Analysis the Robustness of DNN against Images Changes

Chiaai Lin, Qingyu Zhu and Yuxin Jiang

December, 2019

1. Datasets

We use the CIFAR-10 datasets as our dataset to train and test our model. The CIFAR-10 dataset is a collection of images that are commonly used to train machine learning and computer vision algorithms. It is one of the most widely used datasets The CIFAR-10 dataset contains 10 different classes of pictures, including airplanes, cars, birds, cats, deer, dogs, frogs, horses, ships, and trucks. (Vehicles: airplanes, cars, ships, and trucks. Animals: birds, cats, deer, dogs, frogs, and horses.) There are 6,000 images of each class, and the total number of images in the datasets is 60000. So we use 5000 images for each class to train the model and 1000 images for each class to test the model.

Since the images in CIFAR-10 are low-resolution (32x32), we can quickly teach the computer to recognize the object in these images. Also, because we can classify the classes in the CIFAR-10 into two categories as mentioned above(vehicles and animals), we think that our DNN will meet some adversarial examples.

2. Models and Methods

1. self-trained model

Our DNN is formed by 1 input layer, 2 hidden layers, and 1 output layer.
If using our DNN to test the original images, the accuracy is 53% on average.

- pre-trained model

Using the model provided by GLUON, the accuracy is 99% for the original images.

3. Experiments

We use 5000 images as the training set, and use 1000 images as the testing set. The following are our test results, the column shows the testing dataset of classes, the row shows the results in the percentage of each testing dataset.

- Original images testing results(self-trained model vs pre-trained model)

percentage	plane	car	bird	cat	deer	dog	frog	horse	ship	truck
plane	56.1	4.8	5.6	2.1	1.6	1.1	3.0	3.6	18.2	3.9
car	2.2	65.6	0.5	3.4	0.2	1.9	2.0	2.6	5.5	16.1
bird	7.2	2.6	35.0	8.2	6.7	11.4	13.4	10.9	2.1	2.5
cat	3.7	1.9	5.3	30.4	1.7	26.5	14.5	9.1	2.6	4.3
deer	4.5	2.0	12.7	5.0	27.8	7.0	19.5	17.3	2.4	1.8
dog	2.0	1.2	5.4	15.1	4.5	47.8	9.5	10.7	2.6	1.2
frog	0.4	2.5	6.0	7.1	3.6	5.4	69.5	2.5	0.8	2.2
horse	2.9	1.1	3.1	3.5	4.0	8.9	2.7	69.0	1.3	3.5
ship	7.0	9.0	1.1	2.0	1.1	2.4	1.4	1.7	69.4	4.9
truck	2.8	15.4	0.4	3.4	0.4	2.2	2.2	5.9	7.7	59.6

The average accuracy is 53.02%.

And an interesting result is that we found our DNN is more sensitive to the vehicles category than the animal category.

And this result also happened when we changed some types about the image, including blur, contrast and brightness.

- Kernel Convolution(blur)

percentage	plane	car	bird	cat	deer	dog	frog	horse	ship	truck
plane	33.2	6.9	10.3	10	3.6	4.5	7	2.6	20.5	1.4
car	2.5	51.8	0.4	7.1	0.7	2.3	4.5	1.3	13.8	15.6
bird	22.8	5.7	26.9	9.2	5.1	7.3	5.3	2.8	13.2	1.7
cat	18.4	10.6	7.5	21.7	1.6	14.2	3.2	5.4	11.3	6.1
deer	23.5	5.2	18.5	8.4	10.1	2.7	5.9	3.6	21.1	1
dog	19.1	7.1	7.5	13.1	3.5	27.7	2.4	5.3	10.3	4
frog	28.3	27	7.9	5.4	1.4	2.6	7.6	0.9	13.1	5.8
horse	12.3	12.2	4.1	10.8	4	10	1.2	20.9	21.2	3.3
ship	19.5	16.6	2.4	10.3	4.8	6.3	4.6	3.6	28.4	3.5
truck	2.9	27.8	0.9	11.9	0.7	2	2.6	3.4	14.4	33.4

