Fingerprinting in EEG Model IP Protection Using Diffusion Model

Tianyi Wang, Sheng-hua Zhong*
College of Computer Science and Software Engineering, Shenzhen University
wangtianyi2023@email.szu.edu.cn, csshzhong@szu.edu.cn

Motivation

- **Overlooked IP Protection for Sensitive EEG Models**: EEG models, handling highly sensitive and private physiological information, have not received adequate IP protection.
- **Limitations of Watermarking for EEG Models**: Current EEG model protection is limited, with a focus on watermarking that may impair performance.
- **Potential of Diffusion Models for EEG Fingerprinting**: Diffusion model shows promising but underexplored potential for EEG signal generation, addressing data availability and privacy issues.

Conclusion

- **Contribution**:
  - We are the first to propose a fingerprinting method for protecting EEG models.
  - We are the first to apply diffusion models in model protection tasks.
  - The proposed method outperforms existing model protection techniques.

- **Future work**:
  - Further optimize the diffusion model
  - Integrate diffusion model with other sophisticated protection techniques
  - Extend the proposed method to other domains (e.g., CV)

Method

**Stage 1. Fingerprint validation set construction**

- A conditional denoising diffusion probabilistic model (CDDPM) is trained to synthesize high-quality and high-diversity simulated EEG samples.
- Retrieve and combine high-confidence, boundary, and atypical samples to form fingerprint verification set.

**Stage 2. Model fingerprint matching**

- Evaluating the similarity between the outputs of the suspected model and the source model on the fingerprint verification set.

Experimental Results

**Dataset: DEAP**

**Model Protection**

- The AUC results for each model as a source model and the others as irrelevant models.

**Samples generated**

- The AUC of five IP protection methods using EEGConformer as the source model when facing IP attacks

**Experimental Results**

- Low valence real EEG signal
- High valence real EEG signal
- High-confidence real EEG signal
- Boundary sample
- Atypical Sample

**Average of AUC results**

- EEGConformer: 0.93
- EEGNet: 0.81
- CCNN: 0.71
- Sception: 0.96

- The AUC of five IP protection methods using EEGConformer as the source model when facing IP attacks

**PCA Feature 1 vs PCA Feature 2**

- Low-valence real EEG signal
- High-valence real EEG signal
- High-confidence real EEG signal
- Boundary sample
- Atypical Sample