

A Micro-Expression Recognition Method Based on an Uncertainty-Aware Mixing Strategy and Multimodal Fusion

Qian Gao¹, Weijia Feng^{1,*}, Jia Guo¹, Jiayi An¹, Xiaofeng Wang² and Yuanxu Chen³

¹Tianjin Normal University, No.393, Binshui West Road, Xiqing District, Tianjin, 300387, China

²Tianjin Key Laboratory for Advanced Mechatronic System Design and Intelligent Control, Tianjin University of Technology, Tianjin, China

³Ping An Technology Shenzhen, Rm1201, Bld B, Pingan IFC, Xinyuan South Rd, Chaoyang District, Beijing, 100027, China

Abstract

Micro-expression recognition, a critical research direction in affective computing, holds significant value due to its wide-ranging applications in real-world scenarios such as interrogations, clinical diagnostics, and business negotiations. Currently, micro-expression datasets are limited in scale and challenging to annotate. The significant imbalance in sample sizes across different types of micro-expressions causes models to bias toward majority classes during training, while minority class samples receive insufficient attention. This results in overfitting and poor generalization performance in existing recognition methods. Furthermore, most methods rely solely on local information from micro-expression sequences, overlooking certain dynamic features, which adversely impacts recognition performance. To address potential overfitting issues, we propose a micro-expression recognition method based on uncertainty awareness and multimodal fusion. By integrating uncertainty estimation to weight mixed samples, our approach guides the multi-model to focus more on underperforming samples. Additionally, recognition efficiency is further enhanced by incorporating optical flow parameters from micro-expression images. Experimental validation demonstrates that our method achieves significant improvements across multiple key metrics.

Keywords

Micro-expression, uncertainty, Multi-model

1. Introduction

Micro-expression recognition (MER) is a key area in affective computing, holding great significance due to its wide applications in fields such as interrogation, clinical diagnosis, and business negotiations. Micro-expressions are brief and subtle facial expressions that typically occur when individuals attempt to conceal their true emotions. They last for a very short duration (usually no more than 0.5 seconds), which makes MER an especially challenging task [1]. In recent years, with the development of deep learning, this field has made remarkable progress. However, current micro-expression datasets are limited in size and difficult to annotate, making model training prone to overfitting and poor generalization [2]. For instance, datasets like CASME, CASME II, and SMIC contain a limited number of samples, which constrains the performance of deep learning models [3]. Additionally, some micro-expression samples are difficult to classify due to indistinct features or high similarity with other classes [4]. Therefore, developing more effective data augmentation methods and feature learning strategies to improve recognition accuracy and robustness is a key research objective. For example, the MR-UAMF method addresses class imbalance through uncertainty-aware mixing [5] and significantly improves the recognition accuracy of minority classes. Meanwhile, other studies have introduced attention

4DMR@IJCAI25: International IJCAI Workshop on 1st Challenge and Workshop for 4D Micro-Expression Recognition for Mind Reading, August 29, 2025, Guangzhou, China

* Corresponding author.

✉ 2311090038@stu.tjnu.edu.cn (Q. Gao); weijiafeng@tjnu.edu.cn (W. Feng); c04s316@bupt.cn (J. Guo); anjiayi0911@163.com (J. An); wangxiaofeng@tjut.edu.cn (X. Wang); chenyanxu641@pingan.com.cn (Y. Chen)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

mechanisms or improved neural network architectures [6] to enhance focus on critical features and thus improve recognition performance.

Existing micro-expression datasets are small and hard to annotate, causing overfitting and limited model generalization [7]. Many current methods rely on handcrafted features like LBP and LBP-TOP. While simple and effective, they struggle with the subtle dynamics of micro-expressions [6], limiting their ability to leverage deep learning's full potential [3]. Furthermore, existing approaches often fail to adequately extract dynamic features, as they don't make good use of temporal sequence information, which negatively impacts recognition accuracy and robustness.

