ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing

Improving EEG-based Emotion Recognition by Fusing Time-Frequency and Spatial Representations

Kexin Zhu; Xulong Zhang; Jianzong Wang; Ning Cheng; Jing Xiao
Ping An Technology (Shenzhen) Co., Ltd., China

Group 6

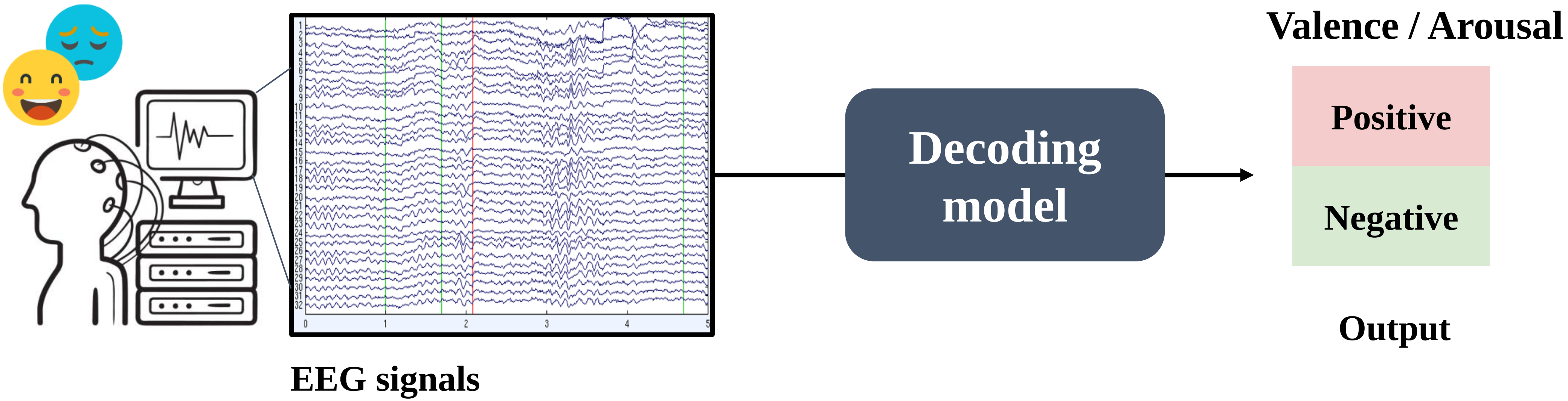
312551077 薛祖恩　312605009 彭宇楨　412551030 黃羿寧

Contents

Summary of Key Contributions

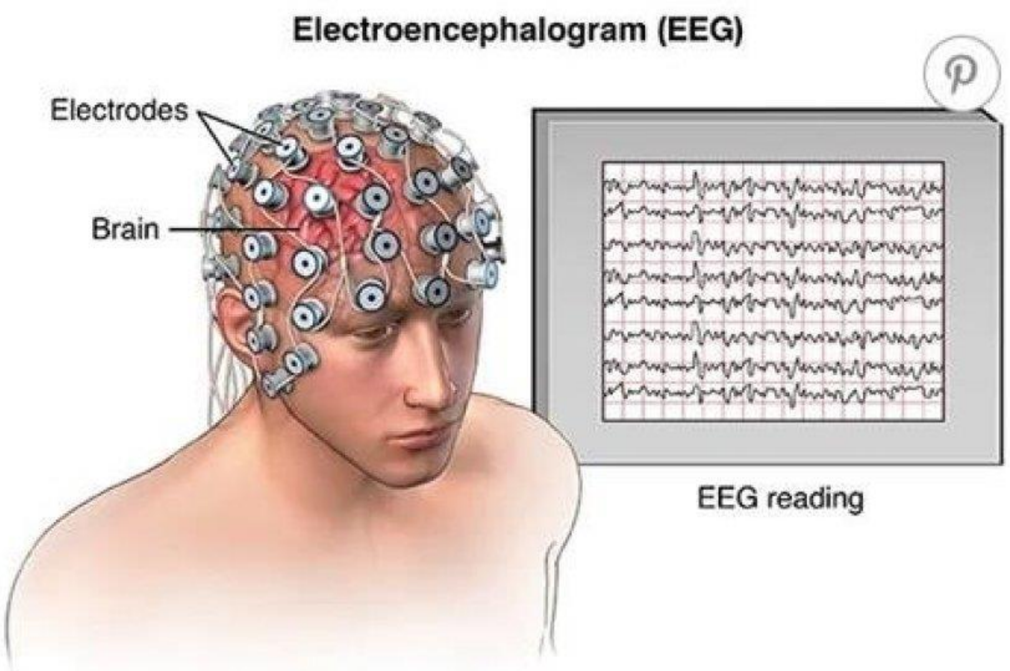
- **EEG Multi-Domain Feature Fusion**
  - Utilize cross-domain attention to fuse spatial representations with time-frequency features for improved feature selection.
  - Developed a two-step fusion method to preserve feature information from both time-frequency and spatial domains.
- **Superior Performance of Subject-independent Emotion Recognition Network**
  - Achieve valence accuracy of 0.859 (with GAT), 0.861 (with GCN)
  - Achieve arousal accuracy of 0.878 (with GAT), 0.884 (with GCN)

