Model Inference and Possible Extrapolations

Andrea Covre, Connor Reitz, Jake Hopkins

CS 7644, CS 4644, CS 4644

Abstract

Over the past years, a large amount of research, innovation, and thought has gone into creating deep neural networks that achieve a high level of performance. However, model theft is a topic of recent popularity: the possibility of replicating an existing model with minimal assumptions and thought, thereby eliminating both intellectual effort and monetary costs associated with this long process. We explore whether inferred models can attain similar performance to their stolen counterparts and what consequences this could impose on this field.

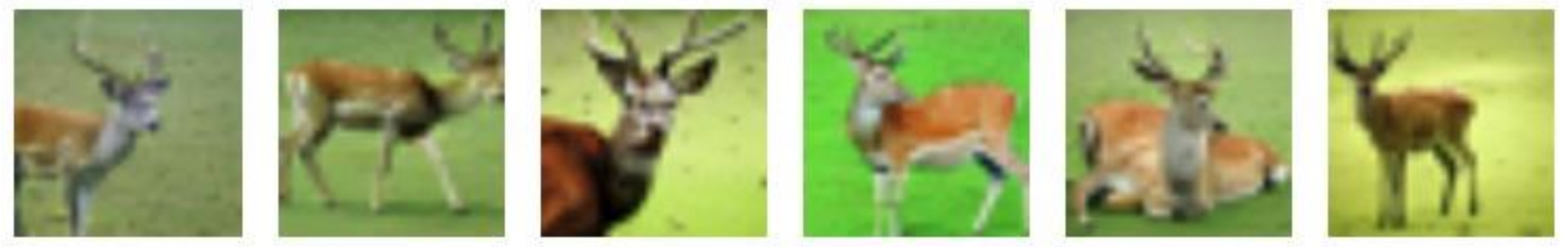
Objectives

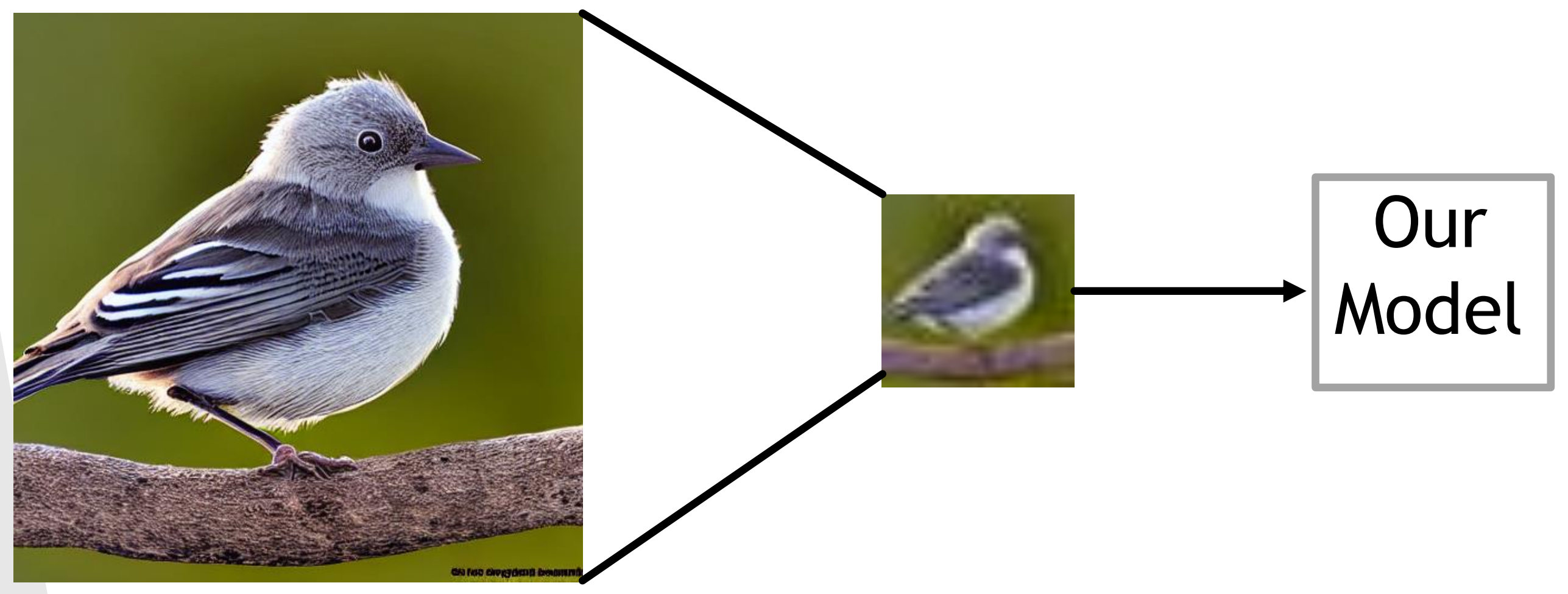
- Model Recreation POC
- Using synthetic data to get real world results

Example Data Created:

