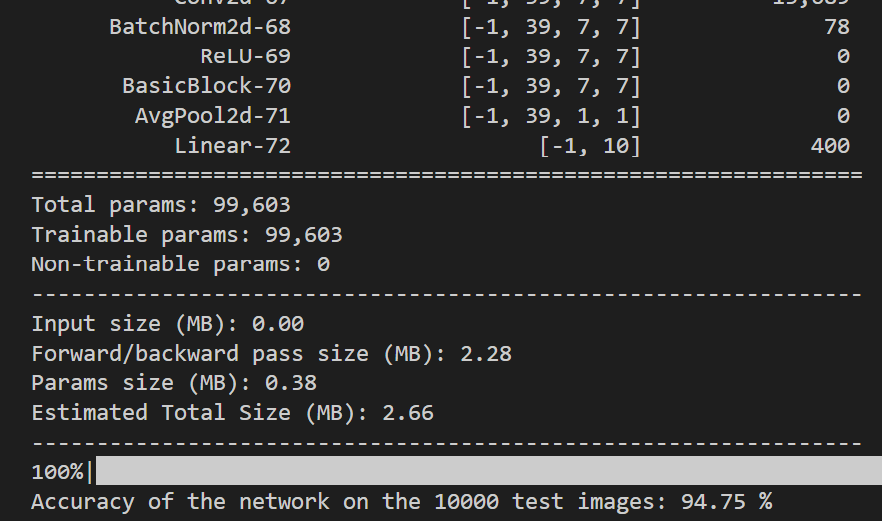
**HW2 Model Compression**

**Torch Summary:**

****

**Compression Method:**

1. **Create a much smaller ResNet-like model as a student network.**
2. **Use the given resnet-50 model as the teacher network.**
3. **Distill the knowledge from the teacher to the student, with KL divergence loss.**
4. **Afterwards, prune the student model based on the group norm importance. The model is pruned iteratively for 30 steps, and fine tuning is done after each step.**
5. **Random cropping, horizontal flip, and erasing are used during the training stage for both knowledge distillation and pruning.**

**References:**

**Student model & Random Erasing method:**

Paper: [[1708.04896] Random Erasing Data Augmentation (arxiv.org)](https://arxiv.org/abs/1708.04896)

Code: [GitHub - zhunzhong07/Random-Erasing: Random Erasing Data Augmentation. Experiments on CIFAR10, CIFAR100 and Fashion-MNIST](https://github.com/zhunzhong07/Random-Erasing)

**KL Divergence Loss:**

Code: [GitHub - haitongli/knowledge-distillation-pytorch: A PyTorch implementation for exploring deep and shallow knowledge distillation (KD) experiments with flexibility](https://github.com/haitongli/knowledge-distillation-pytorch)

**Pruning Method:**

Paper: [[2301.12900] DepGraph: Towards Any Structural Pruning (arxiv.org)](https://arxiv.org/abs/2301.12900)

Code: [GitHub - VainF/Torch-Pruning: [CVPR-2023] Towards Any Structural Pruning](https://github.com/VainF/Torch-Pruning)