

VS-LLM: Visual-Semantic Depression Assessment based on LLM for Drawing Projection Test

Meiqi Wu¹, Yaxuan Kang², Xuchen Li², Shiyu Hu^{2,3}, Xiaotang Chen², Yunfeng Kang², Weiqiang Wang¹, and Kaiqi Huang^{2,3,4}

1 School of Computer Science and Technology, University of Chinese Academy of Sciences; 2 Institute of Automation, Chinese Academy of Sciences; 3 School of Artificial Intelligence, University of Chinese Academy of Sciences; 4 CAS Center for Excellence in Brain Science and Intelligence Technology



Introduction

Motivation

This study aims to address the limitations of traditional self-report scales in depression assessment—such as susceptibility to manipulation, high face validity, difficulty in detecting deep psychological content, and cultural influences—through the automated analysis of intelligent drawing projective tests, thereby providing a more accurate, objective, and universally applicable assessment tool.

Contributions

- We developed a visual-semantic depression assessment system based on LLM (VS-LLM) method, where the visual perception module and the mental semantic caption generation module are respectively used to obtain more detailed and overall information from the painting, enabling more effective analysis of PPAT.
- We first provided an experimental environment for automated analysis of PPAT sketches for depression assessment.
- Our experiments demonstrate the superior performance of our method, which improves by 17.6% compared to the psychologist assessment method.

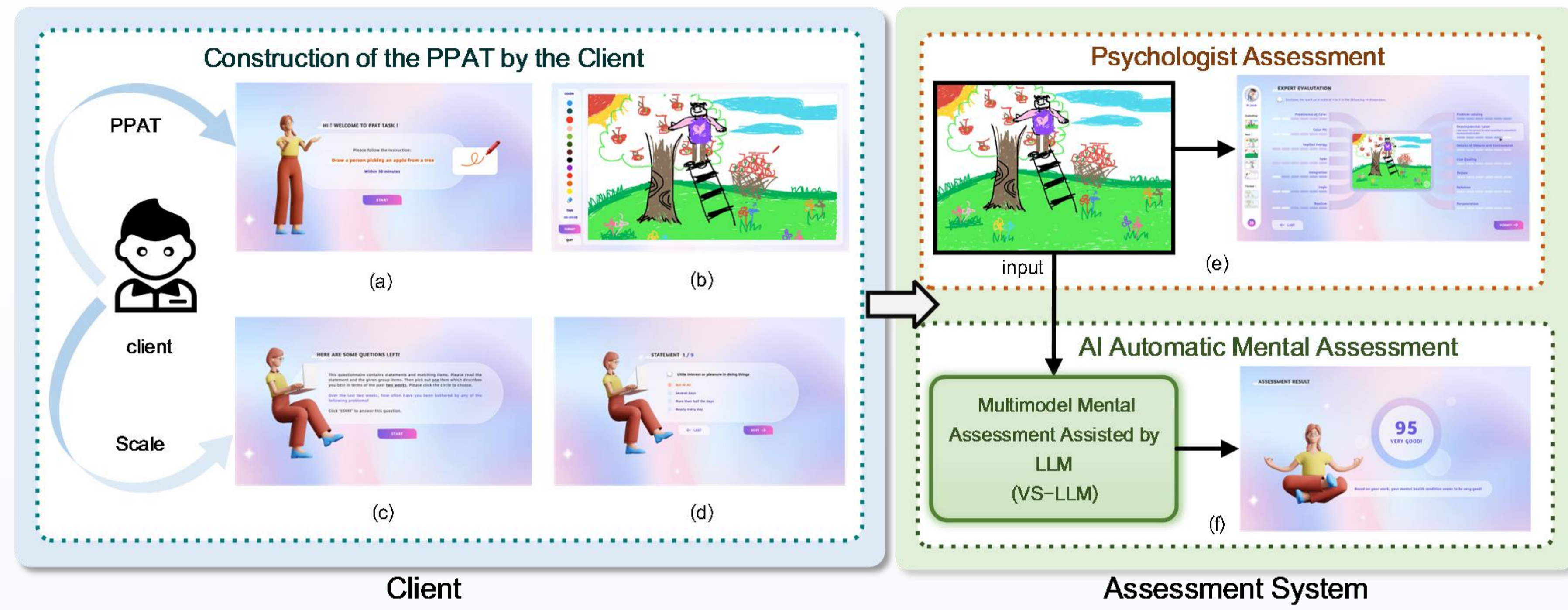
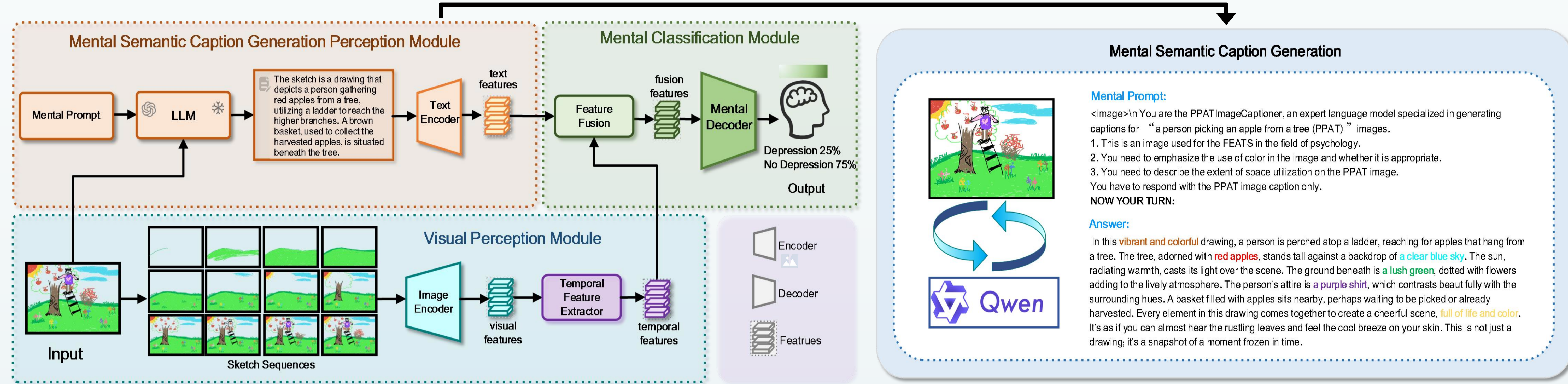