Micro-expression recognition (MER) methods can be classified into three categories: The first method is traditional feature-based method. These rely on handcrafted features such as LBP, LBP-TOP, Histogram of Oriented Gradients (HOG), and optical flow. While these approaches are straightforward and efficient, they struggle with capturing subtle and dynamic changes. For example, Li et al. [8] proposed an LBP-TOP-based method that integrates temporal and spatial features but performs poorly with complex dynamic changes. The second method is deep learning-based method. These mainly utilize Convolutional Neural Networks (CNNs) and their variants like 3D CNNs to automatically learn features. These methods perform well in feature learning but require large amounts of annotated data, making them prone to overfitting due to limited dataset sizes. For instance, Zhang et al. [3] introduced a 3D CNN method using a multi-stream structure to capture spatiotemporal features but encountered overfitting on small datasets. The last method is data augmentation methods. These include Generative Adversarial Networks (GANs) and synthetic data generation techniques. While they expand datasets, issues such as authenticity and potential data bias remain. For example, Wang et al. [9] used GANs to generate synthetic micro-expression data, effectively expanding the dataset, though the realism of the generated data still requires improvement. Some studies have explored the use of attention mechanisms in MER. For instance, Hao et al proposed a hierarchical spatiotemporal attention mechanism that can automatically focus on key regions and time segments of micro-expressions, significantly improving accuracy. Despite progress, research gaps remain. Dataset sizes are still small and annotations difficult, leading to overfitting and limited generalization. Additionally, dynamic features are underutilized, and data imbalance remains a challenge. MR-UAMF, for example, addresses imbalance through uncertainty-aware mixing, improving minority class accuracy [1, 2, 4].

To address the aforementioned challenges, we propose a micro-expression recognition method based on uncertainty awareness and multimodal fusion(MR-UAMF). This approach weights mixed samples based on their uncertainty, encouraging the model to focus more on samples with lower performance, thereby mitigating overfitting. A multimodal model is used to process optical flow and image features separately, fully leveraging the dynamic and spatial information of micro-expressions to enhance accuracy and robustness. By integrating uncertainty quantification, micro-attention mechanisms, and 3D CNN, this approach offers a novel perspective and method for micro-expression recognition, advancing the field. Extensive experiments across multiple datasets validate the effectiveness and superiority of our approach.

The remainder of this paper is structured as follows: Section 2 describes the proposed framework of MR-UAMF, including the overall structure, core algorithms, and implementation details. Section 3 descriptions, and evaluation metrics. Section 4 interprets experimental results and discusses academic implications and potential applications. Section 5 summarizes our contributions and outlines future research directions.

2. Methodology

2.1. Overall Framework

Our proposed framework integrates optical flow feature extraction, an focused uncertainty-aware mixing strategy(FU-MIX), a micro-attention mechanism, and a shallow triple-stream 3D CNN . The goal is to effectively capture both spatial and temporal features of micro-expressions while addressing data imbalance and overfitting.

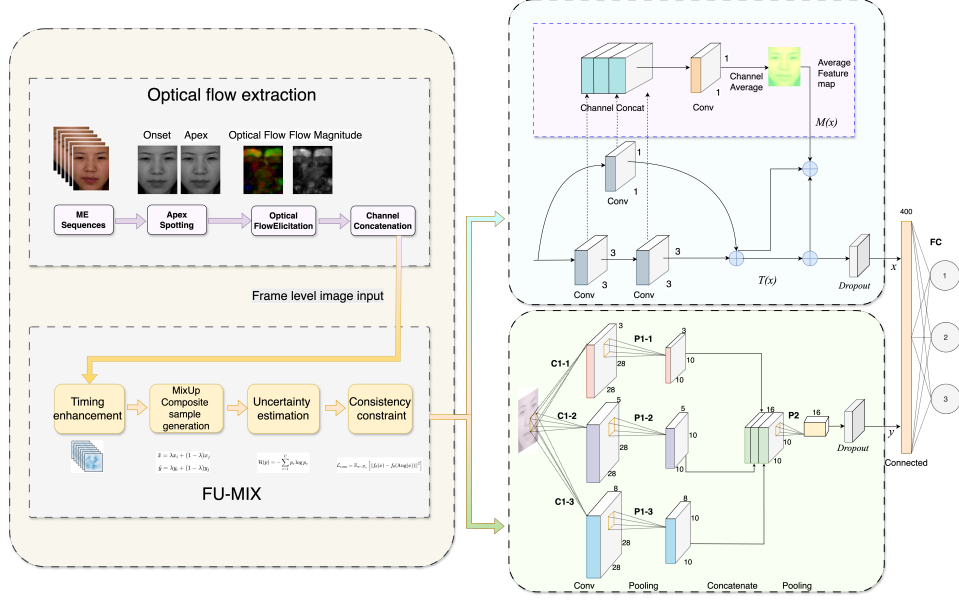


Figure 1: Overall Framework

2.2. Optical Flow Feature Extraction

Given the brief and subtle nature of micro-expressions, we first extract optical flow features from the video sequence to obtain motion information.

