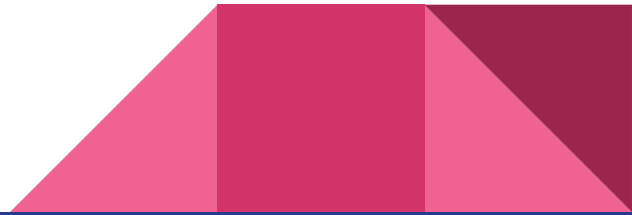
Dataset Generalization via Continual Learning

Mona Abdelrahman and Ariba Khan

Motivation

- Model generalization across different datasets and unseen data performance is a big problem
- Each dataset has its own style, collection methods, etc that qualitatively impact the data
- Problem with Applied ML
  - Can you accurately predict how a model will perform in the wild?
- This is similar to continual learning



Background

- In Continual Learning, data is not available all at once
  - As the data is loaded, the data distribution changes over time
  - In standard ML, all data is available all at once

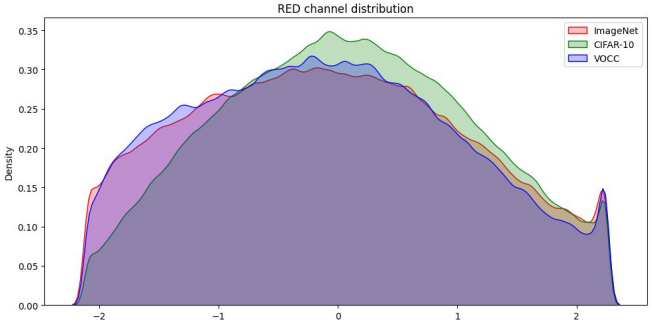
- Goal is to have the model improve over time as it gets more data, potentially also on unseen data

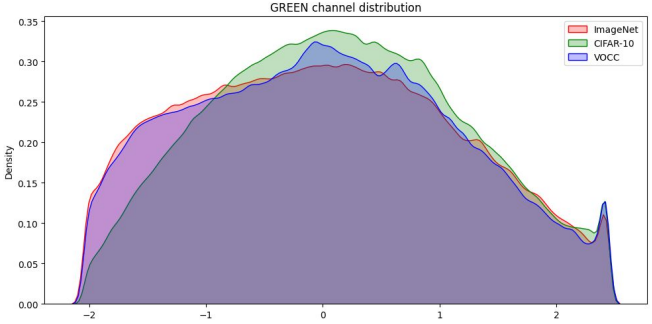
Methods

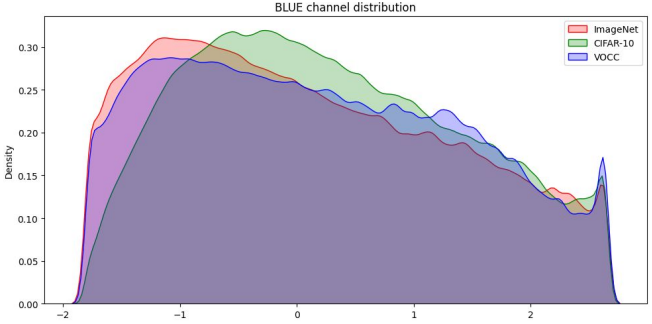
- 3 Datasets with 7 common classes
  - CIFAR-10, ImageNet, PASCAL-VOC 12
- Using a pre-trained EfficientNet followed a continual learning inspired scheme:
  - 20 Epochs on ImageNet
  - 20 Epochs on ImageNet + CIFAR-10
- Also experimented with a VIT
  - Requires large datasets

