Paper Title: “Visualizing features that distinguish 15,000 kinds of plants”

Idea: Let’s use data like this:

<https://www.kaggle.com/c/herbarium-2022-fgvc9/data?select=train_metadata.json>

To evaluation how to visualize interesting features for a fine grained classification problem.

Specific plan:

1. What do the elements of the JSON file in the kaggle, training data, tell us?
2. Genus is probably the type of plant
3. Institution is (i think) the work location where the image was archived
4. What is “Category”?

Next steps:

I suggest you reach out to Xiaotong Liu and Zeyu Zhang (my PhD students who will also help with this project) to ask for best suggestions of how to train a classifier for this, and/or train a metric learning algorithm.

Once we have a good model, Xiaotong Liu has papers on the best ways to visualize the features of those that tell you why and where two images are considered to be similar: <https://arxiv.org/abs/1901.00536>, and Xiaotong can run this through code he has that gives correspondences between features.