CROSS-ATTENTION-GUIDED WAVENET FOR MEL SPECTROGRAM RECONSTRUCTION

IN THE ICASSP 2024 AUDITORY EEG CHALLENGE

Yuan Fang , Hao Li, Xueliang Zhang,
Fei Chen,Guanglai Gao

Inner Mongolia University, China
Southern University of Science and Technology, China

Reporter: Yuan Fang

32209021@mail.imu.edu.cn

Outline

- BACKGROUND
- PROPOSED MODEL
- EXPERIMENT
- CONCLUSIONS

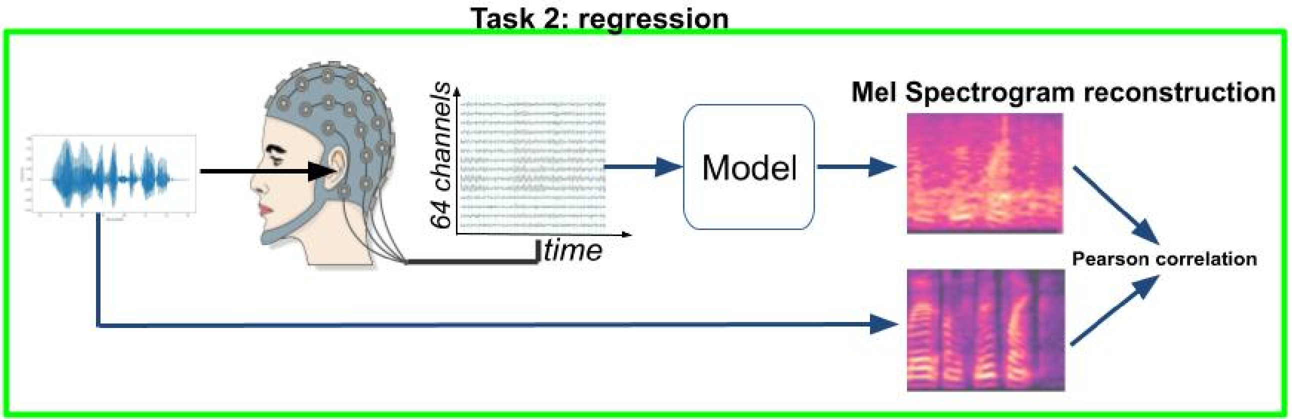

Fig.1 Task 2 of the Auditory EEG Challenge: EEG-to-MEL Spectrogram Reconstruction.

① The ICASSP 2024 Auditory EEG Challenge Task 2 is a regression task.

② Predicting the mel spectrogram based on the input EEG signal.

③ The model is evaluated using Pearson correlation.

1) Inter-individual differences.

2) Low signal-to-noise ratio.

3) EEG to speech is a challenging problem due to its nonlinear nature

Outline

- BACKGROUND
- **PROPOSSED MODEL**
- EXPERIMENTS
- CONCLUSIONS

PROPOSED MODEL

1. Cross-Attention-Guided WaveNet for Mel spectrogram reconstruction.
2. The coarse-to-fine granularity strategy.
3. Cross-attention mechanism is used to fuse two different modalities.
4. A combined loss function is used to optimize multiple outputs.
5. The Mixup augmentation technique to mitigate overfitting and improve generalization performance.

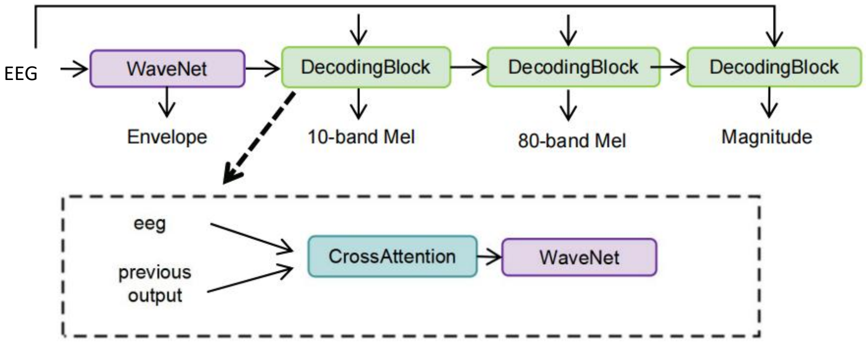

Fig.2 Proposed model.

① In the field of deep learning, multi-objective learning has become a common strategy.

② The coarse-to-fine granularity approach is used to estimate multiple objectives.

