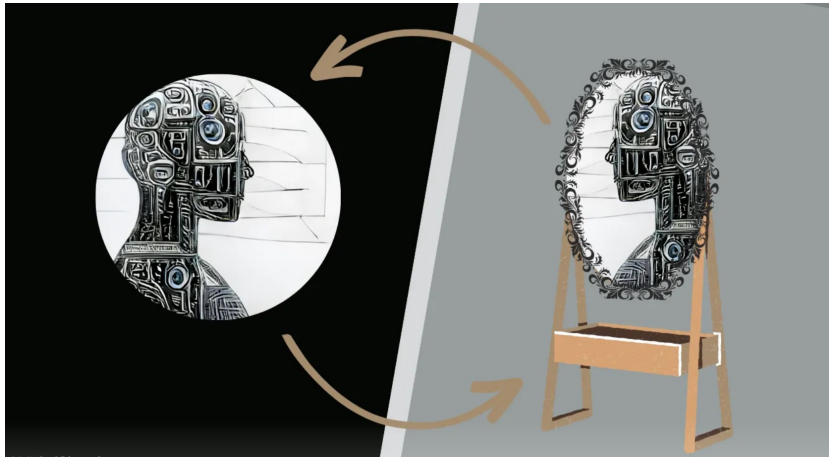


Contrastive Representation Learning for Electroencephalogram Classification

Machine Learning for Health, 2020

Citation: 140

Presenter: Nooshin Taheri

5/21/2024

Why self-supervised learning:

- Advanced machine learning techniques need large, labeled EEG datasets because EEG is complex and is usually contaminated with significant artifacts.
- EEG data collection and labeling are expensive.
- Combining datasets is infeasible due to inconsistent experimental paradigms.
- Publicly available labeled EEG data is limited and existing datasets are small and incompatible.
- Self-supervised learning (SSL) enables learning from varied EEG setups and trials.
- SSL is useful with limited labeled data and costly manual labeling.
- Unlike traditional supervised models that begin with random initial parameters—weights, kernels, and biases—which require extensive labeled data to optimize effectively, SSL utilizes unlabeled data to pre-train these parameters.
- This pre-training enhances model accuracy by providing a more accurate starting point for parameter tuning and accelerates the learning process by reducing reliance on labeled data.

OVERVIEW OF SELF-SUPERVISED LEARNING

- There are two steps in self-supervised learning: Pretext Task and Downstream Task.

- The ‘pretext task’:
  - Creates a good model starting point using both labeled and unlabeled data.
  - This task helps develop initial model parameters and useful data features.
  - These features capture general data characteristics, not specific details.
  - For EEG data, features might capture overall brain activity patterns, not specific conditions like seizures.

- The ‘downstream task’:
  - Refines the model with labeled data.
  - The last layer adjusts to work well with existing layers.
  - Fine-tune the entire model by adjusting all layers.

- Unlike traditional supervised learning, SSL uses initial parameters from the pretext task, not random.
- This inherited setup improves performance, especially with scarce labeled data, by leveraging the broad understanding from the pretext task.

CONTRASTIVE PRETEXT TECHNIQUES:

- Contrastive pretext techniques identify differences between augmentations of input data, labeled or not.
- Augmented inputs are paired with the original data to form contrastive pairs.
- Pairs can include one original and one augmented input or two different augmentations of the same input.
- These pairs train the model using a contrastive loss function.
- The objective is to maximize the agreement between positive pairs (instances from the same sample) and minimize the agreement between negative pairs (instances from different samples)