The average accuracy is 26.17%.

percentage	plane	car	bird	cat	deer	dog	frog	horse	ship	truck
plane	30.4	5.1	11.5	10.7	5	4.2	9.3	1.9	21.1	0.8
car	2.8	45.8	0.5	8.9	1	2.1	6.9	1.3	16.1	14.6
bird	22.3	5.2	28.6	8.6	5	6.2	6.5	2.6	13.3	1.7
cat	20.4	8.8	8.1	22.1	2	12.4	3.8	5.4	11.1	5.9
deer	22.4	8.8	17.2	8.9	8.9	4	4	5	19.2	1.6
dog	18.9	6.2	8.7	13.3	3.6	26.2	2.8	4.6	12.5	3.2
frog	29.3	25.1	9.5	4.5	1.4	2.9	8	1	13.4	4.9
horse	13.7	10.7	5.4	11.9	3.5	9.7	1.8	17.1	23.6	2.6
ship	19.4	12.5	3.3	11.4	5.6	6.6	6.2	3.5	28.3	3.2
truck	4.5	23.1	0.7	13.4	0.9	2.1	3.6	3	17	31.7

The average accuracy is 24.71%.

percentage	plane	car	bird	cat	deer	dog	frog	horse	ship	truck
plane	29.5	3.5	11.8	11.8	5.8	3.4	10.9	1.1	21.5	0.7
car	3.2	35.5	0.8	9.7	2.3	2.3	11.1	0.9	19.7	14.5
bird	23.1	4.6	28.3	8.3	5.6	4.6	7.8	2.6	13.7	1.4
cat	21.7	7.9	10.0	22.1	2.1	10.3	4.3	4.2	12.8	4.6
deer	22.8	6.9	18.1	8.8	10.1	3.2	4.9	3.7	20.2	1.3
dog	19.8	6.0	10.2	13.6	4.1	22.1	3.4	3.6	15.0	2.2
frog	31.4	20.6	11.6	4.1	1.5	2.3	8.6	1.0	14.6	4.3
horse	15.4	8.9	5.8	12.4	5.0	8.3	2.9	12.9	26.5	1.9
ship	18.7	8.6	4.5	13.2	7.0	5.9	8.7	2.4	28.7	2.3
truck	5.2	19.7	1.1	16.1	1.9	2.7	4.5	2.4	19.6	26.8

The average accuracy is 22.46%.

percentage	plane	car	bird	cat	deer	dog	frog	horse	ship	truck
plane	28.1	2.7	12.3	12	6.3	2.6	12.6	0.9	22	0.5
car	3.8	28.4	1.9	11.3	4	2.4	13.5	1	22.4	11.3
bird	24.2	3.9	27.7	7.7	5.7	3.2	9.2	1.9	15.2	1.3
cat	24	6.3	11	20.8	3	8.3	5.2	3.4	14.5	3.5
deer	23.5	5.2	18.5	8.4	10.1	2.7	5.9	3.6	21.1	1
dog	22.3	5.2	11.2	13.9	4.9	18.6	4	1.9	16.9	1.1
frog	35.7	17.5	12.6	4.1	1.5	1.7	7.7	1	15.1	3.1
horse	18	7.2	6.4	14.1	6.1	6.1	4	8	28.3	1.8
ship	19.4	6.3	4.9	14.9	7.7	5.1	10.8	1.9	27.8	1.2
truck	6.4	17.2	2	17.1	2.1	2.8	6	2.4	22.6	21.4

The average accuracy is 19.86%.