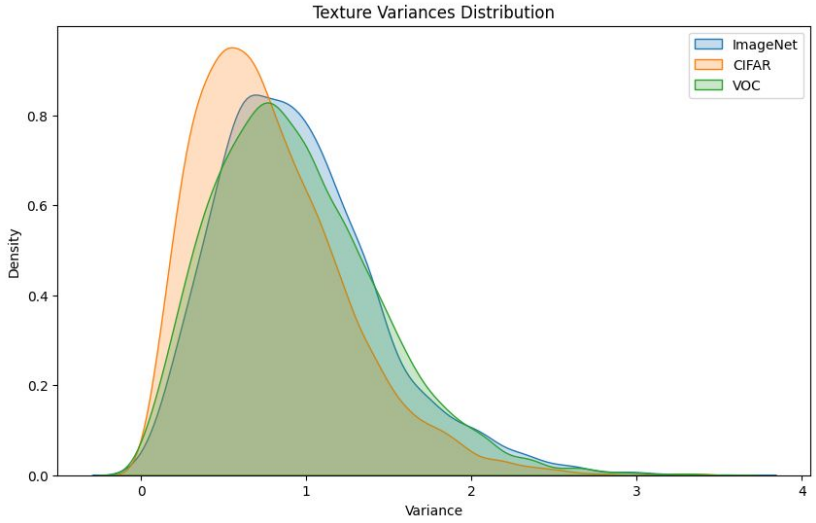
Model Details

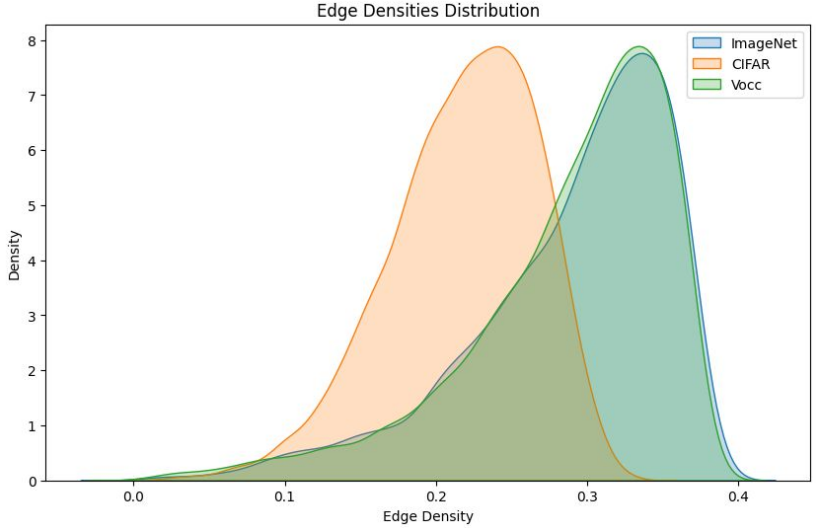
- Classes
  - airplane, automobile, bird, cat, dog, horse, ship
- Image scaled to 120, cropped to 96
- Transforms include:
  - RandomCrop, ColorJitter, RandomHorizontalFlip, RandomAffine, Normalization


