details

The dataset comprised the 1,421 annotated video clips. For the binary classification task (Easy Difficulty vs. High Difficulty), data were split into 553 easy and 868 difficult samples. For the five-class task (NASA-TLX percentiles), samples were distributed from Class 1 to 5 (115, 278, 309, 338, and 381 samples, respectively). Experiments were executed on Ubuntu 22.04.3 using PyTorch 1.13.0 and a single NVIDIA RTX 4090 GPU. Hyperparameters (Learning Rate: 0.001 with a 5-epoch warmup, Batch Size: 32, AdamW optimizer, 20 epochs, Dropout: 0.2) were empirically determined. The trend of accuracy and loss changes during the training process of the two and five-category task models are illustrated in Figs. 6 and 7.

C. Classification Results

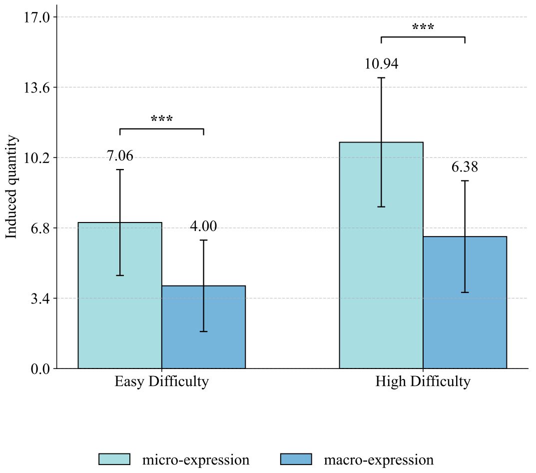
Employing a rigorous 10-fold cross-validation protocol, the netweas listed in Tab. II), the model demonstrated excellent discriminative capability in the binary task (93.51%). The performance drop in the five-class task is an expected outcome of the increased inter-class similarity and finer granularity of cognitive load levels, yet it still indicates meaningful pattern recognition beyond chance levels. The t-SNE feature distribution also demonstrates this conclusion.

VII. Discussion

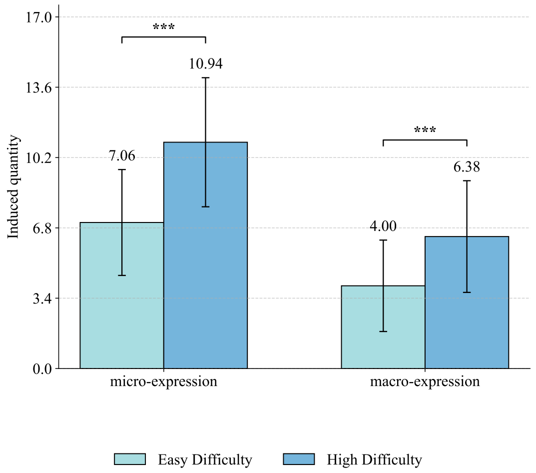
This study proposed and validated a novel approach to eliciting and recognizing micro-expressions driven purely by cognitive load, deliberately isolating the confounding effects of intense emotional stimuli commonly used in traditional paradigms. Our behavioral and computational results robustly support the hypothesis that high cognitive load significantly increases the frequency of micro-expression leakage.

A. Mechanisms of Elicitation: Paradigm Design and Cognitive Load

The core mechanism underlying our findings can be explained through the lens of cognitive resource allocation and the dual-pathway control of facial musculature. During the maze-solving tasks, participants were required to continuously evaluate spatial paths and make rapid decisions, which heavily taxed their working memory. As cognitive demands increased, particularly in the High Difficulty condition, a disproportionate amount of cognitive resources was allocated to the primary navigation task. Consequently, the top-down cortical resources required for secondary executive functions, such as suppressing facial expressions, were severely depleted.



(a) Paired Sample t-test for the number of micro-expressions and macro-expressions induced under the same difficulty level



(b) Paired Sample t-test for the number of elicited responses of the same emotional type under different difficulty levels

Fig. 5: Paired Samples t-tests on the frequency of elicited micro- and macro-expressions across conditions