We compute optical flow-guided features using the onset frame and the apex frame. The optical flow field between the two frames is represented as a tuple:

$$O_i = \{(u(x, y), v(x, y)) \mid x = 1, 2, \dots, X; y = 1, \dots, Y\}$$

where X and Y denote the width and height of the frame, and $u(x, y)$ and $v(x, y)$ are the horizontal and vertical components of O_i , respectively.

We also compute optical strain to approximate the intensity of facial deformation:

$$\varepsilon = \frac{1}{2} [\nabla u + (\nabla u)^T], \quad \text{with } u = [u, v]^T$$

The magnitude of the optical strain is:

$$|\varepsilon_{x,y}| = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 + \frac{1}{2} \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)^2$$

Appending optical strain to the optical flow field O , we form the triplet:

$$\Theta = \{u, v, \varepsilon\} \in \mathbb{R}^3$$

Each video thus yields three types of optical flow-based representations: horizontal component u , vertical component v , and optical strain ε .

2.3. FU-MIX: Focus Uncertainty-Aware Mixing

Uncertainty Estimation To enhance robustness and generalization, we adopt the MR-UAMF strategy, which estimates sample uncertainty using Bayesian sampling from the model's posterior $p(\theta; D)$. For a given sample x_i , its uncertainty u_i is defined as:

$$u_i = \int \kappa(y_i, \hat{f}_\theta(x_i)) p(\theta; D) d\theta$$

where $\kappa(y_i, \hat{f}_\theta(x_i))$ indicates correct classification. We approximate this using Monte Carlo sampling:

$$u_i \approx \frac{1}{T} \sum_{t=1}^T \kappa(y_i, \hat{f}_{\theta_t}(x_i))$$

where θ_t is sampled by minimizing expected risk. In practice, historical training trajectory information is used for approximation.

Weighted Mixed Sample Generation Each sample is assigned a weight w_i proportional to its uncertainty:

$$w_i = \eta u_i + c$$

where η is a hyperparameter and c is a small constant to ensure $w_i > 0$. Mixed samples are generated as:

$$x_{i,j} = \lambda x_i + (1 - \lambda)x_j, \quad y_{i,j} = \lambda y_i + (1 - \lambda)y_j$$

where $\lambda \sim \text{Beta}(\alpha, \alpha)$. The weighted loss function is:

$$\mathbb{E}_{(x_i, y_i), (x_j, y_j)} [w_i \lambda \ell(\theta, x_{i,j}, y_i) + w_j (1 - \lambda) \ell(\theta, x_{i,j}, y_j)]$$

2.4. Micro-Attention Mechanism

We adopt a parameter-efficient residual architecture with self-learned multi-scale features to compute attention maps. Given input $X \in \mathbb{R}^{C \times H \times W}$, three convolutional layers (1×1 , 3×3 , and 5×5) produce feature maps $\{L_{c1}, L_{c2}, L_{c3}\}$.

These are concatenated to form $L'_c \in \mathbb{R}^{(c1+c2+c3) \times H \times W}$, and the average feature map is generated by:

$$M(X) = \frac{1}{c'} \sum_{i=1}^{c'} (L'_c * W_i^*)$$

where W_i^* is a 1×1 convolution kernel. The residual output is:

$$T(X) = L_{c1} + L_{c3}$$

The final output with attention is:

$$O(X) = T(X) \cdot \sigma(M(X))$$

where σ is a normalization function. If $M(X) \approx 0$, the attention influence is minimized.

2.5. 3D CNN

We design a shallow triple-stream 3D CNN to learn from the optical flow cube Θ . The input is resampled to $28 \times 28 \times 3$. Each stream includes:

- One 3D convolutional layer with kernel counts of 3, 5, and 8, respectively;
- One max-pooling layer.