Schema of EEG-based Emotion Recognition



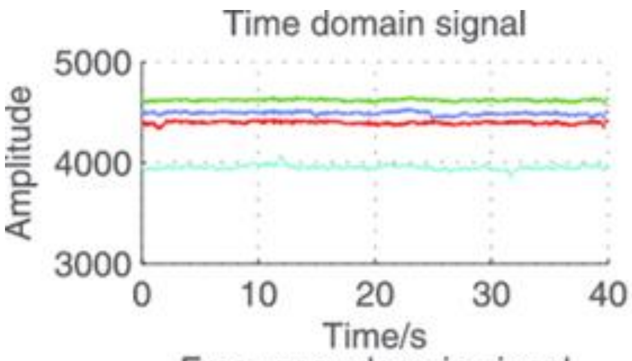
https://eeglab.org/tutorials/06_RejectArtifacts/Scrolling_data.html

Electroencephalography (EEG)

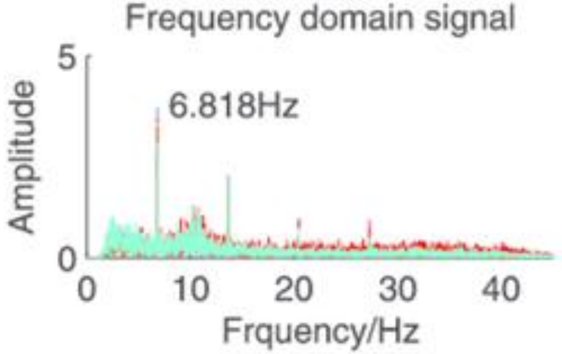


EEG measurement

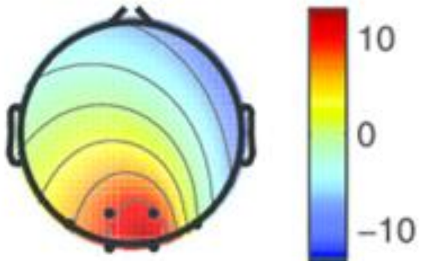
Temporal Domain



Spectral Domain



