Assignment: EN4553 (Machine Vision)

University of Moratuwa

November 30, 2023

1. In this assignment we look at fine-tuning a pre-trained image recognition network. This assignment can be done as groups of 4 or less. If the work of two or more groups is found to be substantially similar, marks for one submission will be equally divided between those groups (so your group will receive only a fraction of the marks you deserve if you let your friends copy from you!). Automated tools and the grader's judgment will be used to determine the similarity of submissions.

   You can use any Python-based deep learning framework such as Tensorflow, PyTorch, or JAX. You will develop a model to identify pet breeds in the Oxford-IIIT Pet Dataset. You could either use the raw images available at https://www.robots.ox.ac.uk/~vgg/data/pets/, or the processed version in https://www.tensorflow.org/datasets/catalog/oxford_iiit_pet. This dataset has 37 different pet breeds (categories) with roughly 200 images for each. Given a pet image, your model will output its breed (one of 37 categories). If you use the raw dataset, randomly divide the dataset into train and test sets by including roughly half of the images in each set. If you use the Tensorflow datasets version, use the provided train and test partitions. The test partition is only used to report the final results, it should not be used to train the network or tune hyperparameters. Specific instructions are intentionally omitted to give students freedom to decide how they want to solve the problems.

   (a) Take any pre-trained network such as ResNet-50, SimCLR, DINO, CLIP image encoder, etc. implemented in your chosen deep learning library. Do k-NN classification with the embeddings produced by the pre-trained network. Report the accuracy on the test set.

   (b) Perform linear classification (multi-class logistic regression) with the pre-trained embeddings and report the accuracy on the test set. (Note that this is equivalent to training only the last fully-connected layer of a classification network with the feature extractor frozen).

   (c) Fine-tune a image classification network end-to-end (while changing the pre-trained weights as well). Report the accuracy on the test set. Use data augmentation and other methods you learned in the class to obtain good test accuracy.

Deliverables: Upload a single .zip file containing a PDF file and your Python source code files (.py files). The report should not exceed three A4 pages and should not include unreasonably small or large fonts. You many include snippets of your code, plots produced, etc. in the PDF. Each group should do only one submission with the names and the index numbers of the all members in the group on the cover page. Also mention the percentage contribution from each member (the total should be 100%). Late submissions will not be accepted.