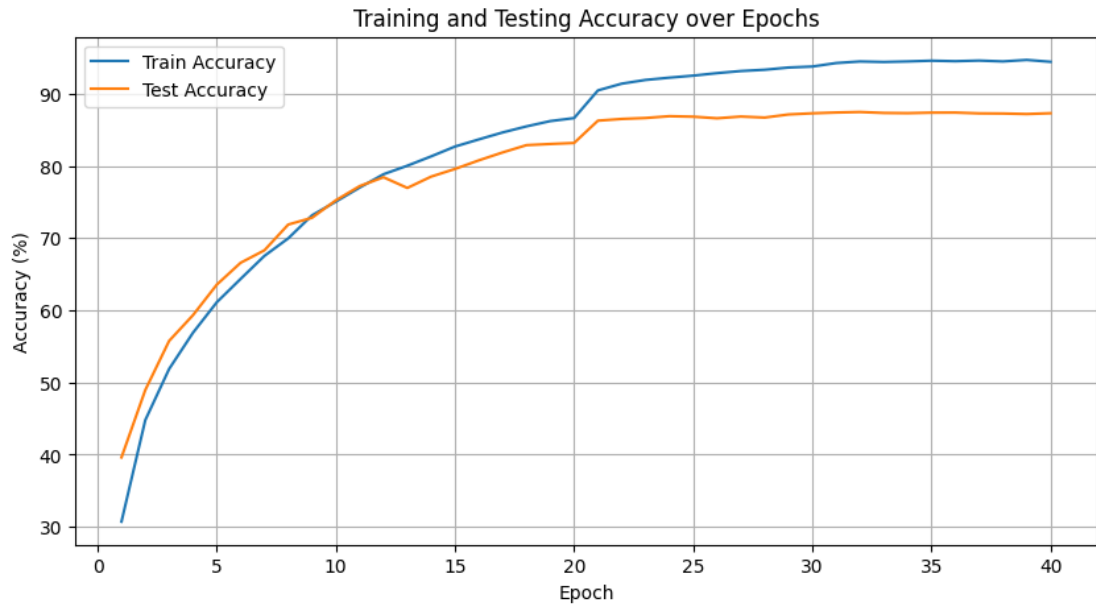
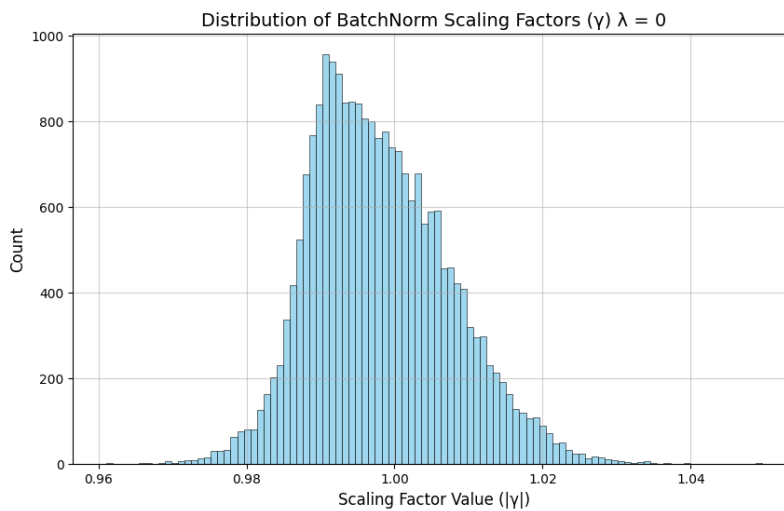


