Rethinking Key-frame-based Micro-expression Recognition: A Robust and Accurate Framework Against Key-frame Errors

Supplementary Material

Abstract

Following the sequence of the paper, we provide more detailed information regarding the causal attention block, method for calculating the direction of OF, the dataset and metrics, the results of robust MER, and the visualizations.

1. Causal Attention Block

Causal Attention Block Length. In Figure 1, we analyze the model performance when the block length is set to 1, 2, 4, and 6. Through comparison, the best block length is 1.

Original Cross Attention. For x_1 (same for x_2), the original cross attention formula is as follows:

$$x_1{}' = \text{Attn}(Q_1, K_2, V_2) + x_1 \tag{1}$$

$$\text{Attn}(Q_1, K_2, V_2) = \text{SM}(\frac{Q_1 K_2^T}{\sqrt{d_k}})V_2 \tag{2}$$

where K_2 and V_2 are obtained through a 1×1 convolution of x_2, while Q_1 is obtained via a 1×1 convolution of x_1. $\sqrt{d_k}$ serves as the scaling factor. SM is the Softmax function.

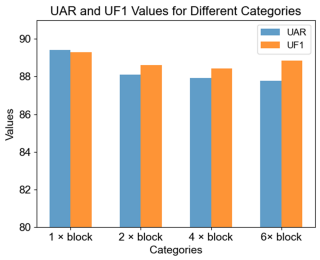


Figure 1. Analysis of length of causal attention block.

2. OF Direction Calculation

Calculation of OF Direction is shown in Algorithm 1. We map the angles to the hue. Note that the OF components are normalized before calculating the OF direction map, as shown in Fig. 3. In the HSV color space, the angle corresponding to the upward direction approximately corresponds to red, and the angle for the downward direction, approximately, corresponds to yellow-green. Fig. 2 are more visualization samples that demonstrate that the movements of AU (Action Unit) are in opposite directions during the onset-apex and apex-offset phases.

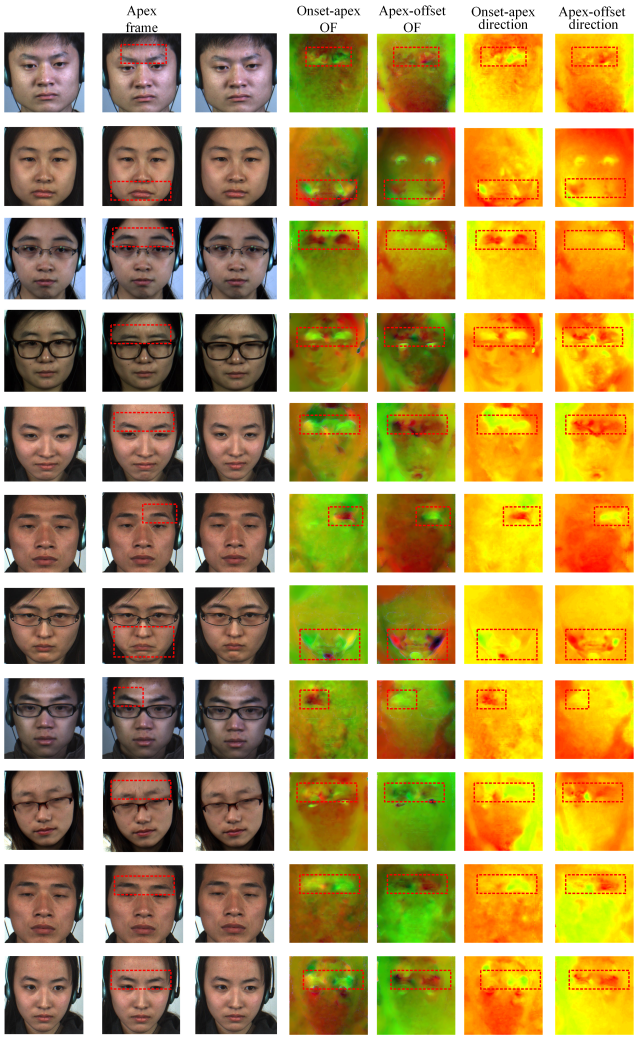
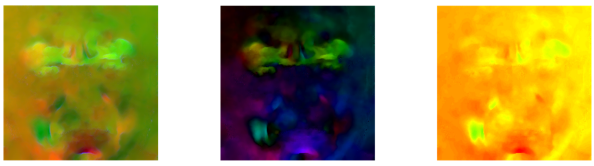


Figure 2. visualizations of OF direction maps. Red box indicates the AU-related areas.