RED channel distribution
ImageNet
CIFAR-10
VOCC
Density

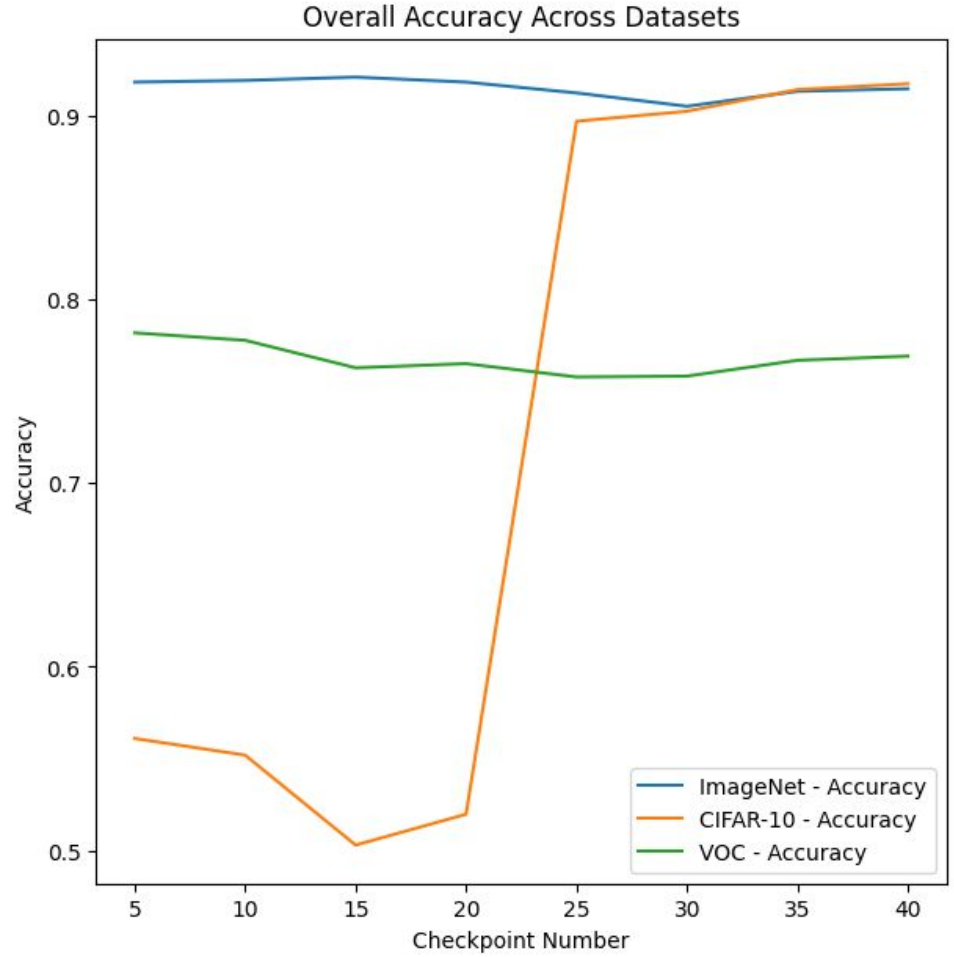

GREEN channel distribution
ImageNet
CIFAR-10
VOCC
Density

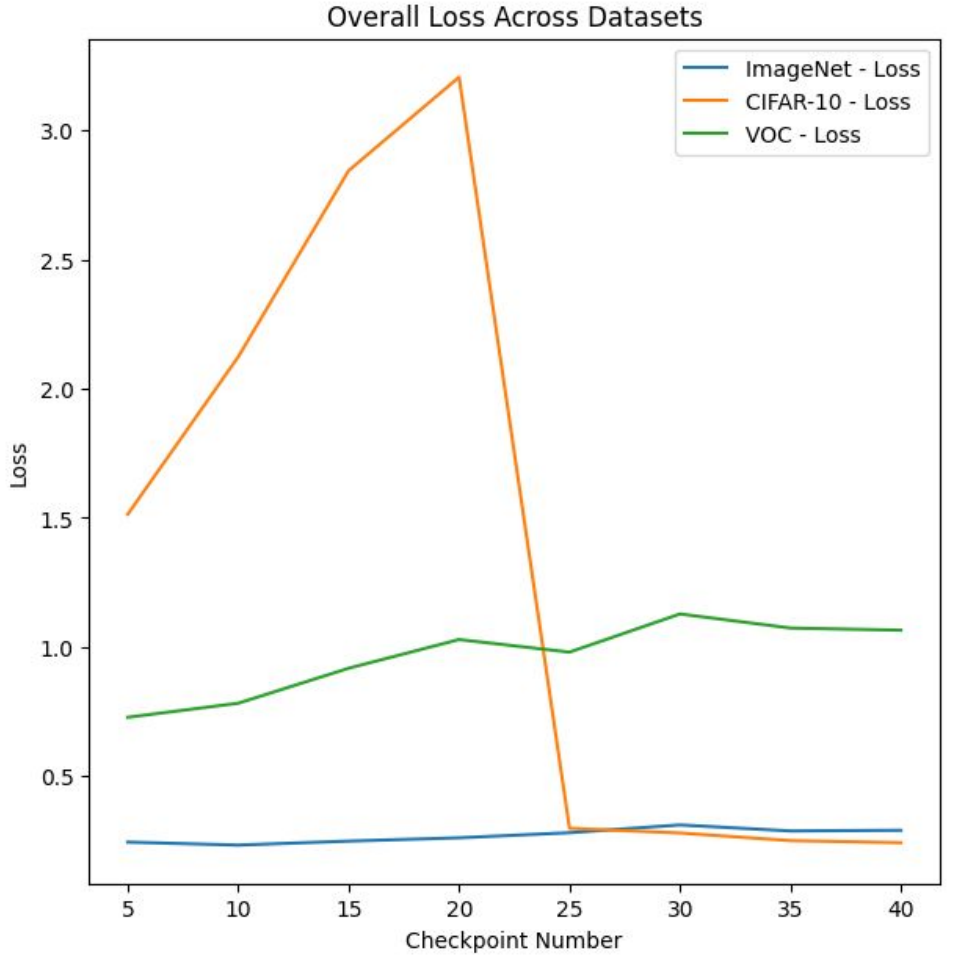

BLUE channel distribution
ImageNet
CIFAR-10
VOCC
Density

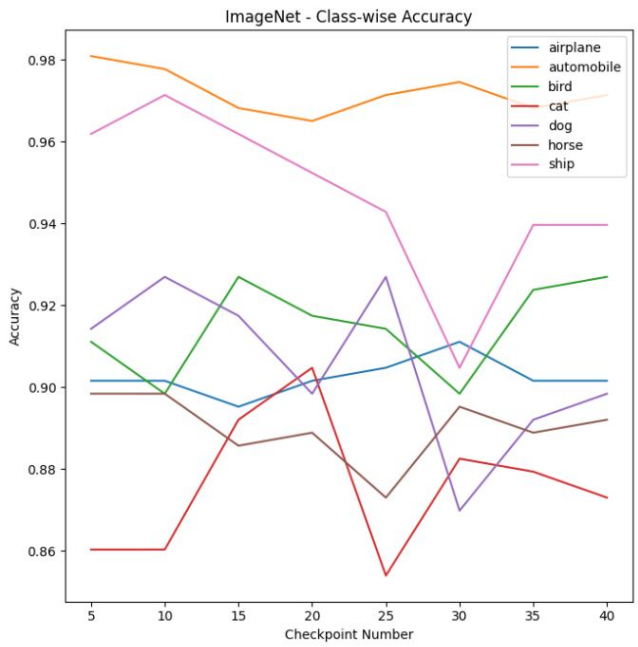

Texture Variances Distribution
ImageNet
CIFAR
VOC
