



서울시립대학교
UNIVERSITY OF SEOUL

EEGPT: Pretrained Transformer for Universal and Reliable Representation of EEG Signals

2025.05.15
박현우

Abstract

Problem: 낮은 SNR (signal-to-noise ratio), 피실험자 간 높은 variance로 인해 좋은 EEG representation을 추출하는 것을 어렵게 함.

Contribution

- Mask-based dual self-supervised learning.
- Spatio-temporal representation alignment
- EEGPT's hierarchical structure separate spatial and temporal info.
- Large mixed multi-task EEG dataset을 이용한 10-million parameter model training.

EEG

EEG (뇌파): 대뇌 피질에서 발생하는 전기 신호를 기록함으로써 뇌의 상태를 동적으로 표시하는 파동

신호 대 잡음비 (SNR / signal-to-noise ratio): 노이즈 전력 대비 신호 전력 세기 비율

$$SNR = \frac{P_{signal}}{P_{noise}}$$

낮은 SNR을 가지는 EEG 데이터의 경우 noise가 심하며 사람들마다 뇌파가 완전히 동일하지 않음에서 발생하는 높은 variance는 실생활로의 접목을 어렵게 만든다.

→ 이러한 문제를 해결하기 위해서 우리는 EEG 신호의 spatio-temporal consistency를 기반으로 한 dual self-supervised learning Method를 제안한다.

Background

Masked autoencoder: input signal에 대해 무작위 patch mask를 적용한 후 encode에 입력하고, decoder는 Masked patch들의 embeddings를 예측한다.

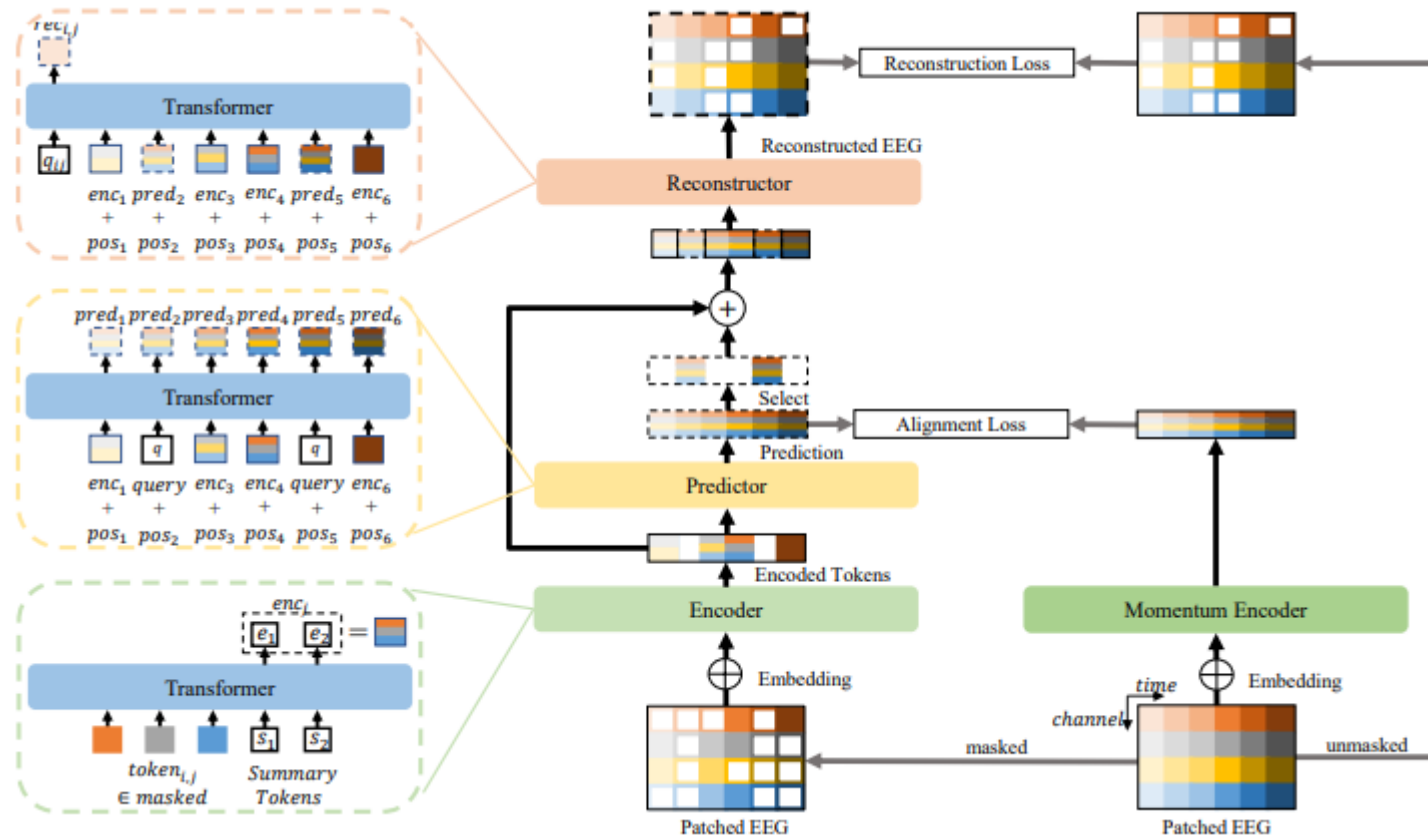
$$\min_{\theta, \phi} \mathbb{E}_{x \sim D} H(d_{\phi}(z), x \odot (1 - M)), \quad z = f_{\theta}(x \odot M)$$

