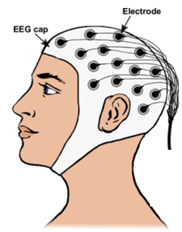
# MBrain: A Multi-channel Self-Supervised Learning Framework for Brain Signals


Donghong Cai[1*], Junru Chen[1*], Yang Yang[1†], Teng Liu[1], Yafeng Li[2]

[1]Zhejiang University [2]Nuozhu Technology Co., Ltd.

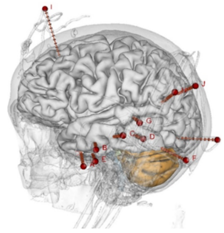
Email: jrchen_cali@zju.edu.cn, yangya@zju.edu.cn; Homepage: https://yangy.org


## Background & Challenges




### Non-invasive Methods (EEG)

- Easy to implement **without** any surgery.
- **Cannot** simultaneously consider temporal and spatial resolution along with the deep brain information.

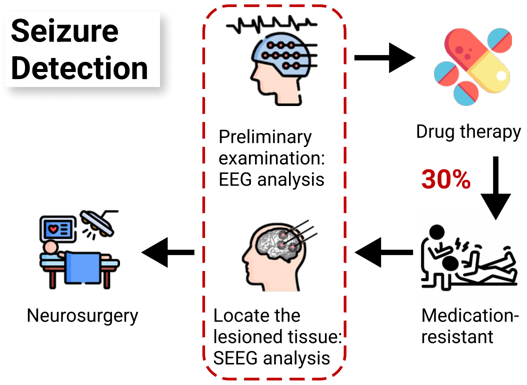


### Invasive Methods (SEEG)

- Require extra **surgeries** to insert the recording devices.
- Have access to **more** precise and **higher** signal-to-noise data.

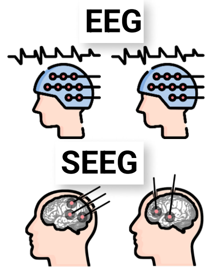
### Why do we want to model EEG and SEEG signals uniformly?

- Share similar physiological mechanisms
- Closely related in healthcare applications




### Challenge 1. Lack of a unified method for handling both signals

- Varying monitoring location for SEEG
- Different signal patterns for EEG and SEEG signals

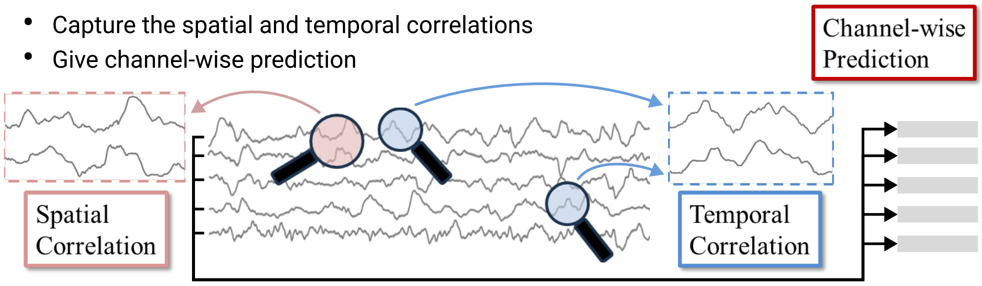



EEG:
1. A gold-standard collection location
2. Collect noisy and rough scalp signals

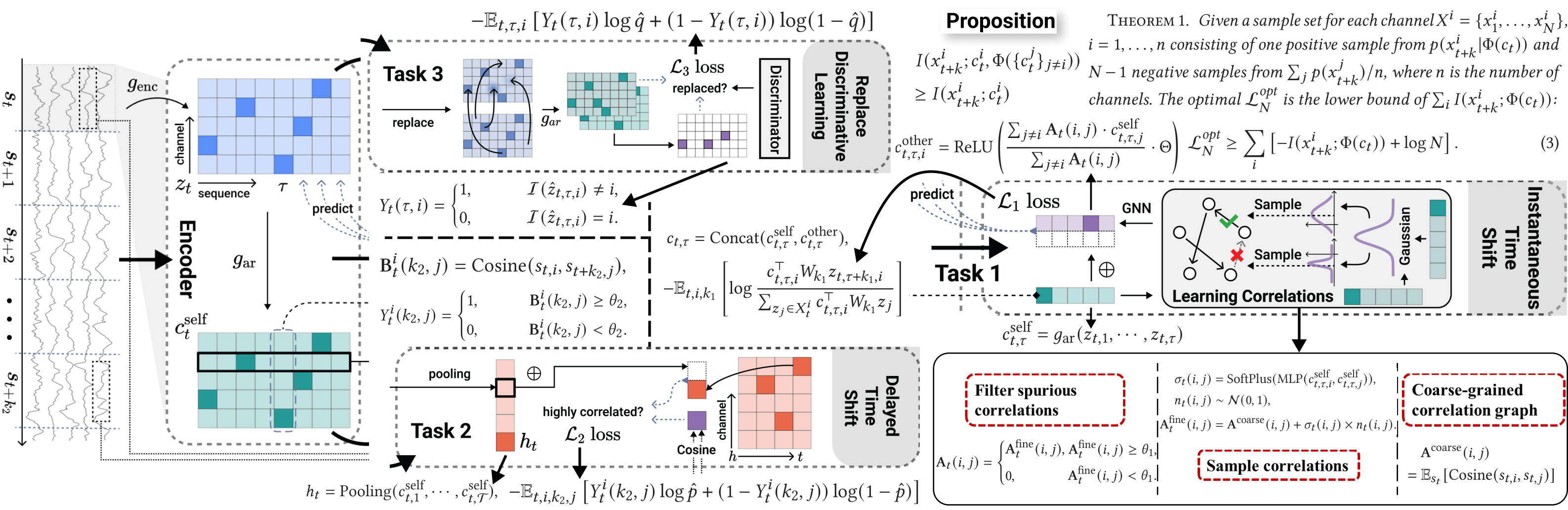
SEEG:
1. Different number and position
2. Collect more stereo and deeper signals

