COMS 4995 NNDL Course Competition: Bird, Dog, or Reptile?

1. Intro

Welcome to NNDL's first-ever course competition!

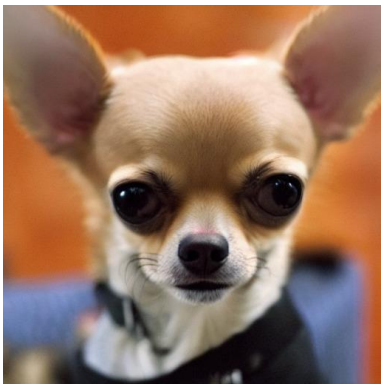
The main topic of this competition is **Multi-label Classification** with image data (Image Classification). Remember the classifiers constructed by linear layers or neural networks we discussed before? It is time for you to practice what you have learned so far in this task.

2. Dataset

a. Details

We prepared a dataset with images from 3 super-classes of entities, including **bird, dog**, and **reptile**. Each of these super-classes also includes sub-classes (e.g., hawk in bird, golden retriever in dog, etc.). The objective is to predict the correct super-class label and sub-class label for each image.

Super-class: bird
Sub-class: rooster



Super-class: dog
Sub-class: chihuahua

Super-class: reptile
Sub-class: mud turtle

The dataset is carefully split into two subsets: training and testing sets. The training set will be shared with you for model training. You can find approximately 6.6k images in this training set, and each image has a resolution of 8 x 8.

b. Task Format

The task format is as below:

Input: Image(s)
Output: Predicted-label(s)

3. Evaluation

We are collaborating with the Columbia AI Initiative to host the evaluation of the testing set via a web-based leaderboard. To test your trained model's performance, you can download the testing images and run your model to produce the prediction results. Then you can further submit the predictions to the leaderboard for evaluation.

a. Metrics

The training and testing sets contain the same set of super-classes (bird, dog, and reptile) but different sub-classes. Some sub-classes may only exist in the testing set. Your model should not only be able to predict the sub-classes but also the sub-classes. You may consider designing different adapters on top of a single model for both tasks or two models for resolving each task.

During inference, when encountering test images of novel/unseen sub-classes, you should design your model to classify those images to an additional sub-class label "novel." Therefore, your models' prediction space for super-class classification should be consisted of {bird, dog, reptile}, and the space for sub-class classification should consist of {all the sub-classes from the training set, novel}

Correspondingly, there are two main evaluation metrics: (1) Classification accuracy of the super-classes, (2) Classification accuracy of the sub-classes.

Further, we also have two sub-evaluation metrics for the sub-class classification: (1) Classification accuracy of the seen sub-classes, (2) Classification accuracy of the unseen sub-classes.

Please refer to the "example_test_submission.csv" file for the submission format.

b. Baseline Performance

You should try your best to achieve high performance and at least ensure your performance is better than our baseline.

Model	Class Accuracy (%)	Sub-Category Accuracy (%)		
		Overall	Seen	Unseen (Novel)
CLIP/B-16	41.93	8.29	12.42	0

4. Download

[Dataset Download Link]

5. Hint

Hints for starting:

- Preprocessing the images (e.g., normalization)
- Converting folder names into class labels

Be Aware!

- The class frequency (number of samples for each class) in the training set may not be the same as in the testing set.

- Sample data distribution of the training set may not be the same as in the testing set. Images of dog breeds like golden retrievers may exist in the training set but not in the testing set, and vice versa.

- **Suggestion:**
  You should aim to train a generalizable model with all the techniques we have discussed so far (e.g., data augmentation, weight decay, etc.) Other tricks may include potentially building your local validation set for testing the model's generalization ability.