③ The effectiveness of this strategy was validated through experimental ablation studies.

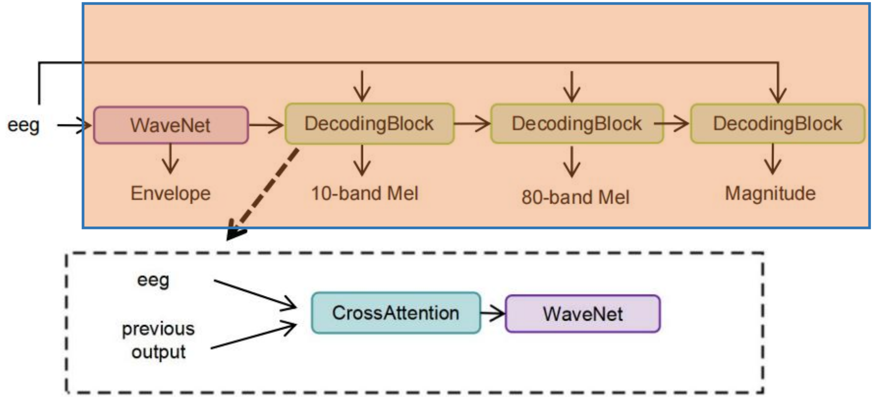

Fig.5 Our coarse-to-fine strategy

PROPOSED MODEL - WaveNet

① WaveNet effectively learns features from sequential data by utilizing dilated convolutions.

② WaveNet showed significant performance in the ICASSP 2023 Auditory EEG Challenge.

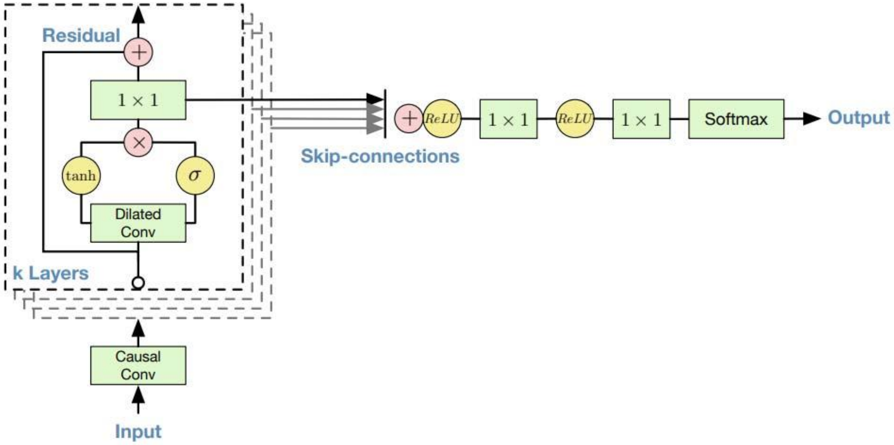

Fig.4 WaveNet Architecture

① Cross-Attention mechanism is a multi-head attention mechanism commonly used in deep learning-based methods as a modality fusion module.

② Cross-Attention mechanism captures dependencies between different scales of features and modalities, facilitating effective information exchange and fusion.

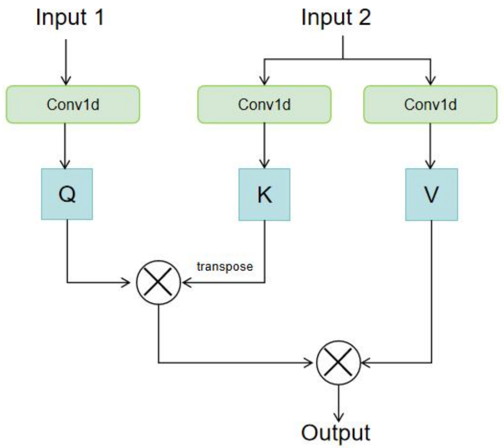

Fig.5 Cross-Attention mechanism

① multiple loss functions jointly to ensure stable training of the model

② L1 norm

③ Negative Pearson correlation coefficient (NP)

④ Kullback-Leibler Divergence (KL divergence)

$$Loss = \alpha * L_1 + NP + KL$$

$$L_1 = L_1(Env) + L_1(Mel10) + L_1(Mel80) + L_1(Mag)$$

$$NP = NP(Env) + NP(Mel10) + NP(Mel80) + NP(Mag)$$

$$KL = KL(Mel10)$$

Considering the constraints of a limited dataset, the Mixup data augmentation technique was adopted to alleviate overfitting and improve performance:

$$x = \lambda x_i + (1-\lambda) x_j$$

$$y = \lambda y_i + (1-\lambda) y_j$$

In the Mixup data augmentation technique, x_i and x_j represent two segments of EEG from different participants, while y_i and y_j represent the corresponding audio signals. The parameter λ is randomly sampled from the range [0,1].

