DLP Lab2 EEG classification

0711529 陳冠儒

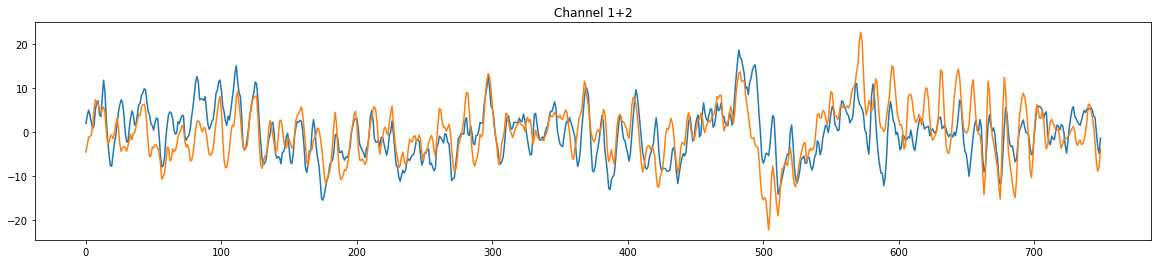
1. Introduction (20%)

使用EEGNet跟DeepConvNet搭配三種不同的activation function (ReLU、LeakyReLU、ELU)，去比較acc的影響。

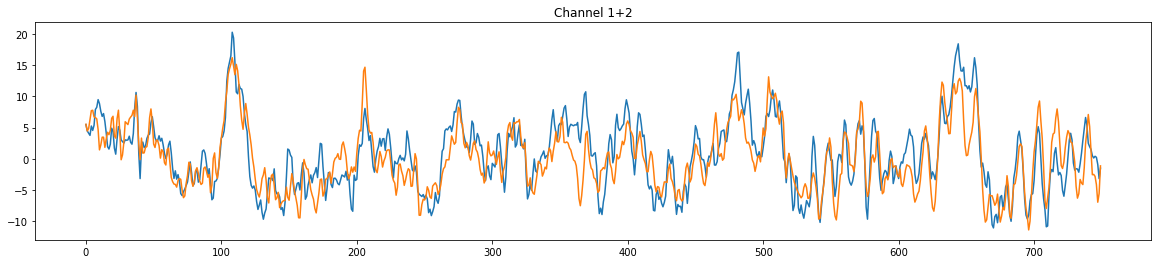
1. Dataset：

這次使用的是BCI Competition III的IIIb dataset，為一個用125 Hz採樣的ERP (Event-Related Potential) EEG dataset，他是一個non-stationary 的classifier (e.g., time-varying)，並用兩個channel的EEG圖來預測是左邊還是右邊。

* Training and Testing data
* X\_train：shape (1080, 1, 2, 750)，1080筆資料，兩個channel，一個channel 750時間單位的訊號。
* y\_train：shape (1080, )，1080筆資料，每個資料為0或是1代表左邊或是右邊。
* X\_test：shape (1080, 1, 2, 750)
* y\_test：shape(1080, )
* Data example
* Label = 0



* Label = 1



* Reference

[1] Dataset IIIb: Non-stationary 2-class BCI data: <http://www.bbci.de/competition/iii/desc_IIIb.pdf>

1. EEGNet的優點

在EEGNet的論文中提到了EEGNet的三個優點：

1. 可以應用於多種不同的 BCI 模式

在paper中他總共使用了四種不同的EEG dataset來進行測試，包含了最主要兩種BCI的資料種類：ERP (Event-Related Potential)和Oscillatory，在這四種上都能達到不錯的成效，且每一個都包含不同數量的數據，使我們能夠探索 EEGNet 在各種訓練數據大小上的功效。