Fig.1 Our mental state assessment on the electronic DPT system

Methodology



Experiment Environment-PPAT Dataset



Table 1. PPAT Dataset Statics.

| Metric | Type | Value |
|--------------|---------------|---------|
| Age | 0-20 | 123/690 |
| | 20-40 | 531/690 |
| | 40-60 | 32/690 |
| | >60 | 4/690 |
| Gender | Male | 302/690 |
| | Female | 388/690 |
| | Neutral | 0/690 |
| Mental State | Depression | 117/690 |
| | No Depression | 573/690 |

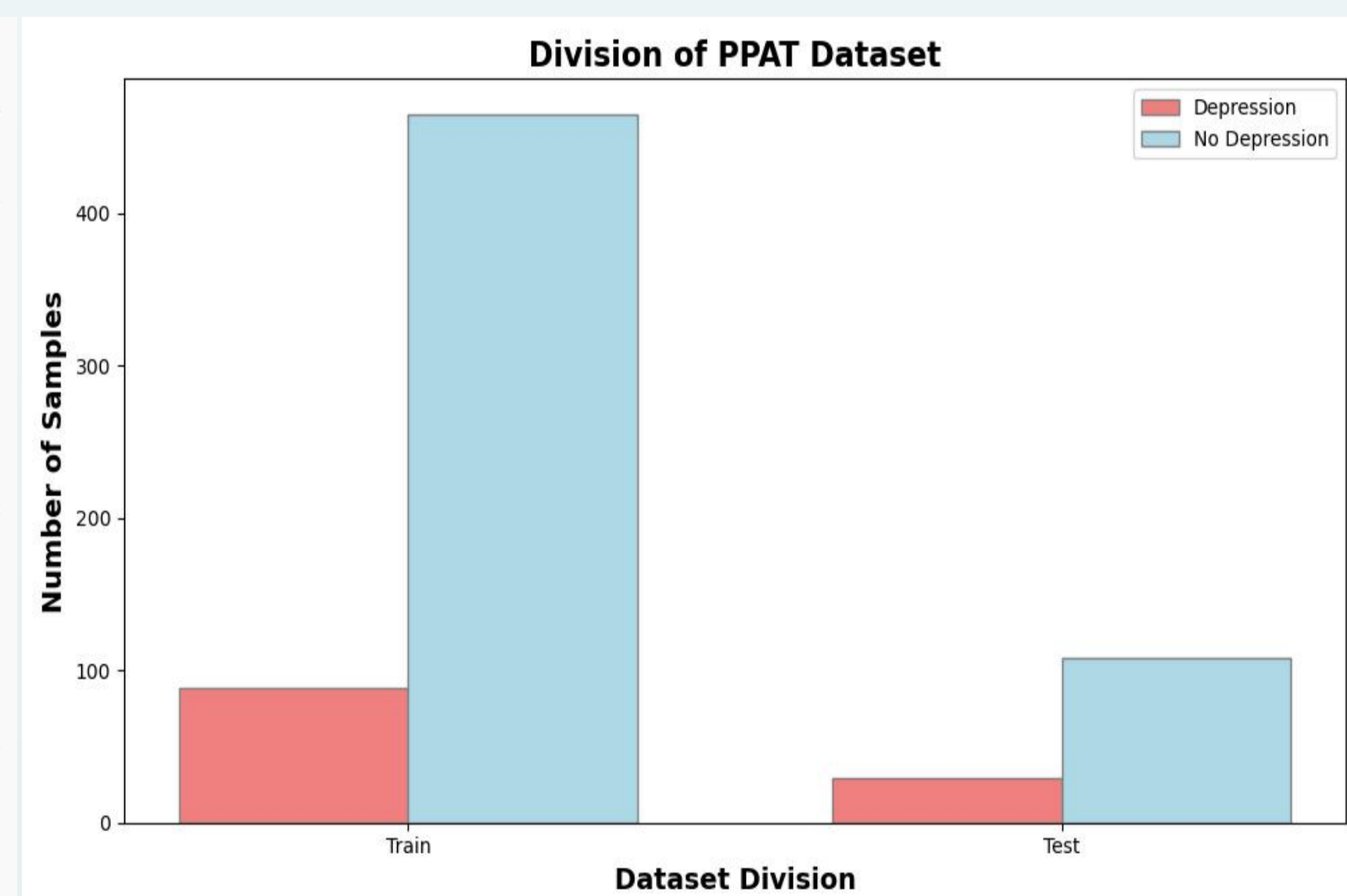


Table 3. Experimental Results of Psychologist Assessment and AI-Automated Assessment on PPAT Dataset. The input for the psychologist assessment method consists of scores from the 14 dimensions of FEATS. The input for the AI automated assessment method is the PPAT sketch. Note that we only calculate the number of trainable parameters of the model.

| Method | Acc(%) | Params(M) | FLOPs |
|-------------------------|--------------------------|-------------|-------|
| Psychologist Assessment | Random Forest [36] | 70.2 | - |
| | SVM [37] | 57.3 | - |
| | Logistic Regression [38] | 56.4 | - |
| | MLP [39] | 64.3 | - |
| AI Automatic Assessment | Resnet18 [42] | 83.3 | 11.18 |
| | Sketch-a-net [9] | 85.7 | 8.41 |
| | VS-LLM (Ours) | 87.8 | 8.87 |

Results

