EmotionKD: A Cross-Modal Knowledge Distillation Framework for Emotion Recognition Based on Physiological Signals

Yucheng Liu, Ziyu Jia, and Haichao Wang

Institute of Automation, Chinese Academy of Sciences, Beijing, China
Tsinghua-Berkeley Shenzhen Institute, Shenzhen, Guangdong, China

Introduction

■ ***Emotion: Recognition***

✓It is an essential aspect of affective computing that allows machines to understand human emotions;

✓Physiological signals are highly reliable indicators of emotion changes within the human body.

■ ***Application of EEG:***

✓Unimodal EEG model: Tsception, AP-CapsNet.

✓Multimodal model using EEG: MFFNN, MSMDFN.

■ ***Difficulty of Application:***

✓Causing the uncomfortable feelings;

✓Subjects' psychological responses may be affected;

✓Harsh data acquisition environment;

✓Cost of facilities is extremely expensive;

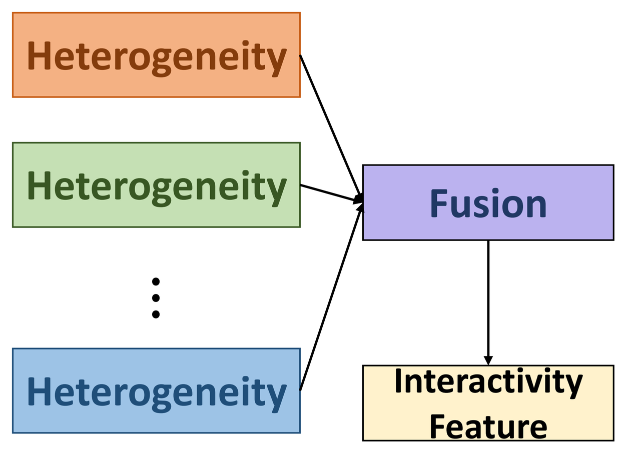
Challenges

■ ***C1: How can capture both two types of feature in multi-modal model?***

There are two kinds of important features in the multi-modal emotion recognition:

✓Heterogeneity: Distinct features within signals of different modalities.

✓Interactivity: correlation between different modalities of human physiological signals.

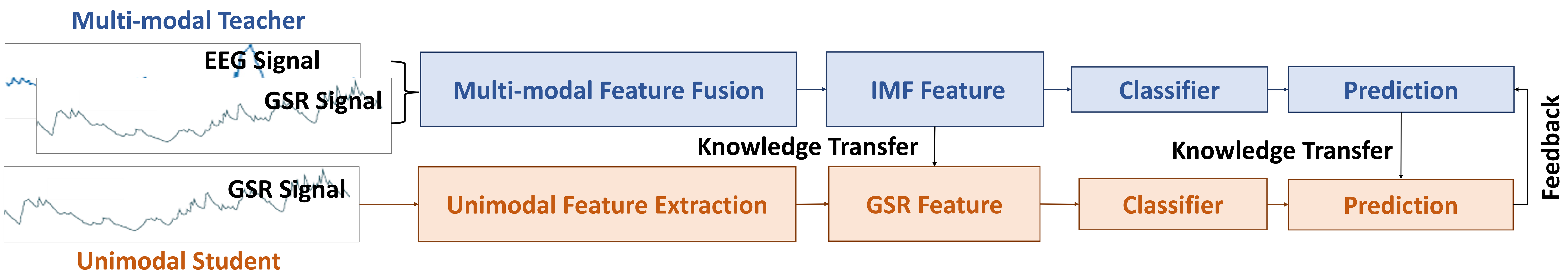


■ ***C2: How to transfer the knowledge flexibly to student model?***

✓In most knowledge distillation methods, the teacher network is fixed.

✓Teacher model cannot adjust the output feature according to the different training stage of the student.

