CBraMod: A Criss-Cross Brain Foundation Model for EEG Decoding

Jiquan Wang[1,2], Sha Zhao ✉ [1,2], Zhiling Luo[3], Yangxuan Zhou[1,2], Haiteng Jiang[4,5,1], Shijian Li[1,2], Tao Li[4,5,1], Gang Pan ✉ [1,2,5]

[1]State Key Laboratory of Brain-machine Intelligence, Zhejiang University

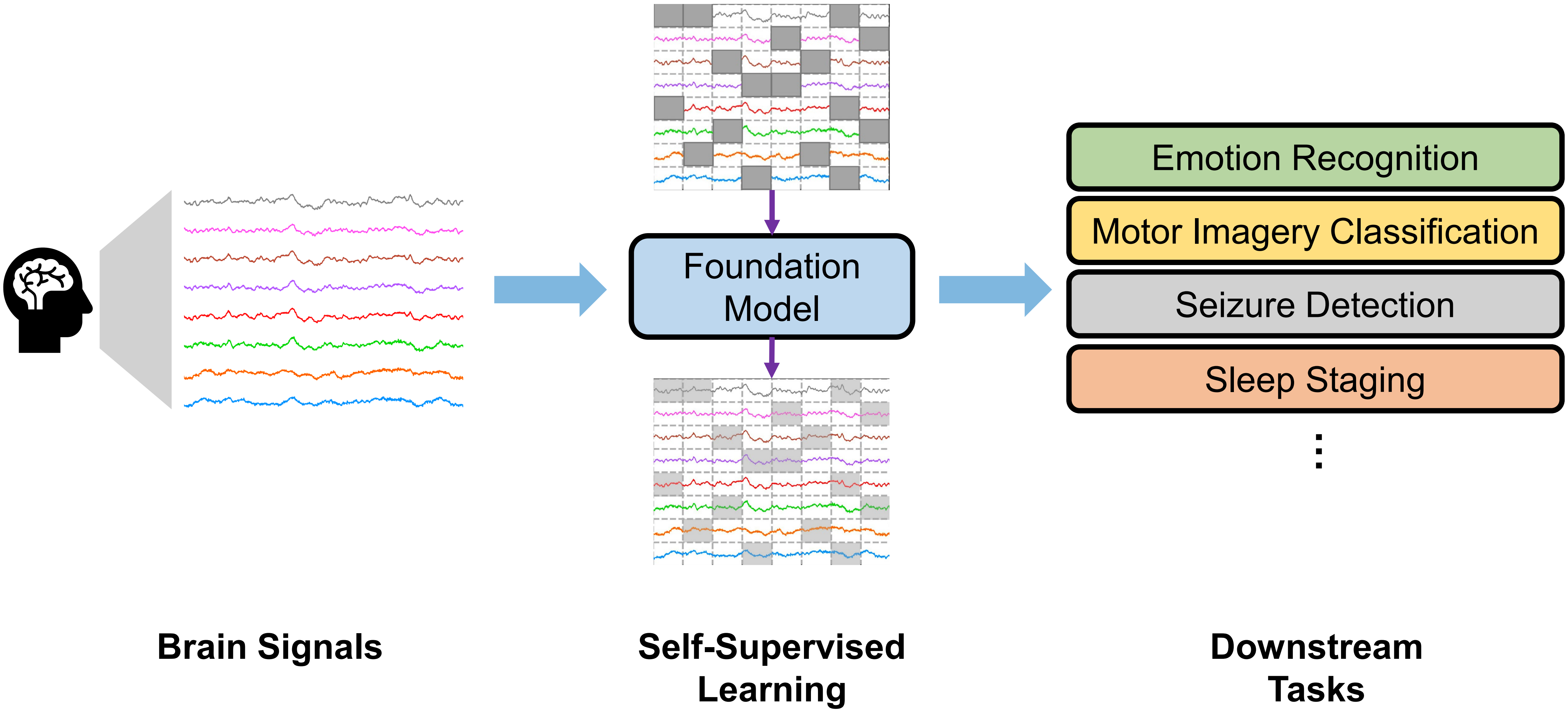
[2]College of Computer Science and Technology, Zhejiang University

[3]Alibaba Group

[4]Department of Neurobiology, Affiliated Mental Health Center & Hangzhou Seventh People's Hospital, Zhejiang University School of Medicine

[5]MOE Frontier Science Center for Brain Science and Brain-machine Integration, Zhejiang University

1 Brain Foundation Model



2 Previous works on Brain Foundation Model

2021
Banville *et al*. [1]
BENDER [2]

2022
MAEEG [3]
BrainBERT [4]

2023
Brant [5]
BIOT [6]

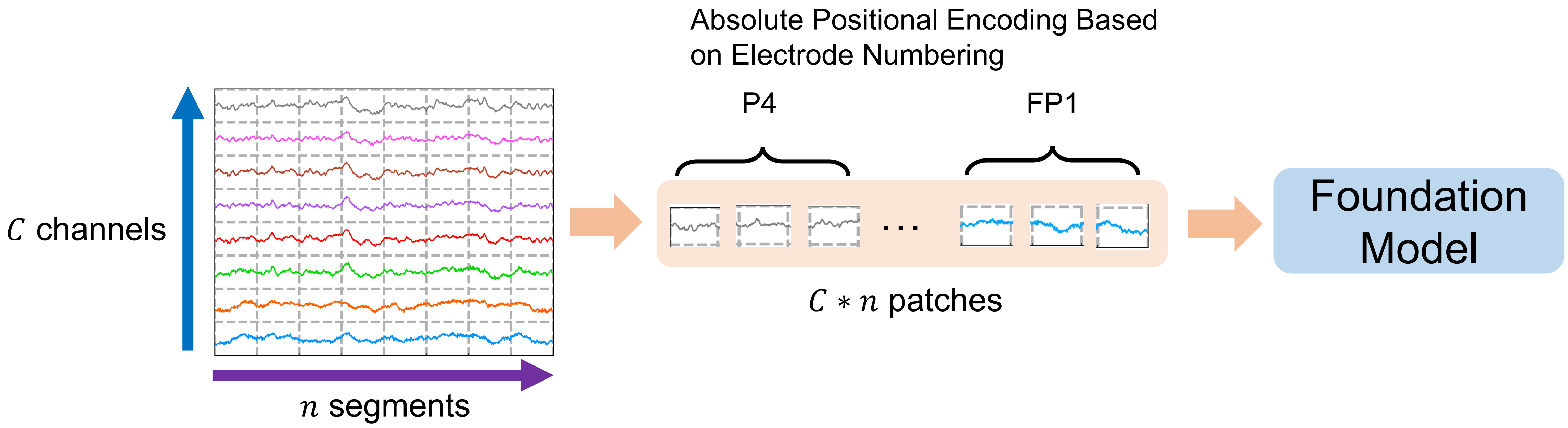
2024
LaBraM [7]
Brant-X [8]
EEG2Rep [9]
EEGPT [10]
NeuroLM [11]