- Brightness

64(brightest)

percentage	plane	car	bird	cat	deer	dog	frog	horse	ship	truck
plane	58.5	1.2	2.8	6.6	0.5	3.5	0.6	0.6	25.4	0.3
car	17.1	26	0.5	6.5	0.6	3.2	1.7	0.8	39.8	3.8
bird	47.3	1.3	13.9	4.7	1.1	4.9	0.5	0.9	25.3	0.1
cat	40.6	2.8	5.7	11.8	1	11.2	1.1	1	22.5	2.3
deer	46.7	1.5	5.7	4.9	1.7	2.8	0.4	1	35.1	0.2
dog	40.8	0.7	7	9.4	0.7	19.2	0.3	1.3	19.8	0.8
frog	53.2	7.1	3.5	2.6	0.1	3.3	1.4	0.2	26.7	1.9
horse	38.7	2.8	3.3	5.8	1.6	5	0	5.5	36.6	0.7
ship	36.3	4.9	0.7	5.1	0.5	4.9	0.4	0.8	45.9	0.5
truck	15.3	17.9	0.6	6.5	0.4	1.7	1	0.3	48.5	7.8

The average accuracy is 19.17%.

32(brighter)

percentage	plane	car	bird	cat	deer	dog	frog	horse	ship	truck
plane	46.6	3.8	5.9	7.8	1.5	4.4	3	1.5	24.7	0.8
car	5.5	42.6	0.7	6.1	0.7	2.7	3	1.5	25.8	11.4
bird	34.8	3.2	19.4	7.1	3.5	6.9	2.5	1.9	19.5	1.2
cat	27.3	5.8	6.5	16.8	1.3	14.4	2	3.5	18.1	4.3
deer	31.4	5.5	10.7	6.2	5.5	4.2	1.3	2.7	31.7	0.8
dog	26.8	3.2	6.8	10.7	1.9	26.5	1.1	3.8	16.2	3
frog	36.6	18.2	5.6	3.6	0.8	3.9	5	0.4	22.2	3.7
horse	22.3	7.9	4	7.1	2.7	8.6	0.3	14.1	30.9	2.1
ship	26.1	11.3	1.1	7.5	1.8	6.8	1.2	2.3	38.9	3
truck	7.5	23.8	0.6	7.9	0.5	2.4	1.5	2	30	23.8

The average accuracy is 23.92%.

-32(darker)

percentage	plane	car	bird	cat	deer	dog	frog	horse	ship	truck
plane	21.1	13.5	10.3	12.5	5.3	5.1	10.6	5.1	13.3	3.2
car	1.5	63.9	0.5	6.5	0.4	1.1	4.2	1.7	4.6	15.6
bird	16.7	12	24.3	14.3	3.6	7.1	5.8	6.2	6.6	3.4
cat	10.6	18	6.4	26.8	1.4	11.3	3.8	8.6	4.1	9
deer	17.4	20.1	14.7	13.4	4.9	4.3	4.7	9.7	8.1	2.7
dog	12.1	16.6	7.1	17	2.7	23.6	2.8	8.2	4.1	5.8
frog	19.7	41.5	7.3	9.7	0.6	2.1	7.9	1.3	4.3	5.6
horse	8.1	19.8	4.4	14.1	1.9	8.6	2.3	25.4	9.4	6
ship	13.2	29.2	1.9	12.8	3.5	5	8.9	4.7	15.8	5
truck	1.5	38.6	0.4	14	0.2	1.7	2	3	4.7	33.9

The average accuracy is 24.76%.

