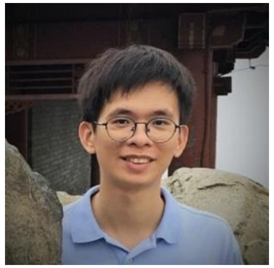
Contrastive Self-supervised EEG Representation Learning for Emotion Classification





Keya Hu[1], Ren-Jie Dai[2], Wen-Tao Chen[1], Hao-Long Yin[1], Bao-Liang Lu[1,3], Wei-Long Zheng[1]

[1]Shanghai Jiao Tong University, [2]Tongji University,
[3]RuiJin-Mihoyo Laboratory, Clinical Neuroscience Center, RuiJin Hospital,
Shanghai Jiao Tong University School of Medicine
{hu_keya, leonard_chen, yinhaolong, bllu, weilong}@sjtu.edu.cn dairenjie123@tongji.edu.cn

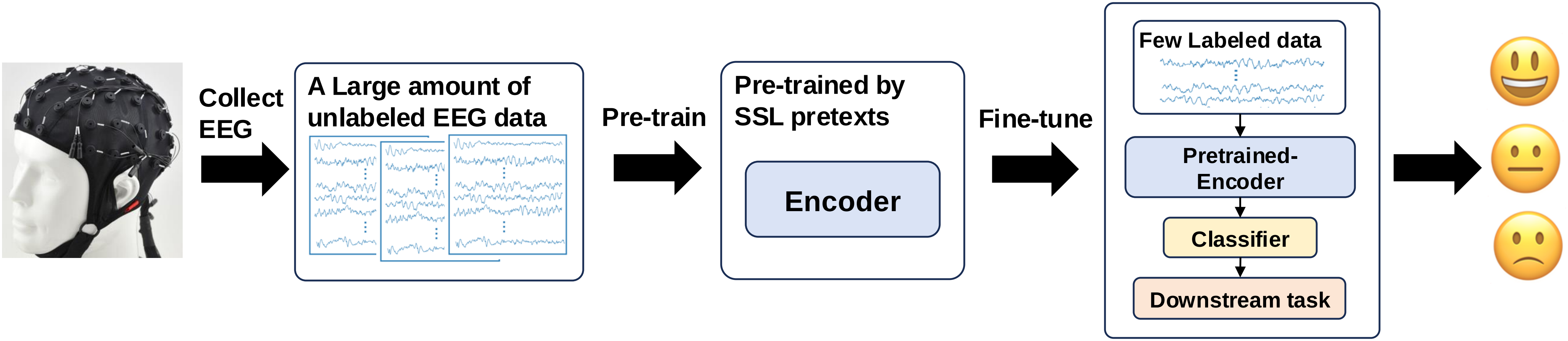
Introduction

Brain-Computer Interface

- **Affective Brain-Computer Interfaces (aBCIs)** offer a technological means to detect human emotions directly from **electroencephalogram (EEG)** signals.

Self-supervised learning

- SSL approaches that utilize **a large amount of unlabeled data** for pre-training and fine-tuning on a smaller labeled dataset have proven to achieve favorable outcomes.

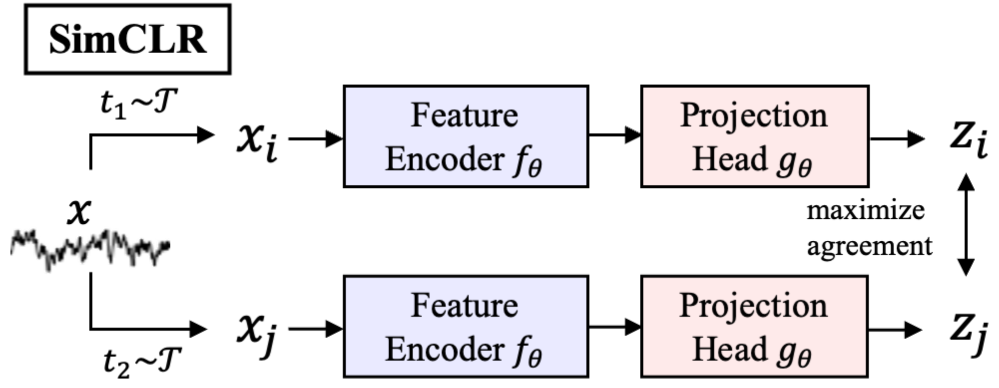


Contribution

To investigate the Contrastive Self-supervised EEG Representation Learning for Emotion Classification, in this paper:

