

Deep Learning for Distinguishing AI Generated Images

Motivation

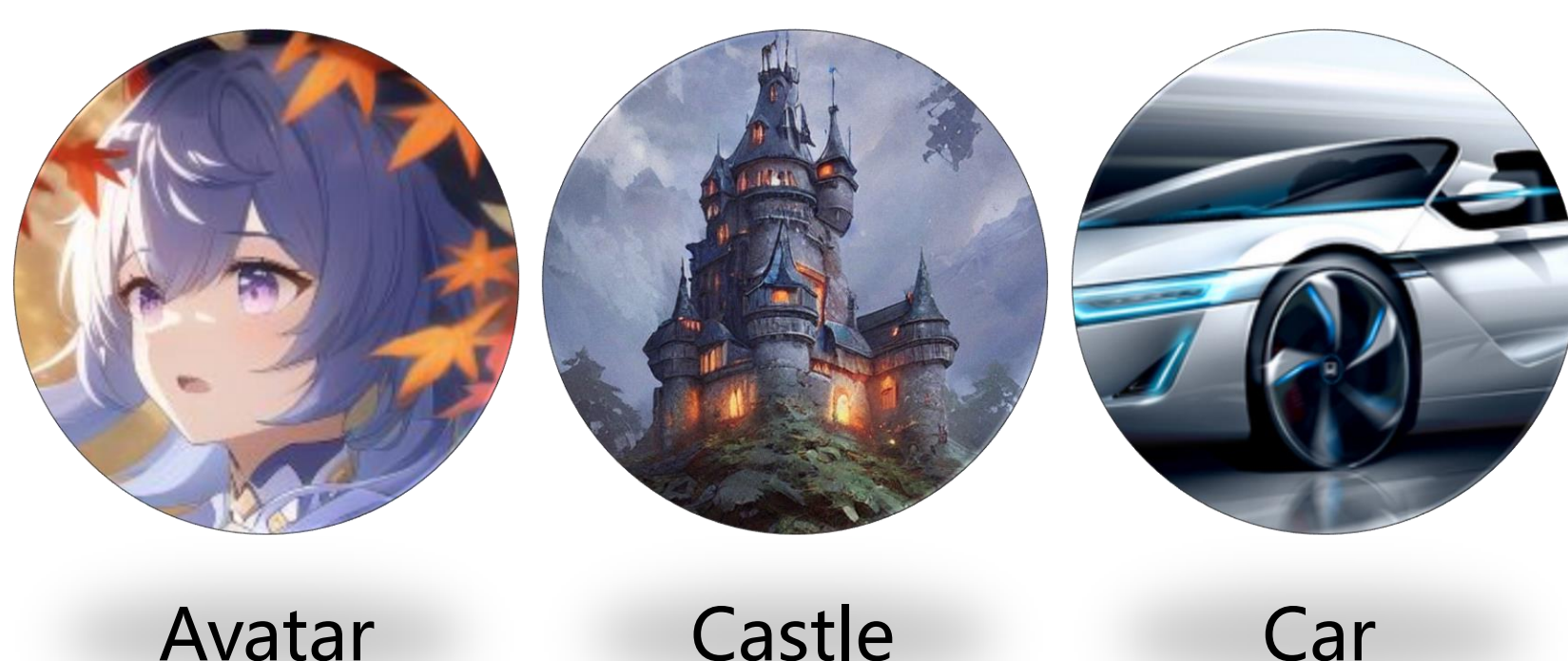
If we want to distinguish AI generated images, maybe neural network itself could be a better classifier than human being?



We might be not able to distinguish that...