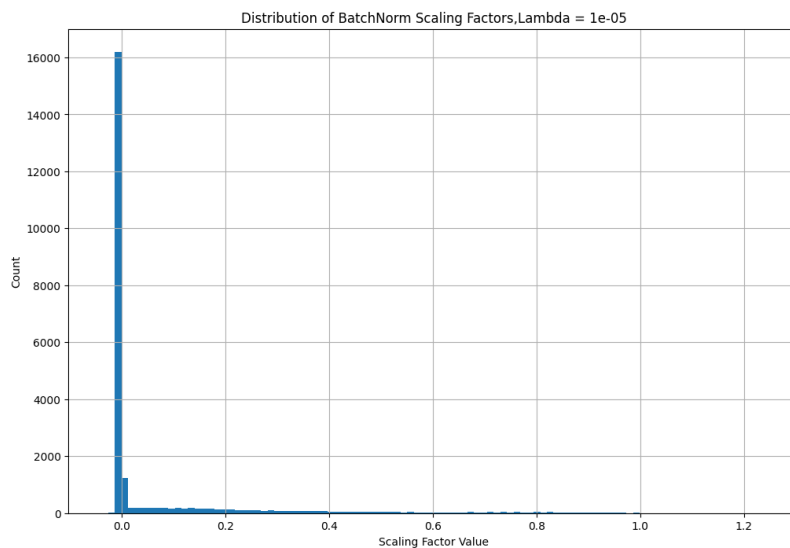
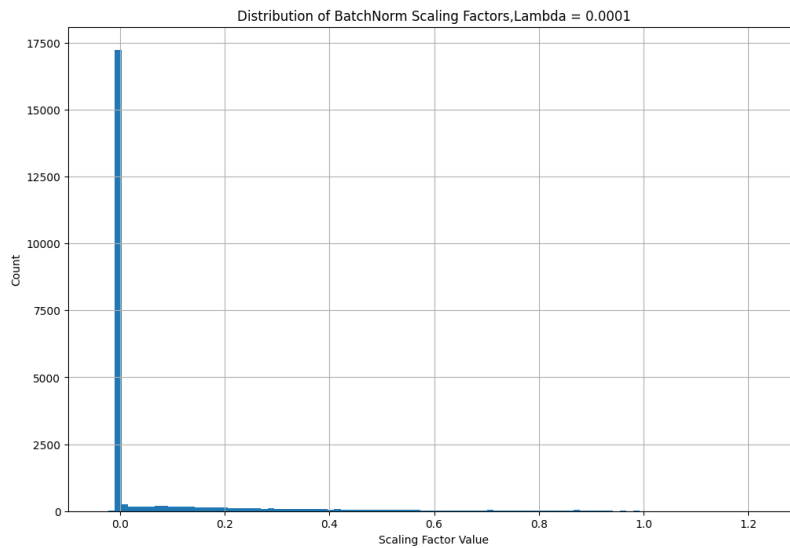
EAI Lab 2 Report

