Deep Learning, SUMMER 2022
LAB2 EGG Classification

K.S. Chen

July 22, 2022

1 Introduction

Electroencephalography (EEG) is a method to record an electrogram of the electrical activity of a brain, while BCI Competition datasets are used to validate classification methods for Brain-Computer Interfaces.

In this lab, we implement two EEG classification models, EEGNet and DeepConvNet, with non-stationary 2-class BCI data from BCI competition dataset III. Furthermore, the performance of each model pairing with different kinds of activation function, including ReLU, Leaky ReLU and ELU is both tested and visualized.

2 Experiment set up

2.1 The detail of your model

2.1.1 EEGNet

EEGNet is a compact convolutional neural network for EEG-based BCIs that can effectively generalizes across different BCI paradigms. It includes the depthwise-separable convolutions in the architecture, which considerably reduce the total number of parameters and multiplications. In addition, EEGNet can perform as well as a more paradigm-specific EEG CNN model, but with two orders of magnitude fewer parameters to fit.[1]

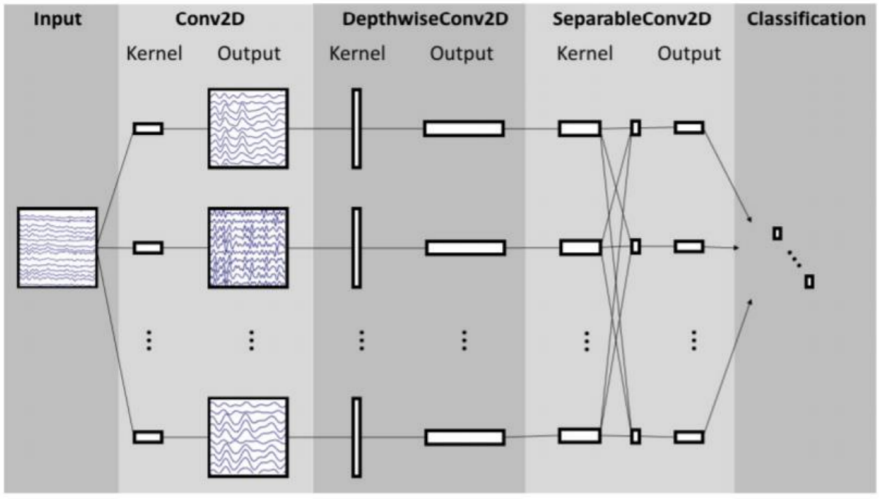

Figure 1: The architecture of EEGNet.

2.1.2 DeepConvNet

DeepConvNet is one of the existing CNN-based models that are used to compare with EEGNet in the original paper[2]. DeepConvNet has ten times more parameters than EEGNet. From the original paper, DeepConvNet was outperformed by EEGNet in three out of four 4-fold with-subject classification performance and had no significant different in the remaining one. https://www.overleaf.com/project/62d8c5ee12eba720fd951a

Layer	# filters	size	# params	Activation	Options
Input		(C, T)			
Reshape		(1, C, T)			
Conv2D	25	(1, 5)	150	Linear	mode = valid, max norm = 2
Conv2D	25	(C, 1)	25 * 25 * C + 25	Linear	mode = valid, max norm = 2
BatchNorm			2 * 25		epsilon = 1e-05, momentum = 0.1
Activation				ELU	
MaxPool2D		(1, 2)			
Dropout					p = 0.5
Conv2D	50	(1, 5)	25 * 50 * C + 50	Linear	mode = valid, max norm = 2
BatchNorm			2 * 50		epsilon = 1e-05, momentum = 0.1
Activation				ELU	
MaxPool2D		(1, 2)			
Dropout					p = 0.5
Conv2D	100	(1, 5)	50 * 100 * C + 100	Linear	mode = valid, max norm = 2
BatchNorm			2 * 100		epsilon = 1e-05, momentum = 0.1
Activation				ELU	
MaxPool2D		(1, 2)			
Dropout					p = 0.5
Conv2D	200	(1, 5)	100 * 200 * C + 200	Linear	mode = valid, max norm = 2
BatchNorm			2 * 200		epsilon = 1e-05, momentum = 0.1
Activation				ELU	
MaxPool2D		(1, 2)			
Dropout					p = 0.5
Flatten					
Dense	N			softmax	max norm = 0.5

Figure 2: The architecture of DeepConvNet.

2.2 Explain the activation function (ReLU, Leaky ReLU, ELU)

1. ReLU (Rectified Linear Unit):

   ReLU is the most commonly used activation function in neural networks. It turns every negative value to zero, while the positive values unchanged. Since ReLU function is defined as max(x,0), the computational cost is extremely low. Furthermore, it is sparsely activated, therefore, some useless units might not be activated at all. Nonetheless,the dying ReLU problem happens when too many units are stucked on the negative sides and always return zeros.

2. Leaky ReLU (Leaky Rectified Linear Unit):

   Leaky ReLU is a varient of ReLU. Instead of setting all negative values to zeros, it times 0.01 and keep the values. By doing so, the dying ReLU problem would be resolved.

3. ELU (Exponential Linear Unit):

   ELU is also a varient of ReLU. It uses exponential curves rather than linear curves on negative values. As a result, it can avoid the dying ReLU problem and saturates for large negative values at the same time.