Real CIFAR-10's *deer* class examples

Fake CIFAR-10's *deer* class examples





		Prediction	
		Fake	Real
Label	Fake	96.29%	3.71%
	Real	0.29%	99.71%

Results

Resnet56 with CIFAR10 94.06%

		Prediction									
		airplane	automobile	bird	cat	deer	dog	frog	horse	ship	truck
Label	airplane	94.39%	0.00%	1.56%	0.62%	0.00%	0.00%	0.00%	0.31%	2.49%	0.62%
	automobile	0.20%	96.61%	0.00%	0.00%	0.00%	0.00%	0.10%	0.00%	1.00%	2.09%
	bird	1.37%	0.00%	92.96%	0.98%	0.88%	1.56%	1.86%	0.20%	0.20%	0.00%
	cat	0.67%	0.10%	1.25%	87.51%	1.54%	6.15%	1.63%	0.48%	0.10%	0.58%
	deer	0.20%	0.00%	1.01%	1.32%	95.54%	0.41%	0.91%	0.61%	0.00%	0.00%
	dog	0.00%	0.00%	0.42%	6.28%	1.15%	91.20%	0.42%	0.52%	0.00%	0.00%
	frog	0.40%	0.00%	1.20%	1.60%	0.20%	0.50%	95.70%	0.00%	0.00%	0.40%
	horse	0.20%	0.00%	0.20%	0.50%	1.89%	0.70%	0.00%	96.22%	0.30%	0.00%
	ship	2.88%	0.77%	0.10%	0.00%	0.19%	0.00%	0.00%	0.00%	95.49%	0.58%
	truck	1.12%	2.45%	0.51%	0.20%	0.00%	0.00%	0.00%	0.00%	0.61%	95.10%

Resnet56 with Fake-CIFAR10 95.81%

		Prediction									
		airplane	automobile	bird	cat	deer	dog	frog	horse	ship	truck
Label	airplane	96.02%	0.00%	3.69%	0.00%	0.00%	0.00%	0.00%	0.00%	0.30%	0.00%
	automobile	5.14%	73.05%	2.98%	1.44%	0.00%	0.00%	0.51%	0.21%	2.47%	14.20%
	bird	0.69%	0.00%	99.21%	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%
	cat	0.00%	0.00%	0.20%	98.53%	0.10%	0.79%	0.39%	0.00%	0.00%	0.00%
	deer	0.00%	0.00%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
	dog	0.10%	0.31%	0.73%	1.24%	0.21%	96.58%	0.62%	0.21%	0.00%	0.00%
	frog	0.30%	0.20%	0.70%	0.00%	0.00%	0.00%	98.70%	0.00%	0.10%	0.00%
	horse	0.30%	0.00%	0.10%	0.10%	0.40%	1.21%	0.00%	97.89%	0.00%	0.00%
	ship	1.03%	0.00%	0.10%	0.00%	0.10%	0.00%	0.10%	0.00%	98.35%	0.31%
	truck	0.00%	1.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	98.90%

Our Model with CIFAR10 43.05%

		Prediction									
		airplane	automobile	bird	cat	deer	dog	frog	horse	ship	truck
Label	airplane	34.40%	14.80%	15.70%	3.10%	0.50%	1.30%	0.80%	1.10%	23.90%	4.40%
	automobile	1.80%	65.70%	9.40%	1.70%	0.90%	2.10%	0.60%	0.90%	3.60%	13.30%
	bird	3.00%	10.20%	41.60%	10.20%	10.30%	10.60%	1.40%	6.30%	6.10%	0.30%
	cat	2.40%	11.80%	8.10%	51.20%	2.90%	9.90%	0.60%	9.50%	2.90%	0.70%
	deer	1.00%	10.80%	2.50%	16.90%	45.00%	5.10%	0.30%	9.70%	8.10%	0.60%
	dog	0.20%	7.60%	10.40%	24.70%	2.30%	40.80%	0.20%	10.70%	2.90%	0.20%
	frog	0.80%	13.90%	7.10%	30.00%	16.70%	10.70%	4.20%	9.90%	5.30%	1.40%
	horse	0.30%	12.10%	5.90%	13.80%	6.40%	9.40%	0.20%	44.00%	4.00%	3.90%
	ship	7.00%	23.10%	4.70%	2.10%	0.30%	0.20%	0.10%	0.70%	56.00%	5.80%
	truck	1.50%	26.50%	3.20%	4.10%	0.70%	2.80%	0.50%	2.00%	11.00%	47.70%

