



Using Vision Transformers to Classify Bird Species

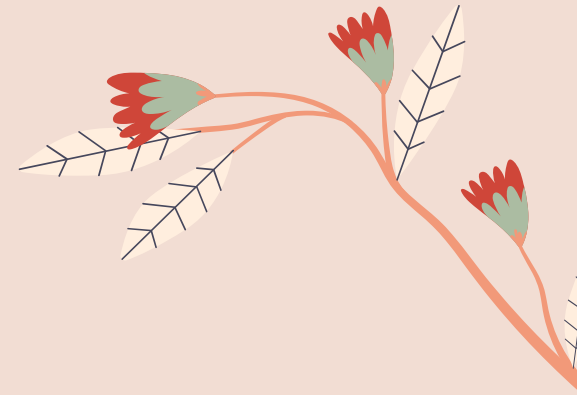
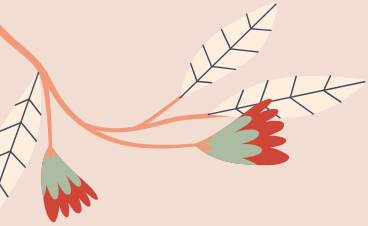


The Team



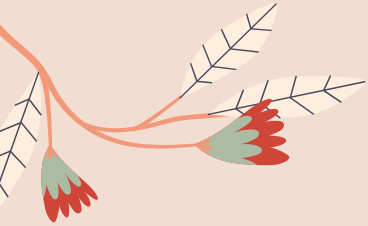
Brendan Lorn

MEng Student
Concentration:
Data Analytics and Artificial Intelligence



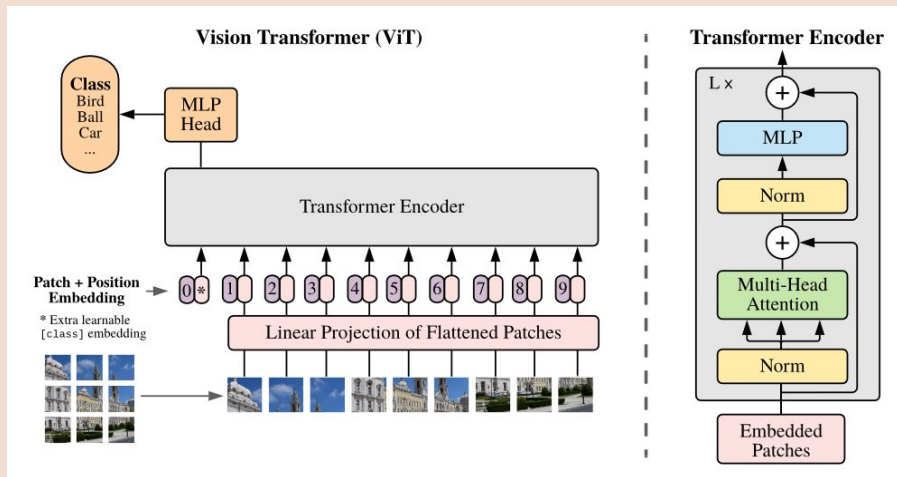
Problem Description

- Motion sensing cameras in natural habitats gather large amounts of data
- Identifying animal species by hand is taxing
- Utilizing AI and Deep Neural Networks can allow this task to be automated



Background

- Emergence of Transformers in NLP
- Convolutional Neural Networks are the go-to
- Vision Transformer (ViT)