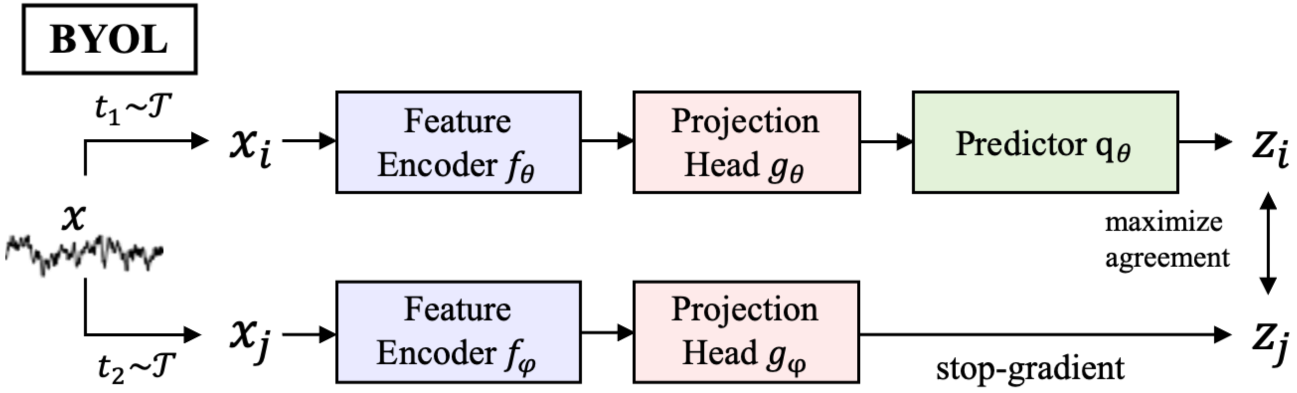
1) We explore the **effective feature encoder** for raw EEG signals.
2) Experiment with and explore **several classic contrastive self-supervised learning methods** to get better pretexts for raw EEG signals.
3) Visualize **key brain regions** for emotion classification.

Pre-training methods



SimCLR[1]

- Use one encoder to generate anchor, positive, and negative samples.
- Maximize agreement between differently augmented views of the same data samples.