[1] Banville H, Chehab O, Hyvärinen A, et al. Uncovering the structure of clinical EEG signals with self-supervised learning[J]. Journal of Neural Engineering, 2021, 18(4): 046020.
[2] Kostas D, Aroca-Ouellette S, Rudzicz F. BENDR: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of EEG data[J]. Frontiers in Human Neuroscience, 2021, 15: 653659.
[3] Chien H Y S, Goh H, Sandino C M, et al. Maeeg: Masked auto-encoder for eeg representation learning[J]. arXiv preprint arXiv:2211.02625, 2022.
[4] Wang C, Subramaniam V, Yaari A U, et al. BrainBERT: Self-supervised representation learning for intracranial recordings[J]. arXiv preprint arXiv:2302.14367, 2023.
[5] Zhang D, Yuan Z, Yang Y, et al. Brant: Foundation model for intracranial neural signal[J]. Advances in Neural Information Processing Systems, 2024, 36.
[6] Yang C, Westover M, Sun J. Biot: Biosignal transformer for cross-data learning in the wild[J]. Advances in Neural Information Processing Systems, 2024, 36.
[7] Jiang W B, Zhao L M, Lu B L. Large brain model for learning generic representations with tremendous EEG data in BCI[J]. arXiv preprint arXiv:2405.18765, 2024.
[8] Zhang D, Yuan Z, Chen J, et al. Brant-X: A Unified Physiological Signal Alignment Framework[C]//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024: 4155-4166.
[9] Mohammadi Foumani N, Mackellar G, Ghane S, et al. Eeg2rep: enhancing self-supervised EEG representation through informative masked inputs[C]//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024: 5544-5555.
[10] Wang G, Liu W, He Y, et al. EEGPT: Pretrained Transformer for Universal and Reliable Representation of EEG Signals[C]//The Thirty-eighth Annual Conference on Neural Information Processing Systems. 2024.
[11] Jiang W B, Wang Y, Lu B L, et al. NeuroLM: A Universal Multi-task Foundation Model for Bridging the Gap between Language and EEG Signals[J]. arXiv preprint arXiv:2409.00101, 2024.

3 Existing Methods

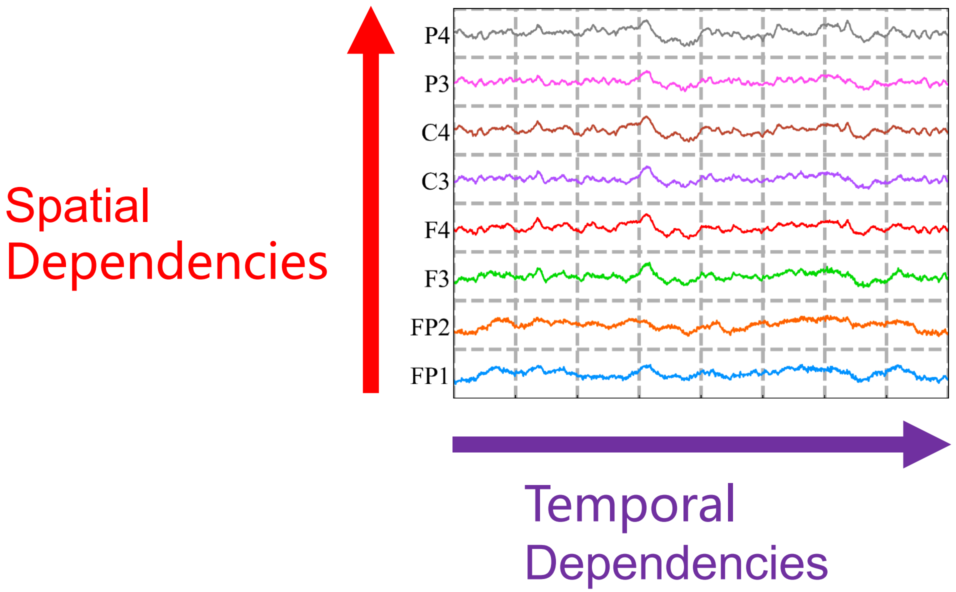
Patching: The EEG signals are divided into patches, which are then flattened into a long sequence. This sequence is subsequently fed into a Transformer-based sequence modeling architecture.

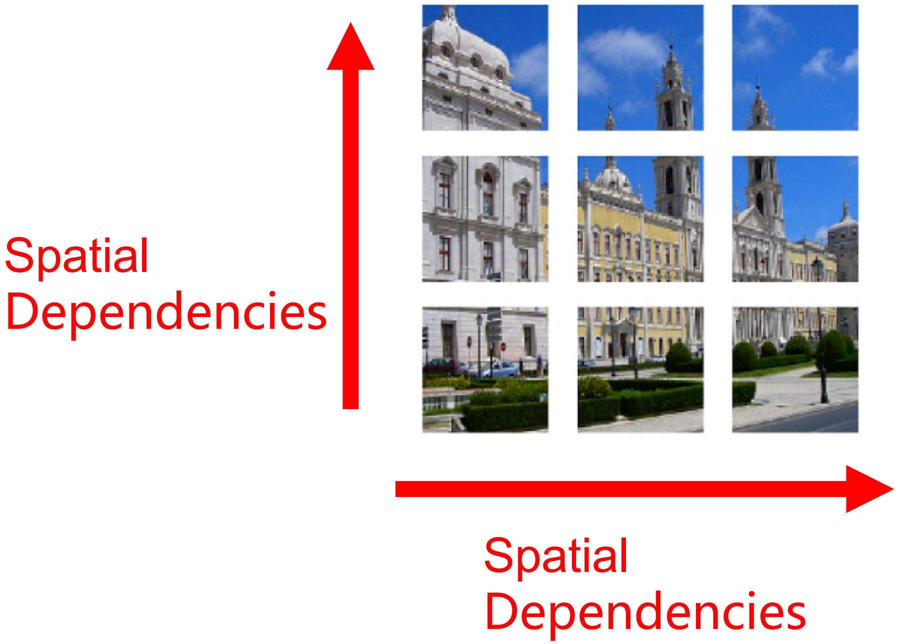
Absolute Positional Encoding: An absolute positional encoding based on electrode numbering is employed as the channel embedding to encode positional information along the channel dimension.



3 Challenge

Unique structural characteristics
EEG signals contain heterogeneous spatial and temporal dependencies, but images contain only spatial dependencies.





The sequence modeling of EEG patches, adapted from computer vision, may overlook the structural properties of EEG signals that jointly incorporate both spatial and temporal dependencies.