Outputs from the three streams are concatenated along the channel axis, followed by a 2×2 average pooling layer. A fully connected layer with 400 nodes abstracts the features, and a softmax layer classifies the output into three compound emotion categories.

2.6. Implementation Details

We use the CASME1, CASME2, SAMM, and SMIC datasets for training and evaluation, splitting each dataset into 80% training and 20% testing. MR-UAMF is employed as a data augmentation technique, assigning greater weights to underperforming samples based on uncertainty estimates. This mitigates overfitting and enhances recognition of minority expression categories.

The 3D CNN extracts optical flow features through three parallel convolutional streams, and the micro-attention mechanism highlights important regions via adaptive residual weighting. These modules work synergistically to improve feature discriminability, thereby enhancing the model’s performance and robustness in micro-expression recognition tasks.

3. Experiments

3.1. Datasets and Evaluation Metrics

The datasets used in this paper are as follows. The first dataset is SMIC. It’s the earliest spontaneous micro-expression dataset, featuring recordings from three camera types: High-Speed (HS) at 100 fps, and Visual (VIS) and Near-Infrared (NIR) at 25 fps. This study uses only HS camera data, with 164 samples from 16 participants across three classes: *Negative*, *Positive*, and *Surprise*.

The second dataset is CASME. It was collected in a controlled lab at 60 fps and has 195 samples from 19 subjects. However, it has class imbalance. We use 154 samples across four classes: *Disgust*, *Repression*, *Surprise*, and *Tense*.

The third dataset is CASME II. It enhances CASME with higher resolution (200 fps, 280×340 pixels), featuring 248 samples from 26 subjects across five classes: *Disgust*, *Happiness*, *Repression*, *Surprise*, and *Others*.

The last dataset is SAMM. The SAMM dataset, gathered in a well-lit, stable setting with a grayscale camera at 200 fps and a 2040×1088 resolution, comprises 159 samples from 32 diverse participants. After excluding classes with under 10 samples, like *Fear* and *Sadness*, we utilize 134 samples spanning five classes: *Anger*, *Contempt*, *Happiness*, *Surprise*, and *Others*.

To comprehensively evaluate the model’s performance, we adopt common classification metrics including **Accuracy**, **Precision**, **Recall**, and **F1-score**. Given the class imbalance inherent in micro-expression datasets, we also report two additional metrics: Unweighted F1-score (UF1): the average of class-wise F1-scores. Unweighted Average Recall (UAR): the average of class-wise recall rates. These metrics offer a more balanced view of model performance, especially in imbalanced settings.

3.2. Comparative Experiments

This section compares the proposed method with existing micro-expression recognition (MER) methods across four widely used datasets: SMIC, CASME, CASME II, and SAMM. The compared methods are categorized as follows: LBP-TOP (Local Binary Patterns from Three Orthogonal Planes), 3DHOG (Three-Dimensional Histogram of Oriented Gradients), HOOF (Histogram of Oriented Optical Flow), OFF-ApexNet, STSTNet, Dual-Inception, MACNN, Micro-Attention, Mini-AORCNN.

All methods were evaluated under identical settings to ensure fairness: the same number of samples, class labels, and K-fold cross-validation protocols were used.

We reproduced results for LBP-TOP, 3DHOG, and HOOF under consistent experimental conditions. We also reproduced the performance of recent deep learning-based MER methods. All classifiers were implemented using Support Vector Machines (SVM). Our method outperforms these handcrafted baselines significantly, demonstrating the advantages of deep learning and uncertainty-aware modeling in handling the subtle and dynamic nature of micro-expressions [1, 2, 4]. These results demonstrate that our focused uncertainty-aware method consistently outperforms or matches the state-of-the-art across all datasets. The simplicity of our design enables robust and discriminative learning even under limited data scenarios—a significant advantage for micro-expression datasets, which are typically small

in size. Moreover, uncertainty modeling improves generalization by enhancing the model’s ability to handle non-linearities and minority class variations in real-world applications.