Our Model with Fake-CIFAR10 98.91%

		Prediction									
		airplane	automobile	bird	cat	deer	dog	frog	horse	ship	truck
Label	airplane	99.00%	0.00%	0.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.80%	0.00%
	automobile	0.00%	99.50%	0.20%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.30%
	bird	0.20%	0.20%	99.40%	0.00%	0.00%	0.10%	0.00%	0.00%	0.10%	0.00%
	cat	0.00%	1.60%	0.20%	97.60%	0.00%	0.40%	0.00%	0.20%	0.00%	0.00%
	deer	0.00%	0.50%	0.00%	0.10%	99.40%	0.00%	0.00%	0.00%	0.00%	0.00%
	dog	0.00%	0.60%	0.10%	0.80%	0.20%	97.90%	0.00%	0.40%	0.00%	0.00%
	frog	0.30%	0.10%	0.90%	0.00%	0.00%	0.00%	98.60%	0.00%	0.10%	0.00%
	horse	0.00%	0.00%	0.00%	0.00%	0.40%	0.30%	0.00%	99.30%	0.00%	0.00%
	ship	0.00%	0.00%	0.20%	0.00%	0.00%	0.00%	0.00%	0.00%	99.50%	0.30%
	truck	0.10%	0.90%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.00%

			Prediction																			
			Fake										Real									
			airplane	automobile	bird	cat	deer	dog	frog	horse	ship	truck	airplane	automobile	bird	cat	deer	dog	frog	horse	ship	truck
Label	Fake	airplane	99.50%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
		automobile	0.00%	98.60%	0.00%	0.10%	0.00%	0.00%	0.10%	0.00%	0.00%	0.10%	0.30%	0.30%	0.00%	0.20%	0.00%	0.00%	0.00%	0.00%	0.20%	0.10%
		bird	0.70%	0.00%	96.70%	0.00%	0.00%	0.00%	1.30%	0.00%	0.00%	0.00%	0.20%	0.00%	1.00%	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%
		cat	0.00%	0.10%	0.10%	94.40%	0.10%	0.10%	0.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	4.90%	0.20%	0.00%	0.00%	0.00%	0.00%	0.00%
		deer	0.00%	0.30%	0.00%	0.00%	98.90%	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	0.60%	0.00%	0.00%	0.00%	0.00%	0.00%
		dog	0.00%	1.70%	0.30%	2.00%	0.40%	69.20%	0.50%	7.60%	0.00%	0.00%	0.20%	0.10%	0.10%	1.10%	0.00%	14.70%	1.90%	0.20%	0.00%	0.00%
		frog	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.90%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%
		horse	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	99.80%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.20%	0.00%	0.00%
		ship	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	95.40%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	4.50%	0.00%
		truck	0.10%	1.40%	0.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	96.80%	0.30%	0.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	1.00%
	Real	airplane	1.30%	0.00%	0.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	88.90%	0.30%	0.50%	2.50%	0.90%	0.40%	0.20%	1.10%	3.50%	0.30%
		automobile	0.00%	0.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.60%	94.50%	0.00%	1.00%	0.30%	0.10%	0.30%	0.40%	1.50%	0.90%
		bird	0.00%	0.00%	0.50%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	10.60%	0.00%	53.90%	7.30%	12.00%	5.50%	4.40%	5.30%	0.50%	0.00%
		cat	0.00%	0.00%	0.10%	0.20%	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	1.30%	0.10%	0.70%	67.80%	6.60%	13.60%	3.90%	4.80%	0.50%	0.30%
		deer	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.30%	0.00%	0.80%	2.00%	88.20%	2.60%	1.70%	3.10%	0.30%	0.00%
		dog	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	0.10%	0.00%	0.00%	0.00%	0.80%	0.10%	0.10%	8.50%	3.80%	78.90%	1.70%	5.60%	0.30%	0.00%
		frog	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.30%	0.00%	0.00%	0.00%	0.80%	0.00%	0.30%	3.90%	3.10%	0.90%	89.80%	0.50%	0.30%	0.10%
		horse	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	0.00%	0.00%	0.80%	0.00%	0.10%	1.90%	1.90%	2.40%	0.00%	92.80%	0.00%	0.00%
		ship	0.10%	0.10%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	0.00%	5.10%	0.90%	0.20%	0.50%	0.30%	0.10%	1.10%	0.40%	90.90%	0.20%
		truck	0.00%	0.30%	0.00%	0.00%	0.00%	0.00%	0.00%	0.10%	0.00%	0.10%	5.30%	8.90%	0.10%	1.40%	0.30%	0.30%	0.70%	3.20%	3.00%	76.30%

Attempts on Improving Our Model with CIFAR10

Using fake generated data to train our model on real data examples only achieves an accuracy of 43%. Here is some further experimentation we did in an attempt to improve this accuracy:

Data Dropout: Optimization of training set by filtering harmful images

This method was an attempt to dropout outlier generated data that was harmful to the loss function in order to decrease overall error. This method was not successful - likely due to machine constraints and over-expected noise influence.

Inclusion of Different AI Generated Images

This was an attempt to improve performance by generating data from different generative models, in the hopes that getting training data from a variety of sources would improve accuracy. This method was not successful as well - although the images generated by different models look similar, they do not contain the features that ResNet-56 looks for.

Conclusions

Ultimately, our research provides promising evidence of the first steps towards model recreation with generated data. While 43% accuracy may not seem too high, the semblance of accurate results proves that it is theoretically possible.

Contact Information

Andrea Covre: andrea.covre@gatech.edu

Connor Reitz: creitz@gatech.edu

Jake Hopkins: jhopkins39@gatech.edu