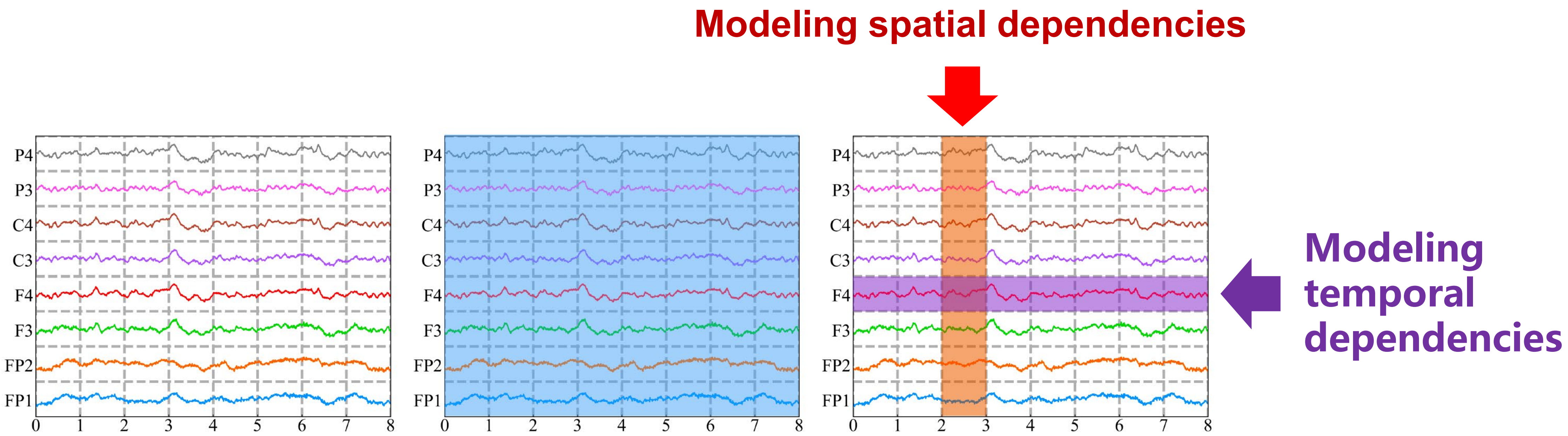
3 Challenge

Channel variation: Existing method mostly adopt absolute positional encoding based on electrode numbering as a channel embedding to encode positional information. However, EEG channels are not solely defined by electrode position but are also influenced by the referencing scheme used (e.g., earlobe, average, REST, or bipolar references).

earlobe		average		REST		bipolar
FP1-LE	≠	**FP1-REF**	≠	**FP1-REST**	≠	**FP1-F7**

4 Our Idea

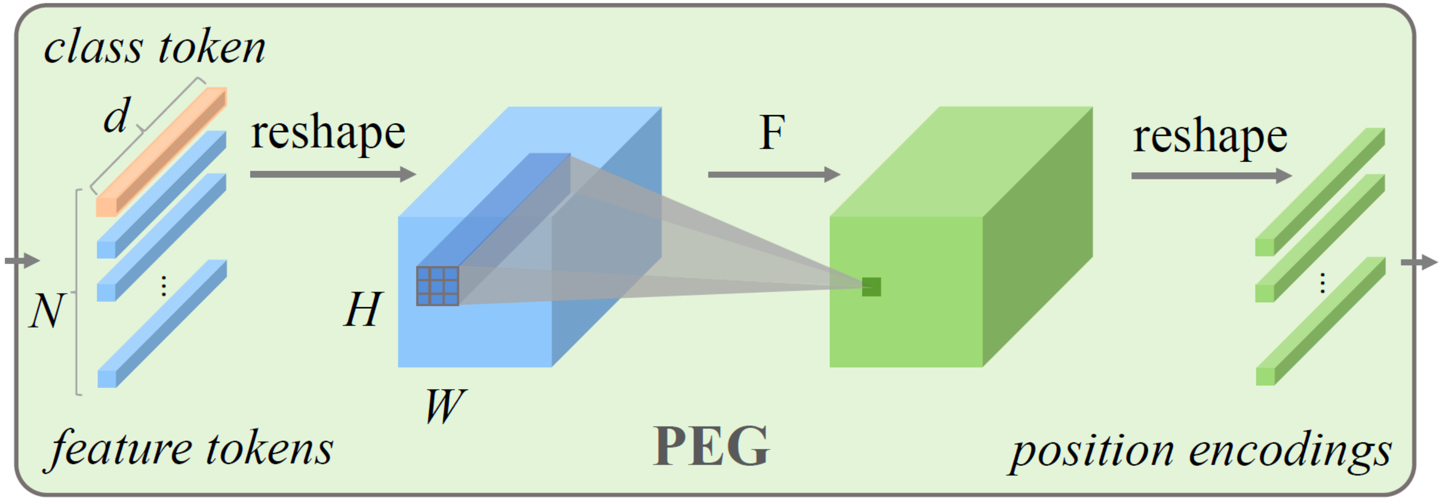
Criss-cross EEG modeling: We propose a criss-cross EEG modeling strategy to thoroughly leverage the structural characteristics of EEG signals, which can model spatial and temporal dependencies in parallel.



(a) EEG Patches (b) Full EEG Modeling (c) Criss-Cross EEG Modeling

4 Our Idea

Conditional positional encoding: Inspired by CPE [1], we propose asymmetric conditional positional encoding (ACPE) as a more flexible approach to positional encoding, which dynamically learn positional relationships among patches.

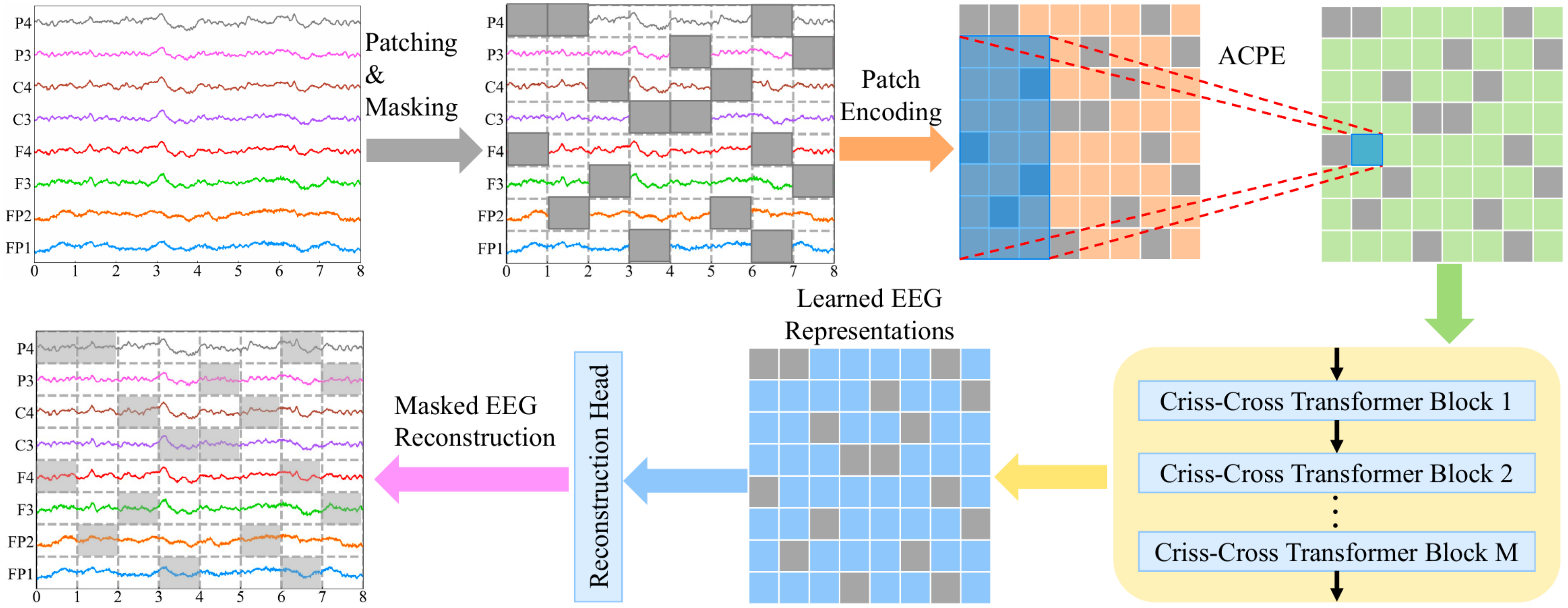


This figure is taken from CPE [1].