Table 1
Comparison of Different Algorithms on SMIC Dataset (Three Classes)

Algorithm	Accuracy	F1-score	Precision	Recall	UF1	UAR
Handcrafted Features						
LBP-TOP [10]	0.4141	0.4104	0.4157	0.4085	0.4049	0.4083
3DHOG [11]	0.3280	0.3464	0.3660	0.3415	0.3441	0.3482
HOOF [12]	0.5085	0.5036	0.5035	0.5061	0.5011	0.4998
Deep Learning Methods						
OFF-ApexNet [13]	0.6167	0.5794	0.5801	0.5793	0.5773	0.5798
STSTNet [14]	0.5810	0.6035	0.6035	0.6037	0.6040	0.6036
Dual-Inception [15]	0.6139	0.5791	0.5802	0.5793	0.5776	0.5738
MACNN [16]	0.7106	0.6389	0.6402	0.6402	0.6449	0.6506
Micro-Attention [17]	0.6907	0.6211	0.6230	0.6220	0.6248	0.6285
Mini-AORCNN [18]	0.6748	0.5894	0.5963	0.5915	0.6013	0.6101
MR-UAMF (Ours)	0.8497	0.8489	0.8476	0.8427	0.8411	0.8394

Table 2
Comparison of Different Algorithms on CASME Dataset (Four Classes)

Algorithm	Accuracy	F1-score	Precision	Recall	UF1	UAR
Handcrafted Features						
LBP-TOP	0.4728	0.3715	0.3812	0.3896	0.3151	0.3304
3DHOG	0.4789	0.3626	0.3553	0.3961	0.2908	0.3006
HOOF	0.4976	0.3073	0.2878	0.3442	0.2374	0.2546
Deep Learning Methods						
OFF-ApexNet	0.6318	0.5359	0.5347	0.5390	0.5326	0.5307
STSTNet	0.6477	0.5672	0.5824	0.5649	0.5203	0.5127
Dual-Inception	0.6175	0.5118	0.5129	0.5195	0.5118	0.5348
MACNN	0.7244	0.6562	0.6580	0.6558	0.6714	0.6639
Micro-Attention	0.7523	0.6778	0.6912	0.6818	0.6656	0.6394
Mini-AORCNN	0.7397	0.6655	0.6771	0.6623	0.6698	0.6625
MR-UAMF (Ours)	0.8117	0.8102	0.8117	0.8105	0.7930	0.8095

Table 3
Comparison of Different Algorithms on CASME II Dataset (Five Classes)

Algorithm	Accuracy	F1-score	Precision	Recall	UF1	UAR
Handcrafted Features						
LBP-TOP	0.3947	0.3172	0.3670	0.3589	0.2554	0.2570
3DHOG	0.4010	0.3701	0.3643	0.3871	0.3009	0.3063
HOOF	0.4282	0.3926	0.3940	0.3992	0.3418	0.3332
Deep Learning Methods						
OFF-ApexNet	0.5370	0.4891	0.4870	0.5000	0.4622	0.4559
STSTNet	0.5987	0.5542	0.5743	0.5685	0.5385	0.5192
Dual-Inception	0.5380	0.5061	0.5051	0.5081	0.4751	0.4693
MACNN	0.6567	0.6506	0.6609	0.6532	0.6588	0.6525
Micro-Attention	0.6519	0.6294	0.6395	0.6371	0.6243	0.6107
Mini-AORCNN	0.6610	0.6387	0.6390	0.6452	0.6262	0.6237
MR-UAMF (Ours)	0.8359	0.8399	0.8357	0.8353	0.8161	0.8133

Table 4

Comparison of Different Algorithms on SAMM Dataset (Five Classes)

Algorithm	Accuracy	F1-score	Precision	Recall	UF1	UAR
Handcrafted Features						
LBP-TOP	0.4403	0.2925	0.2810	0.3955	0.3418	0.4009
3DHOG	0.4508	0.2440	0.2383	0.4030	0.3437	0.3945
HOOF	0.4681	0.2697	0.2568	0.4179	0.3685	0.4272
Deep Learning Methods						
OFF-ApexNet	0.4226	0.3467	0.3360	0.4104	0.3142	0.3660
STSTNet	0.4938	0.3816	0.3668	0.4552	0.3866	0.3935
Dual-Inception	0.4490	0.2804	0.3137	0.3881	0.3568	0.4097
MACNN	0.5899	0.4734	0.4664	0.4925	0.5054	0.5072
Micro-Attention	0.5683	0.4715	0.4574	0.5000	0.4683	0.4671
Mini-AORCNN	0.5665	0.4432	0.4345	0.4925	0.4648	0.4737
MR-UAMF (Ours)	0.8388	0.8271	0.8328	0.8294	0.8458	0.8308

3.3. Ablation Studies

To evaluate the contribution of each component in our proposed micro-expression recognition (MER) framework, we conducted a series of ablation experiments.