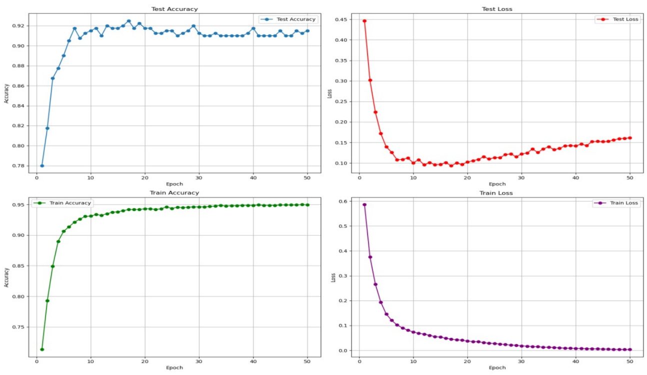


Fig. 6: Trend of accuracy and loss changes for binary classification task

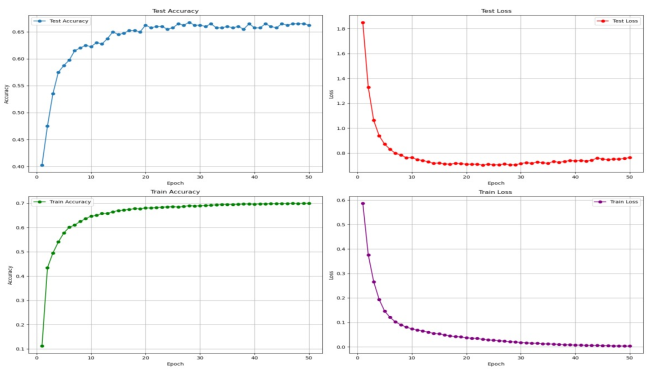


Fig. 7: Trend of accuracy and loss changes for five-classification task

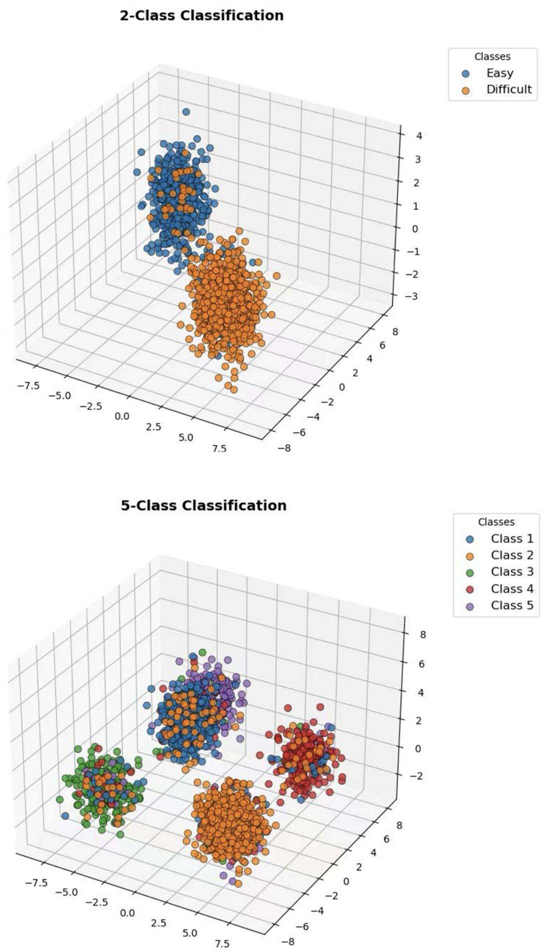


Fig. 8: T-SNE feature distribution

Our experimental paradigm effectively manipulated these resources across three dimensions. The Path Selection Maze demanded minimal cue processing, establishing a baseline of cognitive effort. The Standard Maze increased intrinsic cognitive load through complex topological structures. Crucially, the Countdown Maze introduced a strict time limit, imposing severe extrinsic cognitive load and psychological pressure. The subjective NASA-TLX score distributions perfectly mirrored this design, with the Countdown Maze inducing the highest cognitive stress. Under this peak cognitive exhaustion, the subcortical pathways (responsible for spontaneous, genuine affect like frustration or anxiety) overpowered the weakened cortical suppression mechanism, leading to a surge in involuntary micro-expressions.

B. Sensitivity of Micro-expressions vs. Macro-expressions

A highly notable finding is that micro-expressions significantly outnumbered macro-expressions under high-load conditions. While macro-expressions remained relatively stable across difficulty levels, micro-expression frequencies exhibited high sensitivity to cognitive shifts. This occurs because macro-expressions are subject to conscious control; participants could actively maintain a ”concentration face” to mask their internal struggles. Micro-expressions, however, bypass conscious regulation due to their rapid onset and subcortical origins. This fundamental difference firmly establishes micro-expressions as a far more sensitive and authentic physiological marker of internal cognitive states during non-social, high-load tasks.

Furthermore, the FR deep learning model successfully classified these distinct cognitive states with a 93.51% binary accuracy. This high computational performance validates that the physiological leakage of cognitive stress manifests not as random muscle twitches, but as specific, machine-readable spatiotemporal patterns.

