IPR PROJECT PART2: Implementation Details and results ERTNet: an interpretable transformer-based framework for EEG emotion recognition

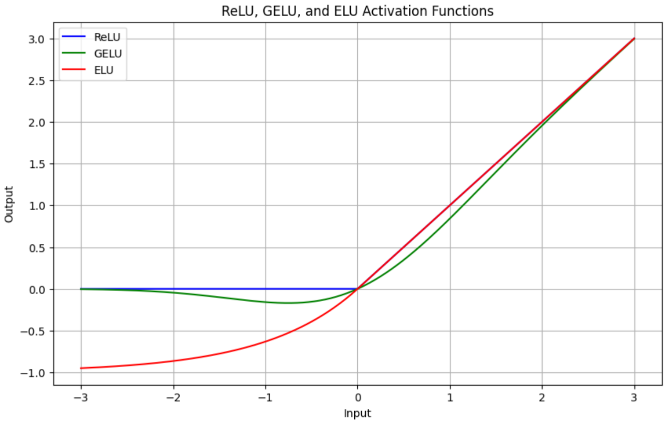
Devansh Ojha, 210322

November 6, 2024

Implementation Details

Changing the Activating Function and the Convolution Type

I changed the activation function from ELU (Exponential Linear Unit) to GELU (Gaussian Error Linear Unit) as proposed by [1], which has been shown to outperform both ELU and ReLU (Rectified Linear Unit) across various tasks, including computer vision, natural language processing, and speech tasks—areas similar to our temporal EEG data.



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

I also incorporated dilation steps into the Separable 2D Convolution, which is the convolution applied before the positional encoding in the transformer block. This adjustment leverages the larger receptive field of dilated CNN networks, as discussed in [2]. For this i just had to update the dilation rate for the Separable Convolution Layer to (2,2), which had the default value (1,1) corresponding to no dilation.

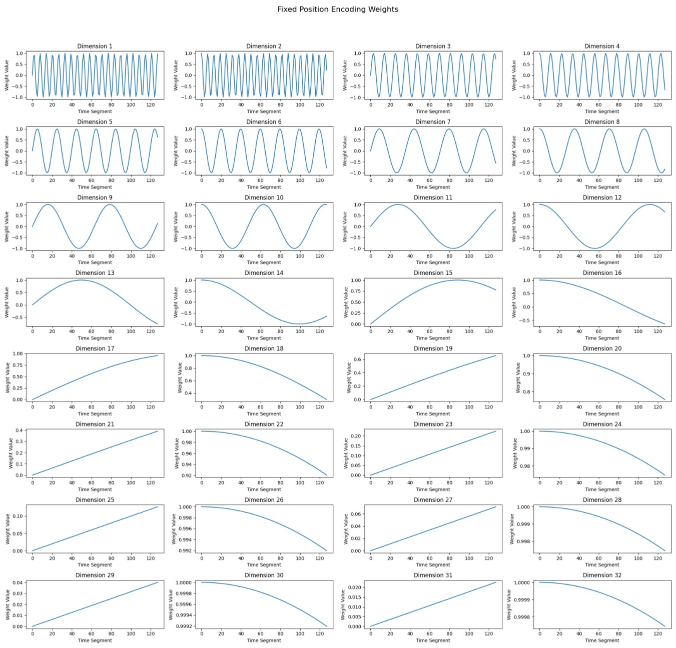
Overlapping Segments

The paper directed us to segment the data into 4s (Sample Rate = 128Hz) non overlapping segments, but I will try to get some overlapping while creating the segments. I added a function to create 4s segments with 25% overlap, i.e., as the length of each segment is 512 they have the overlap of 128. For this I made the set the step size = 512-128 = 384. This made the training data of size 25600, which was earlier of size 19200. So there is a 33% increase in the size of the training data.

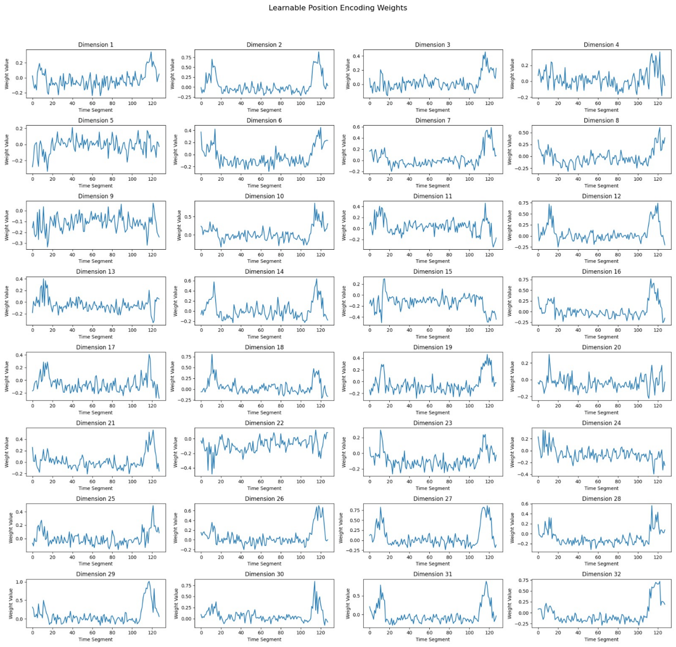
This helps us generate new unique sequences which will better generalize the model and will also help with the later part of Learnable Positional Encoding.

Learnable Positional Encoding

I made the Learnable Positional Encoding as a learnable matrix of size (F_{dim}, max_len), where F_{dim} is the number of output channels from the Separable Convolution Network (32 here) and max_len is the maximum length of input, here it was set as 500 as in the original paper. The positional encoding are just added to the output of the SeparableConvolutionLayer output.



Positional Encoding(as used by the paper)

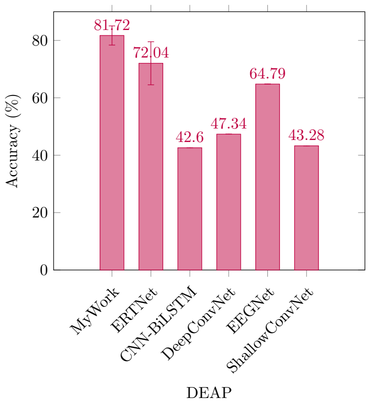


Learnable Positional Encoding(as learned during the training only first 128 values)

Result

Subject Dependent Training Results

There is around 9% accuracy gain in the subject dependent training.



Accuracy of Subject-Dependent DEAP

Subject Independent Training Results

The accuracy of the model on Subject Independent Training did not change much with the new model.

$$ERTNetAccuracy = 33.4 + -(7.4)\%$$
$$NewModelAccuracy = 32.12 + -(3.95)\%$$

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.