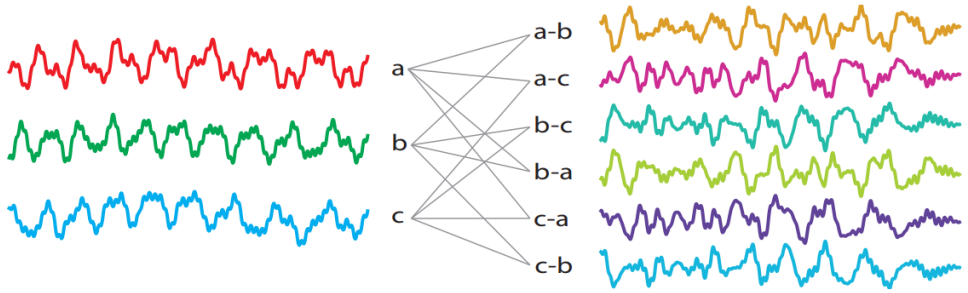
About this paper

- Present a framework for learning EEG signal representations via contrastive learning.
- Modify the SimCLR framework for time-series data to learn EEG representations.
- Extract features from a single channel at a time, allowing recombination of multi-channel recordings and fusion of datasets.
- Apply the pre-trained encoder on three classification tasks:
  - Emotion Recognition (ER) on the SEED dataset.
  - Normal/Abnormal Classification (NAC) on the TUH dataset.
  - Sleep-stage scoring (SSS) on the SleepEDF dataset.

Channel recombination and preprocessing

- **To learn the representation of a single-channel**
  - Combine different datasets to obtain a larger one:
    (1) TUH Normal/Abnormal EEG, (2) SEED dataset,
    (3) Sleep EDF, (4) Texas State University Resting State dataset, and (5) ISRUC-Sleep dataset

  - Recombine channels in a multi-channel recording to obtain more valid channels.



- **Preprocessing :**
  - Resampled all datasets to 200Hz.
  - Applied a fifth-order band-pass Butterworth filter (0.3-80 Hz).
  - Removed high-voltage channels (higher than 500 μVs) as artifacts.
  - Train the encoder with 20-second channel chunks.

Channel augmentations:

- A key ingredient of contrastive learning is a set of augmentations (or transformations) that do not alter the semantic information of data.
- A contrastive learning algorithm learns representations that are maximally similar for augmented instances of the same data point and minimally similar for different data points.
- The objective is to learn features that reflect the high-level content of EEG signals.
- Consulted neurologists on EEG data augmentations.

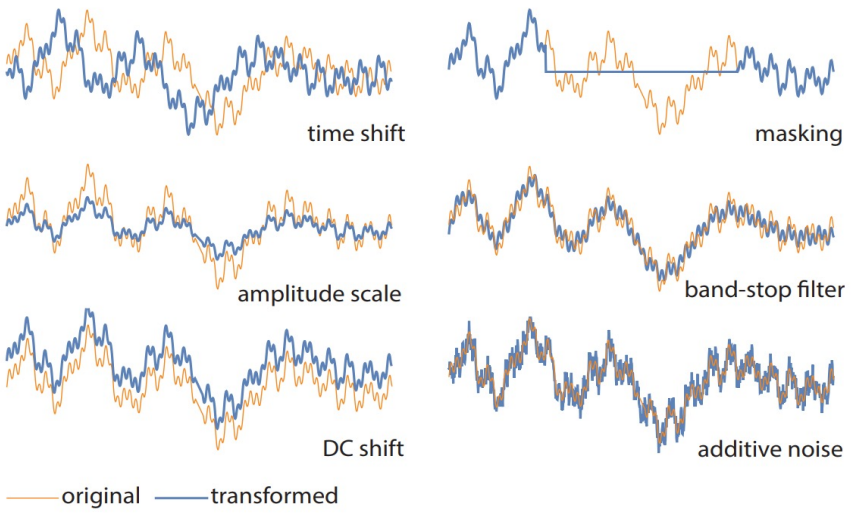


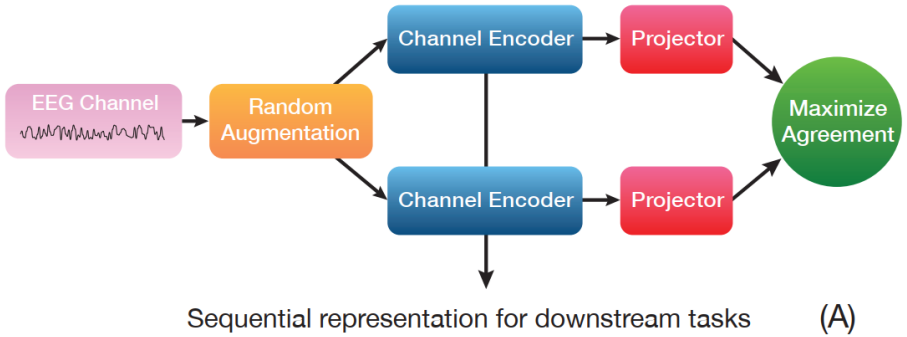
Table 1: Transformation Ranges

Transformation	min	max
amplitude scale	0.5	2
time shift (samples)	-50	50
DC shift (μV)	-10	10
zero-masking (samples)	0	150
additive Gaussian noise (σ)	0	0.2
band-stop filter (5 Hz width) (Hz)	2.8	82.5

