THU-PM-141

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Feature Representation Learning with Adaptive Displacement Generation and Transformer Fusion for Micro-Expression Recognition

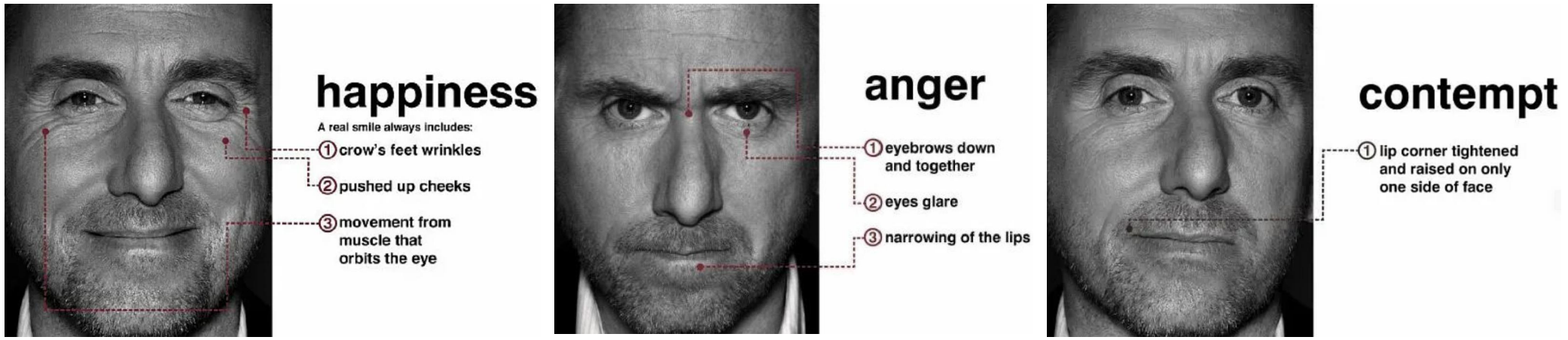
Zhijun Zhai[1], Jianhui Zhao[1*], Chengjiang Long[2], Wenju Xu[3], Shuangjiang He[4], Huijuan Zhao[4]

[1]School of Computer Science, Wuhan University, Wuhan, Hubei, China
[2]Meta Reality Labs, Burlingame, CA, USA
[3]OPPO US Research Center, InnoPeak Technology Inc, Palo Alto, CA, USA
[4]FiberHome Telecommunication Technologies Co., Ltd, Wuhan, Hubei, China

zhijunzhai@whu.edu.cn, jianhuizhao@whu.edu.cn, clong1@meta.com, wenjuxu123@gmail.com

*This work was supervised by Jianhui Zhao.

Micro Expression



Facial muscle movements under mental stress micro-responses.

Traits:

- Subtle and short-lasting for only 1/25th to 1/5th of a second.
- Unconscious reactions that reveal real emotions.

Images from TV series "lie to me".

Application

business negotiation[1]

criminal investigation[2]

General:

- Be aware of the situation, avoiding danger or deception.
- Understand or induce the thoughts of others.

Challenge:

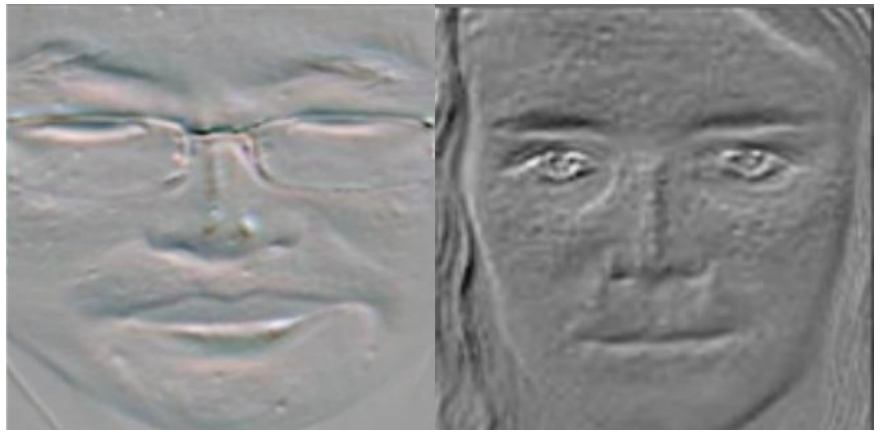
- Intrinsically low intensity and short duration.
- Heavy labor and time cost for labeling datasets.

