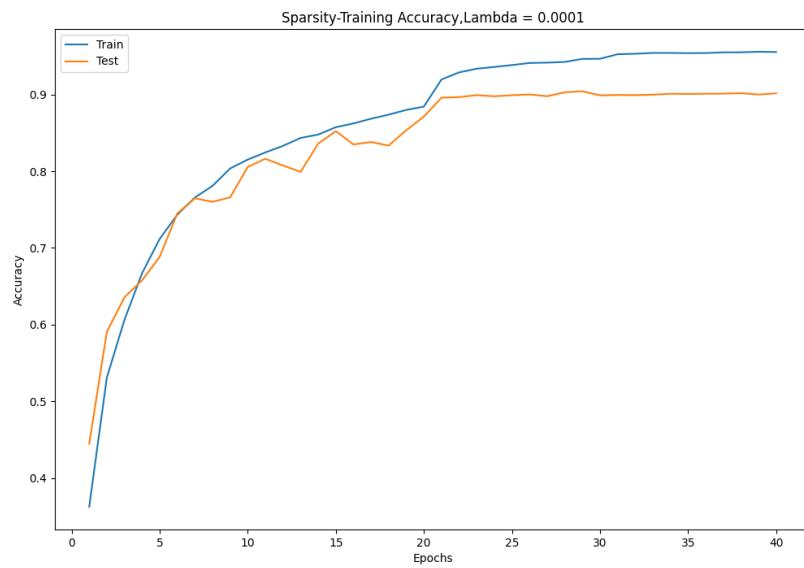


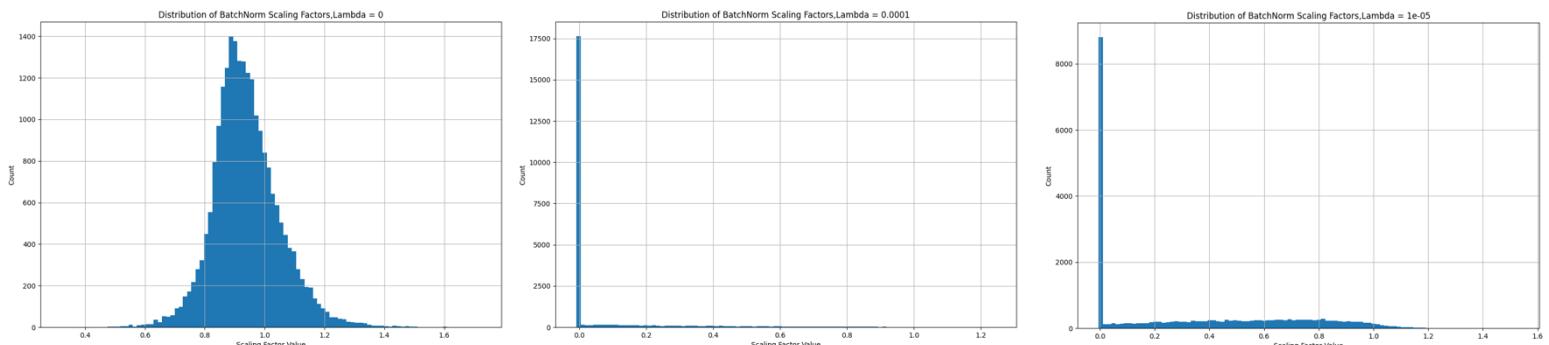
# EAILab 2 Report

學號：NM6131051 姓名：張芷晴

- Plot train accuracy and test accuracy of the original model in sparsity-training over epochs (Only  $\lambda = 1e-4$  are needed.)(5%)



- Plot scaling factor distribution with 3 different  $\lambda$  value (5%)