[1] Chu X, Tian Z, Zhang B, et al. Conditional Positional Encodings for Vision Transformers[C]//The Eleventh International Conference on Learning Representations.

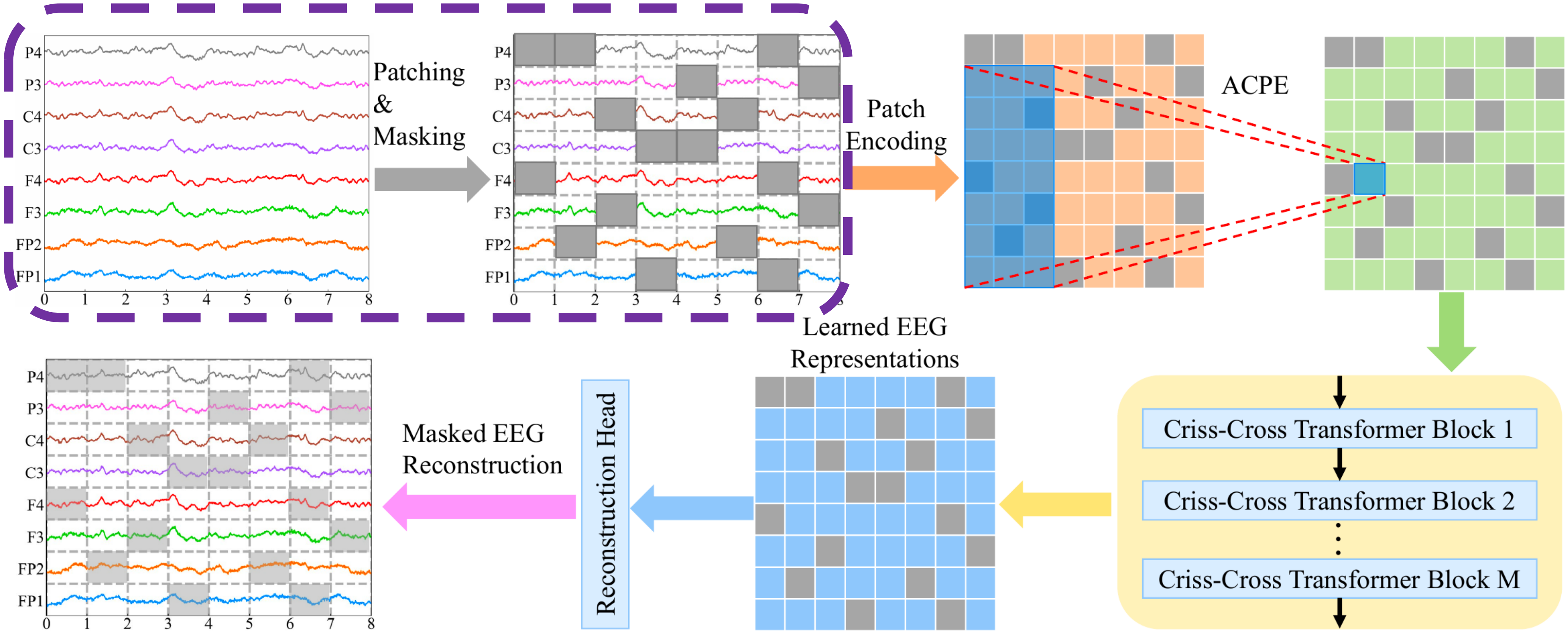
5 Our Method

We propose a novel EEG foundation model, called CBraMod, for EEG decoding on various clinical and BCI application.



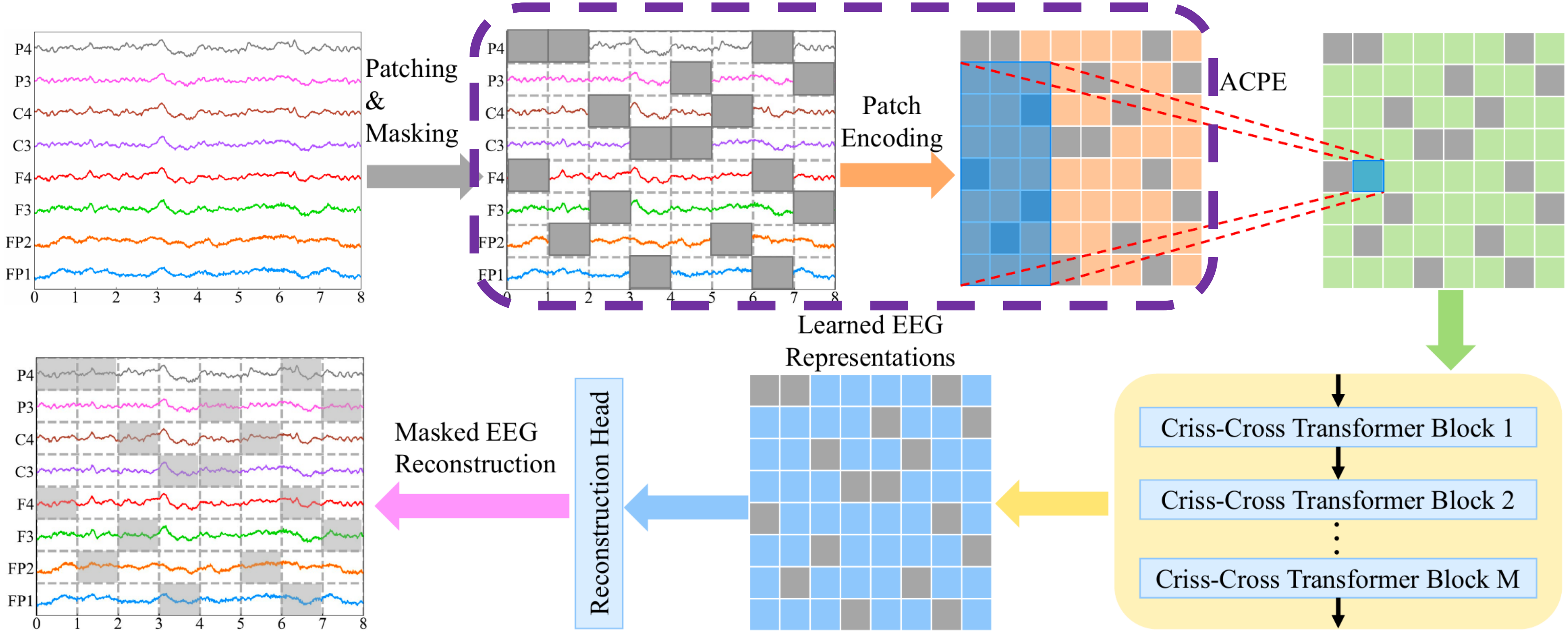
5 Our Method

Patching & Masking: We segment EEG samples into patches using a fixed time window and randomly mask some patches with a mask token.