[1]Images from https://www.youtube.com/watch?v=c4Oed7K7M9s. [2]Images from TV series "lie to me".

Dynamic Feature

Optical Flow[1]

Dynamic Image[2]

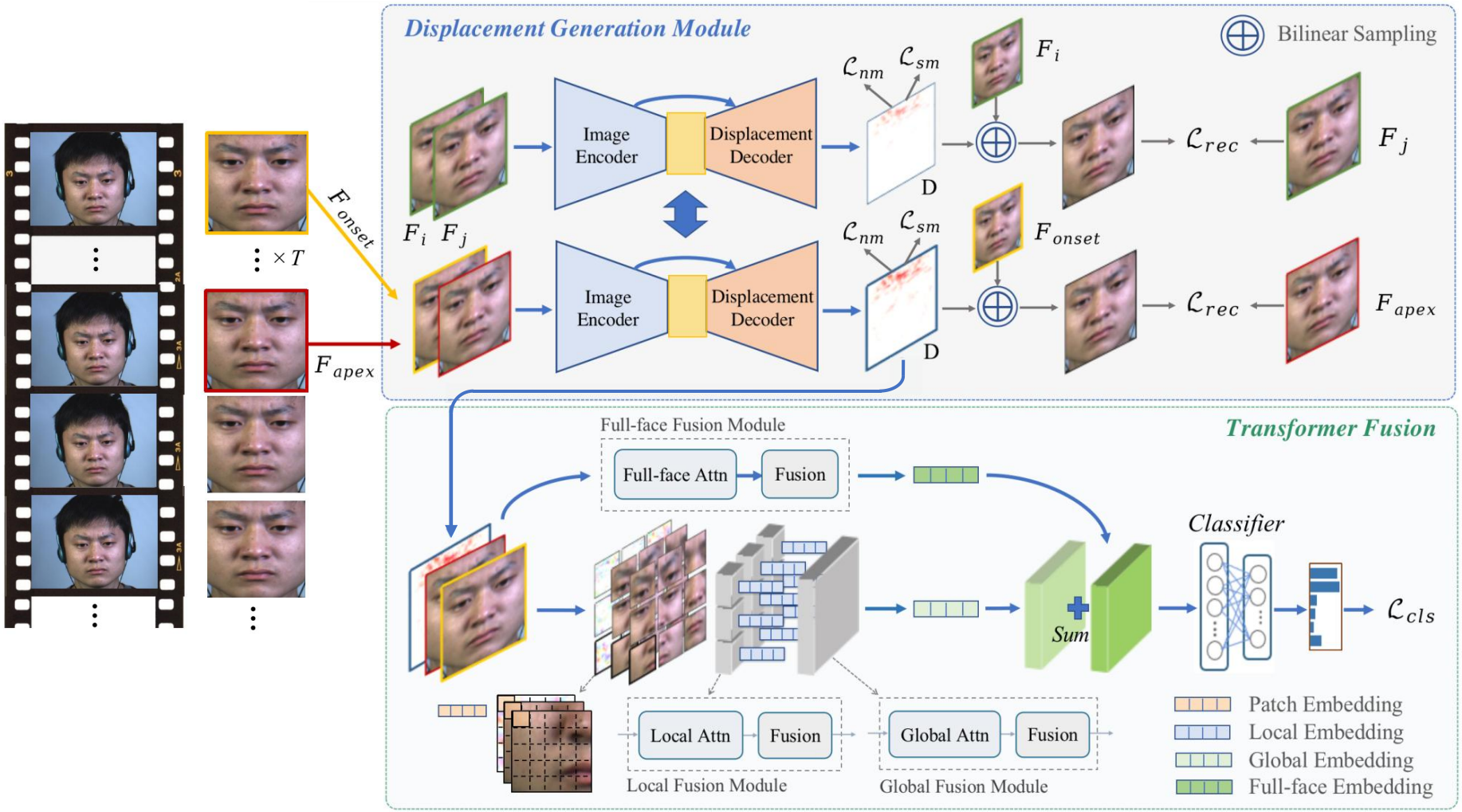


Limitations:

- Not integrated with subsequent neural networks.
- Non-adaptive to a specific task.
- Learning-based approaches are under-explored in ME recognition.