Density
Variance

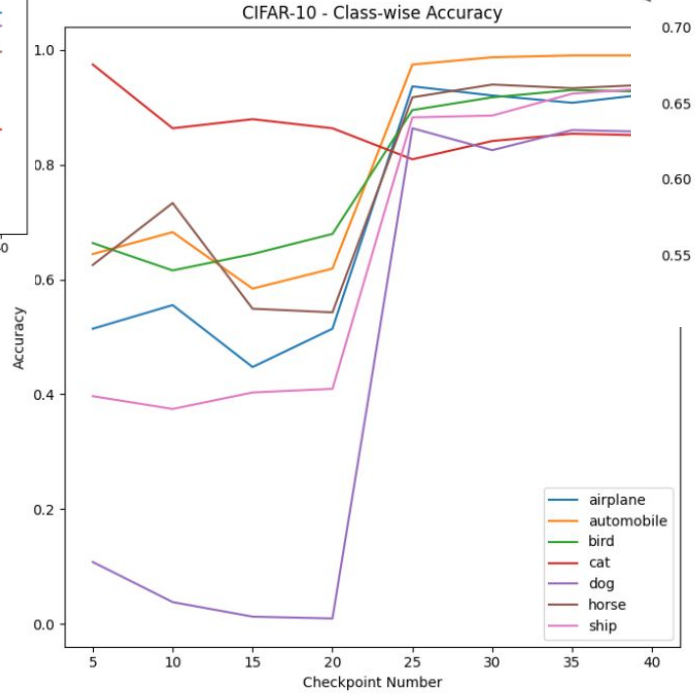

Edge Densities Distribution
ImageNet
CIFAR
Vocc
Density
Edge Density

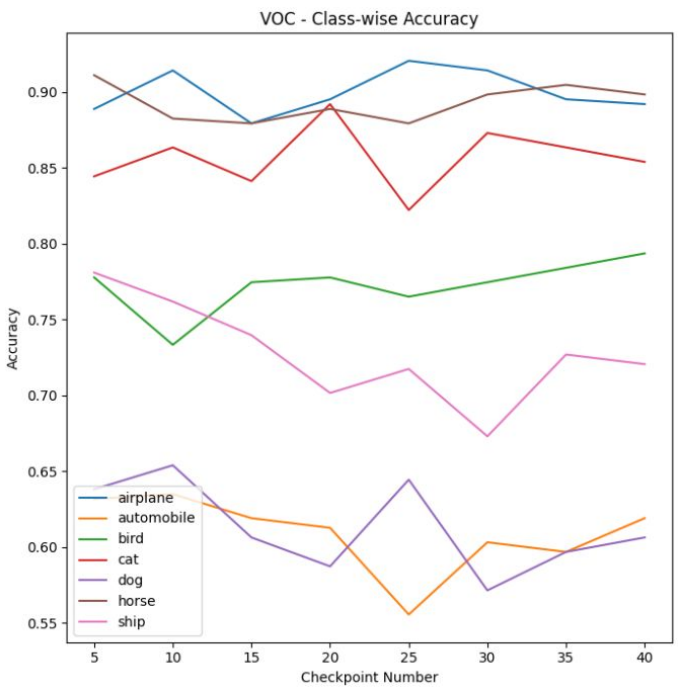
Results

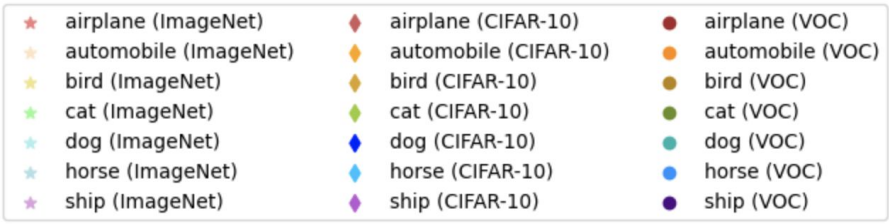

Overall Accuracy Across Datasets
Accuracy
Checkpoint Number
ImageNet - Accuracy
CIFAR-10 - Accuracy
VOC - Accuracy
0.5
0.6
0.7
0.8
0.9