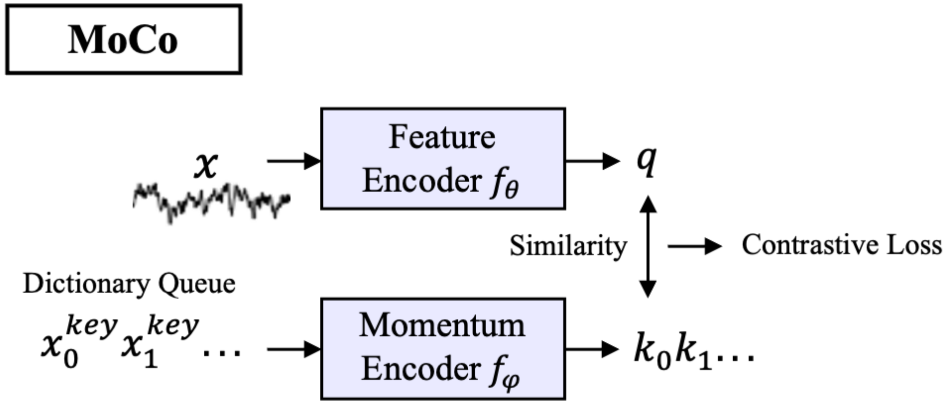


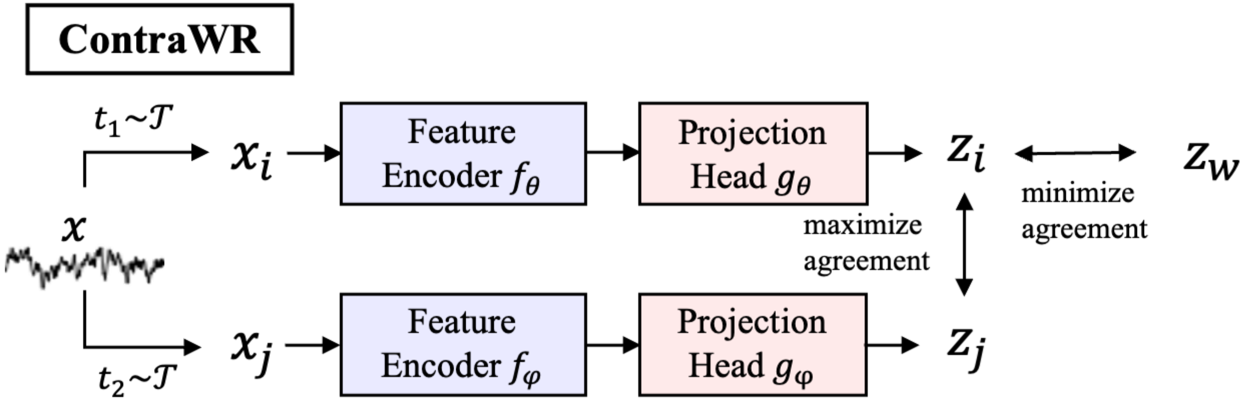
BYOL[2]

- The online network uses augmented data to predict the target network's representation of another augmented view of the same data.

[1] T.Chen,S.Kornblith,M.Norouzi,andG.Hinton,"Asimpleframework for contrastive learning of visual representations," in *International Conference on Machine Learning*, pp. 1597–1607, PMLR, 2020.
[2] J.-B. Grill, F. Strub, F. Altché, et al., "Bootstrap your own latent-a new approach to self-supervised learning," Advances in Neural Information Processing Systems, vol. 33, pp. 21271–21284, 2020.

Pre-training methods





MoCo[3]

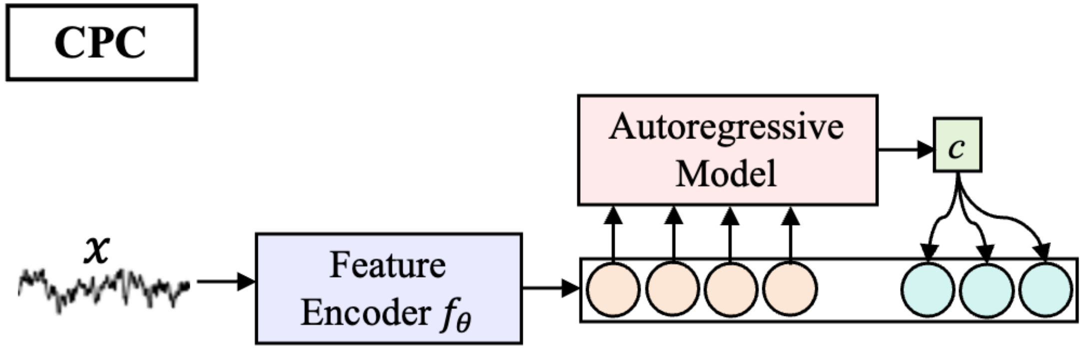
- Maintain a queue of encoded samples as a dictionary.
- The positive samples are obtained by encoding the augmented data, while negative samples are derived from the queue.

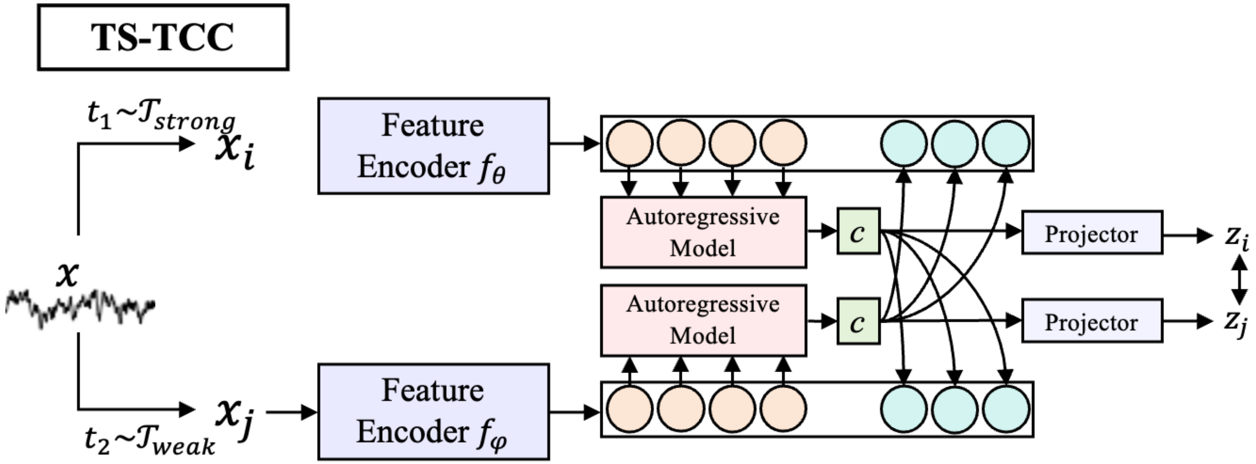
ContraWR[4]

- It replaces negative samples with a single average world representation over the dataset.