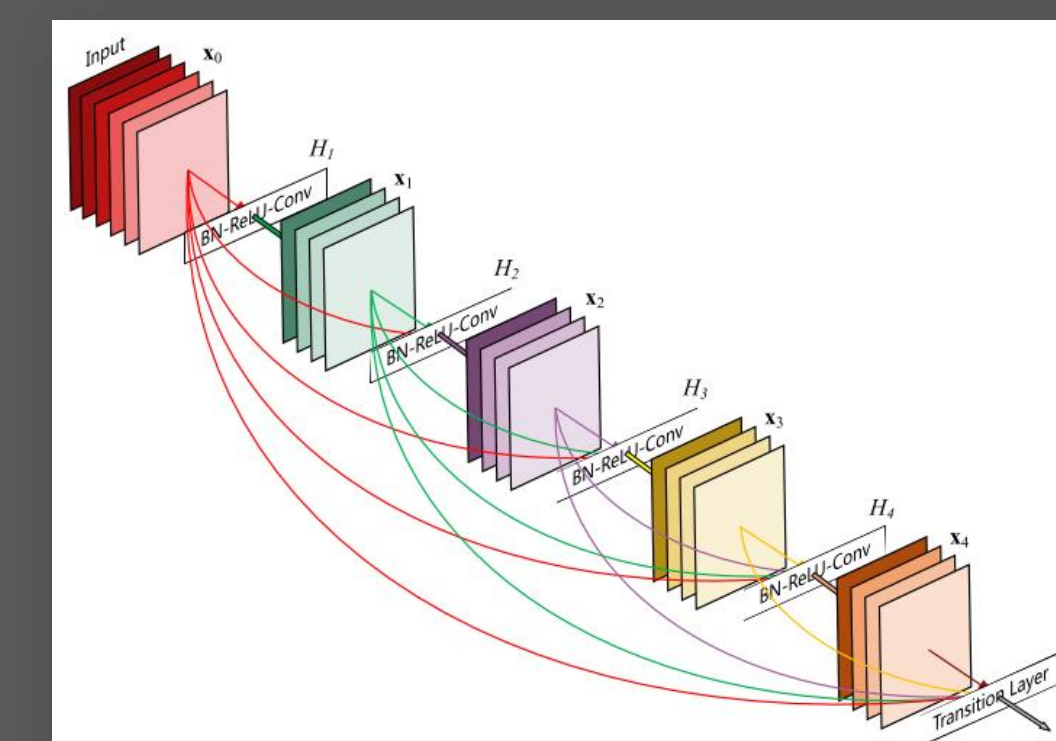
Dataset

- 200 training and 100 testing images for three different kind of images separately
- Mixed all these images together to make a big dataset
- Each image is carefully selected, preprocessed to fit 200 * 200 dimension

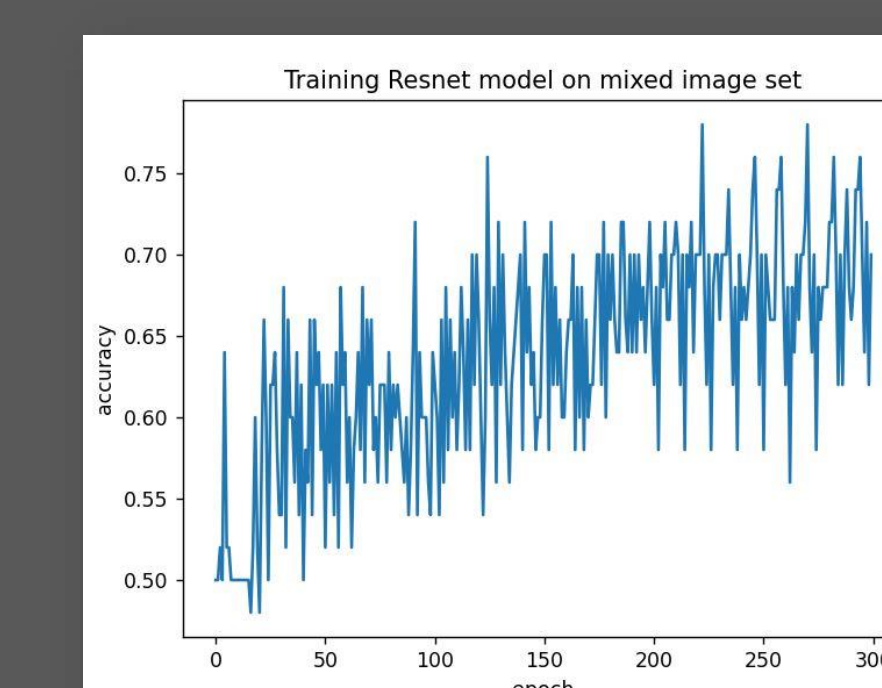
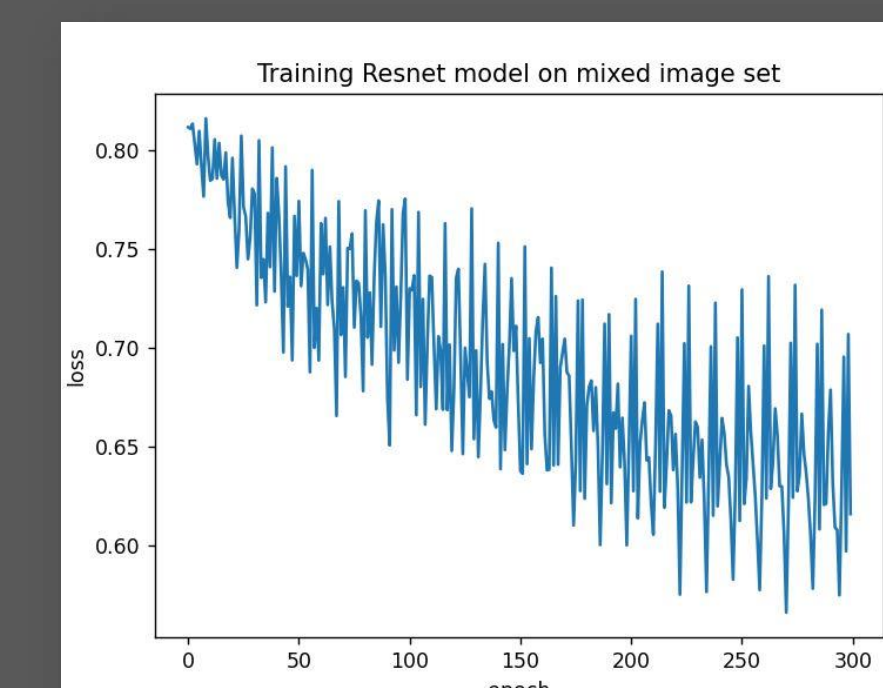