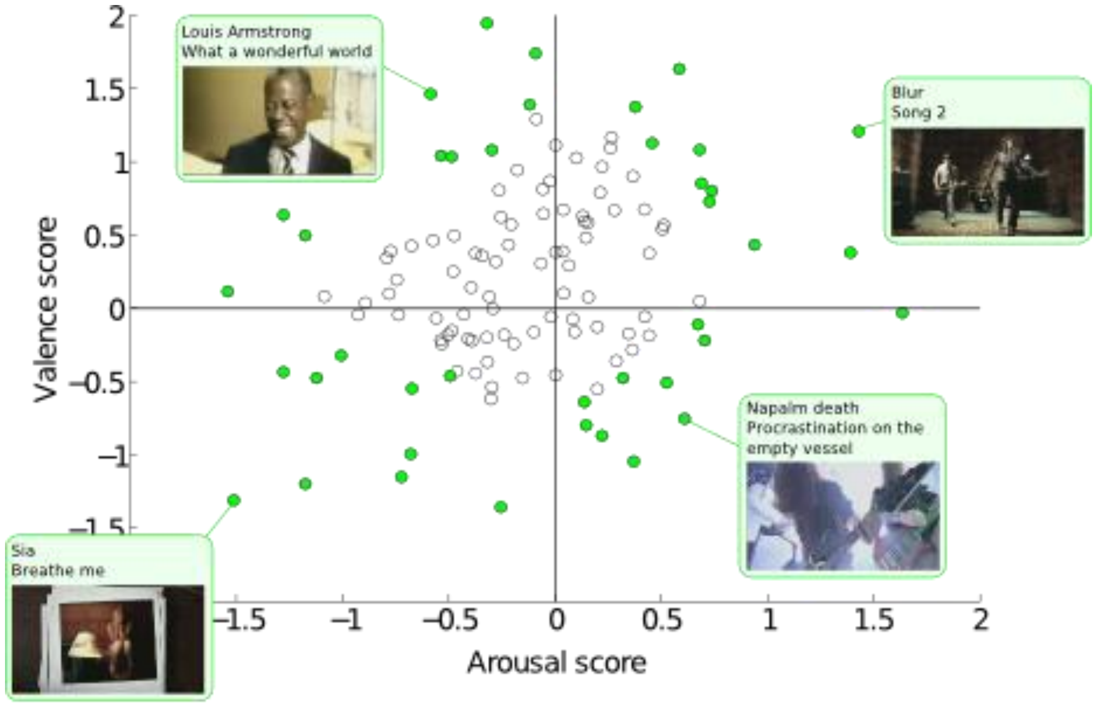
Spatial Domain

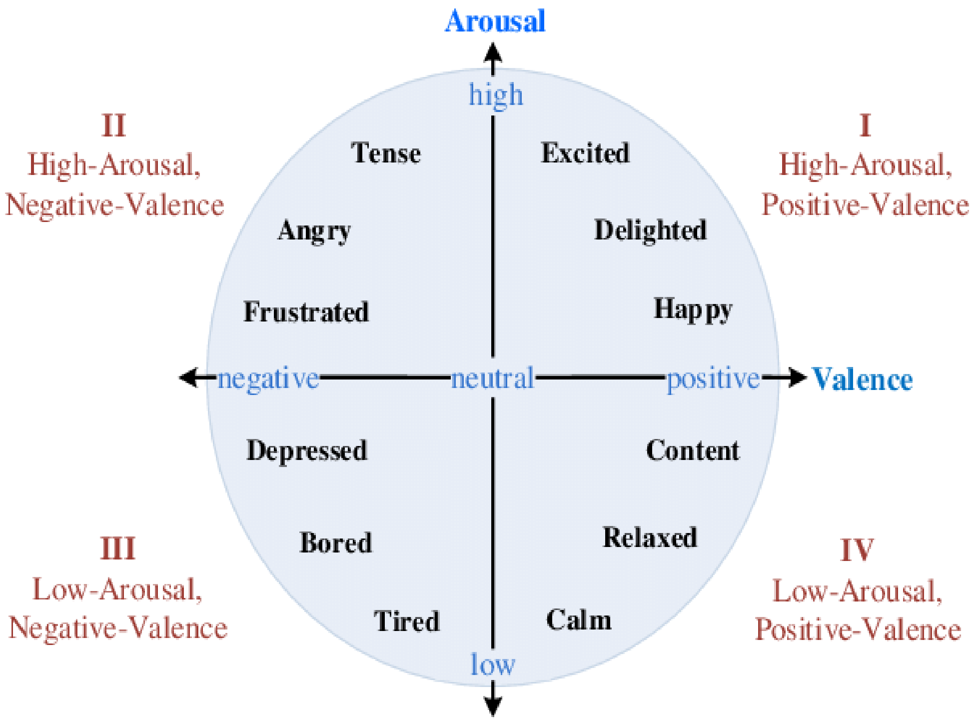


https://www.brightbraincentre.co.uk/electroencephalogram-eeg-brainwaves/
https://www.researchgate.net/publication/298727786_Brain-Computer_Interface_Controlled_Cyborg_Establishing_a_Functional_Information_Transfer_Pathway_from_Human_Brain_to_Cockroach_Brain