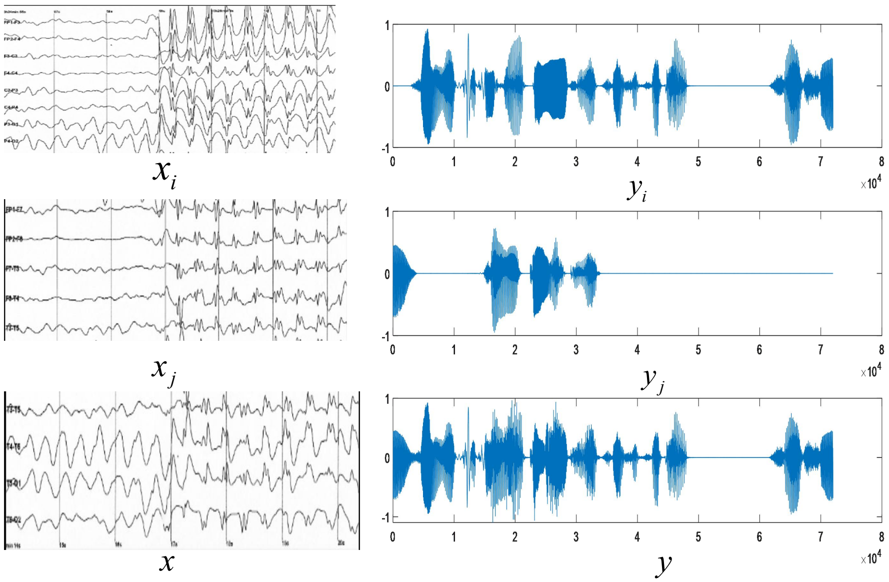

Fig.6 Mixup data augmentation.

Outline

- BACKGROUND
- PROPOSED MODEL
- **EXPERIMENTS**
- CONCLUSIONS

EXPERIMENTS – Dataset

- **Auditory EEG corpus:**
  - Auditory EEG challenge
  - **Train set:**
    - Sub-01 to Sub-26
    - Sub-43 to Sub-85
  - **Val set:**
    - Sub-27 to Sub-42
  - **Test set:**
    - Sub-86 to Sub-104