1. 可以使用非常有限的數據進行訓練

使用 Depthwise 和 Separable Convolutions來構建，讓可訓練參數大大的減少，故只要小的dataset就可以訓練起來。

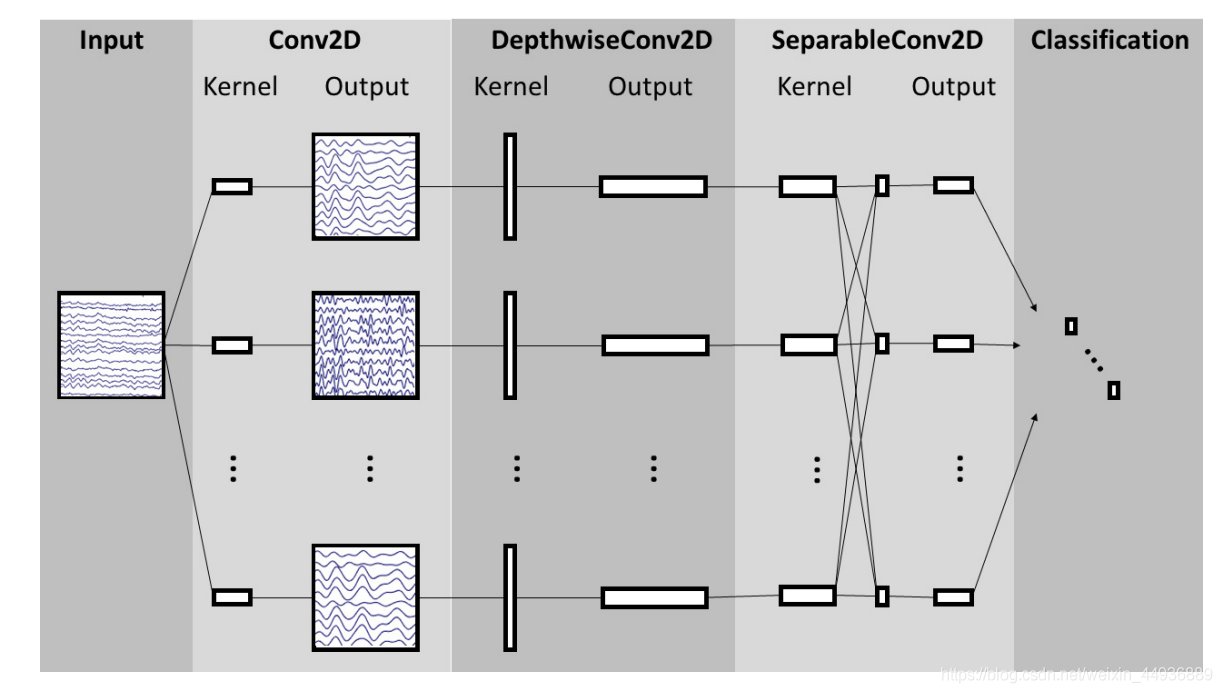
1. 可以產生神經生理學上可解釋的特徵

儘管 CNN 具有強大的自動特徵提取能力，但通常會產生難以解釋的特徵。對於神經科學家來說，深入了解 CNN 衍生的神經生理現象的能力可能與實現良好的分類性能同樣重要。paper中驗證了EEGNet架構在幾個經過充分研究的 BCI 模式上提取神經生理學可解釋信號的能力，以表明網絡性能不是由數據中的noies或人工信號驅動。

1. Experiment setups(30%):
   1. The detail of your model

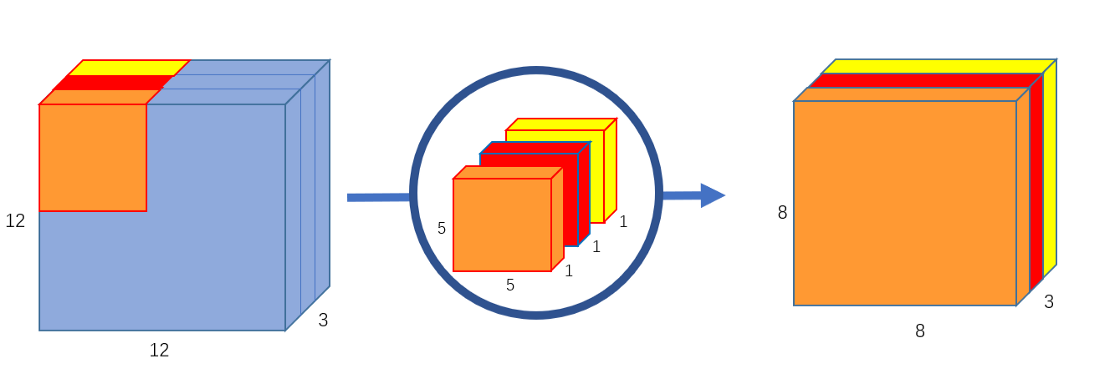
* EEGNet

EEGNet可以主要可以分成三層：Conv2D、DepthwiseConv2D、SeparableConv2D，下面將會對這三層提取的feature進行分析。



* + - * Conv2D：利用時間卷積(temporal convolution)學習頻率過濾器(frequency filters)，得到的各個feature map對應到各特定頻率。
      * DepthwiseConv2D：使用深度卷積(depthwise convolution)，單獨連接到每個feature map，以學習特定於頻率的空間過濾器(frequency-specific spatial filters)，而這裡的depth parameter D是表要為每個feature map學習的spatial filters數。