Emotional EEG Dataset: DEAP Dataset [16]

- 32 subjects (16 males, 16 females), 40 music videos for each participants
- Rate **valence** and **arousal** scores





Existing Models and Approaches

- In recent years, many EEG classification models based on temporal, frequency, and spatial features have been proposed

Authors	Methods	Key Features
Hou and Jia et al. [6]	LSTM for temporal feature extraction GCN for topological structure modeling	Combines temporal and spatial features
He et al. [7]	Channel attention in MLP	Adaptive learning of channel importance
Yin et al. [8]	GCN for spatial feature extraction LSTM for temporal relationship memorization	Integrates spatial and temporal features
He and Zhong et al. [9]	Temporal convolution networks Adversarial discriminative domain adaptation	Addresses domain drift in cross-subject emotion recognition

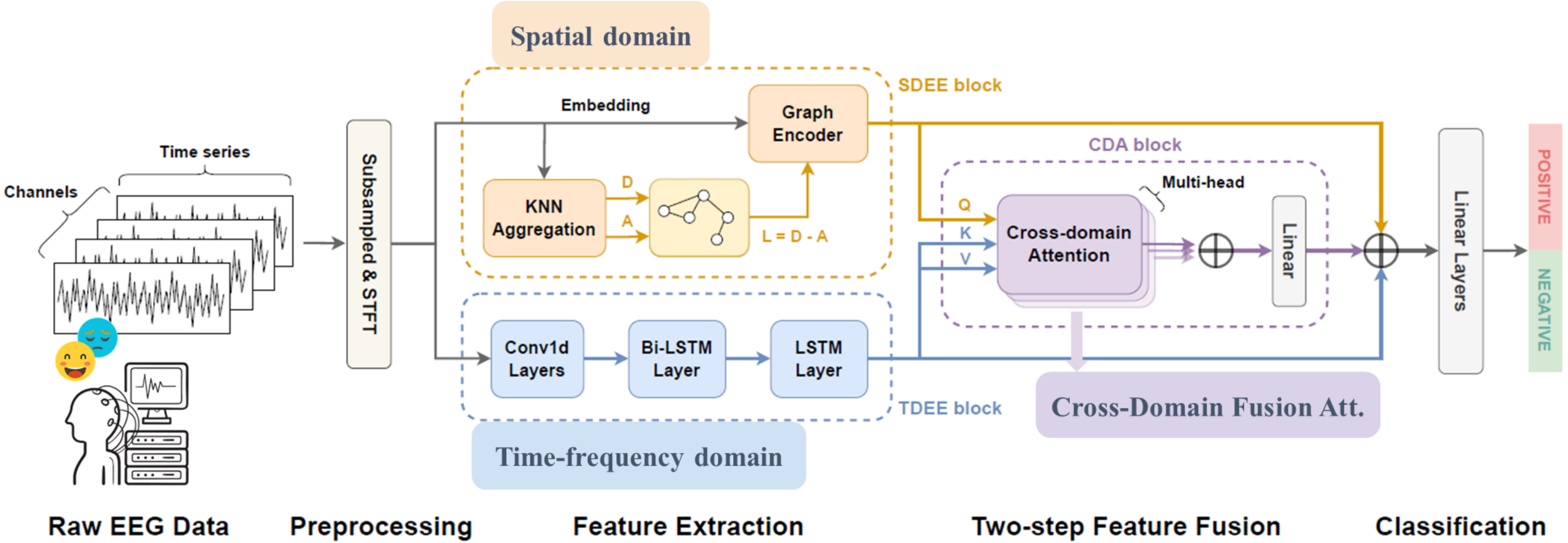
Research Gaps

- **Representation Limitations**
  - Some of the existing works focus on the representations of different domains, lacking the **mapping process of features between representations.**

