

Yanping ZHU, Lei HUANG, Jixin CHEN, Shenyun WANG, Fayu WAN, Jianan CHEN, 2024. VG-DOCOCoT: a novel DO-Conv and transformer framework via VAE-GAN technique for EEG emotion recognition.

Frontiers of Information Technology & Electronic Engineering, 25(11):1497-1514.

<https://doi.org/10.1631/FITEE.2300781>

VG-DOCOCoT: a novel DO-Conv and transformer framework via VAE-GAN technique for EEG emotion recognition

Key words: Emotion recognition; Electroencephalogram (EEG); Transformer; Depthwise over-parameterized convolutional (DO-Conv); Variational automatic encoder-generative adversarial network (VAE-GAN)

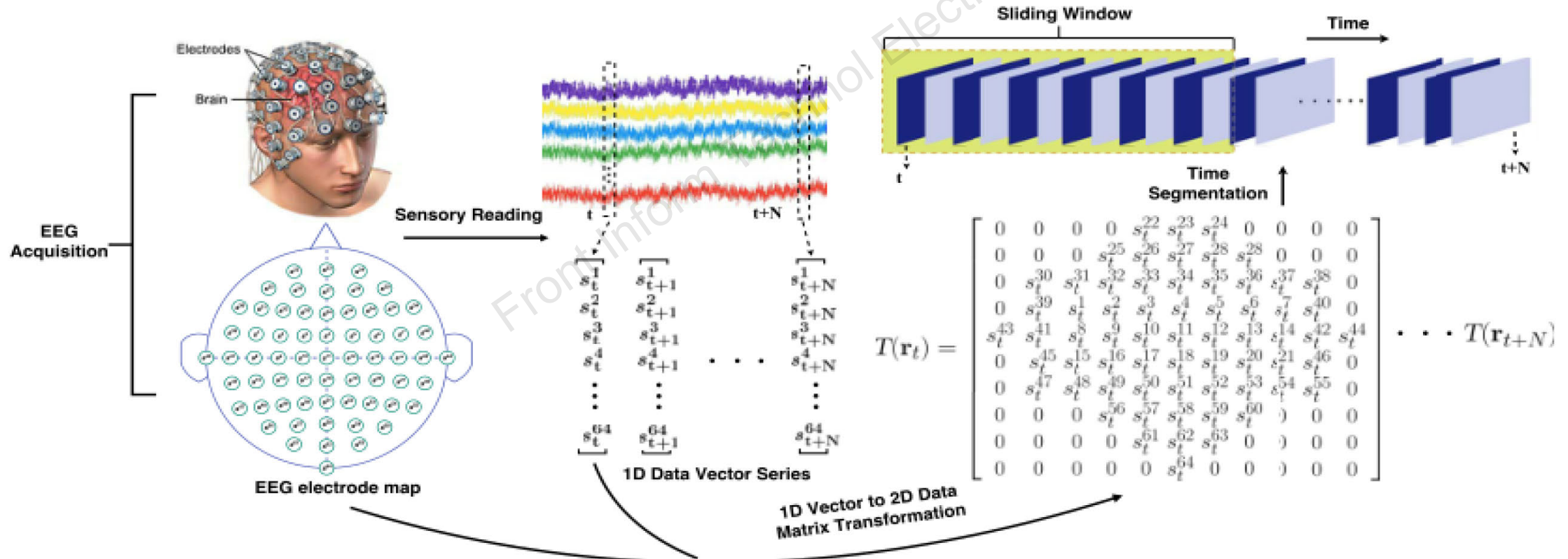
Corresponding author: Yanping ZHU

E-mail: 001520@nuist.edu.cn

 ORCID: <https://orcid.org/0000-0002-7366-1972>

Background

- Emotion recognition has attracted more and more attention in recent years as emotional computing, proposed by Picard (2000), continues to evolve.