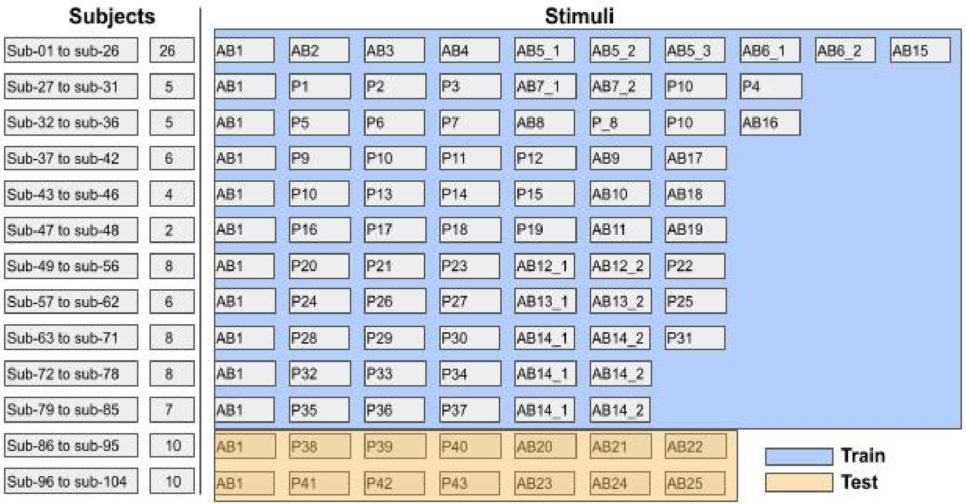

Fig.7 Dataset

EXPERIMENTS – Results

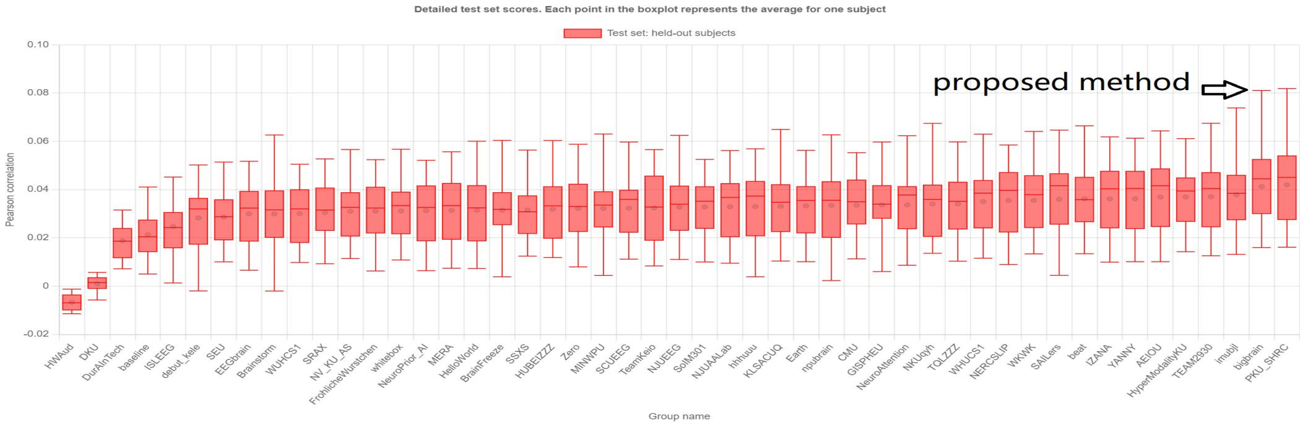

Fig.8 Task 2 of the Auditory EEG Challenge Results of Different Teams

① The proposed model achieved a PCC score of 0.0651, outperforming other baseline models.

② The proposed model ranked second out of 48 teams in the Auditory EEG Challenge 2024 Task 2.

Model	PCC
VLAAI	0.0470
DPRNN	0.0554
Proposed	0.0651

Table 1 Comparative Analysis of Models on validation set

EXPERIMENTS – ablation experiment

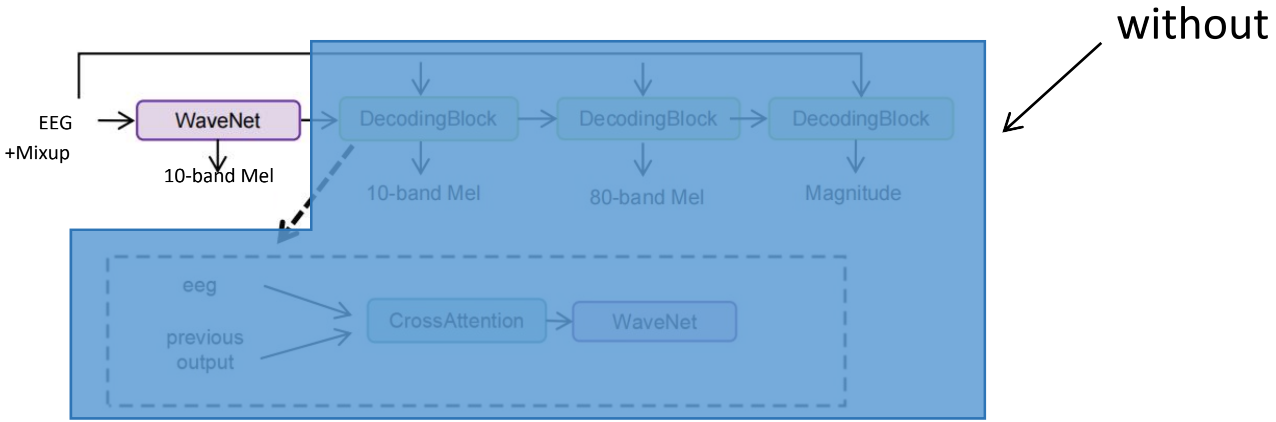

Fig.9 Ablation-1

This ablation method solely utilizes the WaveNet module to reconstruct the Mel spectrogram.