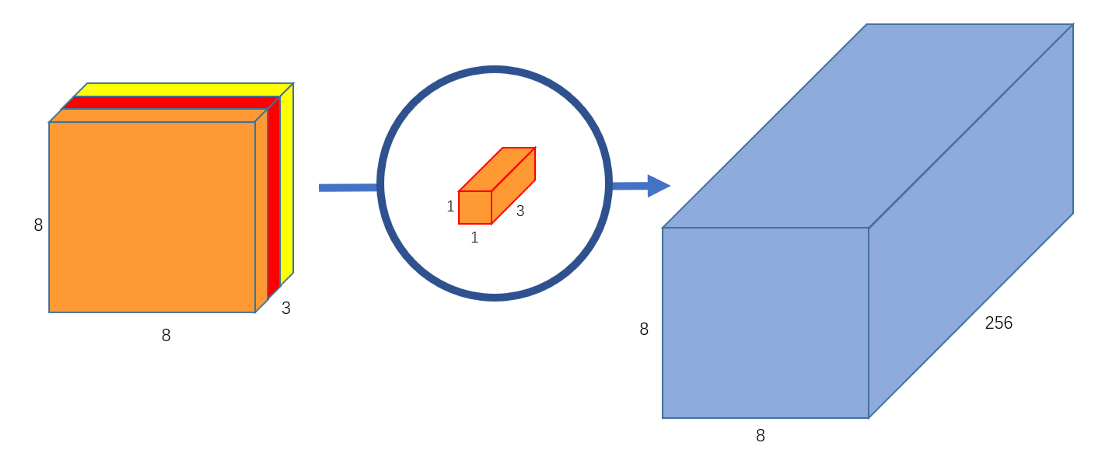
※ DepthwiseConv2D



將input的各個channel單獨學習，一個channel會對應到一個(D=1)或多個(D>1)kernel進行convolution。

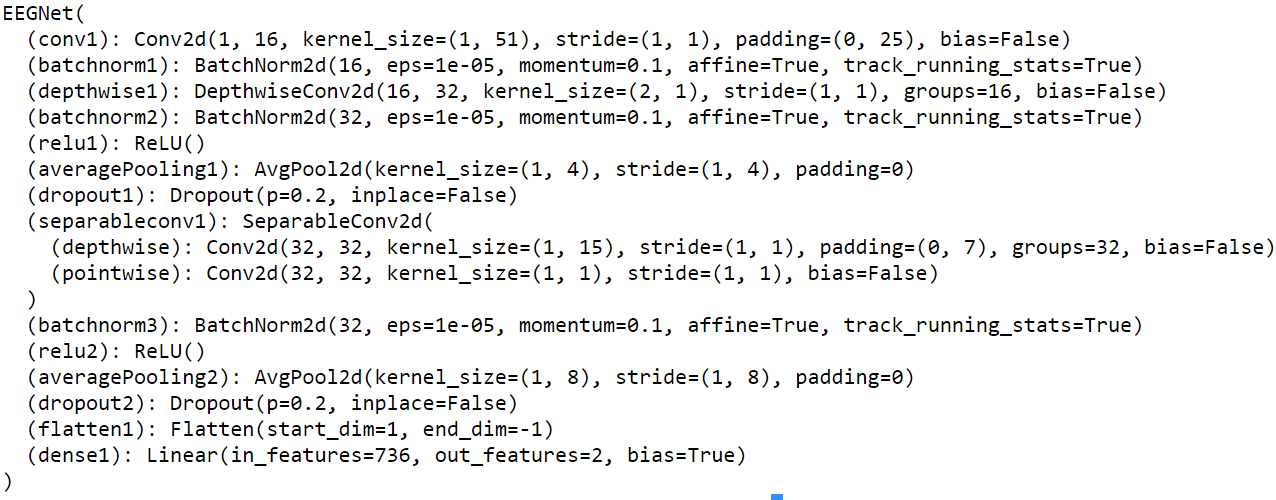
* + - * SeparableConv2D：可分離卷積(separable convolution)是深度卷積與逐點捲積的組合，其中的深度捲積學習每個feature map的時間摘要，而最後再由逐點捲積(pointwise convolution)學習如何將feature map最佳的組合在一起。