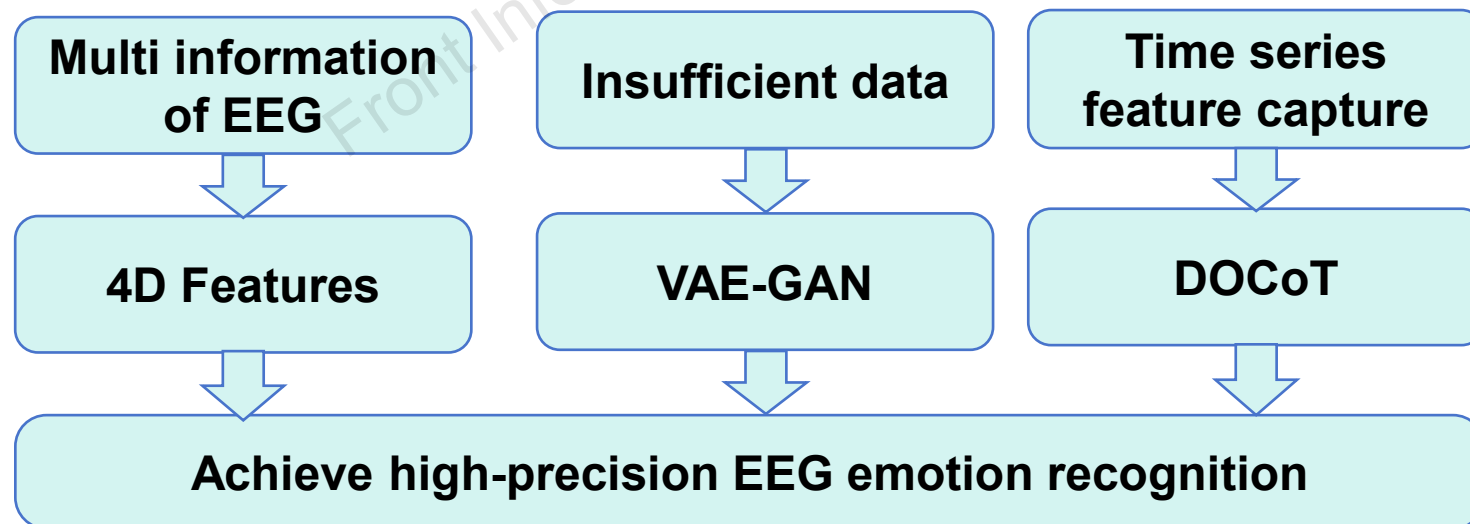


Motivation

- Previous studies often **overlooked** the **temporal** and **wave characteristics** of EEG signals, despite the fact that subjects' specific emotions usually evolve with EEG fluctuations during experiments.
- Considering that the datasets widely recognized in the field of emotion recognition, such as SEED and DEAP, have **less available data**, the improvement of emotion recognition accuracy is **limited** to a certain extent.
- **Attention** mechanism has become a research hotspot in the field of deep learning, which is famous for its good ability to **model context relations** and **capture long-distance dependencies**, and has a good effect on **time series feature capture of EEG**.