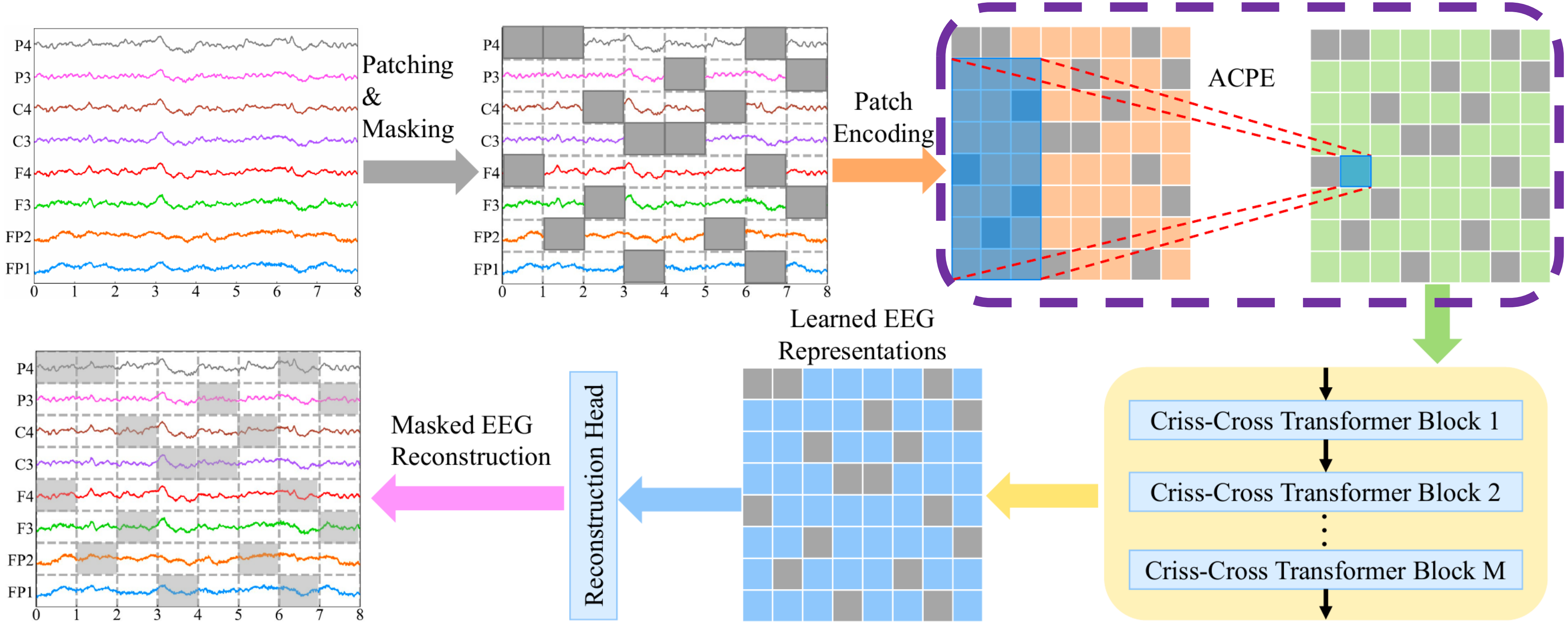
5 Our Method

Time-Frequency Patch Encoding: Each EEG patch is fed into a patch encoding network to obtain corresponding patch embedding.



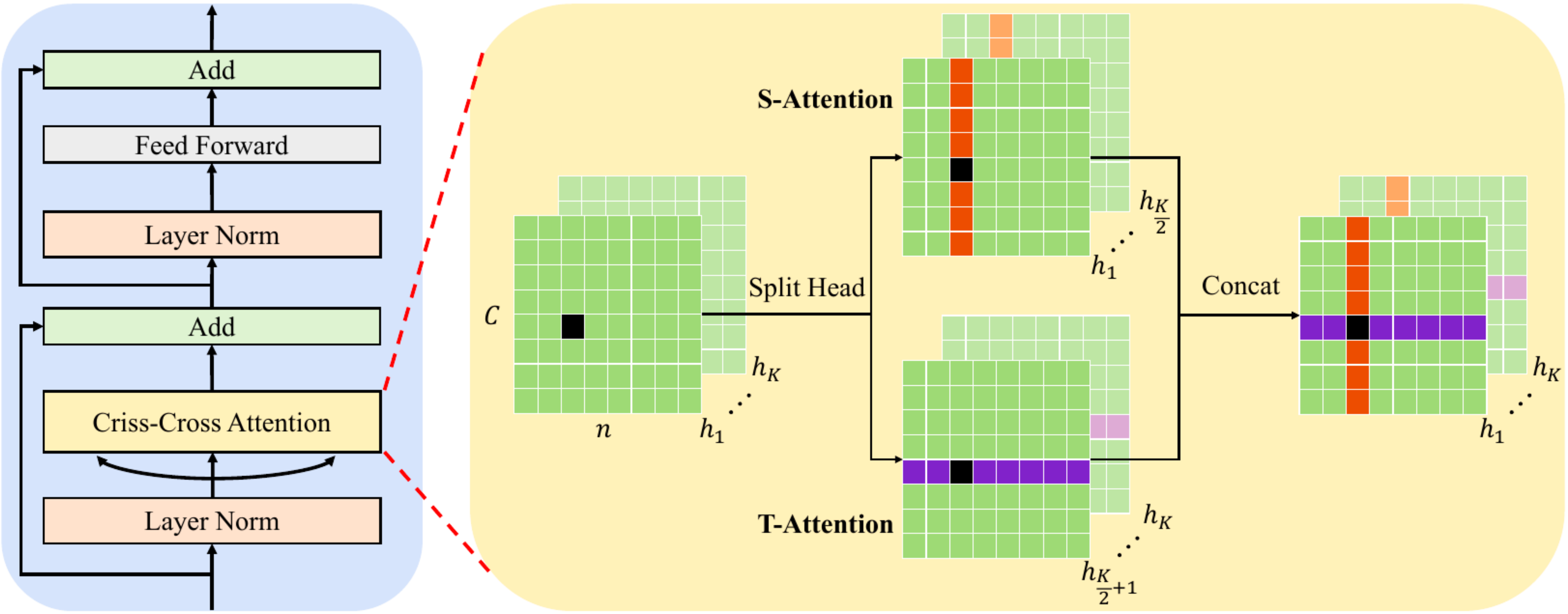
5 Our Method

Asymmetric Conditional Positional Encoding: Spatial-temporal positional embeddings are obtained through an asymmetric conditional positional encoding (ACPE) scheme and added to the patch embeddings.