Method

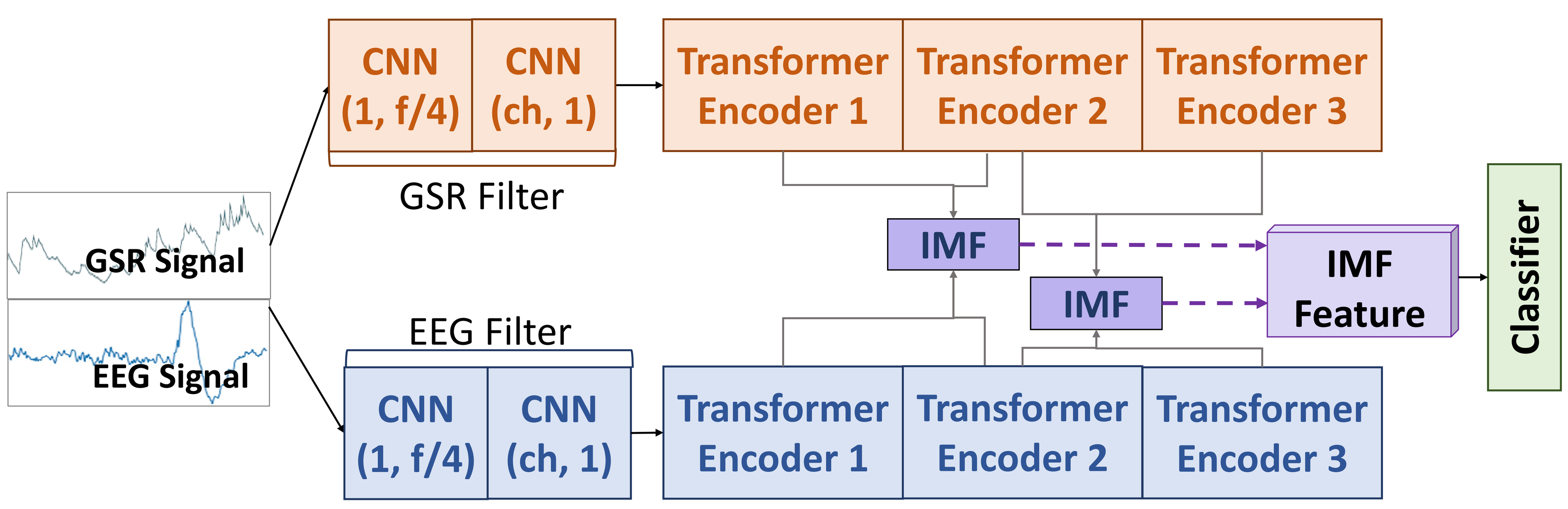


***S1:* Multimodal EmotionNet-Teacher.**

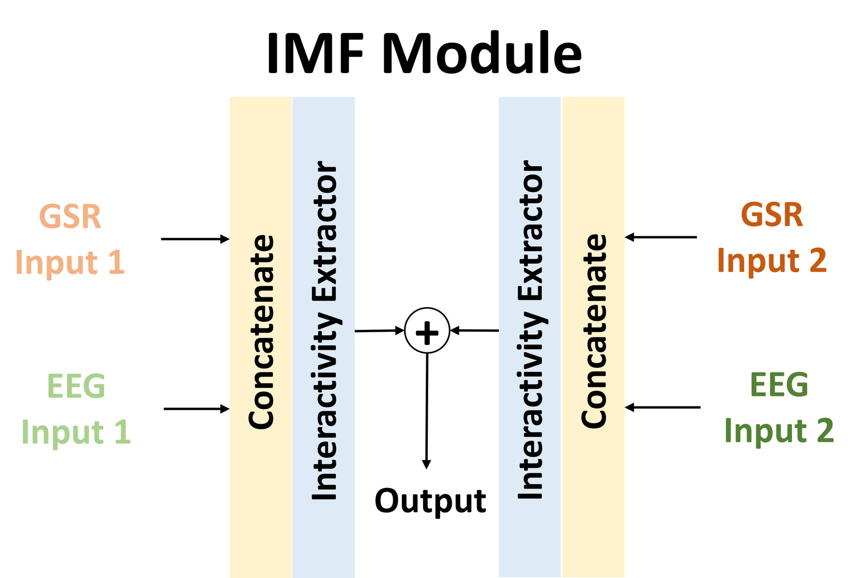
✓CNN filters for each modality.

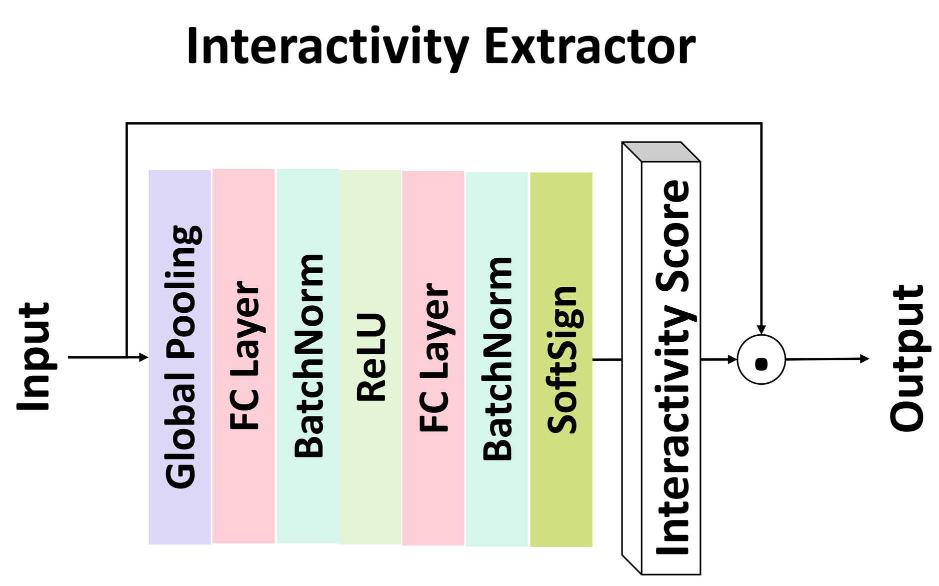
✓Dual-stream transformer structure for Heterogeneity;