-64(darkest)

percentage	plane	car	bird	cat	deer	dog	frog	horse	ship	truck
plane	10.6	22.7	8	18.9	3.2	3.8	12.9	9.5	6.5	3.9
car	0.8	71.7	0.3	8.1	0.2	0.7	3.9	1.6	1	11.7
bird	13.6	21.7	16.3	21.6	1.9	5.2	6.1	7.6	1.5	4.5
cat	6.5	27.7	6.1	31.3	0.6	7.5	2.4	8.7	1.1	8.1
deer	12.6	35.1	9.7	18.8	2.4	2.1	3.5	9.9	2.2	3.7
dog	6.2	25.1	8.1	22.6	1.6	14.7	4.2	10	1.1	6.4
frog	11.3	54.8	5.4	14.5	0.4	1.7	5.3	2	1.2	3.4
horse	4.5	30.7	3.6	21	1.3	4.2	2.9	21.5	3.4	6.9
ship	6.6	45.8	2.2	17.8	0.9	2.2	10.4	4.7	4.6	4.8
truck	0.7	47.2	0.9	17.8	0.1	0.8	1.8	3	0.7	27

The average accuracy is 20.54%.

- Contrast

High contrast

percentage	plane	car	bird	cat	deer	dog	frog	horse	ship	truck
plane	51	4.5	4.5	7.4	0.8	4.3	1.3	1.1	24.3	0.8
car	8.2	43.4	0.3	5.8	0.3	2.7	2.1	1.2	25.5	10.5
bird	40.1	3.7	13.8	6.5	1.6	7.2	0.9	2.3	22.8	1.1
cat	31.1	8.8	4.1	14.1	0.5	14.2	1	3.2	19	4
deer	36.4	6.6	7.1	5.9	2.2	5.2	0.6	2.3	32.8	0.9
dog	30	4.5	5.7	10.5	0.4	24.1	0.5	3.7	18	2.6
frog	40.7	19.7	3.2	3.5	0.1	4.8	1.8	0.8	22.1	3.3
horse	27.2	8.6	2.9	5.5	1.3	8.7	0.3	13	30.7	1.8
ship	29.3	11.7	0.8	6.4	0.5	6.6	0.4	2.2	39.5	2.6
truck	8.4	27.1	0.5	6.6	0.7	2.4	1	1.4	31.6	20.3

The average accuracy is 22.32%.

Low contrast

percentage	plane	car	bird	cat	deer	dog	frog	horse	ship	truck
plane	5.3	21.7	10	22.6	3.3	1	23.4	10.7	1.7	0.3
car	1	70.2	0.4	11.1	0.2	0	12.2	1.9	0.2	2.8
bird	10.2	19.7	18.1	29.3	1.1	0.8	10.4	8.6	0.1	1.7
cat	5.5	23.3	8.2	40	0.9	1.7	7.9	8.7	0.1	3.7
deer	14.2	26.8	13.6	25.6	1	0	9.3	8.4	0.2	0.9
dog	4.8	22	11.1	30.5	2	3.3	14.3	10	0	2
frog	11.2	49.1	7.6	20.7	0.2	0.3	7.3	2	0.1	1.5
horse	4	28.7	7.1	30.4	0.7	0.7	7.2	17.5	0.7	3
ship	3.6	44.2	3.3	22.5	0.5	0.2	19.6	5	0.3	0.8
truck	0.5	58.1	1.4	23.1	0.2	0.1	5.5	2.8	0.2	8.1

The average accuracy is 17.11%.

We have some observation of our test data.

Our average accuracy would decrease in all image changes, including kernel convolution, contrast and brightness. The interesting result is that more blurry, higher contrast and brighter the image, the probability that the image would be classified as “plane” and “ship” increase, and if lower contrast and darker, the probability that the image would be classified as “car” and “cat” increased.

Adversarial examples could be worrisome because they make machine learning models vulnerable to attacks. The intentional feature perturbations in them have already been shown to fool a state-of-the-art model into making a false prediction very easily.

4. Conclusion and Future Work

For the part of future work in our project,

(1) Train a more sophisticated DNN, for example, with more layered of hidden layers so that they can learn features at various levels of abstraction.

(2) Apply cross-validation to test the accuracy on the test set to estimate the process that how well our model had been trained and to estimate other model properties (mean error for numeric predictors, classification errors for classifiers, recall, and precision)

(3) Test DNN’s robustness against more image changes like rotate and saturation.