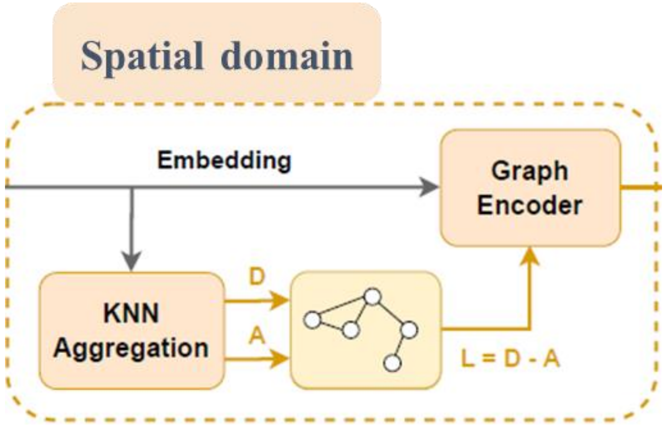
- **Fusion Challenges**
  - Some fusion methods are difficult to **combine different levels of feature information** to comprehensively model EEG signals.

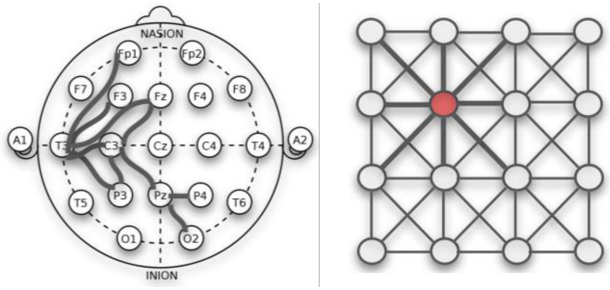
➢ The authors proposed a method for **EEG multi-domain feature fusion** using cross-domain attention, which utilizes information from spatial representations to assist in selecting time-frequency features.

Overview of the proposed model: Multi-domain Feature Fusion Architecture



Spatial-domain Embedding & Encoding Block (SDEE Block)





Graph encoder: Introduce the connection between each channel

1. Mapping brain networks to graph structures
2. Using **KNN** to construct flexible adjacency and degree matrix
3. Graph convolution operations: **GCN [11] & GAT [12]**

GAT [12] improves the normalization constant in the GCN layer into the neighbor node feature aggregation function using attention weight

Spatial-domain Embedding & Encoding Block (SDEE Block)

- **Graph convolution operations**

GCN [11]

$$h_i^{l+1} = \sigma(\sum_j \frac{h_i^l W^l}{c_{ij}}) \tag{1}$$

$$c_{ij} = \sqrt{d_i d_j} \tag{2}$$

h: feature representation vector
i: node
j: neighboring node
W: trainable parameter matrix
d: degree of node
$\vec{a}$: attentional weight vector

GAT [12]

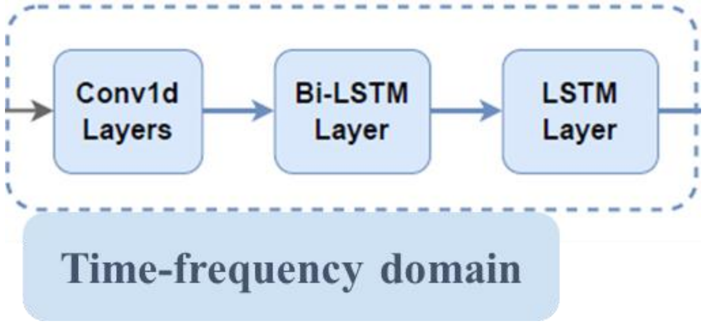
$$h_i^{l+1} = \sigma(\sum_j \alpha_{ij}^l z_j^l) \tag{3}$$