Learning algorithm

- SeqCLR (Sequential Contrastive Learning of Representations) → Similar to SimCLR
- learns features by maximizing the similarity between differently augmented transformations of the same channel through a contrastive loss.
- This model consists of four modules:
  - Channel Augmenter
  - Channel encoder
  - Projector
  - Contrastive loss

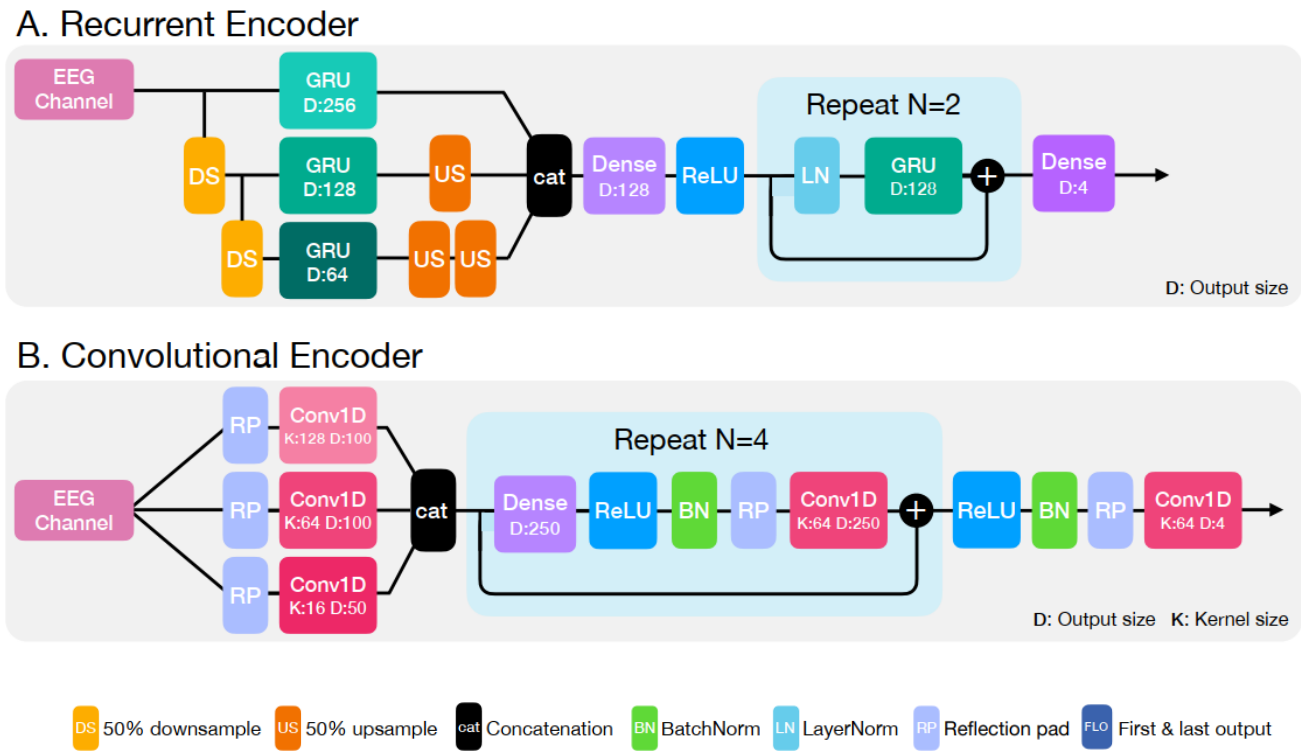
- **Channel Augmenter**
  - Randomly transforms a mini-batch of *N* channels into 2*N* augmented channels.
  - Randomly applies two augmentations to each channel, resulting in a positive pair.



