COMS 4776 NNDL Course Competition: Bird, Dog, or Reptile?

1. Introduction

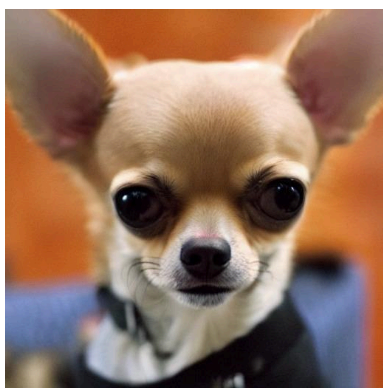
The main topic of this competition is **Multi-label Classification** with image data (Image Classification). Remember the classifiers constructed by linear layers or neural networks we discussed before? It is time for you to practice what you have learned so far in this task.

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, toy terrier 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 **64 x 64.**

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. The leaderboard link will be released shortly.

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 super-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, your model will encounter both **a novel superclass and novel subclasses**. That is, you may see images from a novel super-class (and thus a novel sub-class). Additionally, you may see images from a seen super-class but a novel sub-class.

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

There four two main evaluation metrics: (1) Classification accuracy of the super-classes, (2) Classification accuracy of the sub-classes, *(3) Categorical cross-entropy of the super-classes, (4) Categorical cross-entropy of the sub-classes*. These last two metrics are **new additions**.

Further, we also have sub-evaluation metrics: (1) Classification accuracy of the seen classes, (2) Classification accuracy of the unseen classes. This applies for both the super-classes and sub-classes.

b. Baseline Performance

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

	Cross-Entropy		**Superclass Accuracy (%)**			**Subclass Accuracy (%)**		
	Super class	**Subclass**	**Overall**	**Seen**	**Unseen (Novel)**	**Overall**	**Seen**	**Unseen (Novel)**

CLIP/B-32	1.36	4.44	80.42	99.14	32.91	1310	60.75	0.24

The main focus of this project is thus on improving your model's performance on novel super-classes and sub-classes, relative to the baseline. You should have one more than one version of a model that you develop. The grading will be based on your analysis of the performance of the methods you worked on, and your insight into their strengths and weaknesses.

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 beagles may exist in the training set but not in the testing set, and vice versa (novel subclasses and superclasses).

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