### Challenge 2. A gap between existing methods and applications

- Capture the spatial and temporal correlations
- Give channel-wise prediction




## MBrain




$$-\mathbb{E}_{t,\tau,i}\left[Y_t(\tau,i)\log\hat{q}+(1-Y_t(\tau,i))\log(1-\hat{q})\right]$$

$$Y_t(\tau,i)=\begin{cases}1, & \mathcal{I}(\hat{z}_{t,\tau,i})\neq i,\\ 0, & \mathcal{I}(\hat{z}_{t,\tau,i})=i.\end{cases}$$

$$\mathbf{B}_t^i(k_2,j)=\mathrm{Cosine}(s_{t,i},s_{t+k_2,j}),$$

$$Y_t^i(k_2,j)=\begin{cases}1, & \mathbf{B}_t^i(k_2,j)\geq\theta_2,\\ 0, & \mathbf{B}_t^i(k_2,j)<\theta_2.\end{cases}$$

$$h_t=\mathrm{Pooling}(c_{t,1}^{self},\cdots,c_{t,\mathcal{T}}^{self}),\quad -\mathbb{E}_{t,i,k_2,j}\left[Y_t^i(k_2,j)\log\hat{p}+(1-Y_t^i(k_2,j))\log(1-\hat{p})\right]$$

$$c_{t,\tau}=\mathrm{Concat}(c_{t,\tau}^{self},c_{t,\tau}^{other}),$$

$$-\mathbb{E}_{t,i,k_1}\left[\log\frac{c_{t,\tau,i}^{\top}W_{k_1}z_{t,\tau+k_1,i}}{\sum_{z_j\in X_t^i}c_{t,\tau,i}^{\top}W_{k_1}z_j}\right]$$

$$c_{t,\tau}^{self}=g_{ar}(z_{t,1},\cdots,z_{t,\tau})$$

$$c_{t,\tau,i}^{other}=\mathrm{ReLU}\left(\frac{\sum_{j\neq i}\mathbf{A}_t(i,j)\cdot c_{t,\tau,j}^{self}}{\sum_{j\neq i}\mathbf{A}_t(i,j)}\cdot\Theta\right)$$

$$\sigma_t(i,j)=\mathrm{SoftPlus}(\mathrm{MLP}(c_{t,\tau,i}^{self},c_{t,\tau,j}^{self})),$$
$$n_t(i,j)\sim\mathcal{N}(0,1),$$
$$\mathbf{A}_t^{fine}(i,j)=\mathbf{A}^{coarse}(i,j)+\sigma_t(i,j)\times n_t(i,j).$$

$$\mathbf{A}_t(i,j)=\begin{cases}\mathbf{A}_t^{fine}(i,j), & \mathbf{A}_t^{fine}(i,j)\geq\theta_1,\\ 0, & \mathbf{A}_t^{fine}(i,j)<\theta_1.\end{cases}$$

$$\mathbf{A}^{coarse}(i,j)=\mathbb{E}_{s_t}\left[\mathrm{Cosine}(s_{t,i},s_{t,j})\right]$$

### Proposition

$$I(x_{t+k}^i;c_t^i,\Phi(\{c_t^j\}_{j\neq i}))\geq I(x_{t+k}^i;c_t^i)$$

THEOREM 1. *Given a sample set for each channel $X^i=\{x_1^i,\ldots,x_N^i\}$, $i=1,\ldots,n$ consisting of one positive sample from $p(x_{t+k}^i|\Phi(c_t))$ and $N-1$ negative samples from $\sum_j p(x_{t+k}^j)/n$, where $n$ is the number of channels. The optimal $\mathcal{L}_N^{opt}$ is the lower bound of $\sum_i I(x_{t+k}^i;\Phi(c_t))$:*

$$\mathcal{L}_N^{opt}\geq\sum_i\left[-I(x_{t+k}^i;\Phi(c_t))+\log N\right]. \quad (3)$$