The first is baseline model. The base model without any enhancements, captures only fundamental features and struggles with the complex variations inherent in micro-expressions. It achieves an accuracy of only **69%**, highlighting its limited capability in handling subtle emotional cues.

The second dataset is with uncertainty-aware mixing MR-UAMF. Integrating the focus uncertainty-aware mixing strategy improves the model’s adaptability to complex expression dynamics. The accuracy increases by **10%**, demonstrating the effectiveness of uncertainty-awareness in enhancing model robustness and mitigating overfitting.

The third dataset is with multimodal feature fusion. When the multimodal fusion module is added—combining both optical flow and spatial image features—the accuracy further improves by **5%**, achieving the best performance. This indicates that multimodal fusion enriches feature representations by integrating both temporal and spatial cues.

The last dataset is With Additional Modules. Gradually incorporating auxiliary components such as the micro-attention mechanism and the shallow triple-stream 3D CNN architecture leads to incremental gains in accuracy. These modules help the model focus on salient spatiotemporal features and extract multi-scale representations effectively.

The ablation study clearly shows that: Each module contributes positively to the final performance; The MR-UAMF strategy plays a central role in improving both accuracy and generalization; Multimodal fusion and attention mechanisms further boost model performance by enhancing feature expressiveness. These results validate the rationality and synergy of the components in our framework.

4. Discussion

By examining the table data, it’s evident that the proposed method with MR-UAMF excels on the SMIC dataset, achieving significantly higher metrics than other deep learning methods. On the other three datasets, its performance is comparable to those methods. Notably, the optimization algorithm demonstrates exceptional performance overall. Specifically, the proposed method outperforms other deep learning approaches across all datasets. It achieves an accuracy of 0.8497 on SMIC and the highest accuracy of 0.8388 on SAMM. Table 2 shows it attains the highest accuracy of 0.8117 on CASME. On CASME II, as per Table 3, the accuracy reaches 0.8359. In summary, compared to state-of-the-art methods, the proposed MR-UAMF performs on par with existing algorithms in most cases, validating its effectiveness for micro-expression recognition. Thus, the MR-UAMF is better suited for nonlinear problems like micro-expression image recognition, enhancing the model’s nonlinear fitting ability and

ensuring superior real-world performance.

5. Conclusion

This paper proposes a micro-expression recognition method based on uncertainty awareness and multi-modal fusion (MR-UAMF), aiming to address key challenges in the field of micro-expression recognition, including limited dataset sizes, class imbalance, overfitting, and insufficient utilization of dynamic features. By introducing a focused uncertainty-aware mixing strategy (FU-MIX), our method weights samples based on their uncertainty, guiding the model to focus more on underperforming samples, thereby effectively mitigating overfitting and enhancing recognition performance for minority classes. Furthermore, by integrating optical flow parameters and spatial features from micro-expression images, our approach fully leverages both dynamic and static information, further improving recognition accuracy and model robustness. The synergistic effect of the micro-attention mechanism and a shallow triple-stream 3D convolutional neural network enables the model to efficiently extract multi-scale spatiotemporal features, achieving superior performance on complex micro-expression data. Experimental results demonstrate that MR-UAMF achieves significant performance improvements across four widely used micro-expression datasets: SMIC, CASME, CASME II, and SAMM. Ablation studies further validate the contributions of each component, with the uncertainty-aware mixing strategy and multimodal feature fusion contributing to accuracy improvements of 10% and 5%, respectively, confirming the positive synergistic impact of these components on overall performance.

Moving forward, we plan to explore micro-expression recognition in video data, investigating the impact of temporal sequence information on recognition performance.

Acknowledgements

This study was funded by the NSFC (Grant No. 61602345, 62002263); National Key Research and Development (Grant No. 2019YFB2101900); TianKai Higher Education Innovation ParkEnterprise RD Special Project (Grant No. 23YFZXCYC00046); Tianjin Science and Technology Program Projects (Grant No. 24YDTPJC00630).