- Show the model test accuracy after pruning 50% and 90% channels (5%)

```

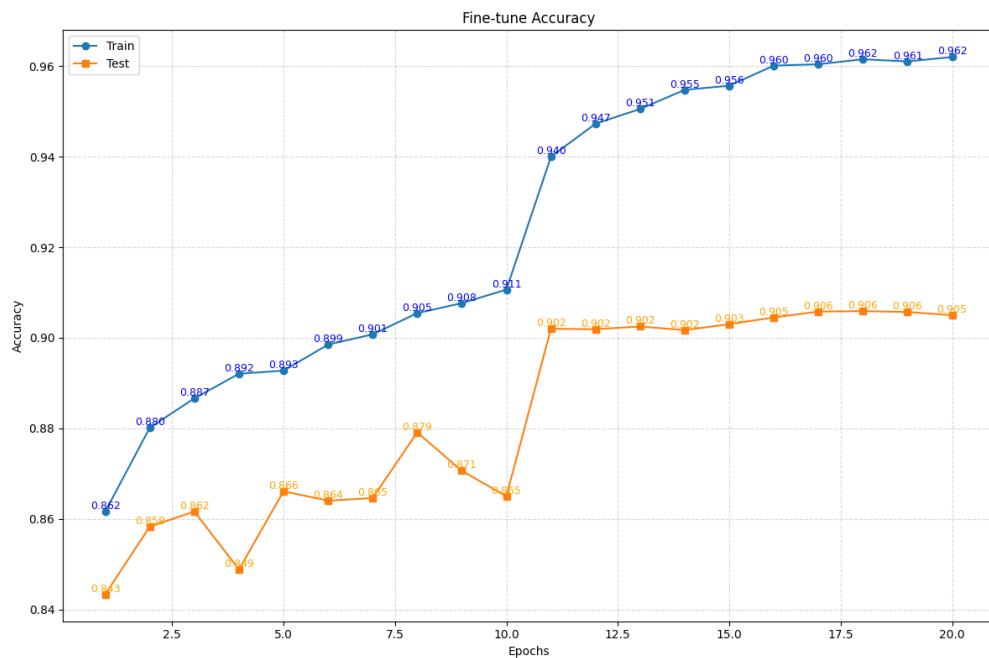
    (TC) > linear(in_features=2048, out_features=10)
)

Test set: Accuracy: 9040/10000 (90.4%) test accuracy after pruning 50%
tensor(0.9040)

Test set: Accuracy: 4725/10000 (47.2%) test accuracy after pruning 90%
tensor(0.4725)

```

4. Plot train accuracy and test accuracy after fine-tuning the pruned 90% model over epochs (5%)



5. Show and compare the test accuracy and model parameters among the original and fine-tuned models (5%)

|                  | Original model | Prune 90% model | Prune 50% model |
|------------------|----------------|-----------------|-----------------|
| Test accuracy    | 0.91           | 0.906           | 0.909           |
| model parameters | 23513162       | 3729889         | 9691903         |

6. Explain how you modified resnet.py (5%)

在 Bottleneck Module 中原本的設計是每層的輸入、輸出通道是固定比例，

調整後改為 `out_channels` (長度為 3 的 list，像是`[64, 64, 256]`) 來控制每個 `conv` 層的通道，因此每個 `bottleneck` 可以有不同的結構。

在 `make_layer` 中，每個 layer 會依據 `cfg` 內容去動態生成對應數量的 `Bottleneck`，並且新增了 `self.current_cfg_idx` 去追蹤目前使用到 `cfg` 的哪個位置。此外也會根據 `stride` 和輸出通道來自動處理是否需要進行 `downsample`。

7. Explain how you copied the original model weights to the pruned model. (5%)

因為要固定 `bn3 layer`，一開始會先用一個計數器 `bn_layer_counter` 來記錄目前走到第幾個 `BatchNorm layer`，透過這個數字來判斷要不要剪那一層。因為在 `ResNet` 的每個 `bottleneck block` 裡，都有三個 `BatchNorm` 層（分別是 `bn1`、`bn2`、`bn3`）所以用了 $(\text{bn\_layer\_counter} - 1) \% 3 == 0$ ，來找出哪些是 `bn3 layer`。此外為了避免某一層剪完後剩下的通道數太少，所以會自動留下絕對值最大的三個通道。

接下來的部分是把原本模型的權重照著剪枝後的通道 `mask` 複製到新模型裡。遇到 `BatchNorm layer` 時，就根據 `mask` 中保留下來的通道，把原本的 `weight`、`bias`、`running mean` 和 `running var` 對應複製到新模型；遇到 `Conv2d layer` 時，如果它後面接著 `BN` 層，就同樣根據前後兩層的 `mask` 篩選出需要的輸入與輸出通道，只複製這些保留下來的部分。

8. Please think about why it prompts you to fix the input and output channel numbers of each bottleneck to the channel numbers before pruning. What could happen if they are not fixed, and why?(5%)

如果不固定的話，會出現維度對不上的問題。如果在剪枝的時候，某一層 `layer` 的 `output channel` 變更，但 `shortcut channel` 沒有跟著改，那兩邊在相加時形狀就會不一樣。所以在複製 `weight` 前要先固定 `input and output channel`。

9. Describe the problems you encountered and how you solved them (5%)

最大的問題就是在剪枝後的模型維度對不上，一開始不知道問題在哪，以為是原始的模型架構有問題或是複製的地方寫錯，但後來先看了討論區發現也有同學有一樣的問題。看助教的回覆發現是沒有固定到 `conv3`，所以才會有問題。

一開始是直接固定 `num_features` in `[256, 512, 1024, 2048]`的話就不剪枝，結果剪枝完參數還是有兩萬多個，才發現這樣會導致 `layer3` 跟 `4` 的 `bn1`、`bn2`

都被排除掉。所以後來才改用 `bn_layer_counter` 去判斷 `bn3` 在哪邊，就不剪枝，遇到 `bn1`、`bn2` 才進行剪枝。