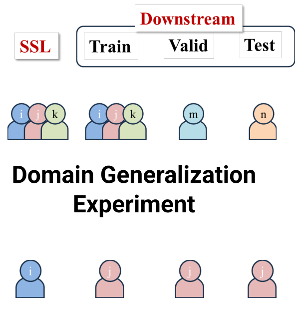
## Experiments

| Models | SEEG | | | | EEG | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Pre. | Rec. | $F_1$ | $F_2$ | Pre. | Rec. | $F_1$ | $F_2$ | AUROC |
| MiniRocket | 5.85±0.20 | **39.18±0.59** | 9.93±0.29 | 17.24±0.37 | **22.86±0.84** | 63.08±1.47 | 33.56±1.11 | 46.66±1.33 | 75.30±0.77 |
| CPC | 22.88±5.06 | 23.92±3.90 | 20.11±3.27 | 21.23±2.49 | 22.81±2.04 | 58.31±7.55 | 32.50±1.24 | 44.02±2.43 | 74.53±1.00 |
| SimCLR | 14.02±3.71 | 26.36±4.99 | 11.07±3.49 | 13.47±4.01 | 12.63±1.62 | 74.88±16.77 | 21.33±1.95 | 36.78±2.61 | 55.86±5.36 |
| T-Loss | 21.38±4.25 | 28.50±4.07 | 23.48±3.30 | 25.90±3.06 | 20.72±1.26 | 69.25±3.99 | 31.82±1.08 | 47.00±0.50 | 75.88±0.49 |
| TST | 8.37±3.96 | 32.48±8.25 | 11.80±3.91 | 15.67±3.69 | 15.65±1.54 | 28.59±12.93 | 19.65±4.36 | 23.87±8.09 | 58.20±4.27 |
| GTS | 24.16±5.91 | 27.99±4.98 | 22.77±2.69 | 24.15±2.79 | 18.86±1.09 | 62.51±5.04 | 28.88±0.88 | 42.54±1.48 | 71.69±1.88 |
| TS-TCC | 24.24±4.51 | 26.61±5.96 | 19.89±5.23 | 22.11±5.08 | 15.55±0.88 | 39.76±11.08 | 21.89±1.20 | 29.60±4.64 | 58.63±1.62 |
| TS2Vec | 27.93±5.23 | 29.49±3.97 | 26.78±3.29 | 27.88±3.52 | 21.40±0.63 | 58.31±6.14 | 31.24±1.18 | 43.24±2.78 | 73.35±1.02 |
| *MBrain* | **30.69±5.92** | 38.94±4.34 | **32.61±3.60** | **35.64±3.04** | 22.13±1.03 | **76.99±4.49** | **34.32±0.90** | **51.34±0.97** | **77.96±0.97** |

21.77% ↑ 27.83% ↑ 4.74% ↑ on average

| Setting | Group *A* | | | Group *B* | | | Group *C* | | | Group *D* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *B*→*A* | *C*→*A* | *D*→*A* | *A*→*B* | *C*→*B* | *D*→*B* | *A*→*C* | *B*→*C* | *D*→*C* | *A*→*D* | *B*→*D* | *C*→*D* |
| DA | 68.55±4.27 | 69.14±6.54 | 68.78±4.12 | 41.08±2.59 | 46.06±3.05 | 46.12±2.04 | 40.04±3.98 | 39.34±2.11 | **48.64±5.48** | 80.82±0.65 | 79.90±1.11 | 80.72±1.31 |
| Max-base | | 62.49±2.30 | | | 39.78±2.04 | | | 33.59±2.23 | | | 75.35±0.79 | |
| Non-DA | | **70.63±1.41** | | | **46.62±2.42** | | | 46.09±2.35 | | | **83.27±0.95** | |

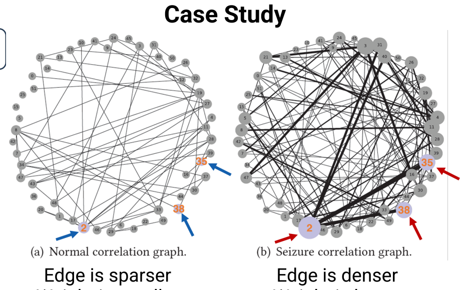
Beat




**Domain Generalization Experiment**

**Domain Adaptation Experiment**

### Case Study



(a) Normal correlation graph. (b) Seizure correlation graph.

Edge is sparser
Weight is smaller

Edge is denser
Weight is larger

**MBrain can learn the correlation patterns of brain signals.**