# A4 Report

M124020042 賴壹誠

1. Train a ResNet18 on the Rotation task

   The best test accuracy achieved on the rotation task was 77.93%, and the lowest test loss was 0.529. Predicting image rotations is a pretext task that encourages the model to learn useful visual representations that can transfer well to other downstream tasks like image classification. For example, it needs to recognize that a dog is still a dog whether it is rotated 90 degrees left or right. By training on rotated images, the model is forced to understand the spatial layout and orientation of objects within the image, rather than just memorizing the appearance of objects. This teaches the model more generalizable features about visual concepts. The rotation task also regularizes the model and prevents overfitting to the dataset specifics.

2. Fine-tuning on CIFAR10 Classification

   With pretrained model, the model achieved 49.44% test accuracy after 20 epochs. With random initialization, the model achieved lower test accuracy of 45.82% after 20 epochs. This shows that the pretraining provides a useful initialization for the classification task, leading to faster convergence and better generalization compared to training from scratch. Overall, pretrained weights learned on self-supervised tasks like rotation prediction provide a useful starting point for downstream tasks compared to random initialization. Fine-tuning pretrained weights acts as a regularization and makes the model generalize better.

3. Full Fine-tuning on CIFAR10 Classification

   With pretrained model, the model achieved 67.94% test accuracy after 20 epochs. With random initialization, the model achieved lower test accuracy of 77.10% after 20 epochs. Surprisingly, the model trained from scratch performs better. This shows that the benefits of rotation pre-training diminish when the entire network is fine-tuned on the target task. The model likely learns new features that replace those learned during rotation prediction. In this case, it is better to train all weights on the target dataset from scratch. The rotation pre-training may even hinder full network training. Still, it provides benefits when only later layers are fine-tuned as seen in the previous section.