$$\alpha_{ij}^l = \frac{exp(e_{ij}^l)}{\sum_k exp(e_{ik}^l)} \tag{4}$$

$$e_{ij}^l = LeakyReLU((\vec{a}^l)^T (z_i^l || z_j^l)) \tag{5}$$

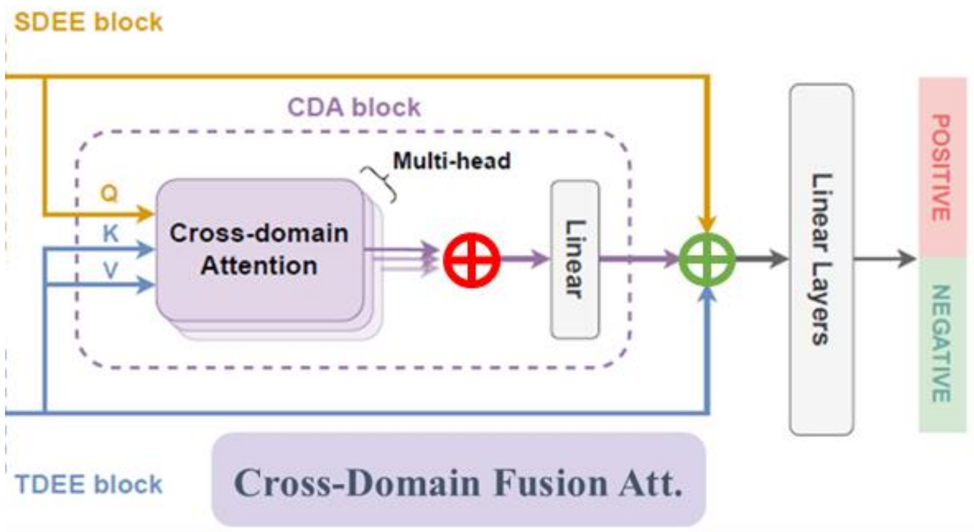
$$z_i^l = h_i^l W^l \tag{6}$$

Time-domain Embedding & Encoding Block (TDEE Block)



- **CRNN [10]**
  1. **Three 1-dimension convolution**: Extract channel correlation information by convolution operation between channels
  2. **Bi-LSTM & LSTM**: Encode the context relationship by LSTM layers

Cross-domain Attention Block (CDA Block)



- **Two-step fusion**

  ⊕ $$X_{CM} = MultiHead(X_\alpha, X_\beta) \quad (13)$$

  ⊕ $$X_{FC} = Concat(X_\alpha, X_\beta, X_{CM}) \quad (14)$$

- **Multi-head cross-domain attention [13, 14]**

$$Attention(Q_\alpha, K_\beta, V_\beta) = Softmax(\frac{Q_\alpha (K_\beta)^T}{\sqrt{d}})V_\beta \quad (10)$$

α : SDEE block output feature vector (i.e. spatial domain)
β : TDEE block output feature vector (i.e. time frequency domain)

Experimental setup

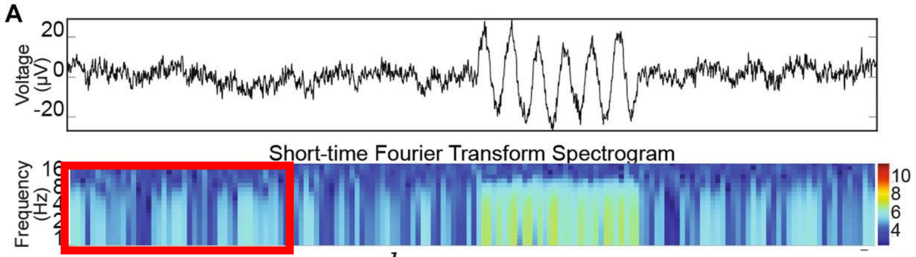
1. Dataset: DEAP Dataset [16] (32 subjects, 32 EEG channels, 40 videos, 60 sec/video)

2. Subject-independent Model: Leave-one-subject-out (LOSO) cross validation

3. Data Preprocessing:

a. **Epoch** with sliding window of a 2 second-width in steps of 0.125 seconds (480 epochs/video).

b. **Short-time Fourier transform (STFT)** is used to calculate differential entropy (DE) as the frequency domain features in five bands, including Delta, Theta, Alpha, Beta, and Gamma.