✓IMF Module for interactivity extraction from feature of transformer



***S2:* IMF Module and Interactivity Extractor for interactivity extraction.**





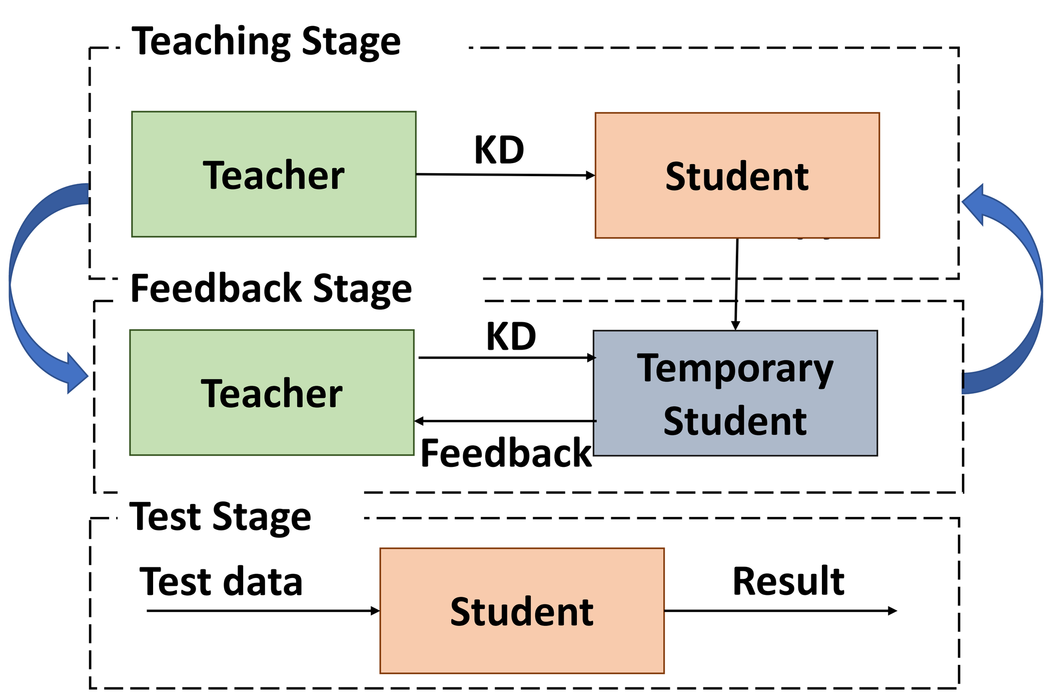
***S3:* Adaptive Feedback Knowledge Distillation**

We adding a feedback stage to the traditional knowledge distillation.

✓Training Stage

✓Feedback Stage

✓Test Stage



Results

✓ We evaluate the performance of EmotionKD on DEAP and HCI-Tagging datasets with SOTA baselines.

✓ As shown in the table, EmotionKD achieves the best overall performance compared with other baseline methods.

Comparison with the unimodal model baselines

Methods	Arousal Acc	Arousal F1-score	Valence Acc	Valence F1-score
DeepConvNet[27]	53.70	50.95	66.45	61.15
CNN+RNN[33]	53.17	36.37	67.97	64.17
CGAN[42]	53.43	46.82	55.17	35.66
CRD[37]	50.86	50.74	61.78	56.10
Visual-to-EEG KD[44]	54.90	52.59	68.66	67.36
EmotionNet-Student	**55.06**	**53.50**	**69.18**	**68.33**

Comparison with the multimodal model baselines

Methods	Arousal Acc	Arousal F1-score	Valence Acc	Valence F1-score
Concatenate	55.53	49.59	62.67	59.26
BDAE[41]	56.53	40.29	56.43	44.43
CNN-SVM[6]	56.85	42.03	62.09	58.00
EmotionNet-Teacher	**62.88**	**60.23**	**66.61**	**66.54**

Conclusion

✓We propose a novel multi-modal EmotionNet-Teacher based on a dual-stream transformer structure with an Interactivity-based Modal Fusion (IMF) module;

✓We design an adaptive feedback mechanism for cross-modal knowledge distillation;

✓The proposed EmotionKD method is the first application of cross-modal knowledge distillation in the field of physiological signal-based emotion recognition to transfer fused EEG and GSR features to the unimodal GSR model.

Webpage: *https://yuchengliu-alex.github.io/EmotionKD/* ***code:*** https://github.com/YuchengLiu-Alex/EmotionKD