Methods

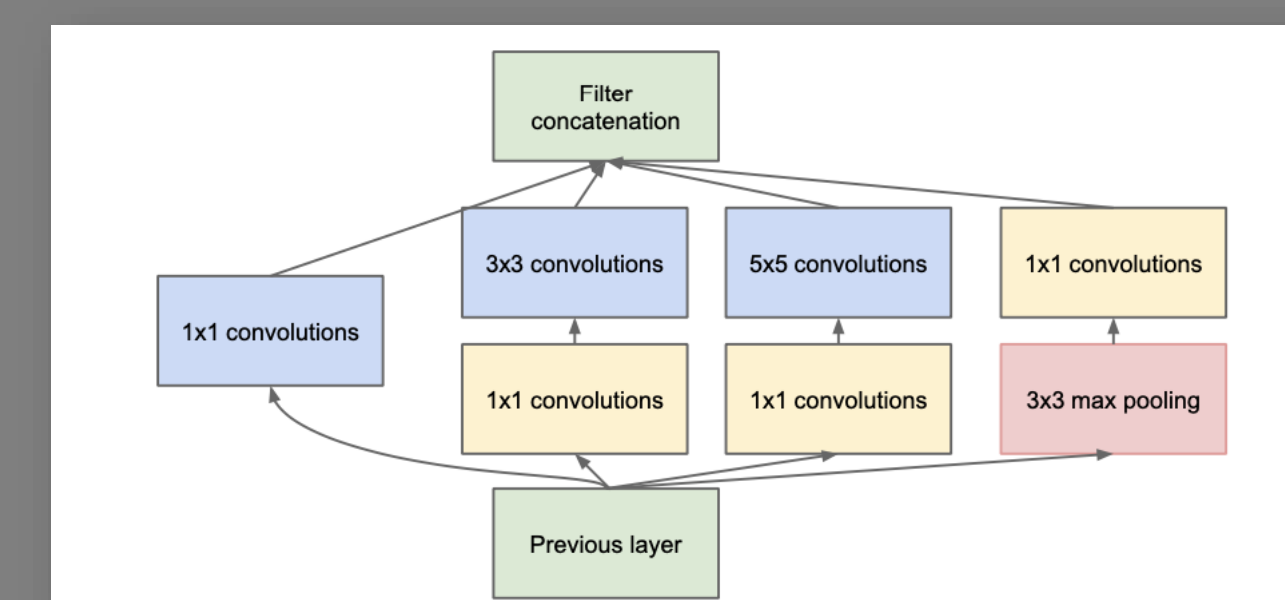
- We'll try to train three different models to complete this task: training a resnet50 model from scratch, fine-tuning pretrained GoogleNet model, using a fine-grained classification model called ViT
- For hyper-parameters, we choose to train 300 epochs in batch-size 100, use cross-entropy loss function and SGD optimizer, set the learning rate 1e-3