Channel encoder:

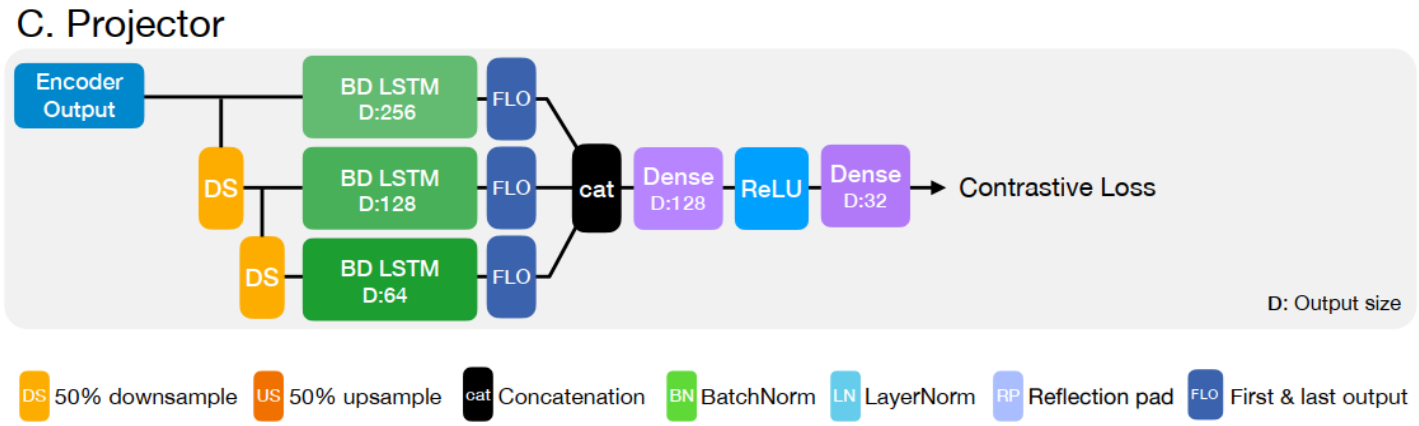
- Transforms an input channel into four feature channels, maintaining the same length for each.
- This feature allows encoding of sequences of varying lengths suitable for different downstream tasks.

- **Encoder Architectures**
  - **Recurrent Encoder**:
    - multi-scale input approach
    - employing downsampling and upsampling to allow GRU units to capture features at various time scales.
    - includes two recurrent residual units.
  - **Convolutional Encoder**:
    - Employs reflection padding to match the kernel size of the convolution layers, ensuring the output length matches the input length.
    - includes four convolutional residual units.



Projector:

- Transform the output of the encoder into a 32-dimensional point
- Uses downsampling and bidirectional LSTM units
- The final outputs of each direction are concatenated and fed into dense layers with a ReLU activation in between.



Contrastive loss:

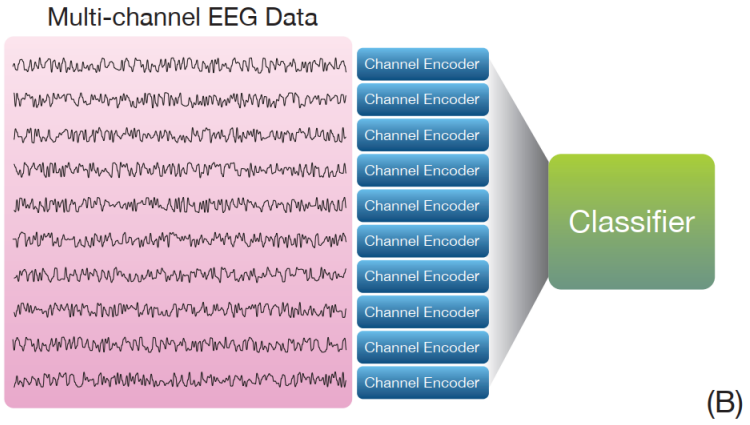
- Assuming that z_i and z_j are the outputs of the projector for the positive pair of x_i and x_j, the NT-Xent loss term for the positive pair is defined as:

$$\ell_{i,j} = -\log \frac{\exp(\text{sim}(\boldsymbol{z}_i, \boldsymbol{z}_j)/\tau)}{\sum_{k \neq i}^{2N} \exp(\text{sim}(\boldsymbol{z}_i, \boldsymbol{z}_k)/\tau)}$$