C. Potential Application Scenarios

The ability to objectively and non-intrusively assess cognitive load via facial micro-expressions offers practical implications for several focused domains. First, in Experimental Psychology and Behavioral Research, this vision-based approach provides a continuous metric for cognitive stress, reducing the reliance on disruptive post-task subjective questionnaires (e.g., NASA-TLX) and cumbersome physiological sensors. Second, in the realm of Human-Computer Interaction and Interface Design, micro-expression frequencies can serve as an objective evaluation tool. Designers can identify specific steps in a software workflow that induce excessive extrinsic cognitive load without requiring users to verbally articulate their frustration. Finally, in Computerized Assessment and Adaptive Testing, monitoring these subtle facial dynamics could help systems gauge a user’s real-time problem-solving bottleneck, enabling the timely adjustment of task difficulty to keep the user engaged rather than cognitively overwhelmed.

VIII. Conclusion and Future Work

A. Conclusion

In conclusion, this research thoroughly delineates the psychological mechanisms governing micro-expression leakage under cognitive stress. We introduce a novel pure cognitive elicitation paradigm (a maze-solving task) designed to isolate cognitive stressors from traditional emotional confounds. Our behavioral analyses demonstrated that elevated cognitive load significantly increases micro-expression frequency. Moreover, our computational validation proved the viability of using deep learning models to objectively infer granular cognitive load states directly from facial data. This cross-disciplinary approach provides robust empirical evidence that micro-expressions serve as a highly reliable visual marker for cognitive assessment.

B. Limitations

While this study presents novel findings, several limitations must be acknowledged. First, the experiment was conducted in a highly controlled laboratory environment with uniform lighting and a relatively fixed frontal posture. In ”in-the-wild” applications, dynamic illumination and unconstrained head poses present significant challenges for vision-based micro-expression recognition. Second, the participant demographic was homogenous, primarily consisting of young adults. Because cognitive load processing capabilities and facial muscular elasticity vary across lifespans, the current findings and computational models may require calibration before generalizing to older populations. Lastly, while the maze paradigm effectively manipulated visuospatial working memory, other types of cognitive tasks (e.g., verbal or arithmetic reasoning) were not explored.

C. Future Directions

Future research should expand the current elicitation paradigm to encompass a broader spectrum of cognitive tasks. For instance, researchers could investigate whether different modalities of cognitive load, such as verbal working memory (e.g., N-back tasks) or arithmetic reasoning, elicit distinct facial expression patterns. Such an analysis would further elucidate the specific neuro-behavioral pathways underpinning facial leakage. Additionally, exploring the role of individual differences, such as personality traits, habitual emotion regulation strategies, or baseline working memory capacity, could provide deeper psychological insights into why certain individuals exhibit higher micro-expression frequencies under identical cognitive stress.

References

[1] P. Antonenko, F. Paas, R. Grabner, and T. Van Gog. Using electroencephalography to measure cognitive load. *Educational psychology review*, 22(4):425–438, 2010.

[2] A. Baddeley. Working memory: looking back and looking forward. *Nature reviews neuroscience*, 4(10):829–839, 2003.

[3] F. I. Craik and R. S. Lockhart. Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, 11(6):671–684, 1972.

[4] A. K. Davison, C. Lansley, N. Costen, K. Tan, and M. H. Yap. Samm: A spontaneous micro-facial movement dataset. *IEEE Transactions on Affective Computing*, 9(1):116–129, 2018.

