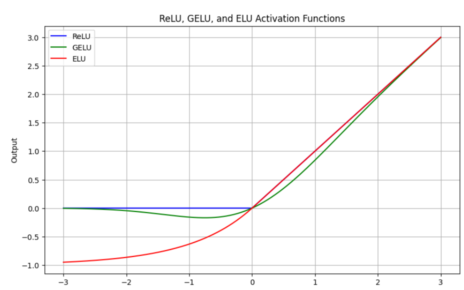
IPR PROJECT PART2 REPORT ERTNet: an interpretable transformer-based framework for EEG emotion recognition

Devansh Ojha, 210322

November 6, 2024

Changing the Activating Function

I propose to change the activation function from **ELU** (Exponential Linear Unit) to **GELU** (Gaussian Error Linear Unit) proposed by [1] where was shown to be superior to ELU and **RELU** (Rectified Linear Unit) on a number of tasks like computer vision, natural language processing, and speech tasks which are also temporal data as the EEG data in our case.



For GELU : $\mu = 0$, $\sigma = 1$
For ELU: $\alpha = 1$

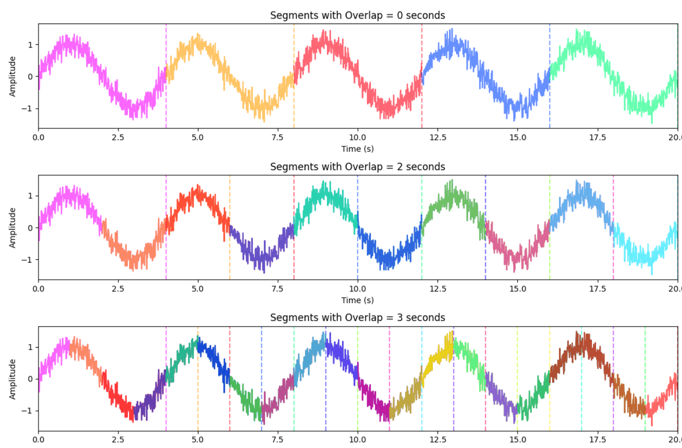
Dilation Convolutions

I propose to add dilation steps to the **Separable 2D Convolution**, which is the convolution before the positional encoding for transformer block. This is because Dilated CNN networks have a larger receptive field than regular CNN [2], without compromising on resolution, coverage and parameter numbers rising.
We add this to the Separable Convolution layer to because it is the main feature extraction module that uses the activations of the earlier spatio-temporal convolution layers. It would benifit from the increased receptive field to observe

long spatio-temporal relations and may allow us to use larger time samples to train the model (instead of the 4s segments proposed by the paper).
We may explore dilated convolution for the initial temporal and spatial convolutions, but EEG data is locally variable, and thus it might not be useful to go for dilated convolution in the initial layers.

Overlapping Segments

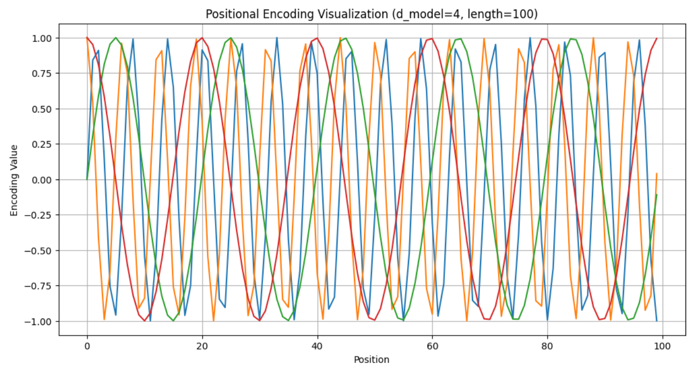
The paper directed us to segment the data into 4s non overlapping segments, but I will try to get some overlapping while creating the segments. Doing overlapping segments will increase the dataset size, but it might also reveal some new EEG contiguous segments that will help the models generalize better.
We have to choose the overlap size carefully because a large overlap will create very huge dataset that may lead to overfitting. On the other hand too small overlap (or no overlap) will not give rise to any new kinds of segments,which restricts us from the advantage of the overalapping segmnets itself.



Segment overlapps and new patterns captured in dataset

Learnable Positional Encoding

The paper is using fixed sinusoidal positional encoding, which does not adjust to the EEG data. It is universal and is very generalized, but we may benefit from a positional encoding that allows for some adjustment and learns the temporal relation in a sample.



Positional Encoding for some of the channels (as proposed by the paper)

Learnable positional encoding might help us achieve better in our specific application, by learning a positional encoding that is best for our case.
We might want to look into **Relative Positional Encodings**, but they might be too complex and computationally expensive for our purpose.
Generally learnable positional encoding requires a bit more training data to genralize, here's were the overlapping in segmentations might help us. We may have to balance the effect of this learnable positional encodings also with the overlap size that affects the number of training and testing cases.

References

[1] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus), 2023.

[2] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions, 2016.