3. Dataset and Metrics

The detailed information of datasets is summarized in Tab. 1 and Tab. 2. The validation metrics used are the unweighted F1 score (UF1), unweighted accuracy (UAR) and accuracy (ACC):

$$UF1 = \frac{1}{C-1}\sum_{i=0}^{C-1} \times \frac{2 \times TP_i}{TP_i + F_{Pi} + F_{Ni}} \tag{3}$$



OF | OF Direction | Normalized OF Direction

Figure 3. visualizations of OF direction maps and normalized OF direction maps.

Algorithm 1 OF Direction Generation

Require:
The optical flow matrix, $flow \in \mathbb{R}^{H\times W\times 2}$, where $flow[...,0]$ is the horizontal component of OF and $flow[...,1]$ is the vertical component of OF.

Ensure:
The visualized BGR image, $bgr \in \mathbb{R}^{H\times W\times 3}$.

1: Initialize an empty HSV image, $hsv \in \mathbb{R}^{H\times W\times 3}$, with data type uint8.

2: Compute the flow direction θ:

$$\theta \leftarrow \text{CartToPolar}(flow[...,1], flow[...,0])$$

3: Map the direction θ to the HSV hue channel h (range $[0, 180]$):

$$h \leftarrow \frac{\theta \times 180}{2\pi}$$

4: Set the saturation channel s and value channel v of the HSV image:

$$s \leftarrow 255, \quad v \leftarrow 255$$

5: Construct the HSV image:

$$hsv[...,0] \leftarrow h, \quad hsv[...,1] \leftarrow s, \quad hsv[...,2] \leftarrow v$$

6: Convert the HSV image to a BGR image:

$$bgr \leftarrow \text{HSVtoBGR}(hsv)$$

7: **Return** bgr.

$$UAR = \frac{1}{C-1}\sum_{i=0}^{C-1}\frac{TP_i}{N_i} \tag{4}$$

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

where C is the number of MEs, N_i is the total number of i-th ME in the dataset.

Table 1. CDE task on the composite dataset of CASME II, SMIC, and SAMM.

Database	CASME II	SMIC	SAMM	Composite
Subjects	24	16	28	68
Samples	145	164	133	442
Negative	88	70	92	250
Postive	32	51	26	109
Surprise	25	43	15	83

Table 2. Task on MMEW.

Class	3 classes	5 classes
Subjects	30	30
Samples	234	279
	109 (Negative)	72 (Disgust)
	36 (Postive)	36 (happiness)
	89 (Surprise)	89 (Surprise)
		16 (Fear)
		66 (Others)

4. More results on MMEW

In Tab. 3, we compare the performance metrics of CausalNet on the five-class classification task of MMEW.

Table 3. Method Comparison on the MMEW for five-class task.

Method	**Pub-Yr**	**MMEW ACC**
SGCN* [5]	NN 24	73.0
EDMDBN [4]	PRL 25	**81.71**
CausalNet	Proposed	**82.32**$_{0.82}$

5. Robust MER

We show the full results of robustness evaluation under key frames with different levels of errors in Tab. 4. Results show that CausalNet achieves the best performance in most indicators, which demonstrates its robustness against key-frame-index errors.

Table 4. Robust MER on the composite dataset using imprecise key frames with different STD values and CASME II using automatic apex-frame spotting algorithm. Results are averaged over three independent runs.