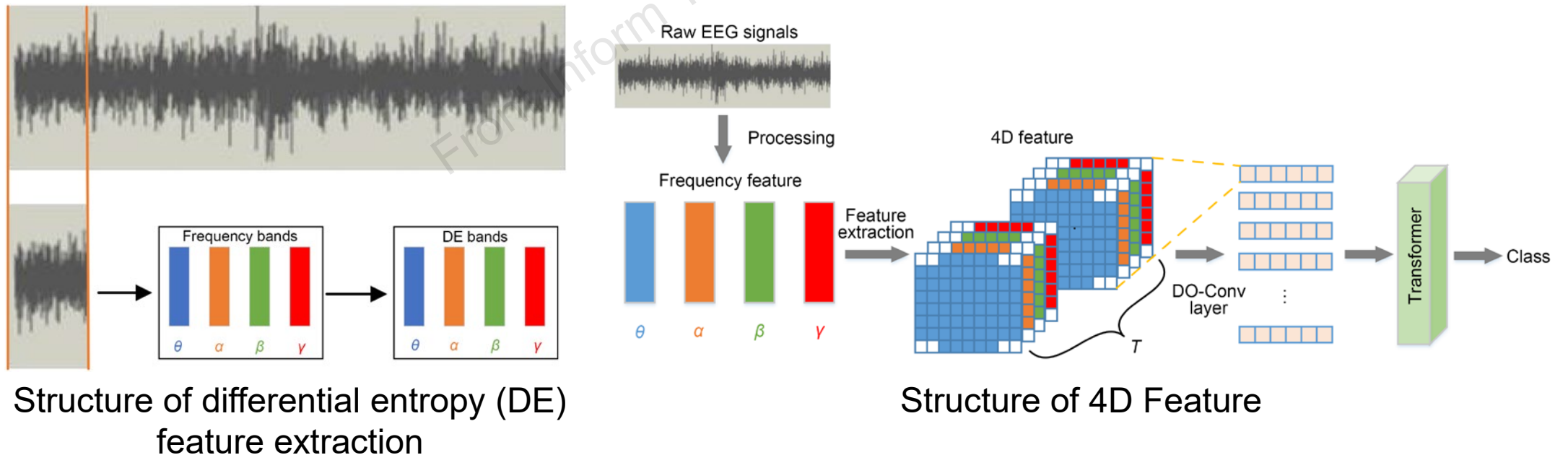
Main idea

- Propose an improved **4D feature structure** that incorporates **frequency**, **spatial**, and **temporal information** of EEG signals. This structure will preserve various data features of signals and provide a more comprehensive representation of the data.
- To address limited data availability, **VAE-GAN** is used for **data augmentation**, enhancing the diversity of the training data. This improves the model's performance, generalization ability, robustness, and accuracy.
- A novel EEG-based emotion recognition model, **VG-DOCoT**, utilizes a **depthwise over-parameterized convolution structure** for local feature extraction across channels and a **transformer structure** to capture temporal dependencies.



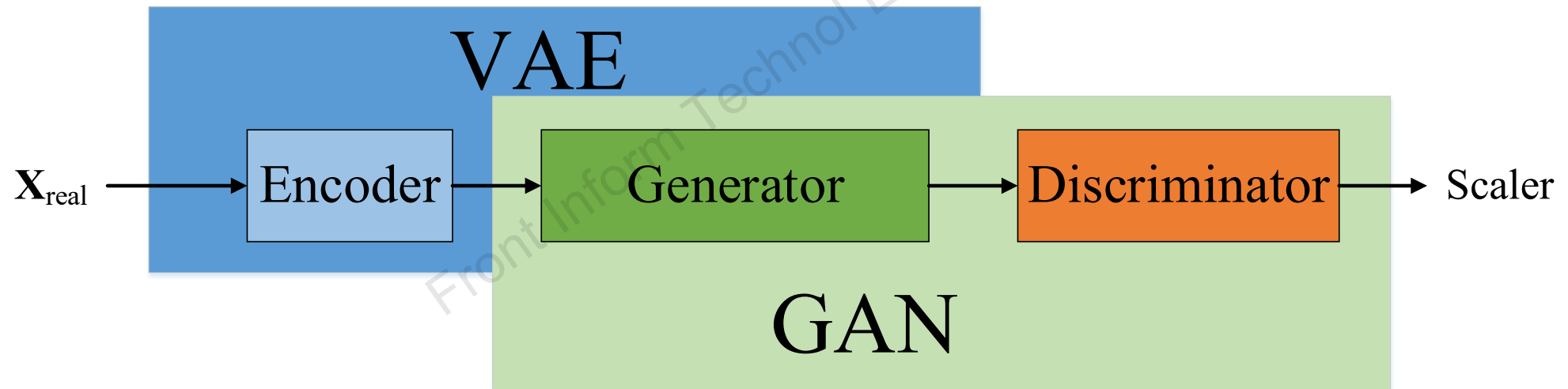
Method (1)

- Use band-pass filters to eliminate power frequency interference.
- Remove the baseline.
- Divide time segments with a 0.5 second **sliding window**.
- **DE features** are calculated for four frequency bands in each time segment to construct a **four-dimensional feature** structure.



Method (2)