※ PointwiseConv2D

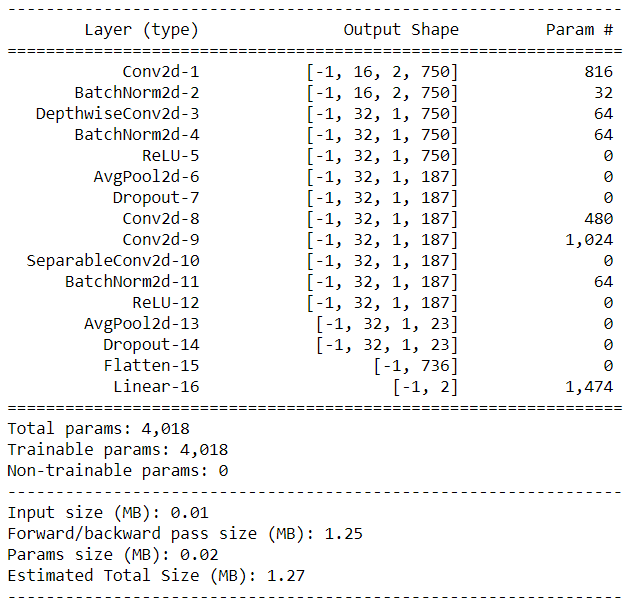


使用1x1 kernel增加image的depth，即增加channel的數量。

* + Model Neural Network



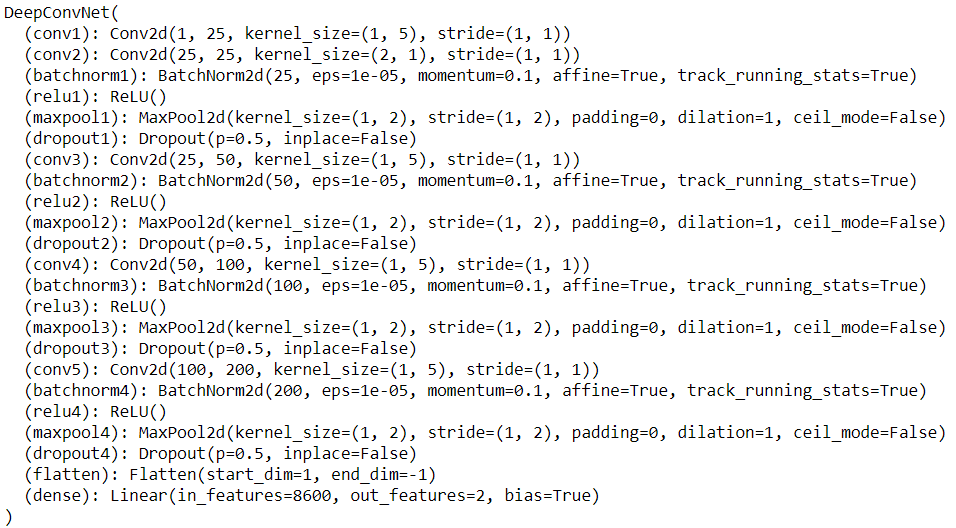
* + Model layer shape and parameters with input shape (-1, 1, 2, 750)



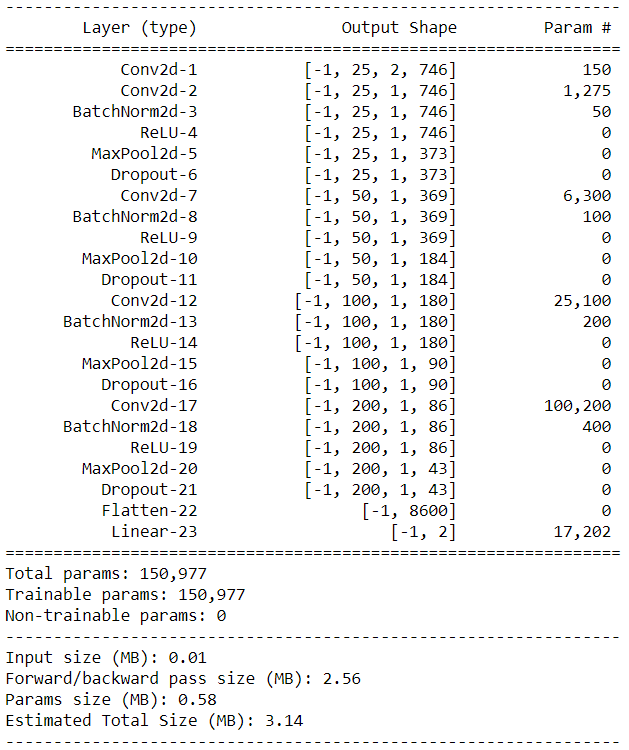
* DeepConvNet

DeepConvNet 架構由五個卷積層和一個用於分類的 softmax 層組成。

* + Model Neural Network



* + Model layer shape and parameters with input shape (-1, 1, 2, 750)