5
10
15
20
25
30
35
40

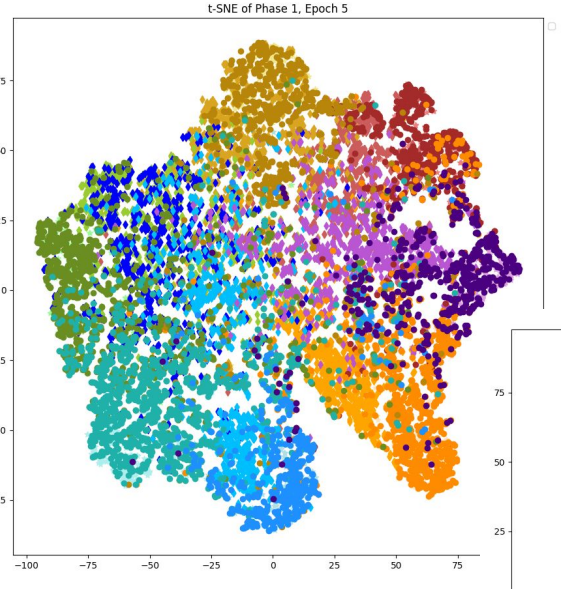

Overall Loss Across Datasets
Loss
Checkpoint Number
ImageNet - Loss
CIFAR-10 - Loss
VOC - Loss
0.5
1.0
1.5
2.0
2.5
3.0
5
10
15
20
25
30
35
40

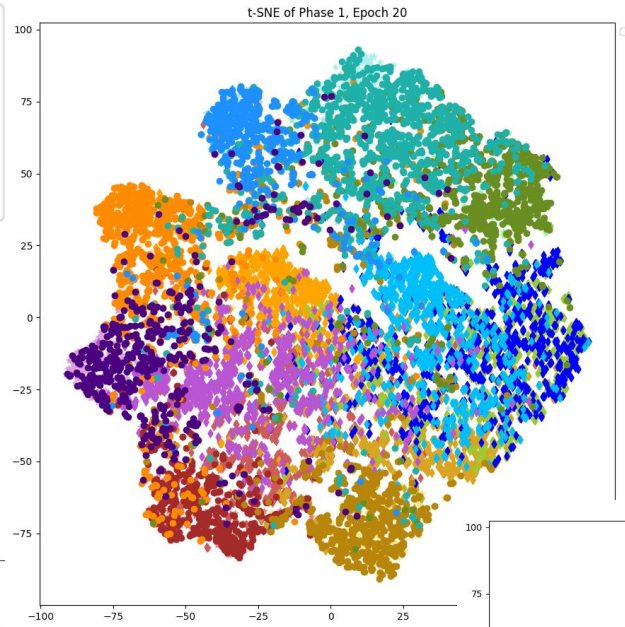

ImageNet - Class-wise Accuracy
airplane
automobile
bird
cat
dog
horse
ship
Accuracy
Checkpoint Number

