DD2424 Deep Learning in Data Science
Project Proposal

Ksenia Biazruchanka
20050523-T068
kseniabi@ug.kth.se

Tobias Haunreiter
19990923-T574
thau@ug.kth.se

Alexander Soukup
20020103-T333
ansoukup@ug.kth.se

Johannes Jeup
20010303-T498
jeup@ug.kth.se

April 30, 2025

Default Project 1 - Transfer Learning

We will use a pre-trained convolutional neural network (e.g., ResNet18 or ResNet34) and fine-tune it for a new image classification task. The dataset to be classified is the Oxford-IIIT Pet Dataset, which consists of 37 categories, each containing approximately 200 images. Each image is annotated with ground truth labels for breed classification, region of interest (ROI), and pixel-level segmentation via a trimap [1].

The first objective is to fine-tune a classifier to distinguish between cats and dogs, with a target accuracy exceeding 99%. Upon achieving this, the output layer will be extended to predict all 37 breed classes (12 cat breeds and 25 dog breeds).

Two strategies will be explored for fine-tuning the extended classifier. In the first approach, l layers are fine-tuned simultaneously, and the impact of tuning an increasing number of layers on performance will be analyzed. The second strategy involves gradually freezing or unfreezing layers to observe how this affects classification accuracy.

Our target accuracy for the full 37-class classification task is approximately 95%. Additionally, we will evaluate the robustness of the model to class imbalance by under-representing one class during training and analyzing its impact. To mitigate imbalance, we will experiment with techniques such as weighted cross-entropy loss and over-sampling of minority classes. Model performance will be evaluated based on classification accuracy and training efficiency on the target dataset.

Extension for B/A Grade

In order to qualify for a higher grade, we plan to explore semi-supervised learning. This means that we will incorporate unlabeled data into the training process while keeping track of the performance of the network. The goal is to come up with a network that produces similar results as the "basic" technique when only labeled data is used for training. It will be examined whether this is true or not using the FixMatch strategy,

developed by Sohn et al. [2]. With this, pseudo-labels are first generated based on weakly-augmented unlabeled images.

Our project will be based on the FixMatch framework, which combines the consistency regularization and pseudo-labeling approaches in a simple yet effective way. Specifically, the method retains pseudo-labels only when the model's prediction on a weakly-augmented image is above a confidence threshold. It then trains the model to predict this pseudo-label when presented with a strongly-augmented version of the same image. Despite its conceptual simplicity, FixMatch achieves state-of-the-art results on several semi-supervised learning benchmarks, such as 94.93% accuracy on CIFAR-10 with just 250 labeled examples. We aim to reproduce and adapt these results for The Oxford-IIIT Pet Dataset, evaluating the method's effectiveness under various conditions and dataset sizes and analyzing key factors contributing to its performance.

References

[1] Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, and CV Jawahar. Cats and dogs. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3498–3505. IEEE, 2012.

[2] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33:596–608, 2020.