

Blue text indicates example text / specifications for the stuff you should fill in. But you will change all font colors to black before submitting. Limit of 2 pages.

Title: Classifying Animal Images Using Vision Transformers

Team: Name 1 (CIS 5190), Name 2 (CIS 4190), Name 3 (CIS 5190)

Kaggle dataset: Animal Image Classification [Note: This is just an example; you should use one of the Kaggle datasets listed in the project instructions.]

Part 1: Implementation contributions: (What are your planned contributions for the implementation requirement. Describe any libraries, frameworks, or codebases you plan to build on (if any), and how you plan to extend them.) Option 1: We plan to compare different neural network architectures and train them on the dataset. In particular, we plan to consider: (i) multi-layer perceptron (MLP), convolutional neural networks (CNNs), and vision transformers. We will implement our approach using Pytorch. For vision transformers, we plan to leverage an existing implementation for ImageNet (https://huggingface.co/docs/transformers/model_doc/vit). We also plan to compare finetuning a pretrained model vs. training one from scratch.

Part 2: Evaluation contributions: (Describe how you will evaluate the success of your proposed algorithm(s). In particular, describe the metric(s) you plan to use, along with the hyperparameters you plan to analyze. Also describe dataset shifts that you plan to evaluate on. Remember, your proposal's success is not necessary or sufficient for a good project grade. A failed proposal, with high-quality analysis of reasons for failure might very well be graded as an outstanding project.) We will measure the accuracy of our approach on a held-out test set. For each architecture, we plan to analyze how accuracy varies depending on the learning rate, number of (convolutional, MLP, or self-attention) layers, presence of dropout and/or batch normalization, and regularization. In addition, we plan to evaluate our approaches on shifts including added Gaussian noise; we will also study a more realistic shift where we train on one set of dog/cat breeds (e.g., chihuahua+sphynx), and test on a held-out set (e.g., greyhound+siamese).

Prior work: (Your starting/reference points. Add 1-3 of the most relevant.)

1. E.g. which prior Kaggle submissions (cite with links) or which papers are you planning to build on top of?

Which parts of the curriculum from this class do you expect to apply?: Our tentative plan is to use deep neural networks that are suitable for image data, since we are using image data and we have a sufficiently large dataset for deep learning. Finally, we expect that reference 2 might provide useful pointers.

Expected challenges and risk mitigation: One challenge we will have to deal with is the amount of time needed to train state-of-the-art neural networks from scratch. While we plan to evaluate both models trained from scratch as well as finetuning pretrained models, we will

consider an alternative where we focus on finetuning, but compare models finetuned on different datasets (e.g., CIFAR10 vs. ImageNet).

Ethical considerations and broader social impact: (What potential good or bad social impacts could your project in the events of success/failure?)

(Supplementary materials if any, are exempted from 2pg limit, but not guaranteed to be considered during evaluation)