[3] K. He, H. Fan, Y. Wu, S. Xie, and R. Gira simpleck, “Momentum con- trast for unsupervised visual representation learning,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9729–9738, 2020.
[4] C. Yang, D. Xiao, M. B. Westover, and J. Sun, “Self-supervised EEG representation learning for automatic sleep staging,” *Journal of Medical Internet Research AI*, 2023.

Pre-training methods





CPC[5]

- Extracts meaningful representations and then feeds them into an autoregressive model to predict future sequences.

TS-TCC[6]

- Generates context vectors by weak and strong augmentations.
- One augmented view's context vector predicts future sequences of the other one.

[5] A. van den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," arXiv, vol. abs/1807.03748, 2018.
[6] E.Eldele,M.Ragab,Z.Chen,M.Wu,C.K.Kwoh,X.Li,andC.Guan, "Time-series representation learning via temporal and contextual con- trasting," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pp. 2352–2359, 2021.

EEG Emotion Dataset

Dataset

SEED[7] and **SEED-IV**[8], The SJTU Emotion EEG datasets are a series of datasets that record the EEG signals of subjects while they are watching emotion videos.

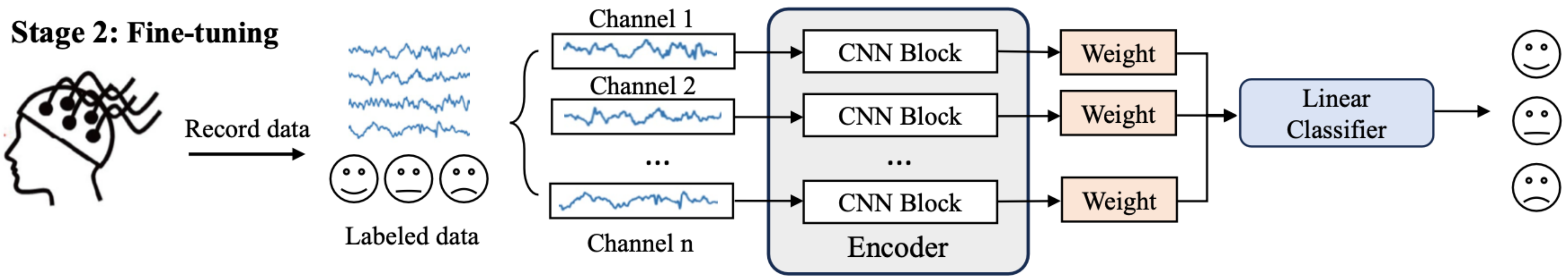
- The SEED dataset includes three emotions: positive, neutral, and negative.
- The SEED-IV dataset includes four emotions: happy, sad, neutral, and fear.
- Each sample is a 1-second non-overlap window raw EEG signals with 62 channels.

[7] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.
[8] W.-L.Zheng,W.Liu,Y.Lu,B.-L.Lu,andA.Cichocki,"Emotionmeter: A multimodal framework for recognizing human emotions," IEEE Transactions on Cybernetics, vol. 49, no. 3, pp. 1110–1122, 2019.