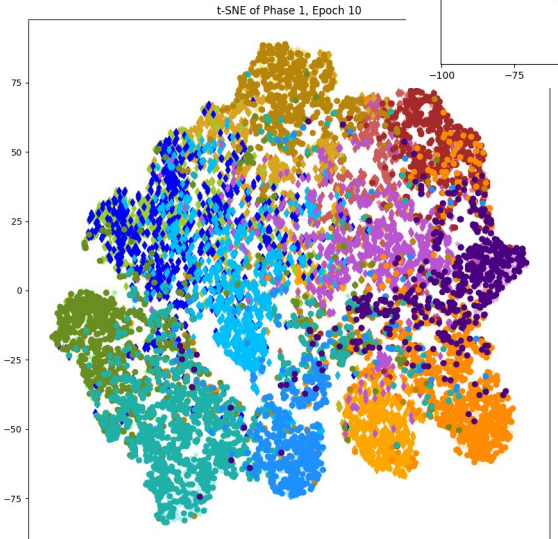

CIFAR-10 - Class-wise Accuracy
airplane
automobile
bird
cat
dog
horse
ship
Accuracy
Checkpoint Number

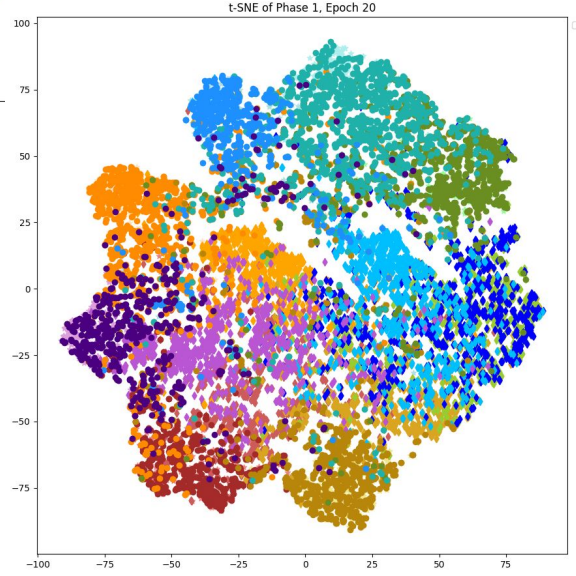

VOC - Class-wise Accuracy
airplane
automobile
bird
cat
dog
horse
ship
Accuracy
Checkpoint Number

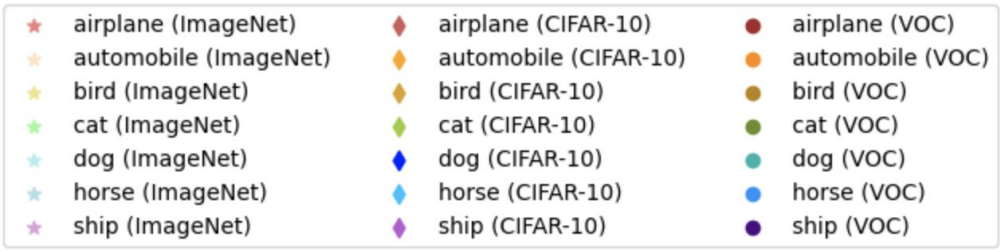

airplane (ImageNet)
automobile (ImageNet)
bird (ImageNet)
cat (ImageNet)
dog (ImageNet)
horse (ImageNet)
ship (ImageNet)
airplane (CIFAR-10)
automobile (CIFAR-10)
bird (CIFAR-10)
cat (CIFAR-10)
dog (CIFAR-10)
horse (CIFAR-10)
ship (CIFAR-10)
airplane (VOC)
automobile (VOC)
bird (VOC)
cat (VOC)
dog (VOC)
horse (VOC)
ship (VOC)