3 Experimental results

3.1 The highest testing accuracy

1. Screenshots of two models

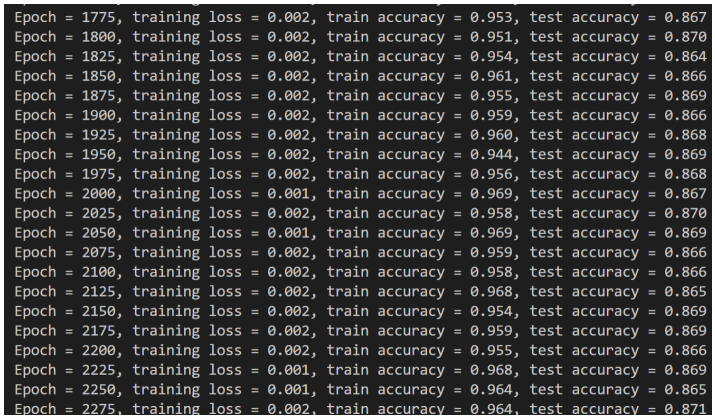

Figure 3: EEG: Highest test accuracy (Leaky ReLU).

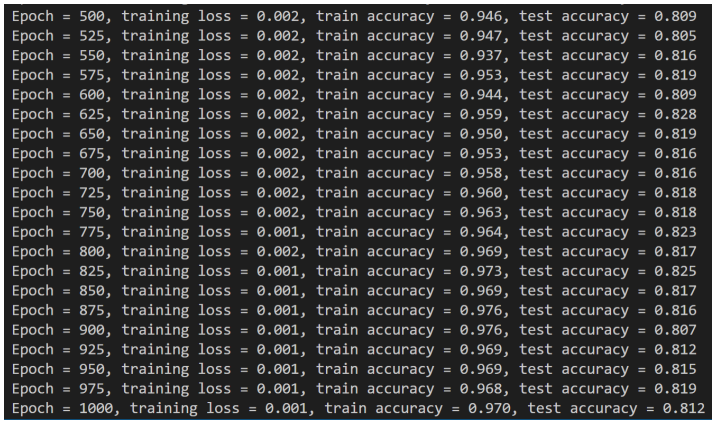

Figure 4: DeepConv: Highest test accuracy (ReLU).

Model	ReLU	Leaky ReLU	ELU
EEGNet	0.853	0.854	0.816
DeepConvNet	0.822	0.808	0.816

Table 1: Test accuracy for 300 epochs.

2. Anyting you want to present

   Setting a random seed for torch is the most important thing for result reconstructions. With the same set of parameters, the test accuracy can somehow range from 80% to more than 87%. Also, note that in this experiment ReLU and Leaky ReLU achieve higher accuracy in general, which is interesting because I always consider ELU as a compromise of both of them. ELU resolves the leaky ReLU problem of ReLU and saturates the large negative value for Leaky ReLU.

3.2 Comparing figures

1. EEGNet
2. DeepConvNet

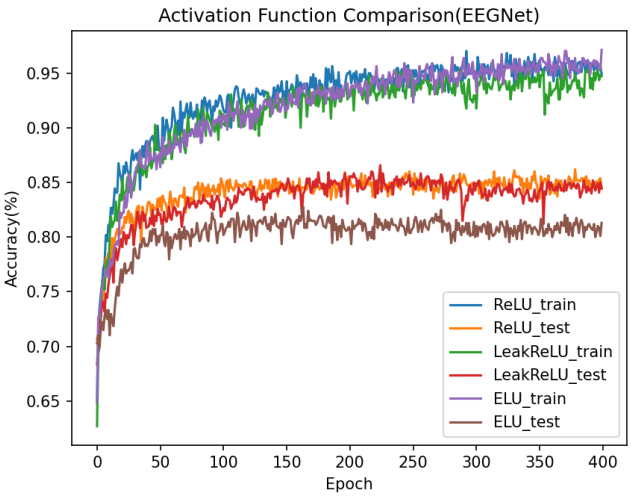

Figure 5: Activation Function Comparison (EEGNet).

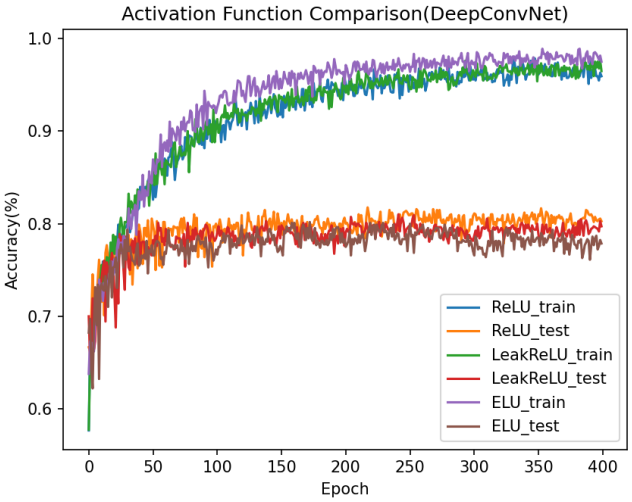

Figure 6: Activation Function Comparison (DeepConvNet).

4 Discussion

I spend my fair share of time studying separable convolutions for this lab. It turns out there are two common approaches, the mathematical approach and the physical approach.

The mathematical approach, spatial separable convolution, simply decomposes the kernel matrix (For example: n*n) into two smaller matrix (For example: n*1 and 1*n). The main problem is that not every matrix can be decomposed.

The physical approach, depthwise separable convolution, focuses on the nature of the network architecture. It first done the convolution separably for each input layer to narrow down the width and height dimensions, while keeping the depth unchanged. Then, extend the depth of the output to the desired dimension by applying many 1*1 kernels, which is called pointwise convolution.

To sum up, both the mathematical and physical approaches are able to effectively reduce the number of parameters of the model. However, due to the technical feasibility, the latter is more prevalent.