[1]videos from https://www.youtube.com/watch?v=5VyLAH8BhF8. [2]M. Verma et, al, IEEE MultiMedia, 2020.

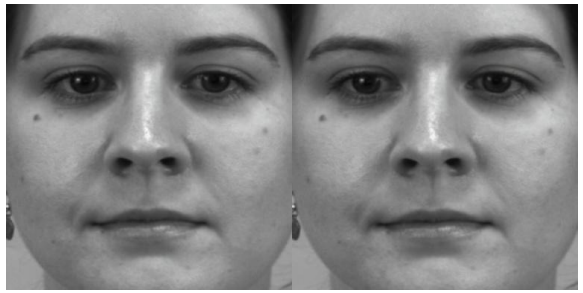
FRL-DGT Framework



Datasets

SMIC

SAMM



CASME Ⅱ

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

- **Negative:** Disgust, Contempt, Anger, Repression, Fear, Sadness
- **Positive:** Happiness
- **Surprise:** Surprise

Experimental Results

Table 1. Performance comparison of the SOTA methods and our proposed FRL-DGT.

Method	Year	Type	Full		SMIC Part		SAMM Part		CASME II Part	
			UF1	UAR	UF1	UAR	UF1	UAR	UF1	UAR
LBP-TOP	2014	Hand-Crafted	0.588	0.579	0.200	0.528	0.395	0.410	0.703	0.743
Bi-WOOF	2018	Hand-Crafted	0.630	0.623	0.573	0.583	0.521	0.514	0.781	0.803
CapsuleNet	2019	Deep-Learning	0.652	0.651	0.582	0.588	0.621	0.599	0.707	0.702
STSTNet	2019	Deep-Learning	0.735	0.761	0.680	0.701	0.659	0.681	0.838	0.869
RCN-A	2020	Deep-Learning	0.743	0.719	0.633	0.644	0.760	0.672	0.851	0.812
GEME	2021	Deep-Learning	0.740	0.750	0.629	0.657	0.687	0.654	0.840	0.851
MERSiamC3D	2021	Deep-Learning	0.807	0.799	0.736	0.760	0.748	0.728	0.882	0.876
FeatRef	2022	Deep-Learning	0.784	0.783	0.701	0.708	0.737	0.716	0.892	0.887
FRL-DGT	2022	Deep-Learning	0.812	0.811	0.743	0.749	0.772	0.758	0.919	0.903
EMRNet*	2019	Deep-Learning	0.789	0.782	0.746	0.753	0.775	0.715	0.829	0.821
FGRL-AUF*	2021	Deep-Learning	0.791	0.793	0.719	0.722	0.775	0.789	0.880	0.871
ME-PLAN*	2022	Deep-Learning	0.772	0.786	0.713	0.726	0.716	0.742	0.863	0.878

best results second best results * use different datasets higher scores

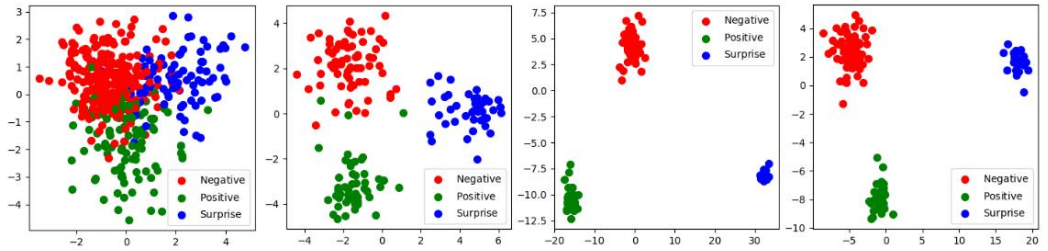
Experimental Results

Table 2. Ablation study of our proposed network.