[5] D. F. Dinges, R. L. Rider, J. Dorrian, E. L. McGlinchey, N. L. Rogers, Z. Cizman, S. K. Goldenstein, C. Vogler, S. Venkataraman, and D. N. Metaxas. Optical computer recognition of facial expressions associated with stress induced by performance demands. *Aviation, space, and environmental medicine*, 76(6: Supplement):B172–B182, 2005.
[6] S. K. D'mello and J. Kory. A review and meta-analysis of multimodal affect detection systems. *ACM computing surveys (CSUR)*, 47(3):1–36, 2015.
[7] S. I. Donaldson and E. J. Grant-Vallone. Understanding self-report bias in organizational behavior research. *Journal of business and Psychology*, 17(2):245–260, 2002.
[8] P. Ekman and W. V. Friesen. Nonverbal leakage and clues to deception. *Psychiatry*, 32(1):88–106, 1969.
[9] M. G. Frank and P. Ekman. The ability to detect deceit generalizes across different types of high-stake lies. *Journal of personality and social psychology*, 72(6):1429, 1997.
[10] K. L. Fraser, P. Ayres, and J. Sweller. Cognitive load theory for the design of medical simulations. *Simulation in Healthcare*, 10(5):295–307, 2015.
[11] A. J. Fridlund. *Human facial expression: An evolutionary view*. Academic press, 2014.
[12] J. Grafsgaard, J. B. Wiggins, K. E. Boyer, E. N. Wiebe, and J. Lester. Automatically recognizing facial expression: Predicting engagement and frustration. In *Educational data mining 2013*, 2013.
[13] J. J. Gross. The emerging field of emotion regulation: An integrative review. *Review of general psychology*, 2(3):271–299, 1998.
[14] E. A. Haggard and K. S. Isaacs. Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy. In *Methods of research in psychotherapy*, pages 154–165. Springer, 1966.
[15] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.
[16] D. Kahneman. Attention and effort. 1973.
[17] N. Lavie, A. Hirst, J. W. De Fockert, and E. Viding. Load theory of selective attention and cognitive control. *Journal of experimental psychology: General*, 133(3):339, 2004.
[18] J. Li, Z. Dong, S. Lu, S.-J. Wang, W.-J. Yan, Y. Ma, Y. Liu, C. Huang, and X. Fu. 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, 2022.
[19] J. Li, S. Lu, Y. Wang, Z. Dong, S.-J. Wang, and X. Fu. Could micro-expressions be quantified? electromyography gives affirmative evidence. *IEEE Transactions on Affective Computing*, 16(4):2959–2974, 2025.
[20] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikäinen. A spontaneous micro-expression database: Inducement, collection and baseline. In *2013 10th IEEE International Conference and Workshops on Automatic face and gesture recognition (fg)*, pages 1–6. IEEE, 2013.
[21] M. Muraven and R. F. Baumeister. Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological bulletin*, 126(2):247, 2000.
[22] A. Newell, H. A. Simon, et al. *Human problem solving*, volume 104. Prentice-hall Englewood Cliffs, NJ, 1972.
[23] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. Van Gerven. Cognitive load measurement as a means to advance cognitive load theory. In *Cognitive load theory*, pages 63–71. Routledge, 2016.
[24] R. W. Picard. *Affective computing*. MIT press, 2000.
[25] S. Porter and L. Ten Brinke. Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. *Psychological science*, 19(5):508–514, 2008.
[26] F. Qu, S.-J. Wang, W.-J. Yan, H. Li, S. Wu, and X. Fu. Cas (me) ^{2}: a database for spontaneous macro-expression and micro-expression spotting and recognition. *IEEE Transactions on Affective Computing*, 9(4):424–436, 2017.
[27] J. M. Richards and J. J. Gross. Emotion regulation and memory: the cognitive costs of keeping one's cool. *Journal of personality and social psychology*, 79(3):410, 2000.
[28] W. E. Rinn. The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expressions. *Psychological bulletin*, 95(1):52, 1984.
[29] J. See, J. Li, A. K. Davison, G. B. Liong, M. H. Yap, W.-H. Cheng, X. Li, X. Hong, and S.-J. Wang. Megc2024: Acm multimedia 2024 facial micro-expression grand challenge. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 11482–11483, 2024.
[30] J. Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2):257–285, 1988.
[31] J. Sweller, J. J. Van Merrienboer, and F. G. Paas. Cognitive architecture and instructional design. *Educational psychology review*, 10(3):251–296, 1998.
[32] A. Vrij, S. A. Mann, R. P. Fisher, S. Leal, R. Milne, and R. Bull. Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and human behavior*, 32(3):253–265, 2008.
[33] Q. Wang, S. Yang, M. Liu, Z. Cao, and Q. Ma. An eye-tracking study of website complexity from cognitive load perspective. *Decision support systems*, 62:1–10, 2014.
[34] R. E. Wood. Task complexity: Definition of the construct. *Organizational behavior and human decision processes*, 37(1):60–82, 1986.
[35] Z. Xia, W. Peng, H.-Q. Khor, X. Feng, and G. Zhao. Deep learning for micro-expression recognition: A survey. *IEEE Transactions on Affective Computing*, 14(2):1024–1045, 2022.
[36] J. Yu, Y. Zhang, G. Zhao, P. He, Z. Zhang, Z. Cai, Q. Liu, J. Sun, and J. Liang. Micro-expression spotting based on optical flow feature with boundary calibration. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 11490–11496, 2024.
[37] J. Zagermann, U. Pfeil, and H. Reiterer. Measuring cognitive load using eye tracking technology in visual computing. In *Proceedings of the sixth workshop on beyond time and errors on novel evaluation methods for visualization*, pages 78–85, 2016.
[38] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang. A survey of affect recognition methods: audio, visual and spontaneous expressions. In *Proceedings of the 9th international conference on Multimodal interfaces*, pages 126–133, 2007.
[39] L. Zhou, Q. Mao, X. Huang, F. Zhang, and Z. Zhang. Feature refinement: An expression-specific feature learning and fusion method for micro-expression recognition. *Pattern Recognition*, 122:108275, 2022.