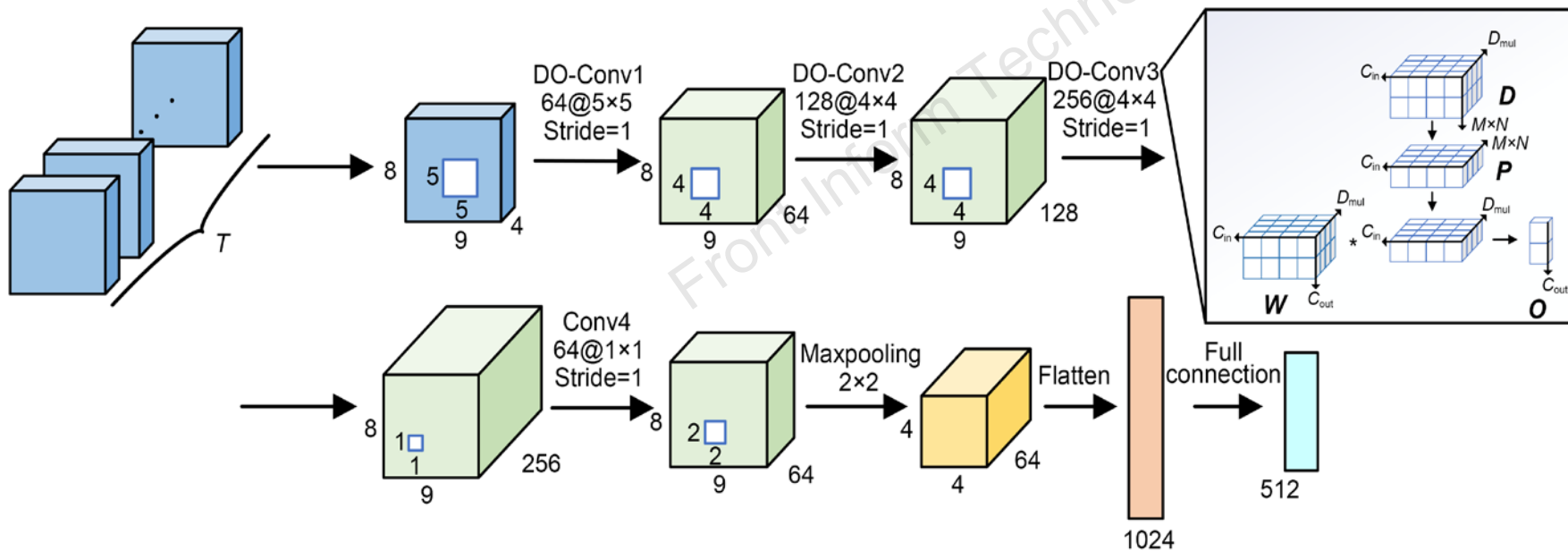
- Receive the preprocessed signal \mathbf{X}_{real} .
- **Expanded** the training set part of the dataset for model training with **VAE-GAN**.



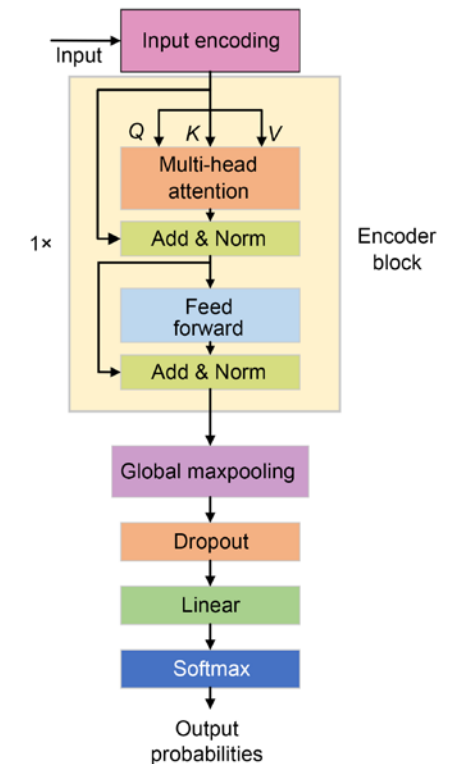
Variational automatic encoder-generative adversarial network (VAE-GAN) model diagram

Method (3)

- DCoT model is composed of **DO-Conv structure** and **Transformer structure**.
- The DO-Conv structure obtains the **spatial** and **frequency** characteristics of EEG signals by **deeply over-parameterized convolutional layers**.
- The transformer architecture relies on **multi-head attention mechanisms** and **location coding** to capture the **temporal** characteristics of EEG signals