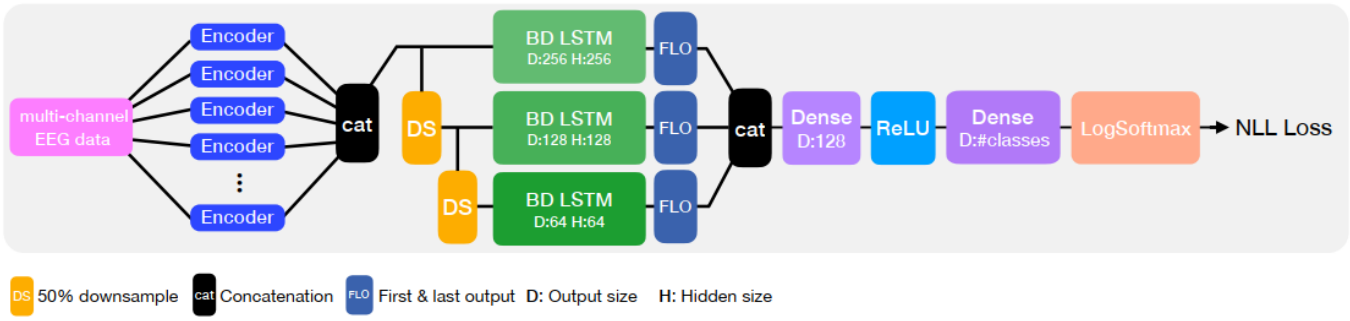
- where sim(***u***; ***v***) is the cosine similarity of ***u*** and ***v*** and τ is the temperature parameter

Classifier:

- For downstream classification tasks, the projector is discarded, and a classifier almost identical to the projector with two differences is used:
  - (1) the output dimension of the last dense layer is set to the number of classes
  - (2) a LogSoftmax layer is added afterward.
- The input is the concatenation of the output of the encoder for all input channels of a multi-channel recording.



(B)



Results on emotion recognition:

- Conducted experiments on SEED dataset
- EEG data of 15 subjects recorded in 62 channels.
- The data was recorded when participants watched emotional videos chosen from movies in three categories of emotions, namely *negative*, *neutral,* and *positive*.
- Pass each channel through the encoder and concatenate the 4-dimensional output sequence.
- The input of the classifier is a 4 * 62-dimensional sequence of length 200.
- Table 2 shows the results of the experiments. The rows marked with SeqCLR-C (convolutional) and SeqCLR-R (recurrent) show the results without fine-tuning where the encoder parameters were frozen during training.
- The proposed method improves other self-supervised algorithms by a large gap. Moreover, when fine-tuned on the entire dataset, SeqCLR achieves 85.77% accuracy, slightly higher than the current state-of-the-art supervised model (BiHDM).

Table 2: Emotion recognition on SEED

Model	Accuracy			
Percentage of labels	1%	10%	50%	100%
RGNN	-	-	-	85.30
BiHDM	-	-	-	85.40
CPC	69.17	76.33	79.98	81.12
RP	67.76	74.29	77.95	80.39
TS	69.73	78.27	81.66	82.10
SeqCLR - C	77.09	81.01	83.73	84.11
SeqCLR - R	76.52	79.04	81.45	83.78
fine-tuned SeqCLR - C	**79.04**	**83.12**	**85.21**	**85.77**
fine-tuned SeqCLR - R	78.18	82.93	84.00	85.25

Ablation study of channel recombination and dataset fusion

- we used channel recombination (CR) and dataset fusion (DF) to obtain a larger training set for self-supervised learning.
- The table shows the effect of removing each of these steps on the accuracy of the classifiers without fine-tuning.
- In particular removing channel recombination had a stronger effect in all three tasks.