5 Our Method

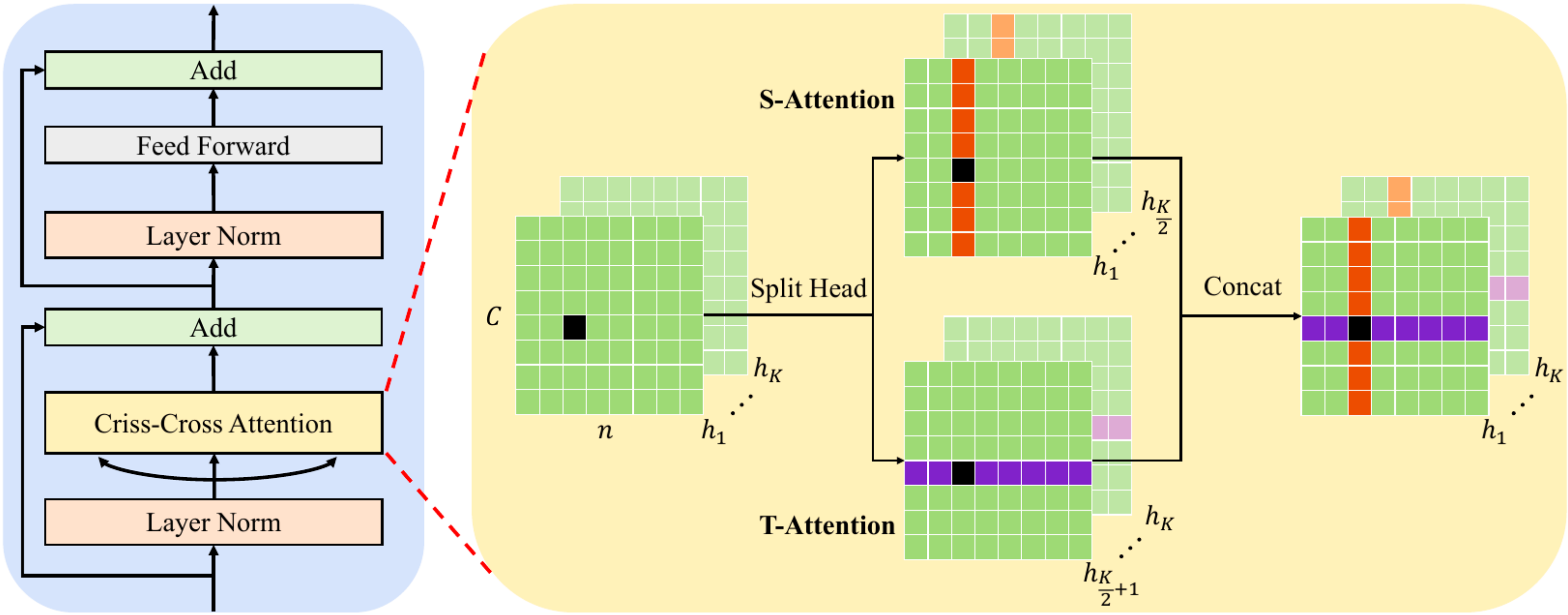
Criss-Cross Transformer: We propose criss-cross transformer as the backbone of CBraMod to capture the heterogeneous spatial and temporal dependencies among EEG patches.



(a) Criss-Cross Transformer Block (b) Criss-Cross Attention Mechanism

5 Our Method

Criss-Cross Attention Mechanism: The criss-cross attention consists of parallel spatial attention (S-Attention) and temporal attention (T-Attention).



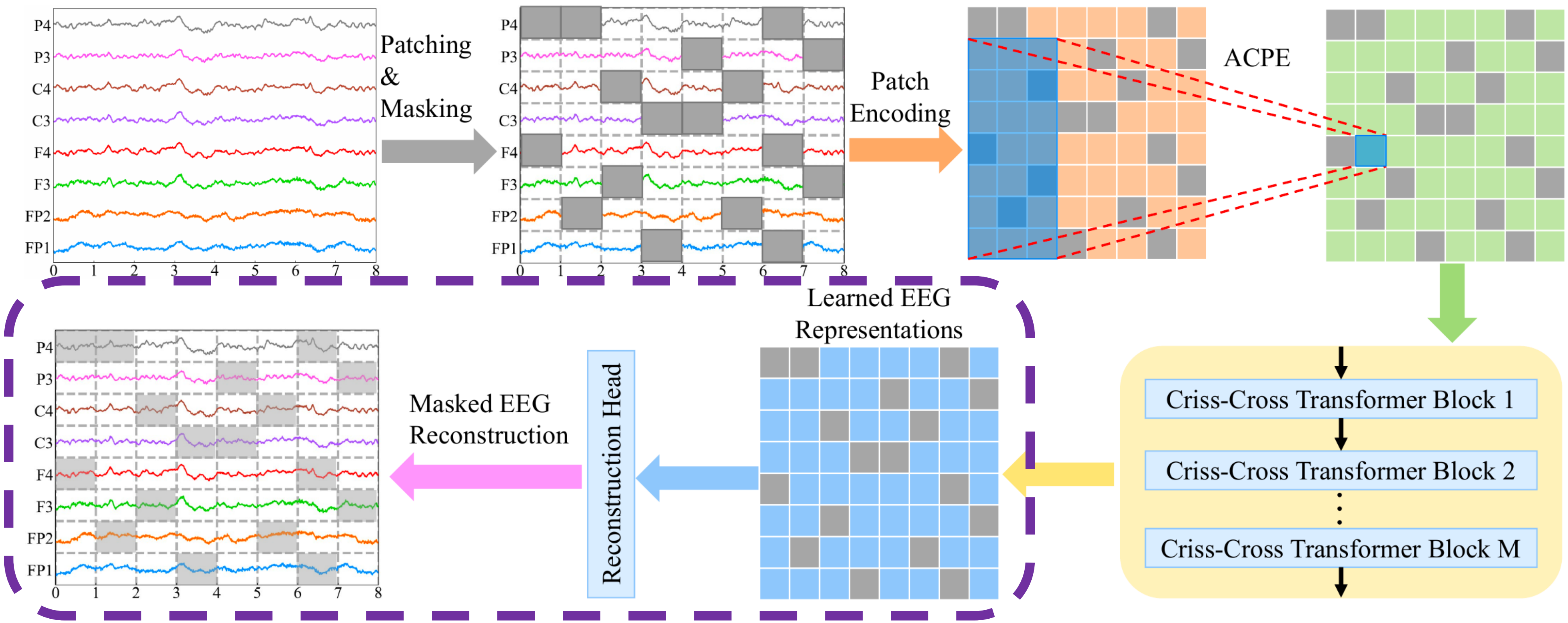
(a) Criss-Cross Transformer Block (b) Criss-Cross Attention Mechanism

$$\text{Criss-Cross-Attention}(\tilde{E}) = \text{Concat}(\text{head}_1, \text{head}_2, ..., \text{head}_\text{K})$$

$$\text{head}_\text{k} = \begin{cases} \text{S-Attention}_k(\tilde{E}), & k \in [1, 2, ..., K/2] \\ \text{T-Attention}_k(\tilde{E}), & k \in [K/2+1, K/2+2, ..., K] \end{cases}$$

