Initial Project Proposal

Group: Qingyu Zhu, Chiaai Lin, Yuxin Jiang

Nov. 2019

1. Project Topic (Choose One of Them)

Plan A: Stealing a DNN model

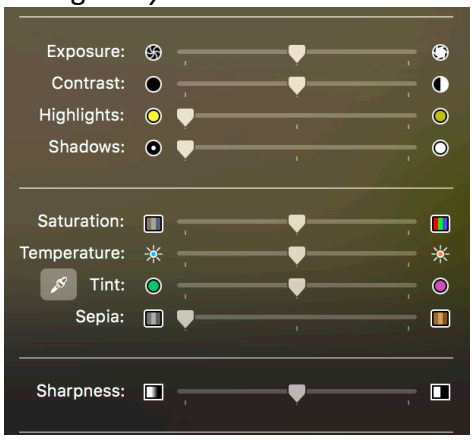
- has some problem on getting the result of the api
- retraining a DNN is little hard

Plan B: Analysis the Robustness of a DNN model

- Have trained a DNN model based on CIFAR10 database with Pytorch
- The trained DNN model involve ten classes:
    - animals: cat, bird, deer, dog, frog, horse
    - vehicles: car, plane, ship, truck
- The trained DNN model can be used to classify a given image as one of the ten classes.
- When we test model, the output is: [predict class: probability of the class]
- the class with the highest probability is considered as the final prediction result
- Some predications are correct, but some are not.

2. To do:

- Train a DNN model based on CIFAR10 database with Pytorch
- Adjust original images color from different aspect.
- Change *only one* variable at a time:



3. Initial Results:

- Robustness of the trained DNN: Some predications are correct, but some may be not.

bird image

```
plane 2.573148250579834
car 0.07344329357147217
bird 0.8978562355041504
cat -0.04024447500705719
deer 0.03017526865005493
dog -0.5593113303184509
frog -3.0086777210235596
horse -1.154160737991333
ship 1.2133240699768066
truck -1.1453443765640259
Predicted: plane
```

Predict as plane(X) with 2.57 score

```
plane 0.55029296875
car -7.597825050354004
bird 1.4370089769363403
cat 2.6692793369293213
deer 3.472820997238159
dog 4.424261093139648
frog -2.2502551078796387
horse 13.335748672485352
ship -6.219541072845459
truck -8.24459457397461
Predicted: horse
```

horse image

Predict as horse(V) with 13.33 score

- Analysis the robustness of the trained DNN model by calculating the rate of change of probability value for each class. Compare the rate of probability change and analyze the sensitivity of the model to changes.

- Initial Result:

Original Picture

No Change

```
plane 1.0106406211853027
car -4.2074198722839355
bird 3.8952109813690186
cat 0.8701410293579102
deer 2.2867722511291504
dog 0.053627416491508484
frog 1.1567344665527344
horse -0.0016703903675079346
ship -1.8023881912231445
truck -3.5313305854797363
Predicted: bird
```

Predicted as bird with 3.895 score

Original Picture

Increase Exposure

```
plane 2.6719863414764404
car -4.269985675811768
bird 4.000802993774414
cat 1.0832431316375732
deer -0.035881638526916504
dog 0.15641002357006073
frog -0.7905476093292236
horse 0.6942473649978638
ship -0.07666075229644775
truck -4.385523319244385
Predicted: bird
```

Predicted as bird with 4.000 score, (4-3.895)

Original Pict

Increase Contrast.

```
plane -1.2911736965179443
car 1.6019207239151
bird -0.2997123897075653
cat 1.4108126163482666
deer 1.3238649368286133
dog 0.6601666808128357
frog 0.4245148003101349
horse 1.882447361946106
ship -4.462400913238525
truck 0.8196033239364624
Predicted: horse
```

Predicted as horse with 1.882 score