Method	STD=10		STD=20		STD=30		Spotting	
	UF1	UAR	UF1	UAR	UF1	UAR	UF1	UAR
Dual-Inception [9]	71.63	69.71	68.45	69.09	67.59	66.22	70.89	68.53
Micro-attention [6]	66.98	65.32	67.12	66.33	62.21	63.11	62.13	63.62
OffApexNet [2]	71.02	73.54	71.66	71.21	70.34	69.32	68.64	66.99
OffTANet [8]	72.32	72.21	71.01	70.28	71.11	69.42	71.04	69.55
SRMCL [1]	78.20	77.22	76.39	76.31	75.40	74.47	80.42	78.32
MMNET [3]	78.26	79.19	79.63	78.03	77.94	75.63	80.90	79.63
HTNet [7]	80.22	80.17	78.41	76.52	77.43	76.32	80.48	81.32
CausalNet	82.92	82.64	81.48	80.73	79.36	78.33	83.52	82.38

6. Feature Visualization

Fig. 4 presents the feature visualization of Dual-Inception [9], OffTANet [8], MMNet [3], SRMCL [1], HTNet [7] and CausalNet on CDE task. As shown in the red box, there is a certain degree of overlap of features from different classes in the methods of (a)-(e). However, this situation is alleviated in CausalNet, demonstrating the strong MER performance of CausalNet. Figure 5 shows the the confusion matrix of CausalNet on the composite dataset in one run.

References

[1] Yongtang Bao, Chenxi Wu, Peng Zhang, Caifeng Shan, Yue Qi, and Xianye Ben. Boosting micro-expression recognition via self-expression reconstruction and memory contrastive learning. *IEEE Transactions on Affective Computing*, 2024. 3, 4

[2] Y.S. Gan, Sze-Teng Liong, Wei-Chuen Yau, Yen-Chang Huang, and Lit-Ken Tan. Off-apexnet on micro-expression recognition system. *Signal Processing: Image Communication*, 74:129–139, 2019. 3

[3] Hanting Li, Mingzhe Sui, Zhaoqing Zhu, and Feng Zhao. Mmnet: Muscle motion-guided network for micro-expression recognition. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 1074–1080. International Joint Conferences on Artificial Intelligence Organization, 2022. Main Track. 3, 4

[4] Bingyang Ma, Lu Wang, Qingfen Wang, Haoran Wang, Ruolin Li, Lisheng Xu, Yongchun Li, and Hongchao Wei. Entire-detail motion dual-branch network for micro-expression recognition. *Pattern Recognition Letters*, 2025. 2

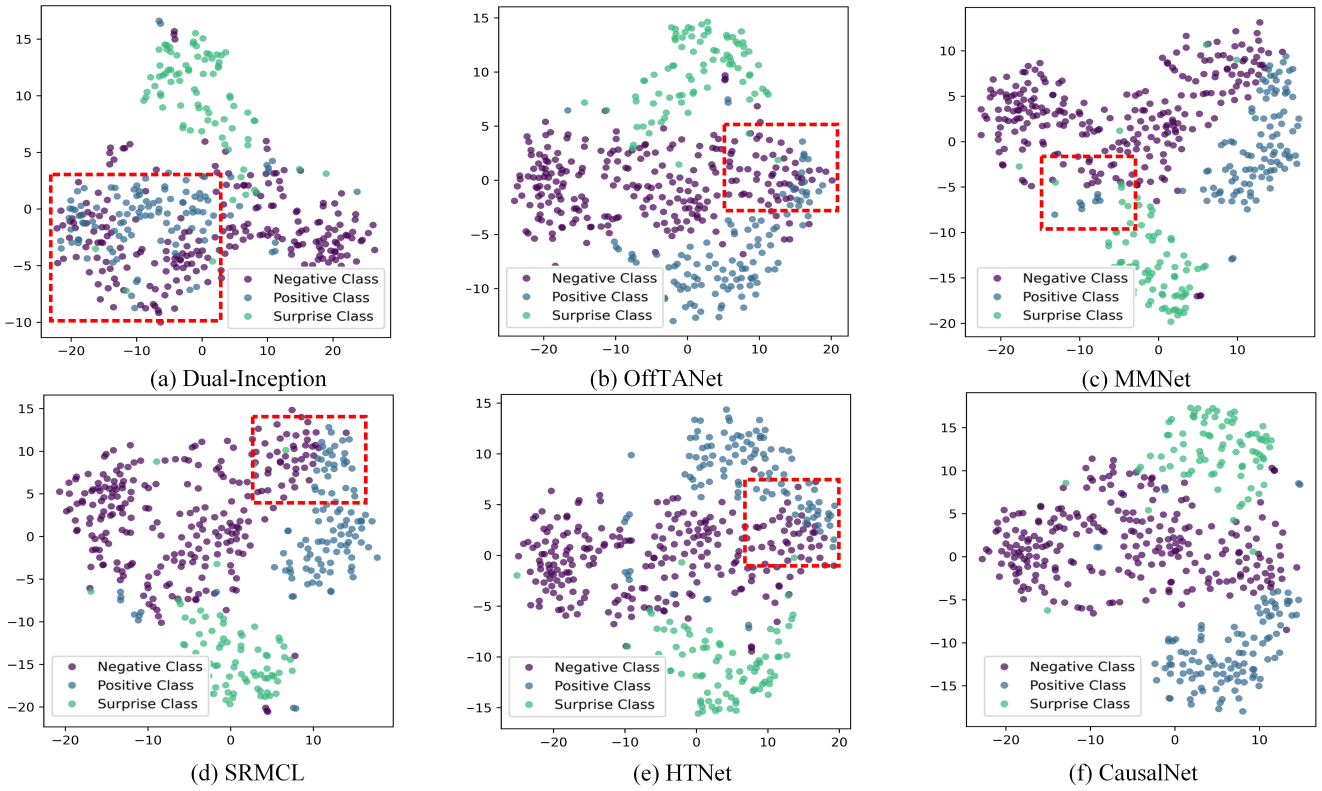
[5] Hui Tang and Li Chai. Facial micro-expression recognition using stochastic graph convolutional network and dual transferred learning. *Neural Networks*, 178:106421, 2024. 2