$$DE = -\int_a^b p(x)\, log(p(x))\, dx$$

32 (channels) × 5 (frequency band) — **SDEE Block**

32 (channels) × 5 (frequency band) × ~512 (time sample) — **TDEE Block**

Classification performance of emotion recognition

Table 1. Comparison between our proposed method and other methods

Study	Feature(s)	Accuracy	
		Valence	Arousal
Li et al. [17]	T-F	0.691	0.710
Wang et al. [5]	SFM	0.712	0.713
Atkinson et al. [18]	mRMR	0.731	0.730
Guo et al. [19]	T-F, FuzzyEn	0.844	0.856
Ours (with GAT)	T-F, Graph	0.859	0.878
Ours (with GCN)	**T-F, Graph**	**0.861**	**0.884**

Note: **GAT has higher computational efficiency** where T-F represents the time-frequency feature, SFM represents spatial-frequency matrices, and mRNR represents minimum-Redundancy-Maximum-Relevance.

Ablation study of blocks

Table 2. Ablation study in proposed method (with GCN) and comparison experiment of different fusion methods.

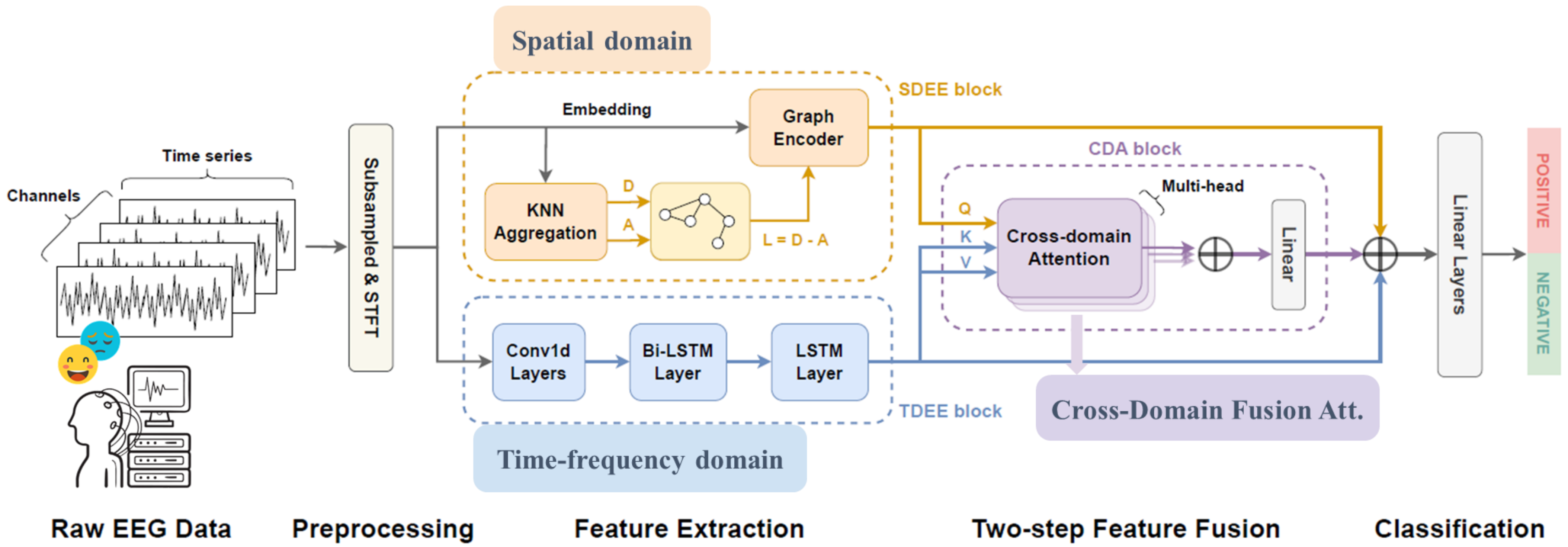
SDEE	TDEE	CDA	Fusion	Accuracy	
				Valence	Arousal
✓			-	0.530	0.512
	✓		-	0.834	0.840
✓	✓		Concat	0.849	0.864
✓	✓	✓	One-step	0.855	0.867
✓	✓	✓	**Two-step**	**0.861**	**0.884**