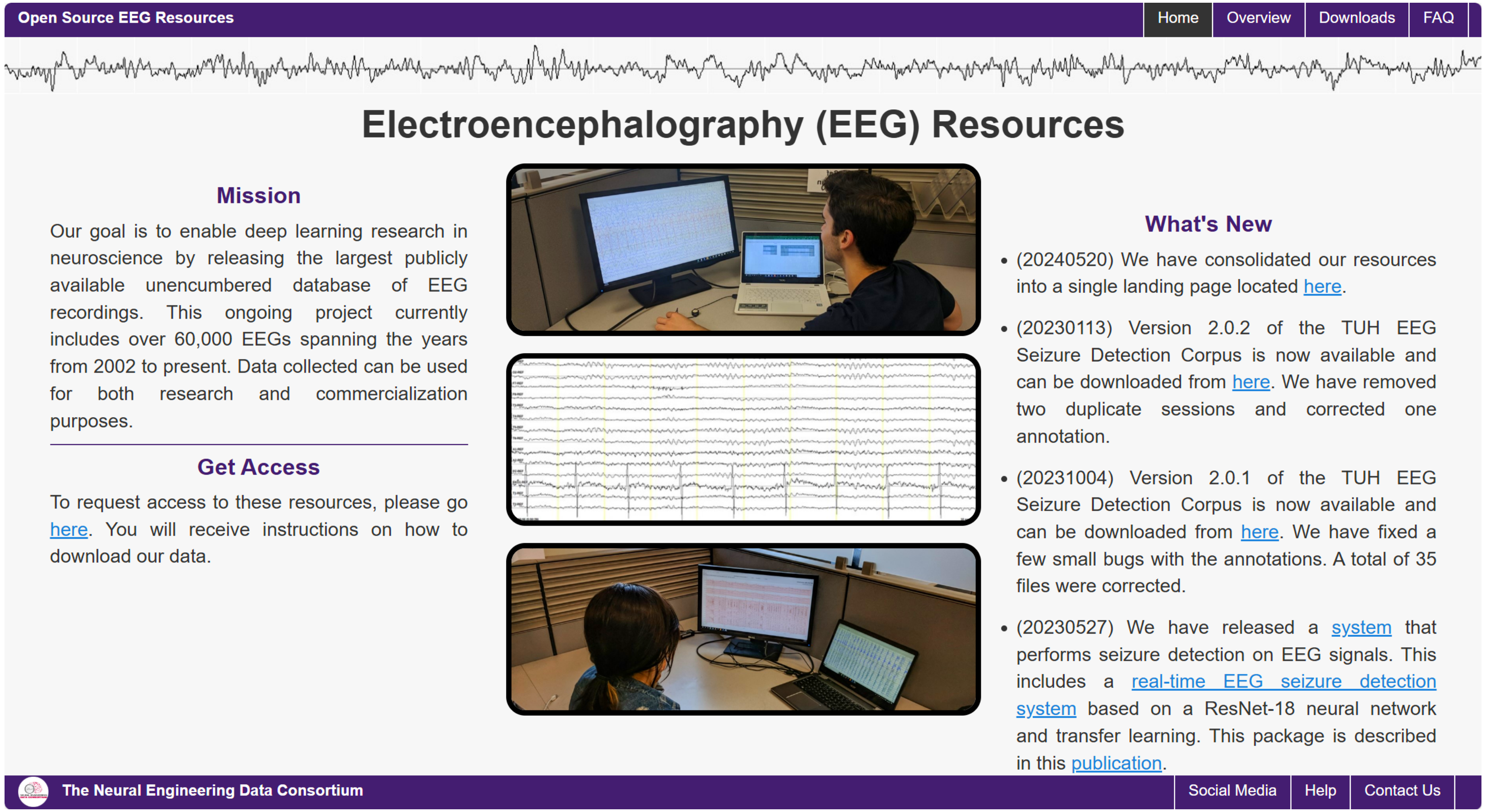
5 Our Method

Masked EEG Reconstruction: A reconstruction head is utilized to reconstruct the masked EEG patches from the learned representations.



6 Experiment Setup

Pre-training: CBraMod is pre-trained on a very large public dataset, Temple University Hospital EEG corpus (TUEG).



From https://isip.piconepress.com/projects/tuh_eeg/index.shtml

6 Experiment Setup

Finetuning: We evaluate the performance of CBraMod on up to 10 downstream BCI tasks using 12 public datasets.

Table 1: Overview of downstream BCI tasks and datasets.

BCI Tasks	Datasets	Rate	# Channels	Duration	# Samples	Label
I. Emotion Recognition	FACED	250Hz	32	10s	10,332	9-class
	SEED-V	1000Hz	62	1s	117,744	5-class
II. Motor Imagery Classification	PhysioNet-MI	160Hz	64	4s	9,837	4-class
	SHU-MI	250Hz	32	4s	11,988	2-class
III. Sleep Staging	ISRUC	200Hz	6	30s	89,240	5-class
IV. Seizure Detection	CHB-MIT	256Hz	16	10s	326,993	2-class
V. Imagined Speech Classification	BCIC2020-3	256Hz	64	3s	6,000	5-class
VI. Mental Disorder Diagnosis	Mumtaz2016	256Hz	19	5s	7,143	2-class
VII. Vigilance Estimation	SEED-VIG	200Hz	17	8s	20,355	regression
VIII. Mental Stress Detection	MentalArithmetic	500Hz	20	5s	1,707	2-class
IX. Event Type Classification	TUEV	256Hz	16	5s	112,491	6-class
X. Abnormal Detection	TUAB	256Hz	16	10s	409,455	2-class

7 Results

Table 2: The results of different methods on emotion recognition.

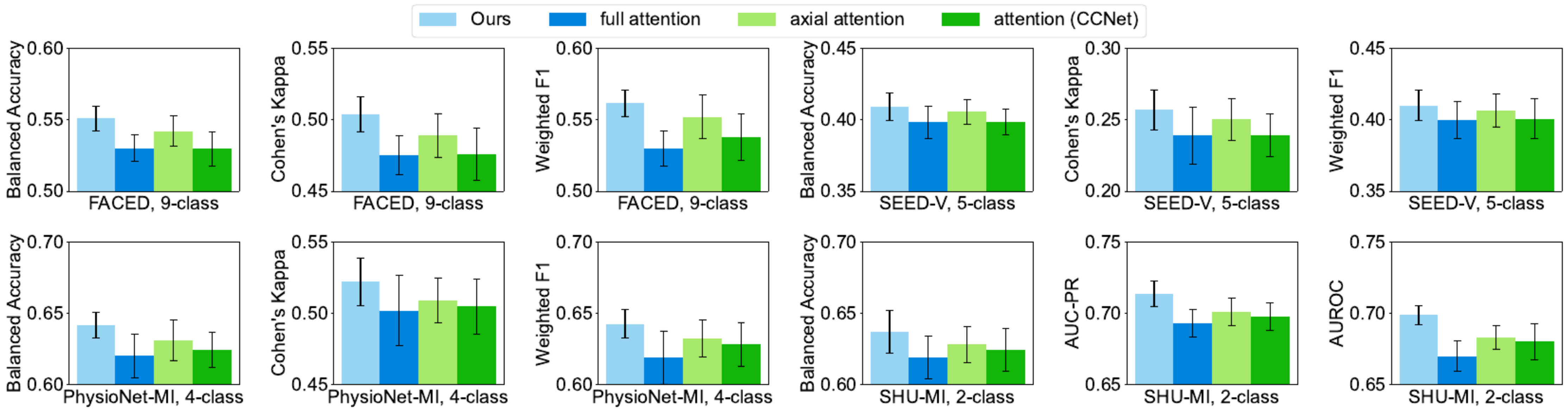
Methods	FACED, 9-class			SEED-V, 5-class		
	Balanced Accuracy	Cohen's Kappa	Weighted F1	Balanced Accuracy	Cohen's Kappa	Weighted F1
EEGNet	0.4090 ± 0.0122	0.3342 ± 0.0251	0.4124 ± 0.0141	0.2961 ± 0.0102	0.1006 ± 0.0143	0.2749 ± 0.0098
EEGConformer	0.4559 ± 0.0125	0.3858 ± 0.0186	0.4514 ± 0.0107	0.3537 ± 0.0112	0.1772 ± 0.0174	0.3487 ± 0.0136
SPaRCNet	0.4673 ± 0.0155	0.3978 ± 0.0289	0.4729 ± 0.0133	0.2949 ± 0.0078	0.1121 ± 0.0139	0.2979 ± 0.0083
ContraWR	0.4887 ± 0.0078	0.4231 ± 0.0151	0.4884 ± 0.0074	0.3546 ± 0.0105	0.1905 ± 0.0188	0.3544 ± 0.0121
CNN-Transformer	0.4697 ± 0.0132	0.4017 ± 0.0168	0.4720 ± 0.0125	0.3678 ± 0.0078	0.2072 ± 0.0183	0.3642 ± 0.0088
FFCL	0.4673 ± 0.0158	0.3987 ± 0.0383	0.4699 ± 0.0145	0.3641 ± 0.0092	0.2078 ± 0.0201	0.3645 ± 0.0132
ST-Transformer	0.4810 ± 0.0079	0.4137 ± 0.0133	0.4795 ± 0.0096	0.3052 ± 0.0072	0.1083 ± 0.0121	0.2833 ± 0.0105
BIOT	0.5118 ± 0.0118	0.4476 ± 0.0254	0.5136 ± 0.0112	0.3837 ± 0.0187	0.2261 ± 0.0262	0.3856 ± 0.0203
LaBraM-Base	0.5273 ± 0.0107	0.4698 ± 0.0188	0.5288 ± 0.0102	0.3976 ± 0.0138	0.2386 ± 0.0209	0.3974 ± 0.0111
CBraMod	$\mathbf{0.5509} \pm 0.0089$	$\mathbf{0.5041} \pm 0.0122$	$\mathbf{0.5618} \pm 0.0093$	$\mathbf{0.4091} \pm 0.0097$	$\mathbf{0.2569} \pm 0.0143$	$\mathbf{0.4101} \pm 0.0108$