Feature encoder

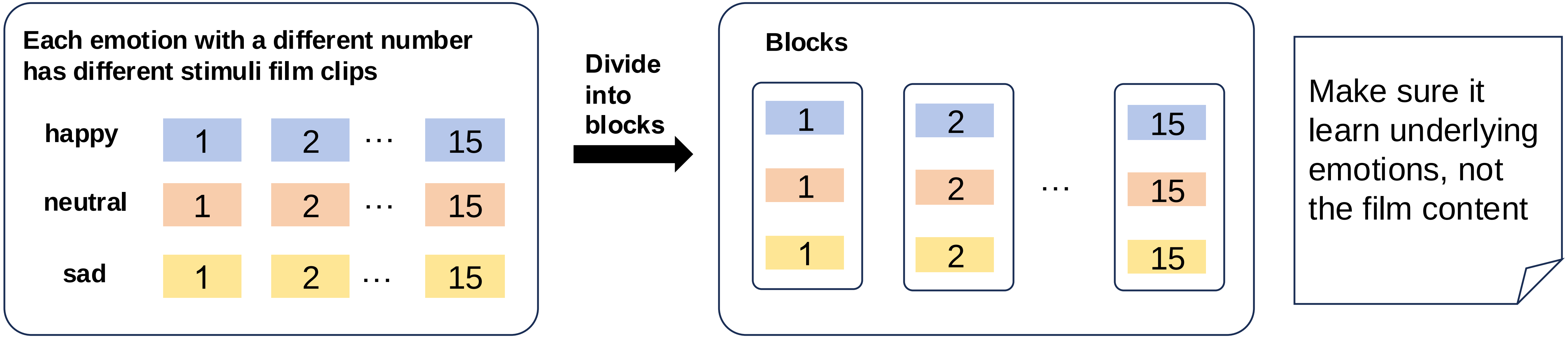
Encoder

- The feature encoder is constructed using 62 dependent 1D CNN models.
- Connect learnable weights to each channel of the feature encoder.
- Connect single-layer linear classifier to the end.



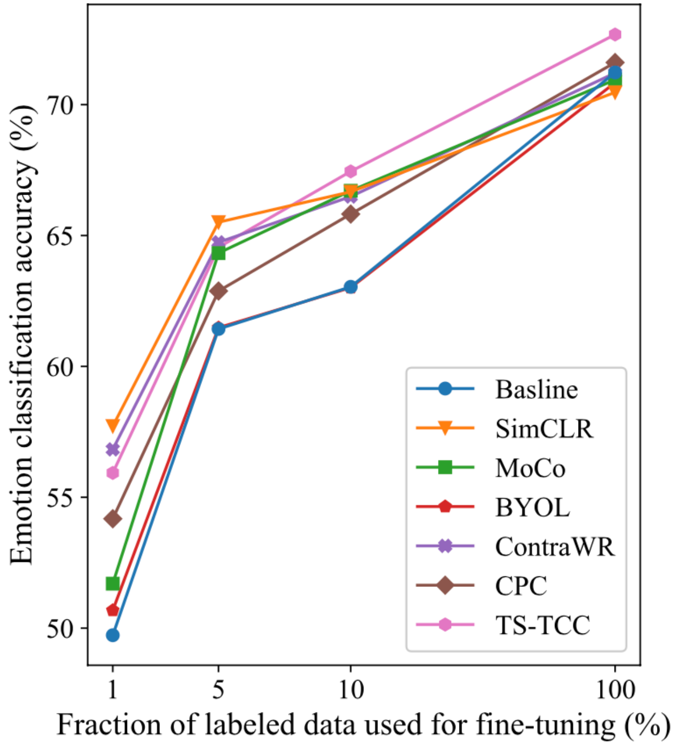
Experiment Setting

- For the SEED dataset, we split it into a 3:1:1 ratio for training, validation, and testing sets, while for the SEED-IV dataset, we use a 4:1:1 ratio.
- We evaluate the models using three different random seeds to obtain three distinct training sets, testing sets, and validation sets.
- Pretrain with all unlabeled data and fine-tune with different portions of labeled data.