Summary of the results

- **The proposed multi-domain feature fusion model significantly improves EEG-based emotion recognition.**
  - Achieve valence accuracy of 0.859 (with GAT), 0.861 (with GCN)
  - Achieve arousal accuracy of 0.878 (with GAT), 0.884 (with GCN)
- **The two-step fusion method and cross-domain attention are crucial for achieving high performance.**

Key Ideas Recap

- **EEG Multi-Domain Feature Fusion**
  - Utilize cross-domain attention to fuse spatial representations with time-frequency features for improved feature selection.



Critical questions

- **Generalization to other tasks:**
  - Can the proposed multi-domain feature fusion method be **generalized** to other EEG-based tasks beyond emotion recognition.
- **Data preprocessing:**
  - Preprocessing steps are necessary to prepare the EEG data for input into the proposed model, but **they were not clearly listed in the paper**
- **Parameter usage:**
  - Lack information in TDEE block: Conv1d (kernal size, stride, padding)
  - k value in kNN was not provided

Thanks for your listening

References

[5] Chen Wang, Jingzhao Hu, Ke Liu, Qiaomei Jia, Jiayue Chen, Kun Yang, and Jun Feng, “EEG-based emotion recognition fusing spacial-frequency domain features and data-driven spectrogram-like features,” in International Symposium on Bioinformatics Research and Application, 2021, vol. 13064, pp. 460–470.

[6] Yimin Hou, Shuyue Jia, Shu Zhang, Xiangmin Lun, Yan Shi, Y. Li, Hanrui Yang, Rui Zeng, and Jinglei Lv, “Deep feature mining via attention-based BiLSTM-GCN for human motor imagery recognition,” ArXiv, vol. abs/2005.00777, 2020.

[7] Yanbin He, Zhiyang Lu, Jun Wang, and Jun Shi, “A channel attention based MLP-Mixer network for motor imagery decoding with EEG,” in IEEE International Conference on Acoustics, Speech and Signal Processing, 2022, pp. 1291–1295.

[8] Yongqiang Yin, Xiangwei Zheng, Bin Hu, Yuang Zhang, and Xinchun Cui, “EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM,” Appl. Soft Comput., vol. 100, pp. 106954, 2021.

[9] Zhipeng He, Yongshi Zhong, and Jiahui Pan, “Joint temporal convolutional networks and adversarial discriminative domain adaptation for EEG-based cross-subject emotion recognition,” in IEEE International Conference on Acoustics, Speech and Signal Processing, 2022, pp. 3214–3218.

[10] Baoguang Shi, Xiang Bai, and Cong Yao, “An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 11, pp. 2298–2304, 2017.

[11] Andac Demir, Toshiaki Koike-Akino, Ye Wang, Masaki Haruna, and Deniz Erdogmus, “EEG-GNN: graph neural networks for classification of electroencephalogram (EEG) signals,” in Annual International Conference of the IEEE Engineering in Medicine & Biology Society, 2021, pp. 1061–1067.

[12] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio, “Graph attention networks,” in International Conference on Learning Representations, 2018.

[16] Sander Koelstra, Christian Mühl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras, “DEAP: A database for emotion analysis using physiological signals,” IEEE Trans. Affect. Comput., vol. 3, no. 1, pp. 18–31, 2012.

[17] Chao Li, Boyang Chen, Ziping Zhao, Nicholas Cummins, and Björn W. Schuller, “Hierarchical attention-based temporal convolutional networks for EEG-based emotion recognition,” in IEEE International Conference on Acoustics, Speech and Signal Processing, 2021, pp. 1240–1244.

[18] John Atkinson and Daniel Campos, “Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers,” Expert Syst. Appl., vol. 47, pp. 35–41, 2016.

[19] G. Guo and Y. Gao, “Multi feature fusion EEG emotion recognition,” in International Conference on Big Data and Information Analytics, 2021, pp. 280–284.