* 1. Explain the activation function (ReLU, Leaky ReLU, ELU)
     + ReLU

|  |  |  |
| --- | --- | --- |
| Function |  | Derivative |
|  |  |  |
| _images/relu.png |  | _images/relu_prime.png |

在0處不可微，但可以直接設為0或是1，ReLU是目前最被廣泛使用的activation function。

* + - * 優
        1. ReLU在的微分值為1，所以他沒有gradient vanishing的問題（gradient vanishing是因為gradient ，梯度更新後的結果將會越來小，最後趨近於0）。
        2. ReLU在的微分值為1，可以降低gradient explosion的問題（gradient explosion是因為gradient而使得梯度更新後其值越來越大，最後超過可計算範圍）。
        3. 沒有複雜的計算，所以計算成本低，很快就能收斂。
      * 缺

1. 「dying ReLU problem」，在處梯度為0，因此在下降期間不會調整權重，在此狀態的神經元將停止響應錯誤和輸入的變化。
2. ReLU只是降低gradient explosion但因為其範圍是[0, inf)，所以還是可能會發生。
   * + Leaky ReLU

|  |  |  |
| --- | --- | --- |
| Function |  | Derivative |
|  |  |  |
|  |  |  |

在負值時有一由控制的洩漏量，通常是設為0.01，並且在0的地方也不可微，通常也會直接設為1或。

* + - * 優
        1. 繼承了ReLU的所有優點。
        2. Leaky ReLU在負值時的洩漏值解決了dying ReLU problem。
        3. 在《Empirical Evaluation of Rectified Activations in Convolution Network》這篇研究中發現Leaky ReLU的性能優於 ReLU。此外，洩漏值較高的 Leaky ReLU的性能優於洩漏值較低的。（但是在不同的model上主要還是要看實驗結果來決定）
      * 缺
        1. α值是在訓練之前定義的，因此在訓練期間無法調整，因此選擇的α值可能不是最佳值。
      * Reference

[1] Empirical Evaluation of Rectified Activations in Convolution Network: https://arxiv.org/pdf/1505.00853.pdf

* + - ELU (Exponential Linear Unit)

|  |  |  |
| --- | --- | --- |
| Function |  | Derivative |
|  |  |  |
| _images/elu.png |  | _images/elu_prime.png |

與Leaky ReLU在負值都有參數控制洩漏值大小，但Leaky ReLU在0的地方不可微，而ELU在所有地方都是連續可微的。

* + - * 優
        1. 繼承所有Leaky ReLU的優點，且微分後仍為non-linear function。
      * 缺
        1. 在function中有exponential，造成其計算成本比ReLU與Leaky ReLU還大，因此收歛較慢。
        2. α值也是在訓練之前定義的，因此在訓練期間無法調整，因此選擇的α值可能不是最佳值。
    - Reference

[1] Activation Functions — ML Glossary documentation: https://ml-cheatsheet.readthedocs.io/en/latest/activation\_functions.html

1. Experimental results (30%)
   1. The highest testing accuracy
      * Screenshot with two models

**Hyper parameters：**

- Optimizer: Adam

- Loss function: torch.nn.CrossEntropyLoss()