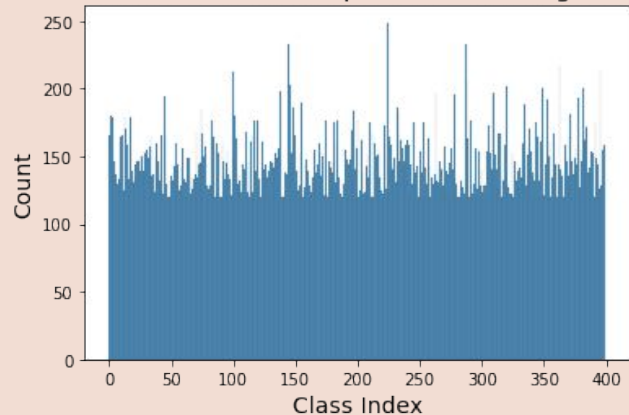


Approach

- Utilizing a pre-trained ViT model to classify bird species
- Fine tuned on Birds 400 Dataset
 - 58388 training images
 - 2000 validation images
 - 2000 test images
 - Example Image:

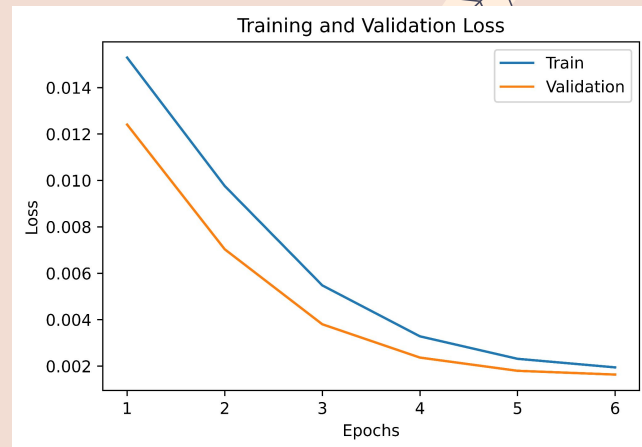


Distribution of Bird Species in Training Data



Results

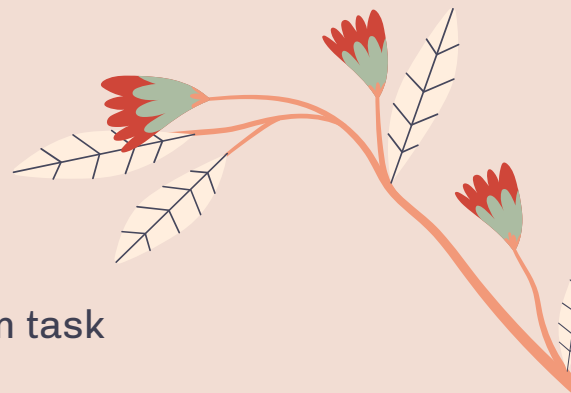
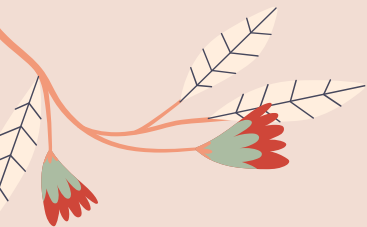
- Hyperparameters:
 - Epochs: 6
 - Batch Size: 64
 - Learning Rate: 0.00001
 - Adam optimizer, linear scheduler
- Test Dataset Evaluation:



Precision	Recall	F1-Score
0.991	0.990	0.989

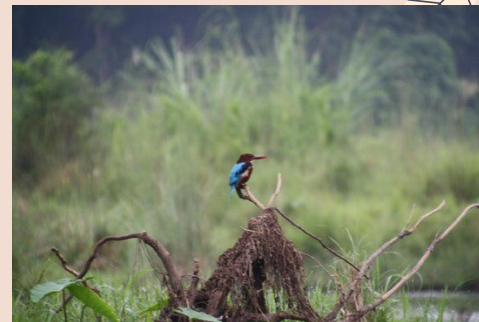
Lessons Learned

- How to utilize a pre-trained vision model for a downstream task
- ViT closely matches F1-score against CNN models
- ViT is a promising new application of transformers



Future Work

- Utilizing a different dataset
 - More data
 - Less clear images
 - More species
- Choosing a different downstream task
 - Detection
 - Segmentation



Questions?