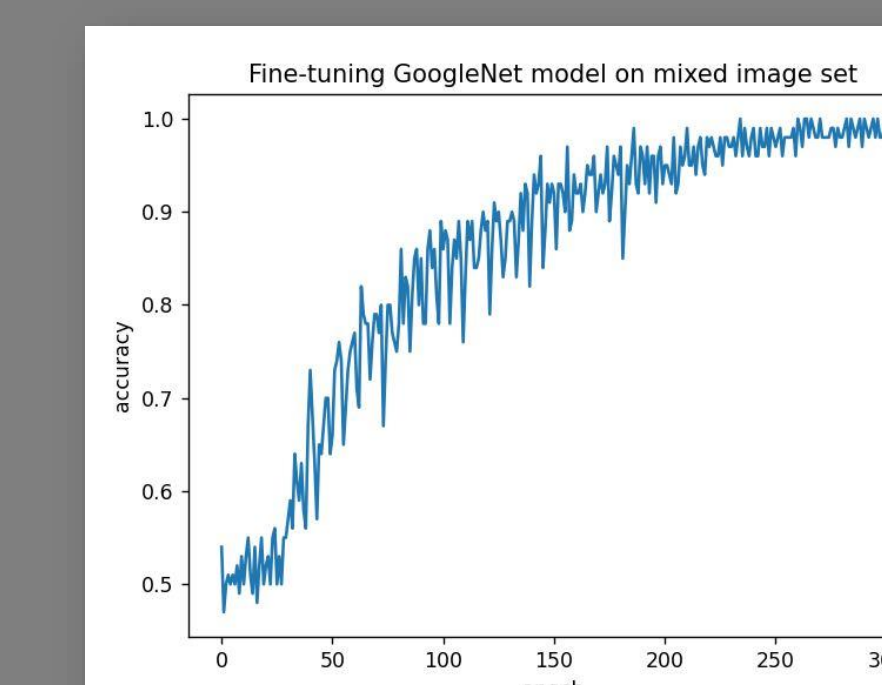
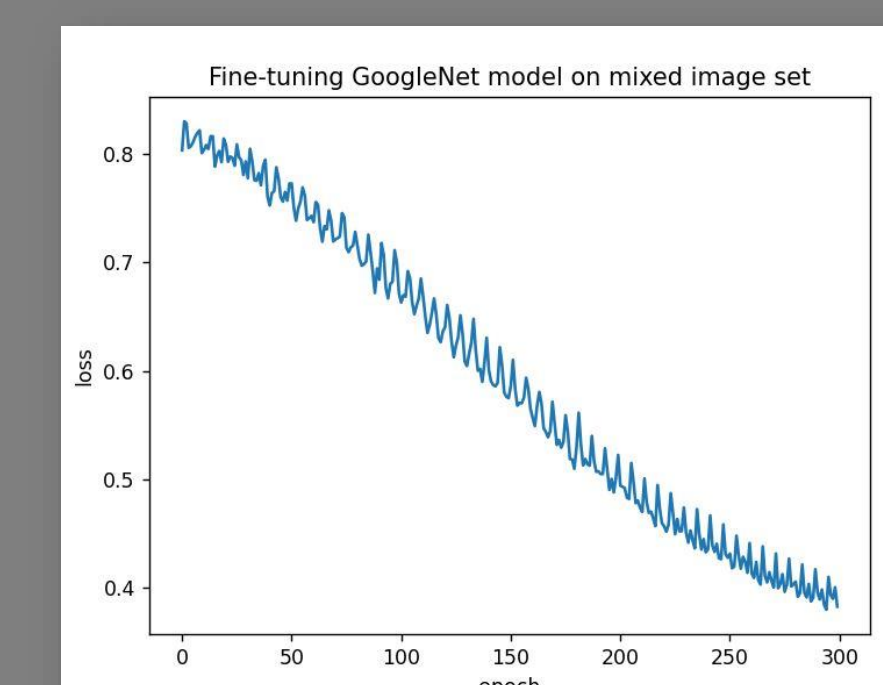
resnet model