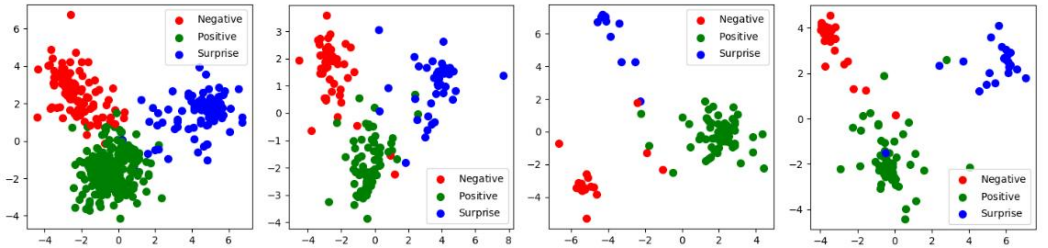
Method	DGM	AU Regions	Full-face Fusion	Global Fusion	Local Fusion	Fu-B-Attn	Full		SMIC Part		SAMM Part		CASME II Part	
							UF1	UAR	UF1	UAR	UF1	UAR	UF1	UAR
M0	→OpticalFlow	✓	✓	✓	✓	✓	0.741	0.718	0.671	0.662	0.695	0.662	0.846	0.834
M1	→OF+NORM	✓	✓	✓	✓	✓	0.758	0.739	0.671	0.667	0.778	0.730	0.869	0.831
M2	→DynamicImage	✓	✓	✓	✓	✓	0.739	0.720	0.684	0.679	0.762	0.745	0.759	0.716
M3	w/o self-supervise	✓	✓	✓	✓	✓	0.778	0.777	0.707	0.718	0.697	0.677	0.914	0.889
M4	✓	✓	✓	✓	✓	→Fu-A-Attn	0.797	0.792	0.746	0.746	0.734	0.719	0.898	0.885
M5	✓	→3x3 image patches	✓	✓	✓	✓	0.765	0.765	0.665	0.673	0.754	0.734	0.894	0.876
M6	✓	✓	×	✓	✓	✓	0.773	0.774	0.689	0.698	0.758	0.704	0.876	0.881
M7	✓	✓	✓	✓	×	✓	0.781	0.765	0.741	0.745	0.725	0.672	0.848	0.838
M8	✓	✓	✓	×	✓	✓	0.782	0.773	0.701	0.706	0.711	0.671	0.904	0.886
M9	✓	✓	✓	✓	✓	✓	0.812	0.811	0.743	0.749	0.772	0.758	0.919	0.903

→ **X:** replace the corresponding component with X
OF+NORM: normalized OpticalFlow
Fu-B-Attn: linear fusion before attention
Fu-A-Attn: linear fusion after attention

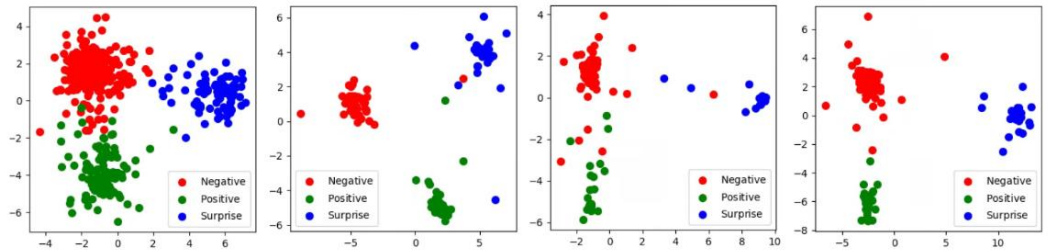
Visualization



(a) EMRNet (From left to right: Full, SMIC, SAMM, and CASME II)