- epoch of EEGNet: 1000, epoch of DeepConvNet: 2000

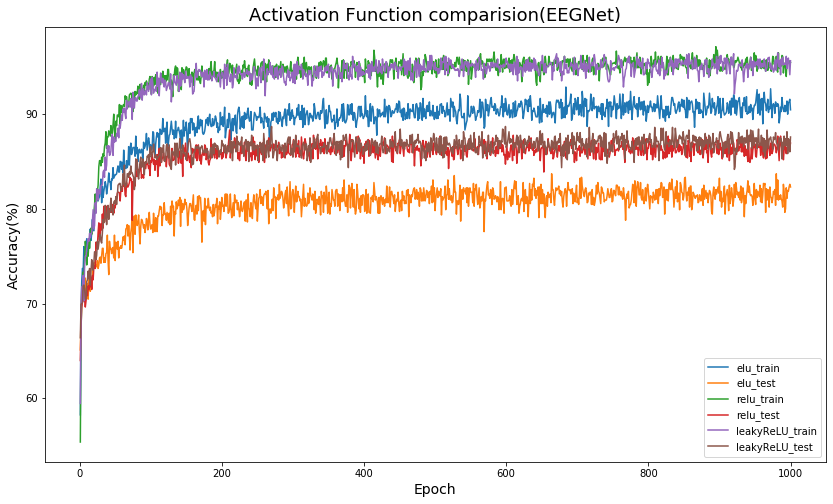
- batch: 128

- learning rate: 1e-3

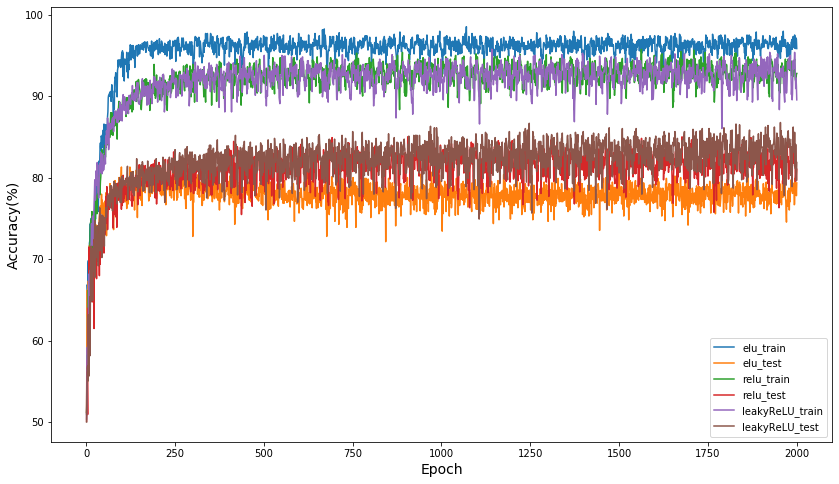
- weight decay: 0.02

|  |  |  |  |
| --- | --- | --- | --- |
|  | ReLU | LeakyReLU | ELU |
| EEGNet | 89.074074% | 88.240741% | 82.685185% |
| DeepConvNet | 83.703704% | 83.425926% | 82.129630% |

* 1. Comparison figures
     + EEGNet



Testing acc: ReLU > Leaky ReLU > ELU

* + - DeepConvNet  
        
      Testing acc: ReLU > Leaky ReLU > ELU

1. Discussion (20%)
   1. Anything you want to share
      * 為什麼在這個Lab使用DeepConvNet與EEGNet進行比較而不是ShallowConvNet？  
        因為DeepConvNet 是一種通用架構，不限於特定的特徵類型，因此它可以作為與 EEGNet 的更有效的比較；而ShallowConvNet 架構是專門為振盪信號分類而設計的（通過提取與對數帶功率相關的特徵）；因此，它可能不適用於基於 ERP (此dataset)的分類任務。

* DeepConvNet與EEGNet比較結果

論文[1]中說DeepConvNet 和 EEGNet 的分類性能在所有cross-subject分析中都相似，而 DeepConvNet 在幾乎所有within-subject分析中的性能都較低。這種差異的一種可能解釋是用於訓練模型的訓練數據量；在cross-subject分析中，訓練集大小大約是within-subject分析的 10-15 倍。這表明與 EEGNet 相比，Deep-ConvNet 的數據密集程度更高，鑑於 DeepConvNet 的模型大小比 EEGNet 大兩個數量級，這一結果並不令人意外。

從本文第二點的Detail of Model可以得知本次實驗用的EEGNet和DeepConvNet的可訓練參數分別為4018跟150977，可以看出這兩個網路大小相差很大，且這次的dataset數據量也不算大，因此得到EEGNet都優於DeepConvNet的結果是不意外的。

* ReLU、Leaky ReLU與ELU結果比較

實驗結果不論是在EEGNet或是DeepConvNet的testing accuracy都是ReLU >= LeakyReLU > ELU，我覺得是因為parameter與資料集大小的關係，這個結果大小順序正好與trainable parameter量相反，因為這次的資料集大小不算大，所以越多的parameter會越難訓練起來。

* + - Reference

[1] EEGNet: A Compact Convolutional Neural Network for EEG-based Brain-Computer Interfaces: <https://arxiv.org/pdf/1611.08024.pdf>

[2] Deep Learning With Convolutional NeuralNetworks for EEG Decoding and Visualization: <https://onlinelibrary.wiley.com/doi/epdf/10.1002/hbm.23730>