반면 저자는 spatio-temporal representation alignment 항을 추가하여, 아래와 같이 표현한다.

$$\min_{\theta, \phi} \mathbb{E}_{x \sim D} H(d_{\phi}(z), x \odot (1 - M)) + H(z, f_{\theta}(x)), \quad z = f_{\theta}(x \odot M)$$

이 방법은 인코딩된 representation이 더 넓은 semantic (의미론적인) information을 포괄하도록 유도하며, 이는 Minimal sufficient representation과 유사하다. 이를 통해 품질과 일반화 성능이 향상된다.

Background



Background

앞의 그림은 모델 구조를 정의하며, Patching / Embedding Masking / Encoder / Predictor / reconstructor / 의 연산을 포함한다.

1. EEG 입력 신호 $x \in \mathbb{R}^{M \times T}$ (M: channel / T: times)를 patches로 나눈 후, 각 patches를 local spatio-temporal embeddings를 통해 tokens로 임베딩함.
2. 이후 tokens를 masking된 부분과 masking되지 않은 부분으로 분할한다. 이후 dual self-supervised learning을 이용해 model을 pretrain하는데, 여기에는 spatio-temporal representation alignment와 mask-based reconstruction을 포함한다.
3. Linear-probing을 통해 downstream task를 학습한다.

Spatio-Temporal Representation Alignment

예측된 features를 momentum encode의 output과 align함으로써, 인코더가 강력한 feature를 추출하는 능력을 향상시키고, encode의 output이 고품질의 global features를 포함하도록 한다.

Encoder: masking된 patches로부터 spatial info를 통합한다. $enc_j = ENC(\{token_{i,j}\}_{(i,j) \in M})$

Predictor: 인코더로부터 얻은 masking된 부분의 feature와 temporal position embeddings를 예측한다.

$$\{pred_t\}_{t=1,2,\dots,N} = PRED(\{enc_j + pos_j\}_{\exists i, (i,j) \in M})$$

Momentum encoder: 구조는 encoder와 동일 $menc_j = MENC(\{token_{i,j}\}_{(i,j) \in M \cup \bar{M}})$

Alignment loss: $\mathcal{L}_A = -\frac{1}{N} \sum_{j=1}^N \|pred_j - LN(menc_j)\|_2^2$ LN: layer norm

Mask-based Reconstruction

Reconstructor는 encoder가 encoding한 masking된 부분 M 의 feature와 입력 정보를 이용해 복원된 patch를 생성