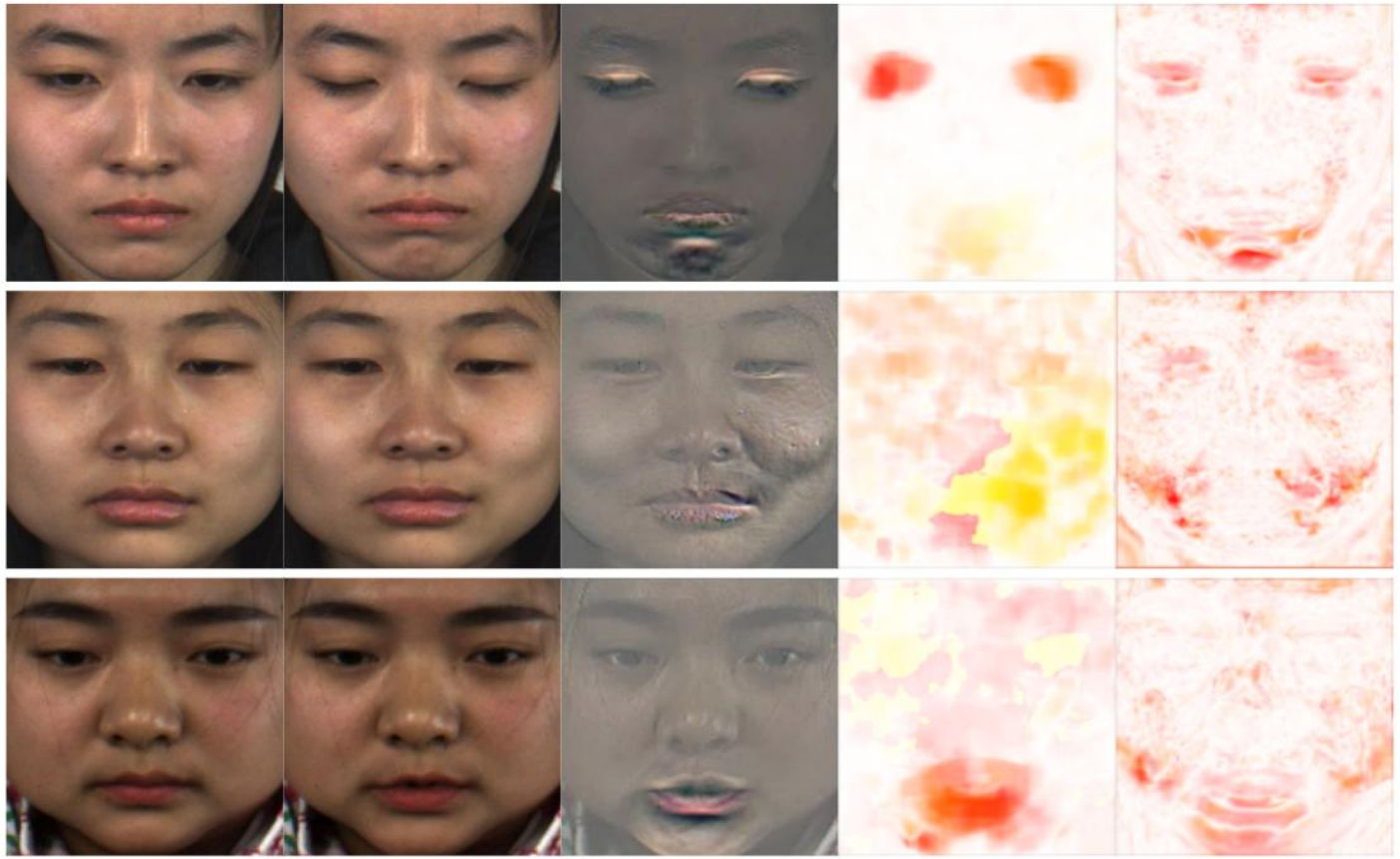
(b) FGRL-AUF (From left to right: Full, SMIC, SAMM, and CASME II)



(c) FRL-DGT (From left to right: Full, SMIC, SAMM, and CASME II)

Visualization of feature distributions.

Visualization



Visualization of dynamic features.

Conclusion & Future Work

Conclusion

- Propose a novel end-to-end **FRL-DGT** for ME recognition from onset-apex pairs.
- Design a convolutional **DGM** with self-supervised learning for targeted dynamic feature extraction, making full use of the subsequent classification supervision information.
- Design a multi-level **Transformer Fusion** module with linear fusion before attention mechanism for effective learning and integration.

Future Work

- Add an apex detection module to extend our method to ME segments with unknown apex index.
- Explore more efficient fusion mechanisms.

Paper QR Code:

https://www.chengjianglong.com/publications/FRLDGT_CVPR.pdf

Thank you!