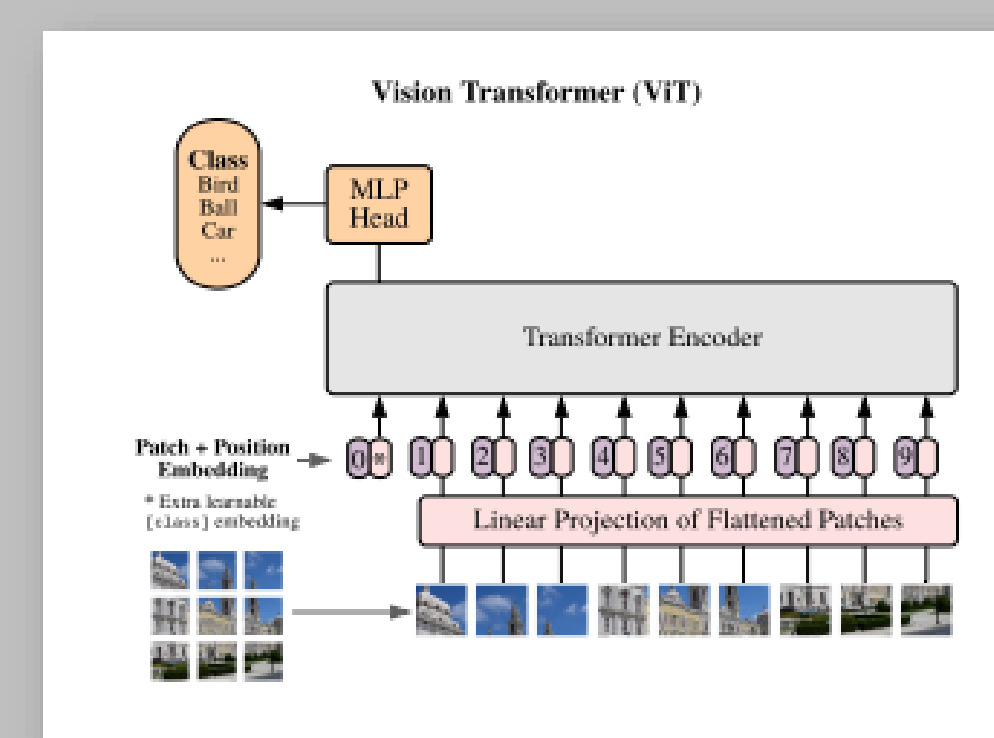
Large disturbance of loss, converged badly