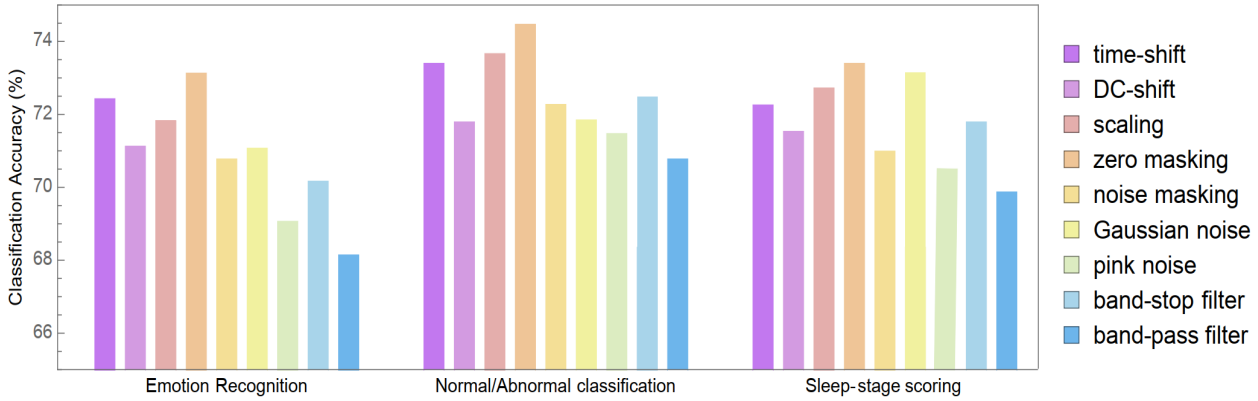
Table 5: Ablation study of CR and DF

Channel recombination	Dataset fusion	SEED	TUH	SleepEDF
✗	✗	78.93	79.12	77.72
✓	✗	83.01	83.78	81.10
✗	✓	80.23	83.44	79.59
✓	✓	**84.11**	**86.27**	**83.05**

Choosing effective augmentations:

- Set up a classification task with the convolutional SeqCLR architecture,
  - only using a single augmentation at a time.
  - trained nine encoders and tested them on the three classification tasks.
  - froze the encoder parameters, for training the classifiers

- We observed that the six augmentations, namely (1) zero-masking, (2) amplitude scaling, (3) time-shift, (4) Gaussian noise, (5) DC-shift, and (6) band-stop filter perform significantly better in extracting useful features for the downstream tasks.



Ablation study of augmentations:

- masking and scaling are the most effective augmentations across the three classification tasks.
- Additive noise and DC shift have the least effect on the performance of the classifiers

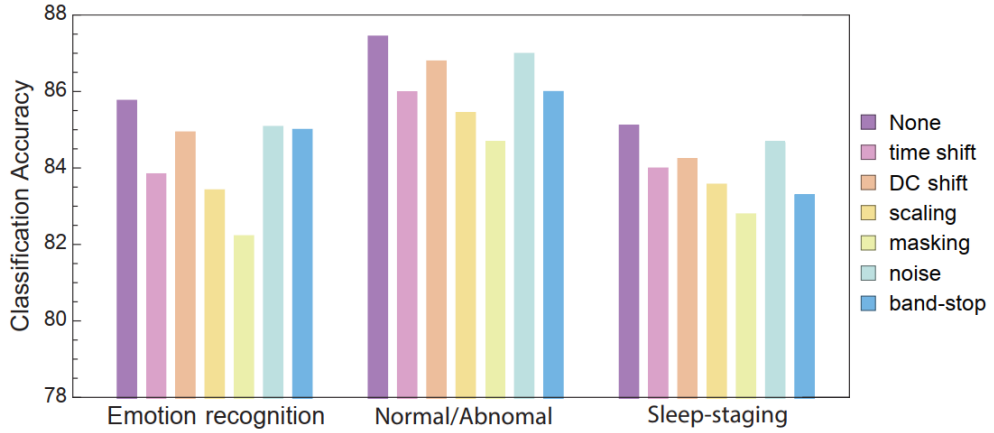


Figure 5: Ablation study of augmentations. Each bar shows the accuracy of the classifier when that augmentation is removed.

Conclusion

- Presented a self-supervised framework, SeqCLR, for learning EEG signal representations.
- Adapted the SimCLR framework to time-series data to boost sample-efficiency and classification accuracy across three specific tasks:
  - Emotion recognition on the SEED dataset
  - Normal/abnormal classification on the TUH dataset
  - Sleep-stage scoring on the SleepEDF dataset
- Achieved improved performance over other baseline self-supervised models and, with fine-tuning, surpassed current state-of-the-art supervised models in emotion recognition and sleep staging.
- Implemented six augmentations, identifying masking and scaling as particularly critical for feature extraction in downstream tasks.
- Demonstrated that self-supervised and contrastive learning is effective for deriving valuable representations from EEG data.