EXPERIMENTS – Ablation experiment

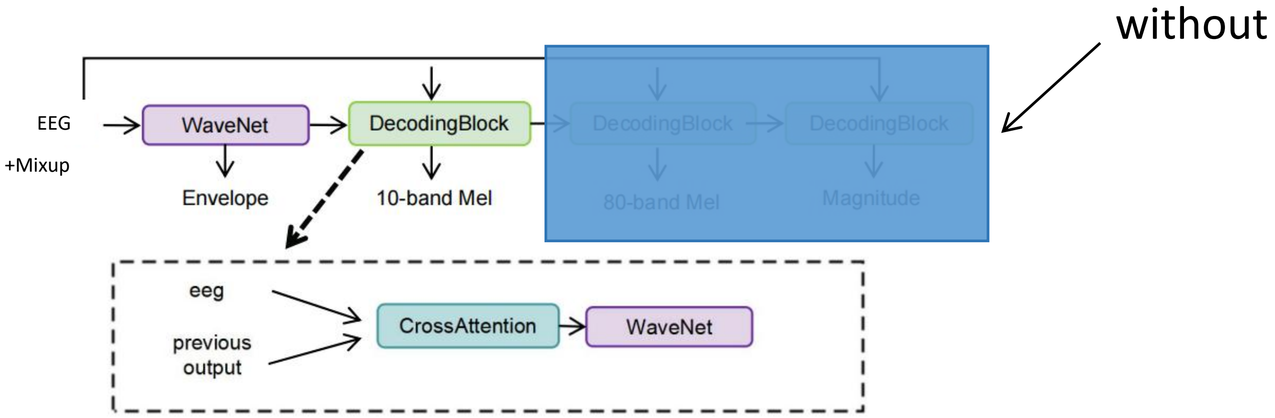

Fig.10 Ablation-2

This ablation method involves removing the last two decoding blocks. The purpose is to examine the influence of the coarse-to-fine granularity strategy.

EXPERIMENTS – Ablation experiment

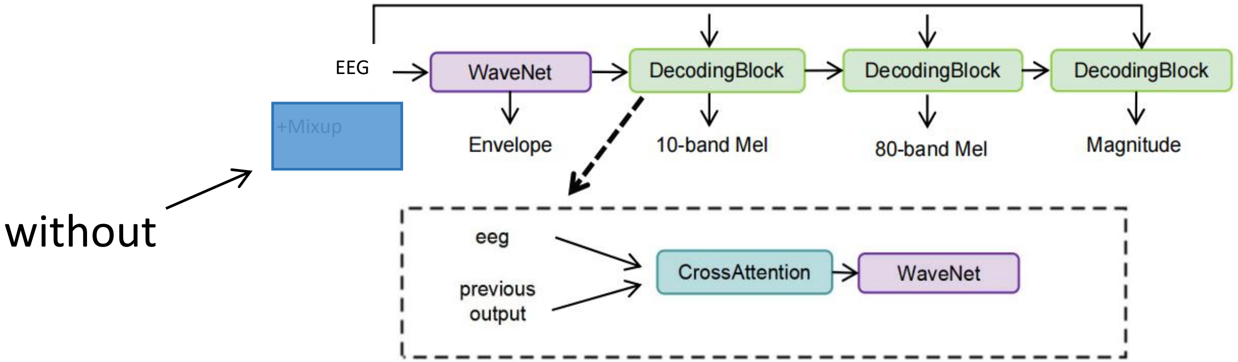

Fig.11 Ablation-3

This ablation method omits the mixed data augmentation technique. The purpose is to evaluate the impact of data augmentation operations on the model's performance.

① Each module of the model has made a significant contribution to the overall performance.

② The coarse-to-fine granularity strategy improved the performance by 0.002.

③ The decoding block and coarse-to-fine granularity strategy led to a 0.0071 improvement.

④ Mixup contributed an improvement effect of 0.0039.

Model	PCC
Ablation-1	0.0580
Ablation-2	0.0631
Ablation-3	0.0612
Proposed	0.0651

Table 2 Ablation experiments results

Outline

- BACKGROUND
- PROPOSED MODEL
- EXPERIMENTS
- **CONCLUSIONS**

Conclusions

- ✓ The proposed CAT-guided WaveNet model leverages CAT to bridge the gap between different modalities and utilizes WaveNet with a coarse-to-fine granularity to construct the Mel spectrogram.
- ✓ Compared to baseline, the proposed method demonstrates stronger performance and improved generalization ability on unseen data.
- ✓ The code has been uploaded to GitHub.

https://github.com/IMU-FangYuan/Multi-Stage-Multi-Target-WaveNet-for-the-ICASSP-2024-Auditory-EEG-Challenge-2024

内蒙古大學

INNER MONGOLIA UNIVERSITY

THANK YOU

Speech Signal Processing Group, Inner Mongolia University