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



2. **Plot** scaling factor distribution with 3 different λ value





3. **Show** the model test accuracy after pruning 50% and 90% channels

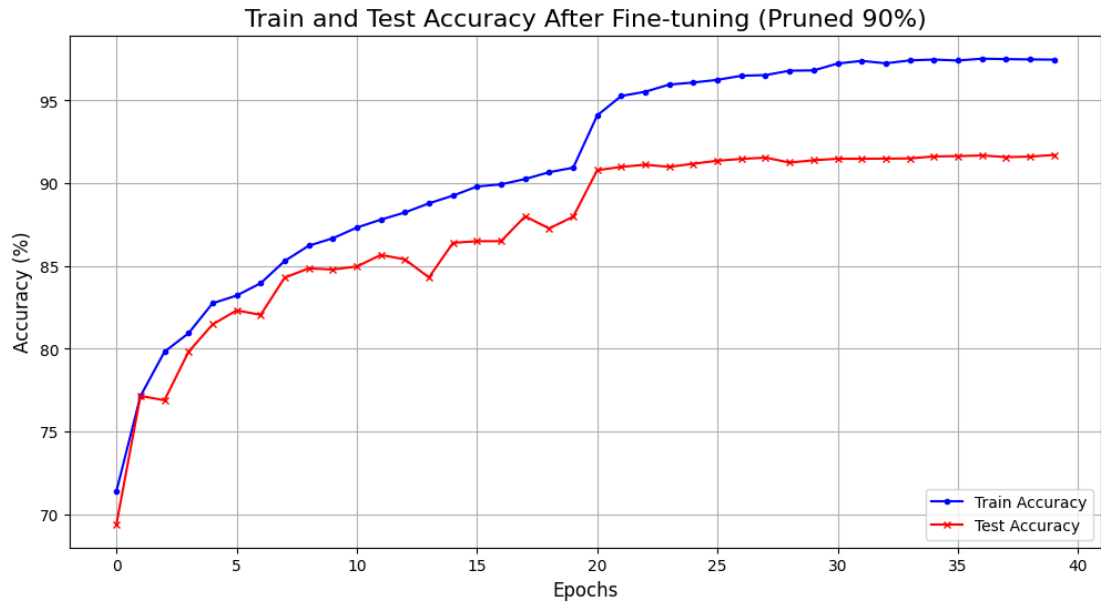
Prune 50%

```
Test set: Accuracy: 8997/10000 (90.0%)
tensor(0.8997)
```

Prune 90%

```
Test set: Accuracy: 2945/10000 (29.5%)
tensor(0.2945)
```

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



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

Model	Original	Prune 50% model	Prune 90% model
Test accuracy	0.91	0.91	0.903
Parameters	23513162	9824205	3665475

6. **Explain** how you modified resnet.py

這邊主要是讓模型能根據 `cfg` 動態建立：

Bottleneck 的部分：設定好 `in`, `out` channels，`conv1`, `bn1`, `conv2`, `bn2`, `conv3`, `bn3` 的輸出通道數分別被設定為 `out_channels[0]`, `out_channels[1]` 和 `out_channels[2]`。

`Make_layer` 的部分：從 `cfg` 列表中「切片」出當前 Bottleneck 所需的 3 個通道數。

7. **Explain** how you copied the original model weights to the pruned model.

建立 `cfg` 和 `mask`：首先遍歷舊模型 (`model`)，根據 `threshold` 建立 `cfg_mask` 和 `cfg`。同時迭代 `old_modules` (舊模型) 和 `new_modules` (新模型)。

`nn.BatchNorm2d`：使用 `end_mask.nonzero().view(-1)` 找出輸出通道的索引 `idx1`。只複製 `idx1` 對應的 `weight`, `bias`, `running_mean`, 和 `running_var`。

`nn.Conv2d`：同時使用 `start_mask` (得到 `idx0`) 和 `end_mask` (得到 `idx1`)。權重複製 (`[out_ch, in_ch, k, k]`) 必須同時在輸入和輸出維度上進行索引：`w = m0.weight.data[idx1, :, :, :][:, idx0, :, :].clone()`。

`nn.Linear`：Linear 層 (`[out_features, in_features]`) 只剪枝輸入維度。因此只用 `start_mask` 取得 `idx0`，並執行 `m1.weight.data = m0.weight.data[:, idx0].clone()`。bias 則直接複製。

另外在剪枝的時候為了要讓剪完以後的模型在 `out_channels` 以及下一階段的 `in_channels` 維度是對的上的，`bn3` 不能動到，因此需要特別寫一部分來判斷當前 layer 是否為該 conv 的 `bn3`。

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?

若一開始沒有固定的話會出現後續剪枝維度對不上的狀況，此外 Bottleneck 會依賴兩條路徑的維度相加，因此當通道數不是固定相同時，就會導致維度不匹配。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

一開始會遇到的問題是剪枝完的維度對不上，原先認為可能是 resnet 的架構有問題，但看到討論區有一樣的問題後，助教的回覆是沒有固定到 `conv3`，而我一開始設定是把 `bn3` 以及他下一層的 layer 的固定住，所以才會有問題，後續這部分就解決了。