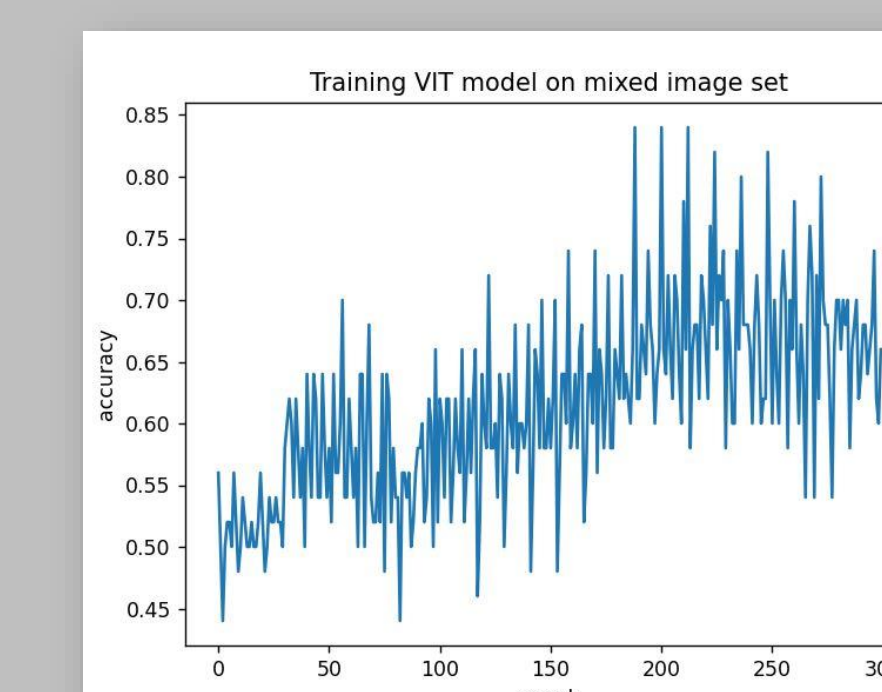
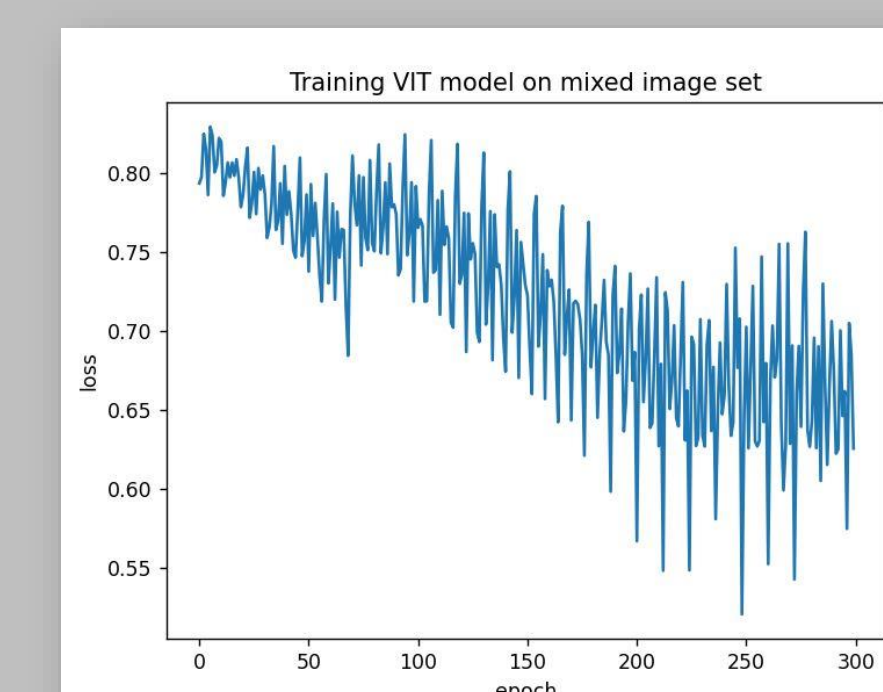
GoogleNet model