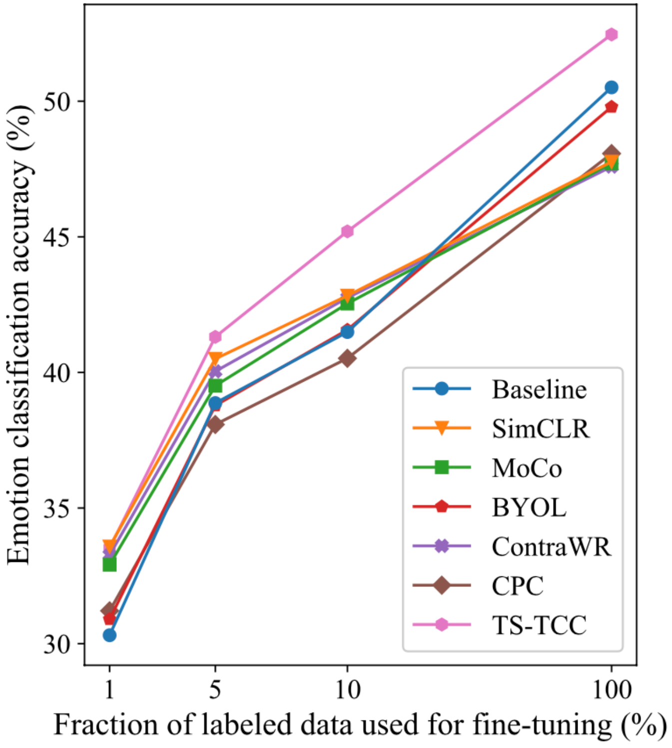


Experimental Results

Fine-tune on SEED, SEED-IV, and pre-train on itself



(a) SEED

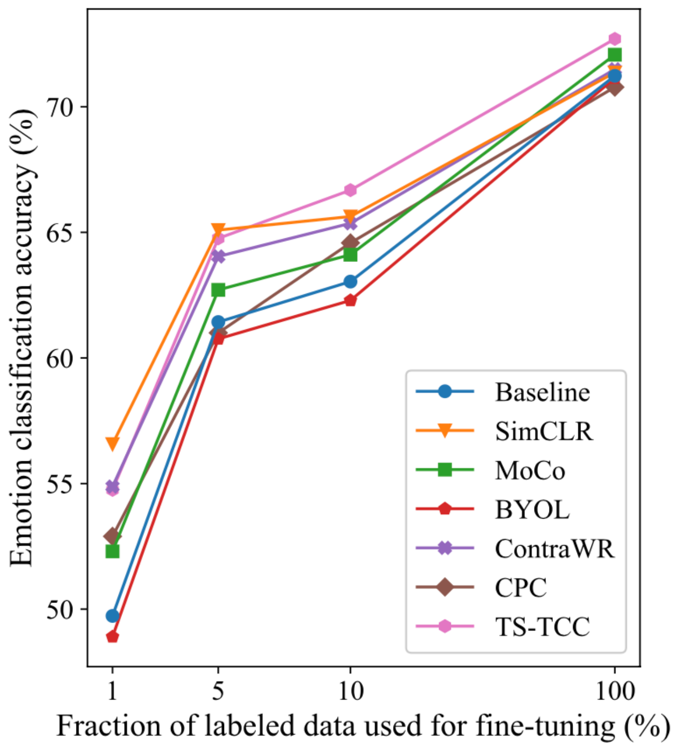


(b) SEED-IV

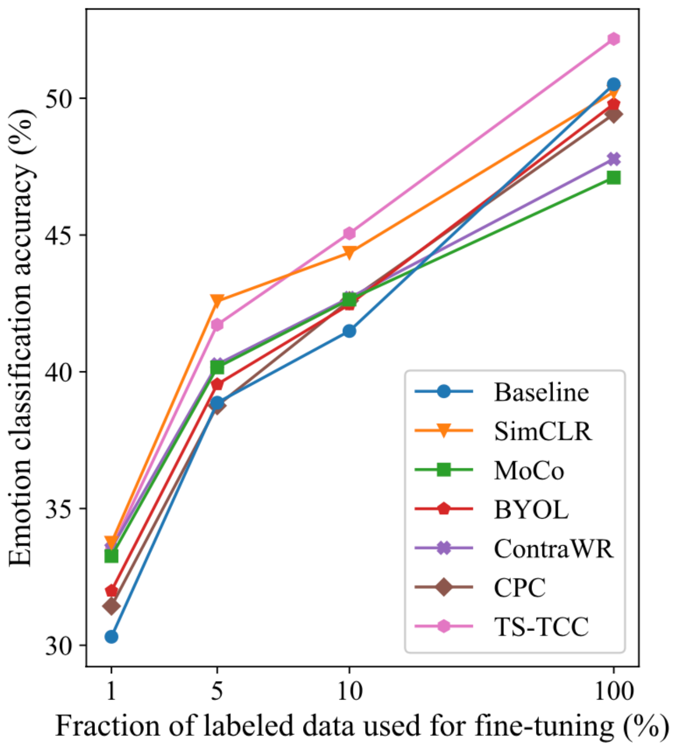
- With no more than 10% of the data volume, most self-supervised methods maintain a good lead in accuracy during fine-tuning.
- SimCLR, ContraWR, and TS-TCC methods consistently show great performance and achieve accuracy of 65.51%, 64.74%, and 64.55% with only 5% data.