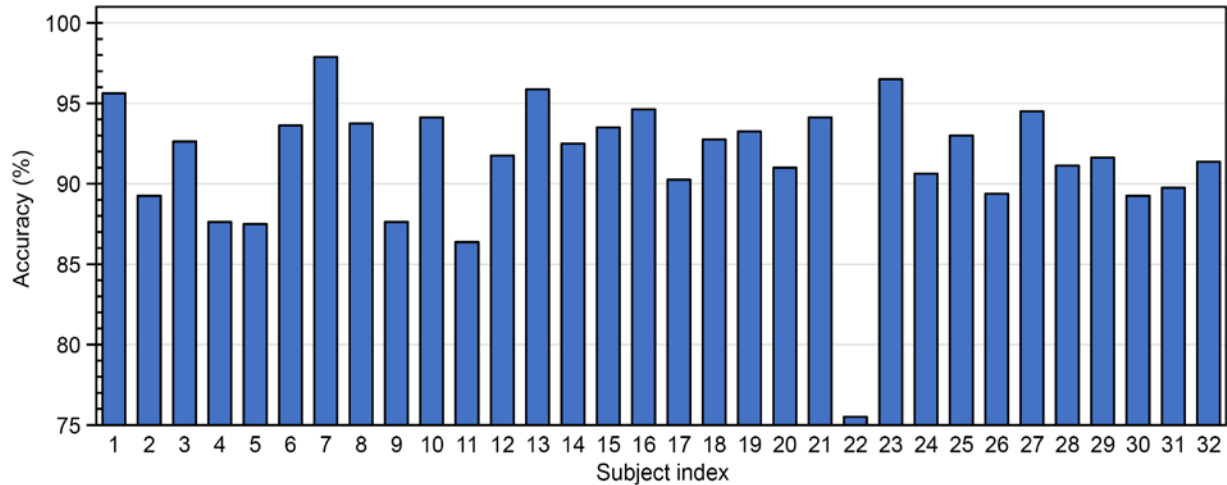


Structure of DO-Conv model

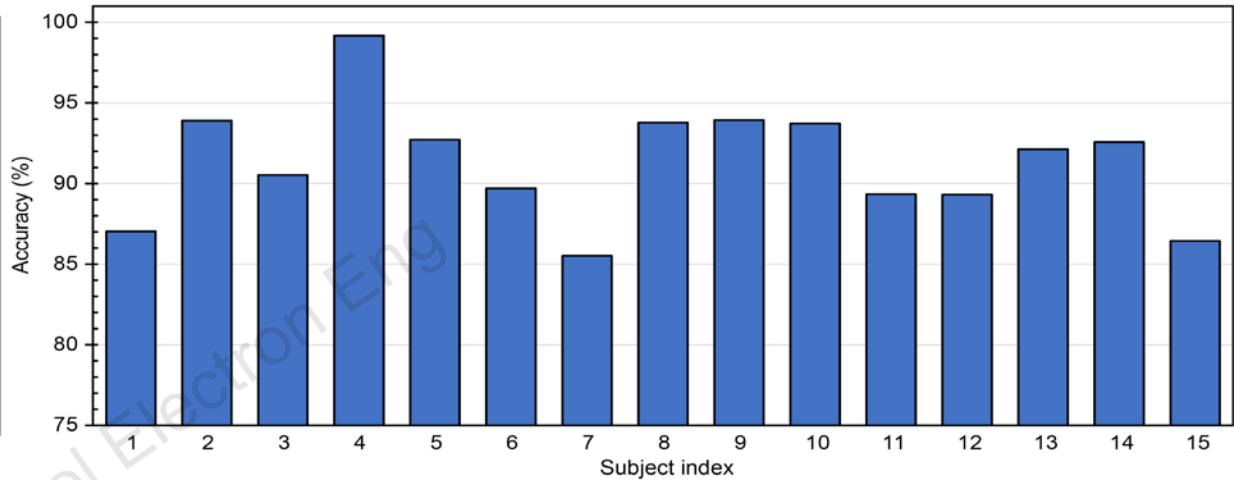


Structure of the transformer model

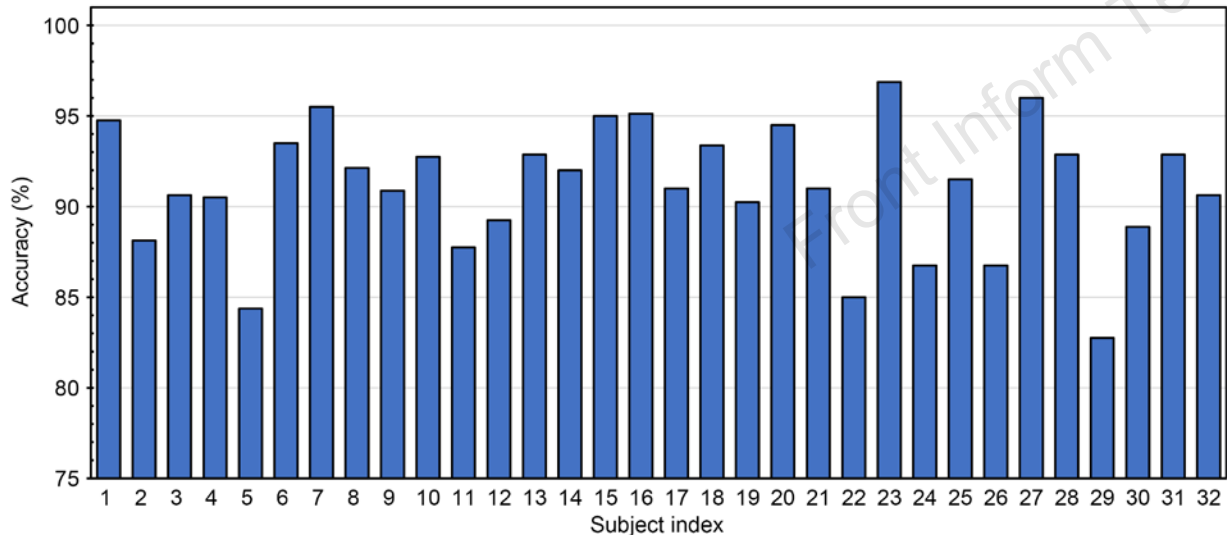
Major results



Accuracy of arousal on the DEAP dataset



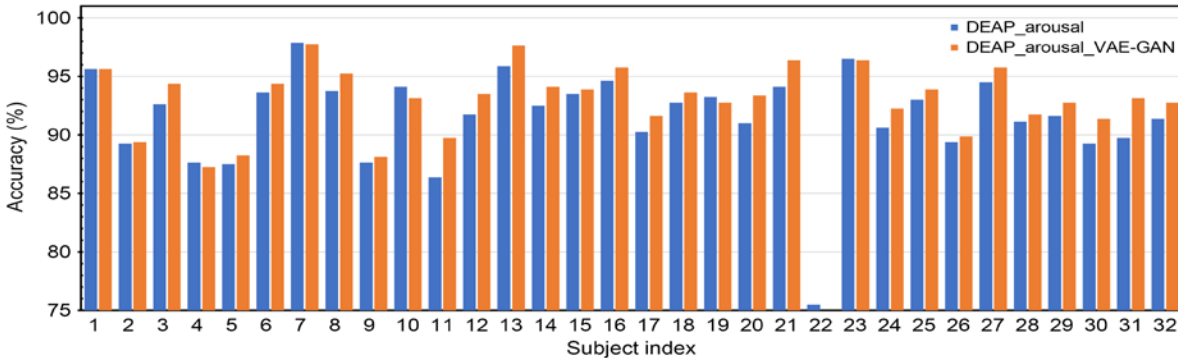
Accuracy on SEED



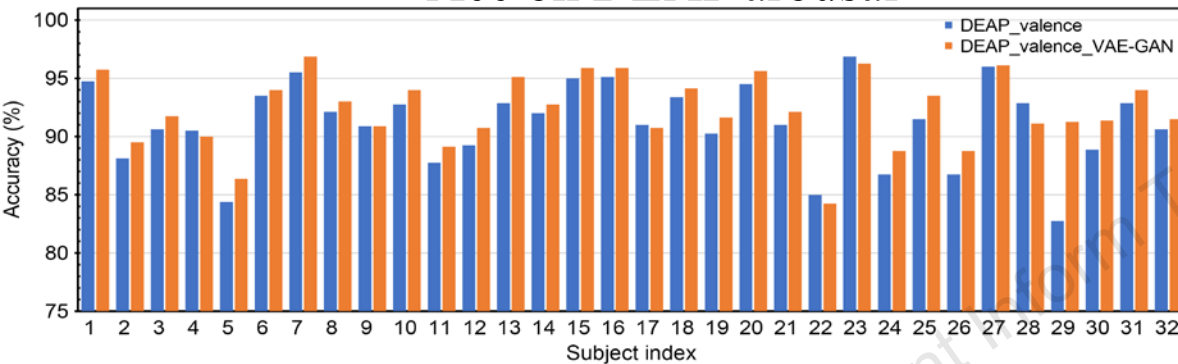
Accuracy of valence on the DEAP dataset

- Our model performs binary classification tasks on the DEAP dataset with an average accuracy of **91.51% for arousal** and **91.13% for valence**.
- SEED dataset was divided into neutral emotion, positive emotion and negative emotion, and the prediction **accuracy of 91.32%** was obtained.