Table 3: The results of different methods on motor imagery classification.

Methods	PhysioNet-MI, 4-class			SHU-MI, 2-class		
	Balanced Accuracy	Cohen's Kappa	Weighted F1	Balanced Accuracy	AUC-PR	AUROC
EEGNet	0.5814 ± 0.0125	0.4468 ± 0.0199	0.5796 ± 0.0115	0.5889 ± 0.0177	0.6311 ± 0.0142	0.6283 ± 0.0152
EEGConformer	0.6049 ± 0.0104	0.4736 ± 0.0171	0.6062 ± 0.0095	0.5900 ± 0.0107	0.6370 ± 0.0093	0.6351 ± 0.0101
SPaRCNet	0.5932 ± 0.0152	0.4564 ± 0.0234	0.5937 ± 0.0147	0.5978 ± 0.0097	0.6510 ± 0.0062	0.6431 ± 0.0082
ContraWR	0.5892 ± 0.0133	0.4527 ± 0.0248	0.5918 ± 0.0116	0.5873 ± 0.0128	0.6315 ± 0.0105	0.6273 ± 0.0113
CNN-Transformer	0.6053 ± 0.0118	0.4725 ± 0.0223	0.6041 ± 0.0105	0.5975 ± 0.0169	0.6412 ± 0.0076	0.6323 ± 0.0082
FFCL	0.5726 ± 0.0092	0.4323 ± 0.0182	0.5701 ± 0.0079	0.5692 ± 0.0252	0.5943 ± 0.0172	0.6014 ± 0.0168
ST-Transformer	0.6035 ± 0.0081	0.4712 ± 0.0199	0.6053 ± 0.0075	0.5992 ± 0.0206	0.6394 ± 0.0122	0.6431 ± 0.0111
BIOT	0.6153 ± 0.0154	0.4875 ± 0.0272	0.6158 ± 0.0197	0.6179 ± 0.0183	0.6770 ± 0.0119	0.6609 ± 0.0127
LaBraM-Base	0.6173 ± 0.0122	0.4912 ± 0.0192	0.6177 ± 0.0141	0.6166 ± 0.0192	0.6761 ± 0.0083	0.6604 ± 0.0091
CBraMod	$\mathbf{0.6417} \pm 0.0091$	$\mathbf{0.5222} \pm 0.0169$	$\mathbf{0.6427} \pm 0.0100$	$\mathbf{0.6370} \pm 0.0151$	$\mathbf{0.7139} \pm 0.0088$	$\mathbf{0.6988} \pm 0.0068$

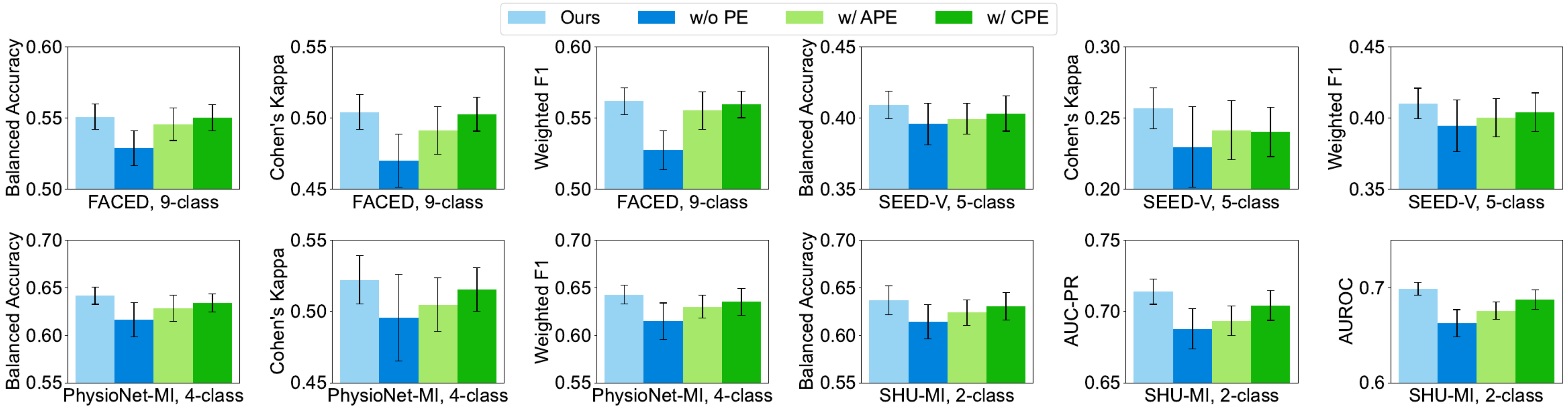
Our CBraMod achieve the state-of-the-art performance across multiple downstream datasets.

7 Results



Attention Mechanism Comparison: Our criss-cross attention employs dual parallel spatial and temporal attentions to capture spatial and temporal dependencies, which is more suitable for EEG modeling.

7 Results



Positional Encoding Comparison: Our ACPE dynamically learn positional relationships among patches, performs better than existing methods.

9 Conclusion

- In this paper, we propose an EEG foundation model called CBraMod, which can learn generic representations of EEG signals through patch-based masked EEG reconstruction.
- We devise a criss-cross transformer as the backbone of CBraMod to model the spatial and temporal dependencies between EEG patches in parallel, and an asymmetric convolutional positional encoding scheme to encode spatial-temporal positional information of EEG signals with diverse formats.
- CBraMod is pre-trained on TUEG, a very large corpus of EEG.
- CBraMod achieves the state-of-the-art performance across up to 10 downstream BCI tasks (12 public datasets), proving its strong capability and generalizability.
- We hope that the proposed modeling approach and positional encoding scheme can provide meaningful insights for building EEG foundation models, thereby advancing the development of real-world BCI systems.

浙江大學 ZHEJIANG UNIVERSITY

脑机智能全国重点实验室
The State Key Laboratory of Brain–Machine Intelligence

ICLR

Thank you!

wangjiquan@zju.edu.cn