Tiny disturbance of loss, converged nicely

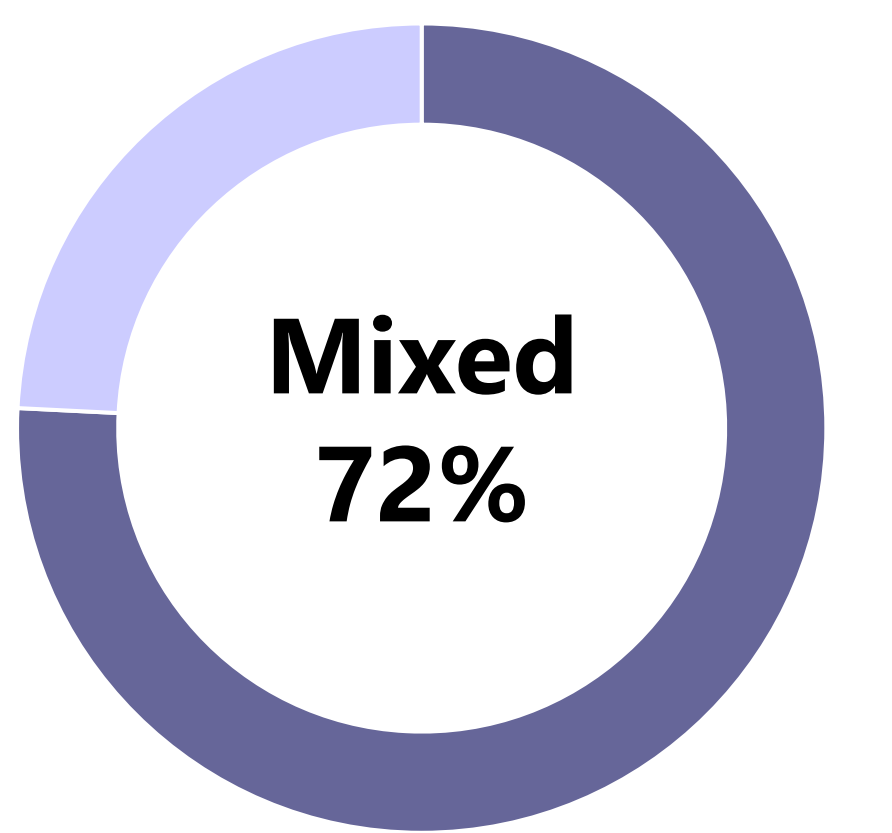
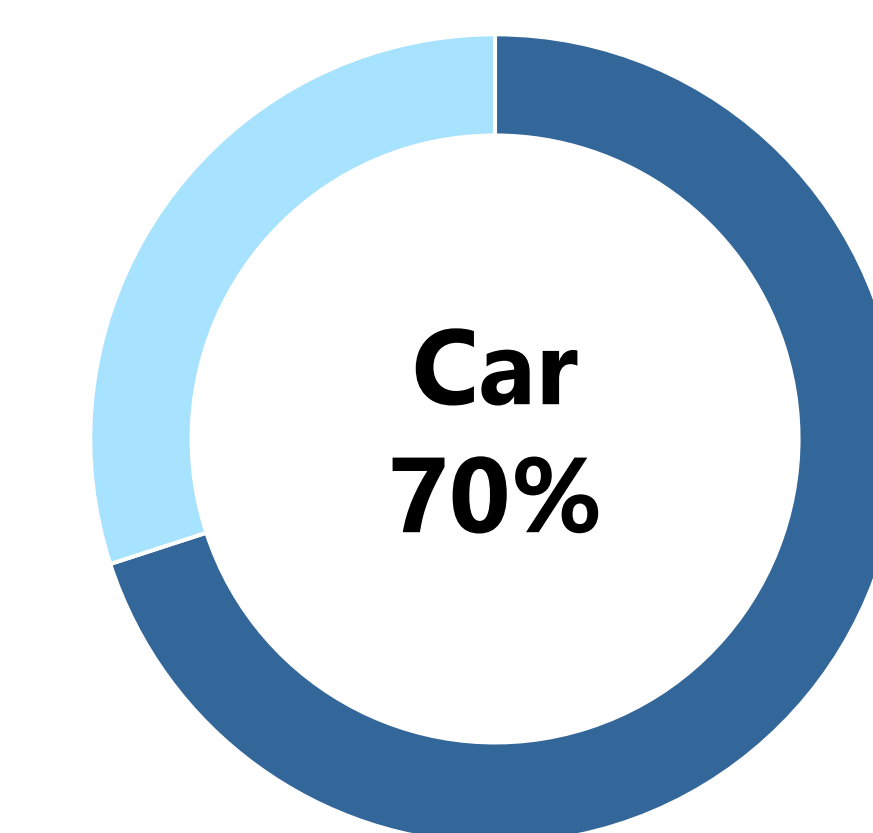
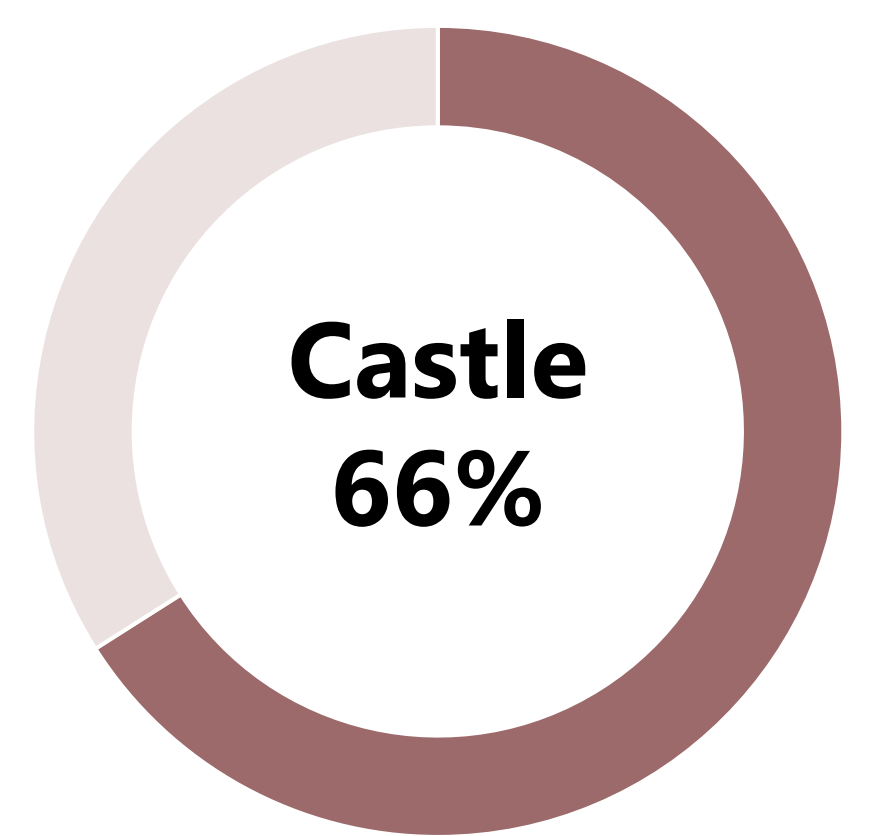
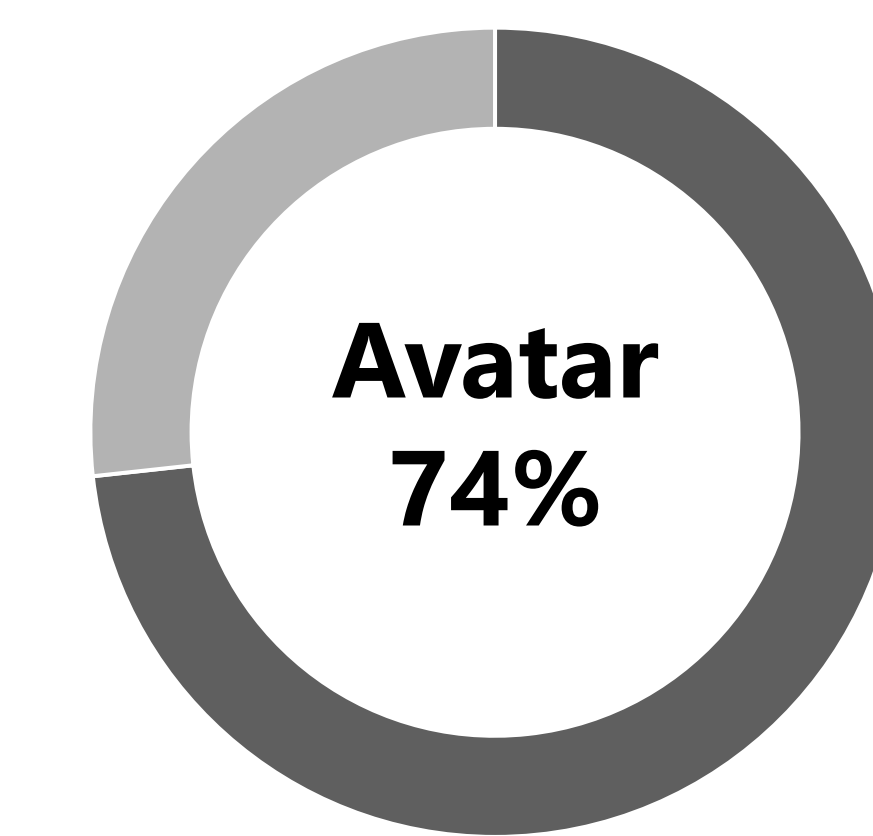


ViT model



Large disturbance of loss, converged badly

Results



Conclusions

1. Neural network could achieve more than 70% accuracy on this task after training, **which means it could really distinguish AI generated image to a certain degree**
2. Fine-tuning GoogleNet outperforms other models on all these dataset
3. Neural network perform best on avatar distinguishing (It may be affected by the quality of the dataset)