Declaration of Generative AI

During the preparation of this work, the author utilized KIMI and Deepseek for grammar and spell-checking, as well as for text translation and paraphrasing (including translation and rewriting). After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

- [1] Y. Li, J. Wei, Y. Liu, J. Kauttonen, G. Zhao, Deep learning for micro-expression recognition: A survey, *IEEE Transactions on Affective Computing* 13 (2022) 2028–2046.
- [2] F. Zhang, L. Chai, A review of research on micro-expression recognition algorithms based on deep learning, *Neural Computing and Applications* 36 (2024) 17787–17828.
- [3] Z. Wang, K. Zhang, W. Luo, R. Sankaranarayana, Htnet for micro-expression recognition, *Neuro - computing* 602 (2024) 128196.
- [4] L. Zhang, X. Hong, O. Arandjelović, G. Zhao, Short and long range relation based spatio-temporal transformer for micro-expression recognition, *IEEE Transactions on Affective Computing* 13 (2022) 1973–1985.
- [5] Z. Han, Z. Liang, F. Yang, Umix: Improving importance weighting for subpopulation shift via uncertainty-aware mixup, *arXiv preprint arXiv:2209.08928* (2022). URL: <https://arxiv.org/abs/2209.08928>.

- [6] B. Song, K. Li, Y. Zong, J. Zhu, W. Zheng, Recognizing spontaneous micro-expression using a three-stream convolutional neural network, *IEEE Access* 7 (2019) 184537–184551.
- [7] B. Xia, W. Wang, S. Wang, E. Chen, Learning from macro-expression: A micro-expression recognition framework, *Proceedings of the 28th ACM International Conference on Multimedia* (2020) 2936–2944.
- [8] Y. Liu, et al., A main directional mean optical flow feature for spontaneous micro-expression recognition, *IEEE Transactions on Affective Computing* 7 (2015) 299–310.
- [9] M. Peng, Z. Wu, Z. Zhang, T. Chen, From macro to micro expression recognition: Deep learning on small datasets using transfer learning, *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (2018) 657–661.
- [10] G. Zhao, M. Pietikainen, Dynamic texture recognition using local binary patterns with an application to facial expressions, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29 (2007) 915–928.
- [11] S. Polikovsky, Y. Kameda, Y. Ohta, Facial micro expressions recognition using high speed camera and 3d-gradient descriptor, in: *IET Conference*, 2009, p. 5.
- [12] Y.-J. Liu, J.-K. Zhang, W.-J. Yan, S.-J. Wang, G. Zhao, X. Fu, A main directional mean optical flow feature for spontaneous micro-expression recognition, *IEEE Transactions on Affective Computing* 7 (2016) 299–310. doi:10.1109/TAFFC.2015.2485205.
- [13] Y. S. Gan, S.-T. Liong, W.-C. Yau, Y.-C. Huang, L.-K. Tan, Off-apexnet on micro-expression recognition system, *Signal Processing: Image Communication* 74 (2019) 129–139. doi:10.1016/j.image.2019.02.005.
- [14] S.-T. Liong, Y. S. Gan, J. See, H.-Q. Khor, Y.-C. Huang, Shallow triple stream three-dimensional cnn (ststnet) for micro-expression recognition, in: *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, IEEE, 2019, pp. 1–5.
- [15] L. Zhou, Q. Mao, L. Xue, Dual-inception network for cross-database micro-expression recognition, in: *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*, IEEE, 2019, pp. 1–5.
- [16] Z. Lai, R. Chen, J. Jia, Y. Qian, Real-time micro expression recognition based on resnet and atrous convolutions, *Journal of Ambient Intelligence and Humanized Computing* (2020). doi:10.1007/s12652-020-01779-5.
- [17] C. Wang, M. Peng, T. Bi, T. Chen, Micro-attention for micro-expression recognition, *Neurocomputing* 410 (2020) 354–362. doi:10.1016/j.neucom.2020.06.005.
- [18] L. Feng, Z. Jiahao, Q. Jiayin, Lightweight micro-expression recognition architecture based on bottleneck transformer, *Computer Science* 49 (2022) 370–377. doi:10.11896/j.sjcx.210500023.