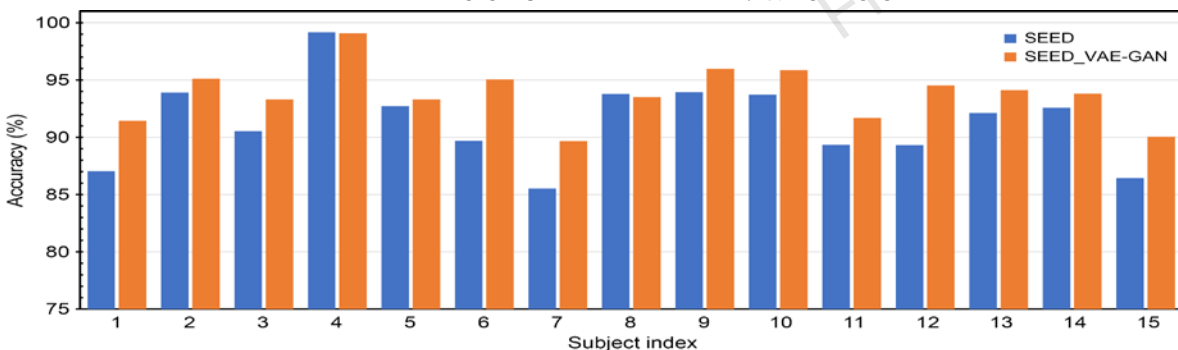
Major results(Cont'd)