[6] Chongyang Wang, Min Peng, Tao Bi, and Tong Chen. Micro-attention for micro-expression recognition. *Neurocomputing*, 410:354–362, 2020. 3

[7] Zhifeng Wang, Kaihao Zhang, Wenhan Luo, and Ramesh Sankaranarayana. Htnet for micro-expression recognition. *Neurocomputing*, 602:128196, 2024. 3, 4

[8] Jiahao Zhang, Feng Liu, and Aimin Zhou. Off-tanet: A lightweight neural micro-expression recognizer with optical flow features and integrated attention mechanism. In *Pacific Rim International Conference on Artificial Intelligence*, pages 266–279. Springer, 2021. 3, 4

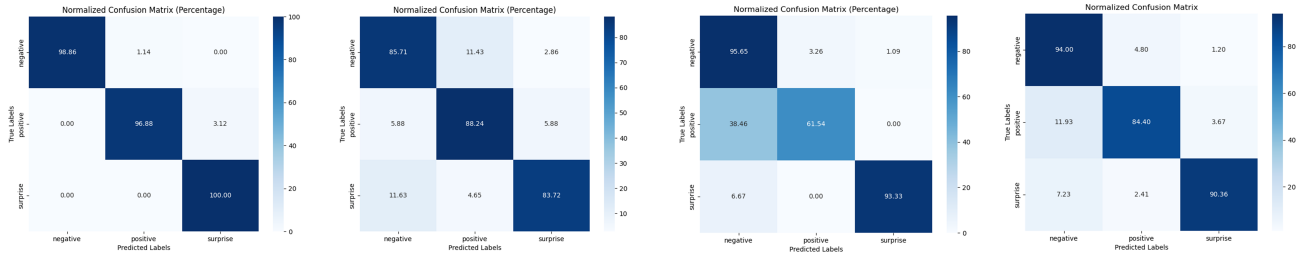
[9] Ling Zhou, Qirong Mao, and Luoyang Xue. Dual-inception network for cross-database micro-expression recognition. In *2019 14th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2019)*, pages 1–5, 2019. 3, 4



(a) Dual-Inception (b) OffTANet (c) MMNet

(d) SRMCL (e) HTNet (f) CausalNet

Figure 4. Different view of feature visualization of Dual-Inception [9], OffTANet [8], MMNet [3], SRMCL [1], HTNet [7] and CausalNet on CDE task. The red box indicates that the separation degree of features of different classes is relatively low here.



(a) CASME II (b) SMIC (c) SAMM (d) OVERALL

Figure 5. Confusion matrix of CausalNet on CDE task.