Experimental Results

Fine-tune on SEED, SEED-IV, and pre-train on each other



(a) SEED-IV → SEED



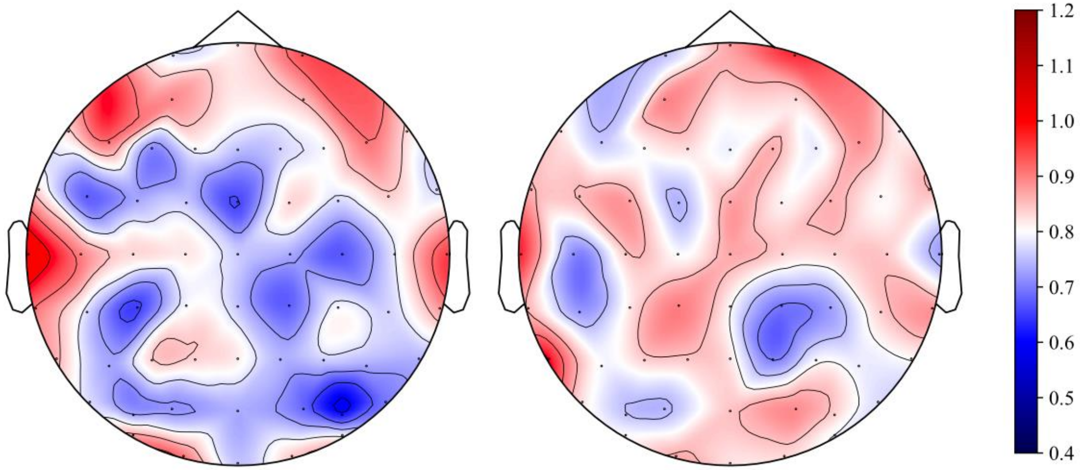
(b) SEED → SEED-IV

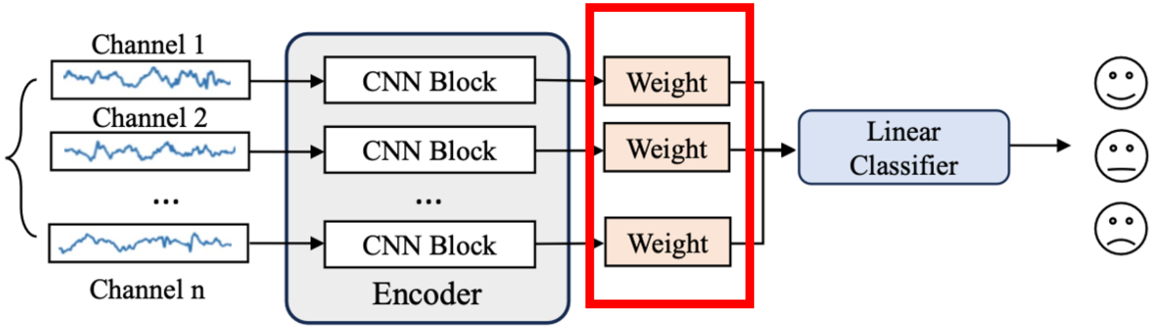
- The pre-trained feature extractors exhibit good transferability, capturing EEG features that are not specific to the training dataset.
- TS-TCC demonstrates the highest stability, accuracy, and best transferability. It is essential to design self-supervised methods that prioritize temporal relationships.

Key brain regions for emotion

Topography map

- The topography maps represent the **average weights** of each channel.
- The regions with high weights are relatively consistent with previous studies, including the **prefrontal lobe, temporal lobe, and occipital lobe**.





Conclusion

- Pre-training on the raw EEG signals, followed by fine- tuning on a labeled dataset for emotion classification, leads to an **improvement** in classification accuracy, particularly in scenarios with **limited labeled training data**.
- The pre-trained feature extractors exhibit good **transferability**.
- For signals like EEG, it is essential to design self-supervised methods that prioritize **temporal relationships**.

Thanks & QA

Keya Hu[1], Ren-Jie Dai[2], Wen-Tao Chen[1], Hao-Long Yin[1], Bao-Liang Lu[1,3], Wei-Long Zheng[1]

[1]Shanghai Jiao Tong University, [2]Tongji University,
[3]RuiJin-Mihoyo Laboratory, Clinical Neuroscience Center, RuiJin Hospital,
Shanghai Jiao Tong University School of Medicine
{hu_keya, leonard_chen, yinhaolong, bllu, weilong}@sjtu.edu.cn dairenjie123@tongji.edu.cn