Acc on DEAP arousal



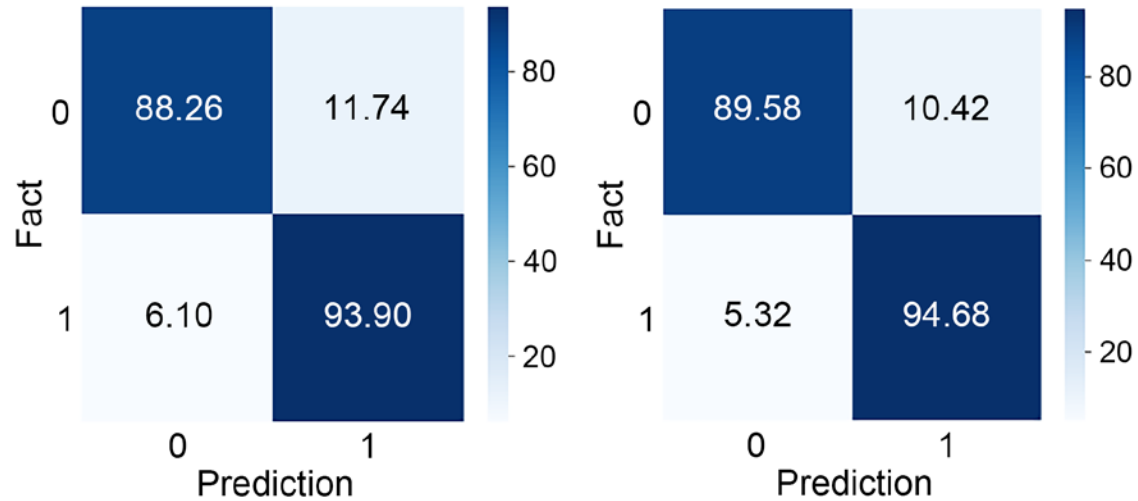
Acc on DEAP valence



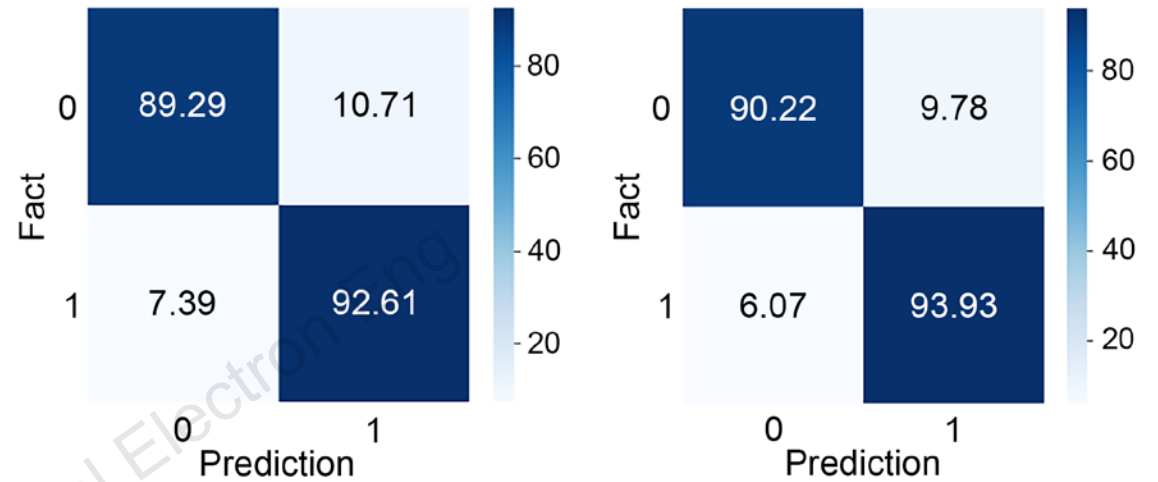
Acc on SEED

- It is evident that the utilization of the VAE-GAN model to expand the training set leads to an average wakefulness prediction accuracy of 92.52% on the DEAP dataset, which is 1.01% higher than that achieved without the VAE-GAN model. However, subject No. 22 has an accuracy rate of 74.5% and is not included in the table above due to its low accuracy. The average valence prediction accuracy in the DEAP dataset reaches 92.27%, which is 1.14% higher than of the previous model.
- On the SEED dataset, the average prediction accuracy reaches 93.77% with the VAE-GAN model, which is 2.45% higher than of the previous model. These results demonstrate the suitability of the VAE-GAN model for expanding EEG datasets.

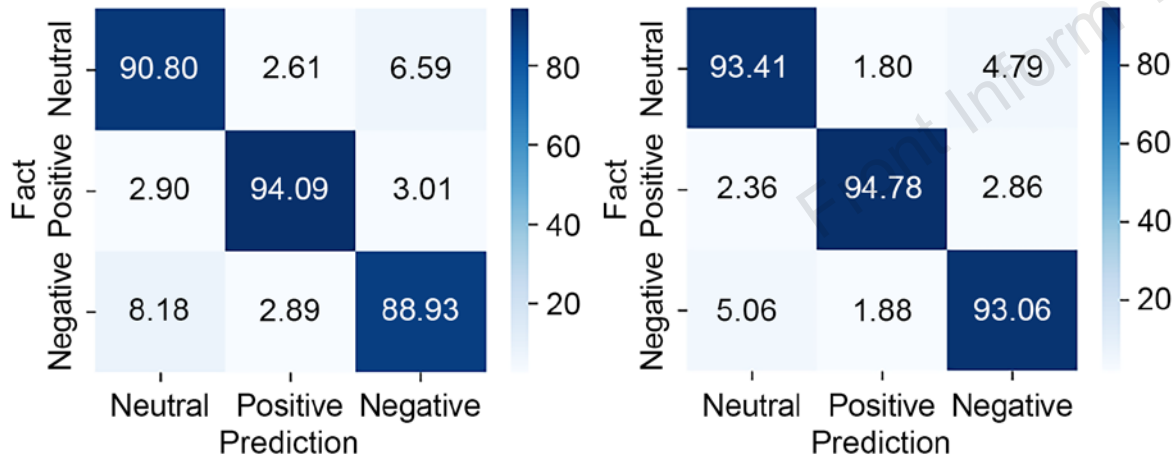
Major results(Cont'd)



Confusion Matrix on DEAP arousal



Confusion Matrix on DEAP valence



Confusion Matrix on SEED

- In the DEAP dataset, the prediction accuracy of the **low valence and high valence are 89.29% and 92.61%**, respectively, and the prediction accuracy of the high valence is higher as well.
- In the SEED dataset, the prediction accuracy of neutral, **positive, and negative emotions are 90.8%, 94.09%, and 88.93%**, respectively. Positive emotions are the most easily detected, while negative emotions and neutral emotions were more easily confused.
- After the VAE-GAN structure is added, the accuracy of emotion classification in all datasets is improved to some extent.

Conclusions

1 Use VAE-GAN to expand the training set part of the dataset.

The application of data augmentation techniques effectively addresses the common challenge of limited EEG data, enhancing the model's performance and reliability.

2 Propose a novel DOCoT structure tailored for EEG-based emotion recognition.

The novel DOConv structure is designed to extract both distinctive and interrelated EEG channel information. Meanwhile, the transformer architecture excels at capturing global dependencies within EEG signals, providing a holistic analysis of emotional states.

3 Offer a scalable, efficient, and highly accurate method with high robustness

On DEAP dataset our method achieved an average accuracy of 92.27% for valence classification and 92.52% for arousal classification. For SEED's three classification tasks, we attained an impressive average accuracy of 93.77%.



Yanping Zhu received the Ph.D. degree in information and communication engineering from Nanjing University of Science and Technology, China, in 2014. She also worked as an academic visitor in University of Sheffield, UK, in 2018. She is currently working as an associate professor at Nanjing University of Information Science and Technology, Nanjing, China. Her research interests include emotion recognition algorithm based on EEG brainwave signals, UWB localization and close-range microwave imaging.