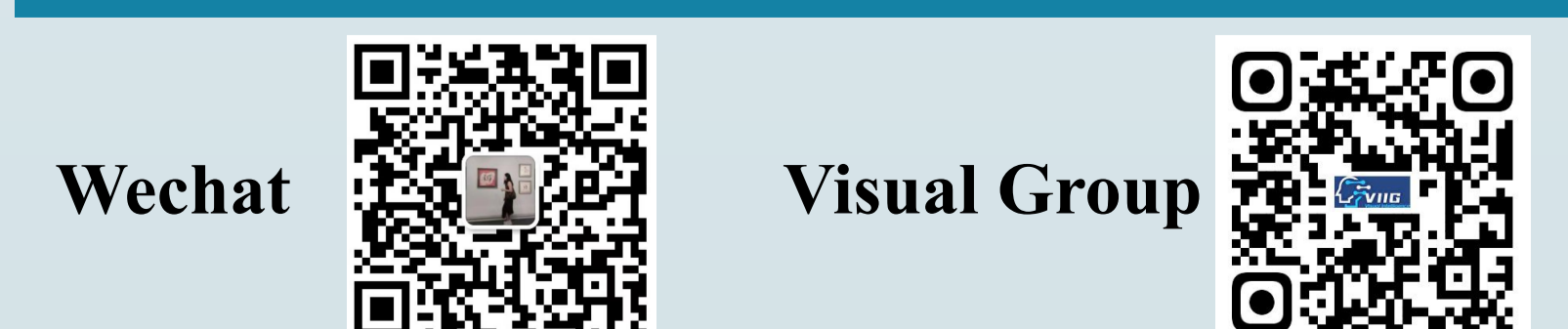
Table 4. Ablation Studies on PPAT Dataset. FL is the abbreviation of focal loss and CEL is the abbreviation of cross-entropy loss.

| Number | Mental Sematic Caption Generation | Visual Perception Module | | Loss | Acc(%) |
|----------------|-----------------------------------|--------------------------|----------------------------|------|-------------|
| | | Image Encoder | Temporal Feature Extractor | | |
| 1 | ✗ | Resnet18 | ✓ | FL | 84.1 |
| 2 | ✓ | Resnet18 | ✗ | FL | 83.3 |
| 3 | ✓ | Sketch-a-Net | ✓ | FL | 87.1 |
| 4 | ✓ | Resnet18 | ✓ | CEL | 86.4 |
| 5(Ours) | ✓ | Resnet18 | ✓ | FL | 87.8 |

Conclusion

In this paper, We developed a visual-semantic depression assessment based on LLM (VS-LLM), where the visual perception module and the mental semantic caption generation module are respectively used to obtain more detailed and overall information from the sketch, enabling more effective analysis of PPAT. We provided an experimental environment for automated analysis of PPAT sketches for depression assessment. Our experiments demonstrate the superior performance of our method and confirm the importance of incorporating mental descriptions assisted by LLM.

Contact us



My Homepage: <https://wmeiqi.github.io/>