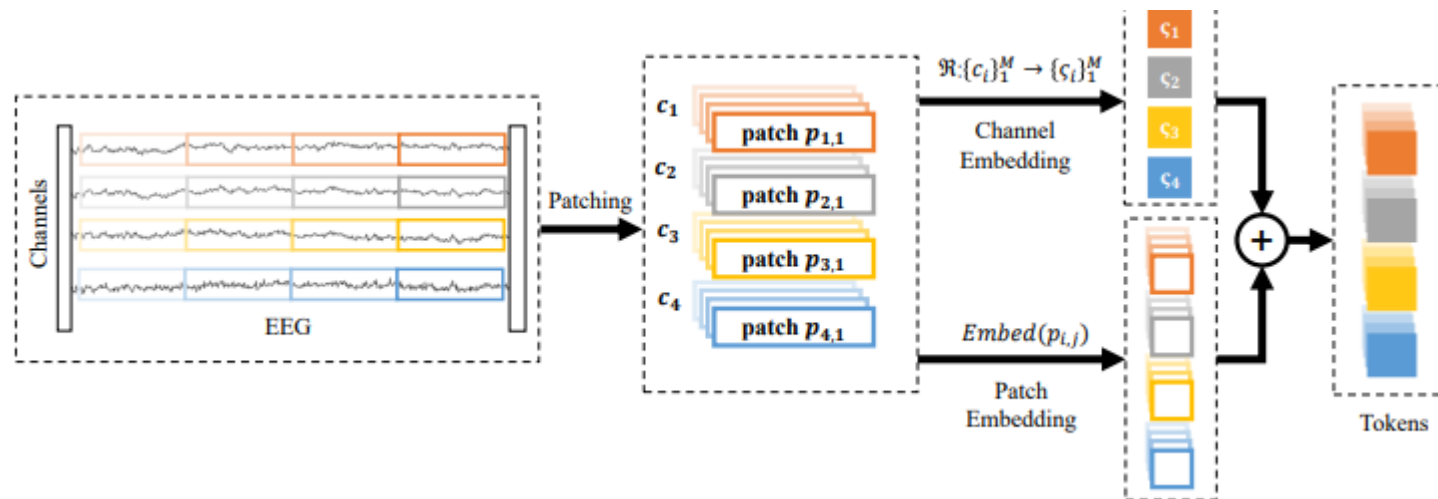
$$\{rec_{u,t}\}_{(u,t) \in M} = REC \left(\{enc_j + pos_j\}_{\exists i, (i,j) \in M} \cup \{pred_j + pos_j\}_{\forall i, (i,j) \in M} \right)$$

$$\mathcal{L}_R = -\frac{1}{|M|} \sum_{(i,j) \in M} \|rec_{i,j} - LN(p_{i,j})\|_2^2$$

$$L = \mathcal{L}_A + \mathcal{L}_R$$

Local Spatio-temporal embeddings

EEG 신호는 시공간 차원에서 동일한 크기의 patches로 나뉘며 각 패치는 특정 채널의 시간 구간을 나타낸다.



$$p_{i,j} = x_{i,(j-1)d:j d}$$
$$token_{i,j} = Embed(p_{i,j}) + \varsigma_i$$

Linear-Probing Method

Downstream task에서는 pretrained encoder에 추가 모듈을 연결하여 분류 과제를 해결한다.

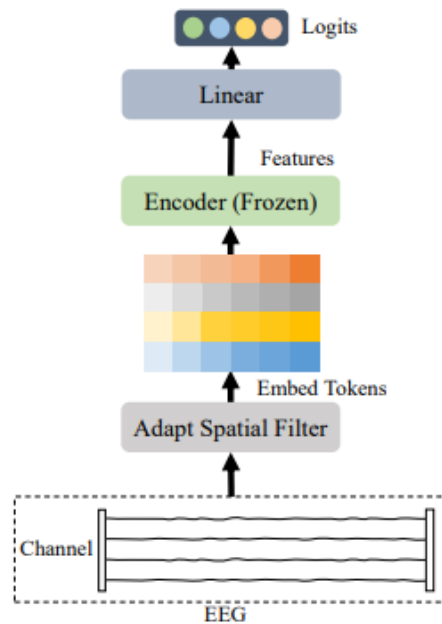


Figure 3: Linear-probing method.

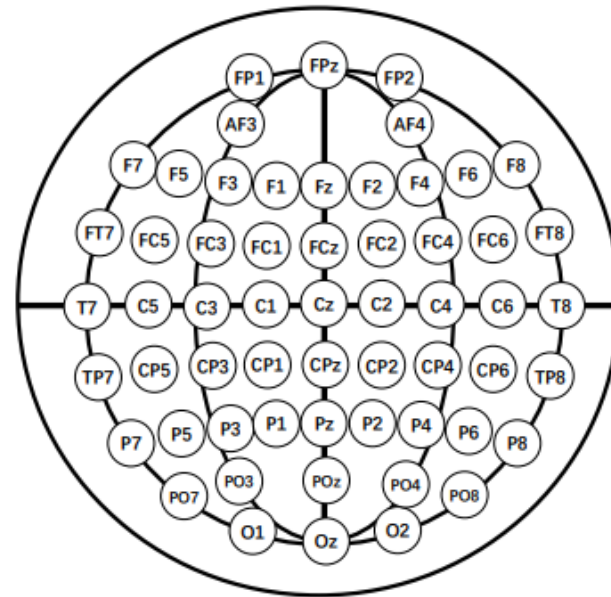


Figure 4: Electrode locations.

