Project Proposal
DD2424 - Deep Learning in Data Science

Francesco Olivieri, Inês Mesquita, Leandro Duarte

1. Project Title: Comparing ResNet and Vision Transformers: Supervised vs Semi-Supervised Learning for Pet Breed Classification
3. Project Type and Goal: Default + extension project from E $\rightarrow$ B/A
4. Problem Description and Approach: This project aims to compare the performance of ResNet-50 and ViT models under two different training paradigms: fully supervised and semi-supervised learning (using a specific semi-supervised technique for both). The goal is to evaluate how each model performs across these settings and gain insights into which architecture is better suited depending on the specific conditions. This study was decided upon reading the project description and identifying a shared interest in these topics.
5. Data: The Oxford-IIIT Pet Dataset, 37 pet breed categories with roughly 200 images per class.
6. Deep Learning Software Packages: numpy, matplotlib, scikit-learn, pandas, pytorch, torchvision, hugging face, fastai
7. Implementation: For our project, we will implement the core components ourselves, including the model setup, data preprocessing, training loops, and evaluations procedures. We may consult online tutorials, papers and open-source implementations to gain a deeper understanding of the material.
8.Experiments and Baselines For the E part, the experiments being done are: compute the accuracy when the final layer of the pre-trained ConvNet is replaced, with the Pet Dataset's training data. For the baseline, we have the accuracy mentioned in the assignment, (around 99%) on the binary test data. Then we pass to the multi-class classification problem, where we will fine-tune according to the two strategies given: Fine-tune l layers simultaneously and gradual un-freezing. The expected accuracy is around 95%. Then we do fine tuning with the case of imbalanced classes. Our baselines will be the previous experiments. In the extension we will perform experiments where we decrease the percentage of the labeled training data to 50%, 10% and 1% and record the accuracy. The same set of experiments will then be repeated using a pre-trained Vision Transformer model, and its performance will be compared to the ones obtained with the first model.
9. Milestones for Grading:

- **E grade:** Use a ResNet pre-trained model and finetune it to do binary classification (Dog vs Cat) and multi-class classification (37 breeds). Apply both fine tune and gradual unfreezing techniques and analyze.
- **D-C range:** Incomplete implementation of a B/A extension
- **B-A range:** Apply a vision transformer model and fine tune it to our dataset. Apply semi-supervised learning with unlabelled data to our dataset and track the performance results. Apply these semi-supervised learning techniques in both ResNet and ViT and compare results.

10. Learning Goals:

- **Francesco:** Learn how to adapt models to specific tasks, gain hands-on experience with Vision Transformers, and explore a technique in semi-supervised learning
- **Inês:** Understand and implement semi-supervised learning techniques and impact assessment as well as the core principles of Transfer Learning.
- **Leandro:** Learn how to efficiently implement Vision Transformers models and fine tune them, using good practices and SOTA techniques that can transport results to real-world implementation.

11. Target Grade: A
12. References: SSL techniques: [link] · ResNet-50: [link] · SSL ViT: [link1], [link2] · Fine-tuning ViT: [link]