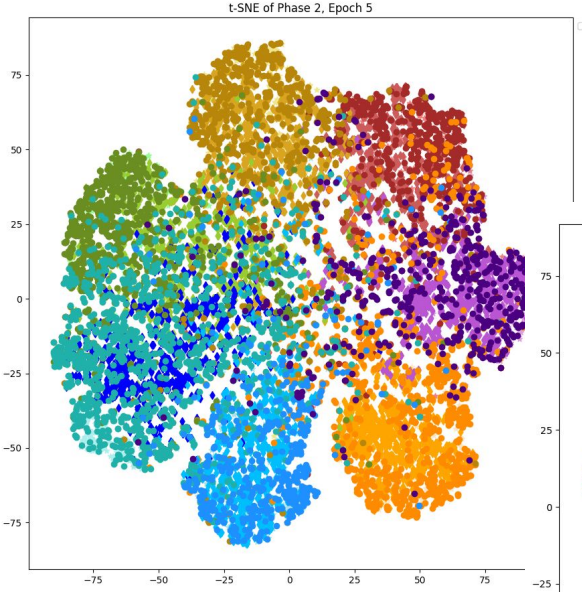

t-SNE of Phase 1, Epoch 5

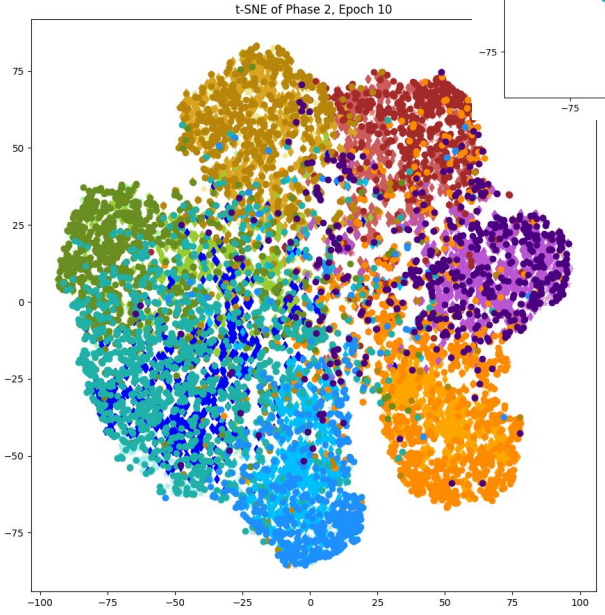

t-SNE of Phase 1, Epoch 20

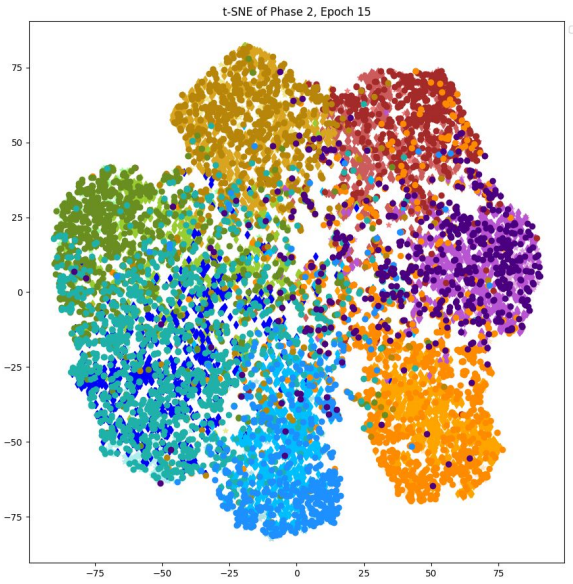

t-SNE of Phase 1, Epoch 10

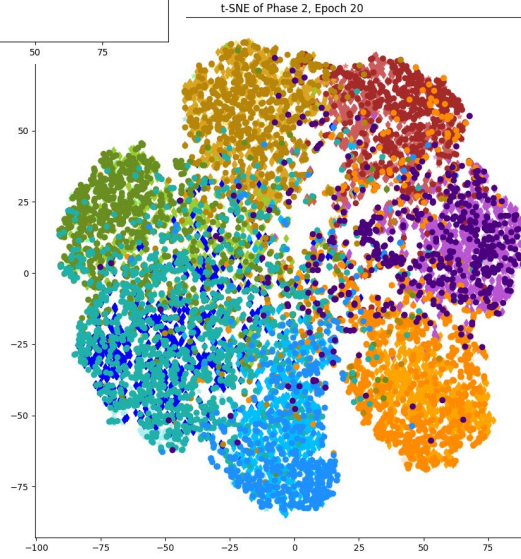

t-SNE of Phase 1, Epoch 20

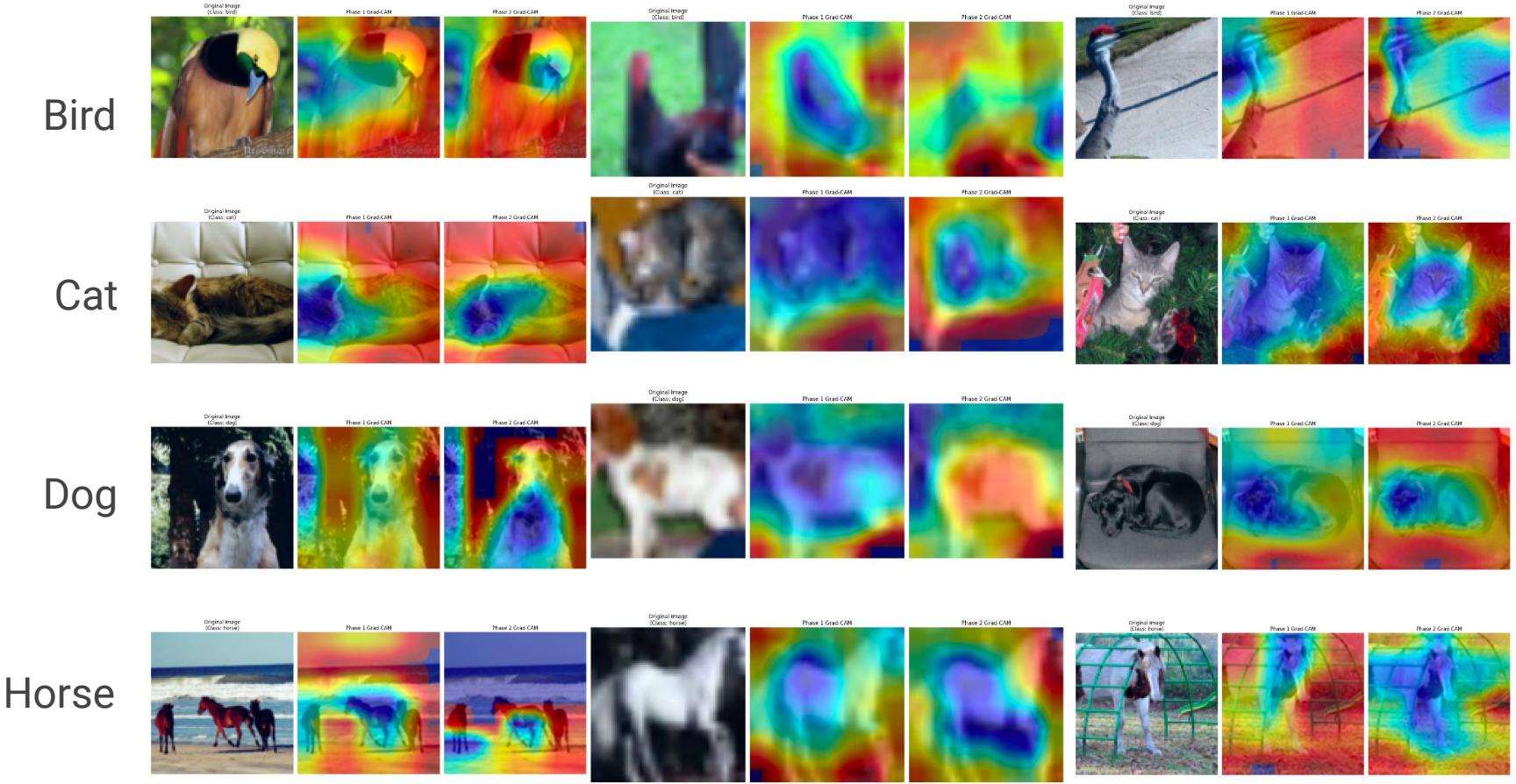

airplane (ImageNet)
automobile (ImageNet)
bird (ImageNet)
cat (ImageNet)
dog (ImageNet)
horse (ImageNet)
ship (ImageNet)
airplane (CIFAR-10)
automobile (CIFAR-10)
bird (CIFAR-10)
cat (CIFAR-10)
dog (CIFAR-10)
horse (CIFAR-10)
ship (CIFAR-10)
airplane (VOC)
automobile (VOC)
bird (VOC)
cat (VOC)
dog (VOC)
horse (VOC)
ship (VOC)


t-SNE of Phase 2, Epoch 5


t-SNE of Phase 2, Epoch 10


t-SNE of Phase 2, Epoch 15


t-SNE of Phase 2, Epoch 20


ImageNet
CIFAR-10
VOC12
Bird
Cat
Dog
Horse

Analysis - Generalization Principle

- More clear separation in class accuracy in VOC dataset
- Model seems to have decent performance on the VOC without ever training on it, generalizes well to it
- Some classes on CIFAR also performed well before it was included in training
  - Indicates good generalization from the model on this unseen data

Analysis - Emergence Principle

- Over time, can start to see less separation of classes by dataset
- This demonstrates some emergent properties of the model, and of the datasets
  - Potential differences in style that may be subtle, but make a big difference with prediction
- Ties into the Manifold View

Future Improvements

- Larger datasets or augmentation techniques to increase size
  - Helps with generalization?
  - Helps with VIT performance?
- See how generalization performance changes with different dataset schemes
  - Starting off with one dataset over another