Dataset

Table 1: Datasets for pretraining and downstream tasks

	Datasets	Paradigms	Subjects	Targets
pretraining Datasets	PhysioMI	MI&ME	109	5
	HGD	MI	14	4
	TSU	SSVEP	35	40
	SEED	EMO	15	3
	M3CV	MULTI	106	-
Downstream Datasets	BCIC-2A	MI	10	4
	BCIC-2B	MI	10	2
	Sleep-EDFx	SLEEP	197	5
	KaggleERN	ERN	26	2
	PhysioP300	P300	9	2
	TUAB	Abnormal	2383	2
	TUEV	Event	288	6

Downstream result

Table 2: The results of different methods on TUAB.

Methods	Model Size	Balanced Accuracy	AUROC
SPaRCNet [44]	0.79M	0.7896 \pm 0.0018	0.8676 \pm 0.0012
ContraWR [45]	1.6M	0.7746 \pm 0.0041	0.8456 \pm 0.0074
CNN-T [46]	3.2M	0.7777 \pm 0.0022	0.8461 \pm 0.0013
FFCL [47]	2.4M	0.7848 \pm 0.0038	0.8569 \pm 0.0051
ST-T [48]	3.5M	0.7966 \pm 0.0023	0.8707 \pm 0.0019
BIOT [15]	3.2M	0.7959 \pm 0.0057	0.8815 \pm 0.0043
Ours-Tiny	4.7M	0.7959 \pm 0.0021	0.8716 \pm 0.0041
Ours	25M	0.7983 \pm 0.0030	0.8718 \pm 0.0050

Downstream result

Table 3: The results of different methods on TUEV.

Methods	Model Size	Balanced Accuracy	Weighted F1	Cohen's Kappa
SPaRCNet [44]	0.79M	0.4161 \pm 0.0262	0.7024 \pm 0.0104	0.4233 \pm 0.0181
ContraWR [45]	1.6M	0.4384 \pm 0.0349	0.6893 \pm 0.0136	0.3912 \pm 0.0237
CNN-T [46]	3.2M	0.4087 \pm 0.0161	0.6854 \pm 0.0293	0.3815 \pm 0.0134
FFCL [47]	2.4M	0.3979 \pm 0.0104	0.6783 \pm 0.0120	0.3732 \pm 0.0188
ST-T [48]	3.5M	0.3984 \pm 0.0228	0.6823 \pm 0.0190	0.3765 \pm 0.0306
BIOT [15]	3.2M	0.5281 \pm 0.0225	0.7492 \pm 0.0082	0.5273 \pm 0.0249
Ours-Tiny	4.7M	0.5670 \pm 0.0066	0.7535 \pm 0.0097	0.5085 \pm 0.0173
Ours	25M	0.6232 \pm 0.0114	0.8187 \pm 0.0063	0.6351 \pm 0.0134

Downstream result

Table 4: The results of universal EEG models on various datasets.

Datasets	Methods	Balanced Accuracy	Cohen's Kappa	Weighted F1 / AUROC
BCIC-2A	BENDR	0.4899±0.0070	0.3199±0.0094	0.4836±0.0076
	BIOT	0.4590±0.0196	0.2787±0.0261	0.4282±0.0289
	LaBraM	0.5613±0.0052	0.4151±0.0069	0.5520±0.0052
	Ours	0.5846±0.0070	0.4462±0.0094	0.5715±0.0051
BCIC-2B	BENDR	0.7067±0.0011	0.4131±0.0022	0.7854±0.0029
	BIOT	0.6409±0.0118	0.2817±0.0236	0.7095±0.0141
	LaBraM	0.6851±0.0063	0.3703±0.0125	0.7576±0.0067
	Ours	0.7212±0.0019	0.4426±0.0037	0.8059±0.0032
Sleep-EDFx	BENDR	0.6655±0.0043	0.6659±0.0043	0.7507±0.0029
	BIOT	0.6622±0.0013	0.6461±0.0017	0.7415±0.0010
	LaBraM	0.6771±0.0022	0.6710±0.0006	0.7592±0.0005
	Ours	0.6917±0.0069	0.6857±0.0019	0.7654±0.0023
KaggleERN	BENDR	0.5672±0.0020	0.1461±0.0037	0.6030±0.0044
	BIOT	0.5118±0.0089	0.0297±0.0224	0.5495±0.0167
	LaBraM	0.5439±0.0029	0.0944±0.0066	0.5693±0.0052
	Ours	0.5837±0.0064	0.1882±0.0110	0.6621±0.0096
PhysioP300	BENDR	0.6114±0.0118	0.2227±0.0237	0.6588±0.0163
	BIOT	0.5485±0.0325	0.0968±0.0647	0.5308±0.0333
	LaBraM	0.6477±0.0110	0.2935±0.0227	0.7068±0.0134
	Ours	0.6502±0.0063	0.2999±0.0139	0.7168±0.0051