



Lab: Cifar10 Aside

Ok, I lied! Before we go over `feature_extraction.py` I want you to do something first ...

Before you try feature extraction on pretrained models I'd like you to take a moment and run the classifier you used in the Traffic Sign project on the Cifar10 dataset. Cifar10 images are also (32, 32, 3) so the main thing you'll need to change is **the number of classes from 43 to 10**. Cifar10 also doesn't come with a validation set, so you can randomly split training data into a training and validation.

You can easily download and load the Cifar10 dataset like this:

```
from keras.datasets import cifar10
(X_train, y_train), (X_test, y_test) = cifar10.load_data()
# y_train.shape is 2d, (50000, 1). While Keras is smart enough to handle this
# it's a good idea to flatten the array.
y_train = y_train.reshape(-1)
y_test = y_test.reshape(-1)
```

You can then use sklearn to split off part of the data into a validation set:

```
from sklearn.model_selection import train_test_split
X_train, X_valid, y_train, y_valid = train_test_split(X_train, y_train,
                                                    test_size=0.3, random_state=42, stratify = y_train)
```

The Cifar10 dataset contains 10 classes:

Here are the classes in the dataset, as well as 10 random images from each:

airplane



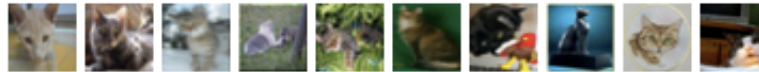
automobile



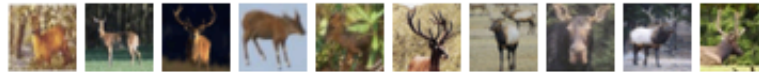
bird



cat



deer



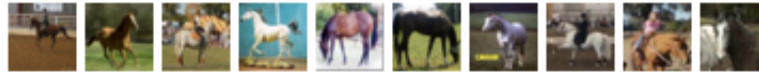
dog



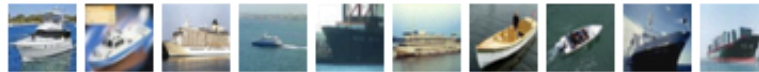
frog



horse



ship



truck



[Overview of the Cifar10 dataset.](#) Source: Alex Krizhevsky.

While the German Traffic Sign dataset has more classes, the Cifar10 dataset is harder to classify due to the complexity of the classes. A ship is drastically different from a frog, and a frog is nothing like a deer, etc. These are the kind of datasets where the advantage of using a pre-trained model will become much more apparent.

Train your model on the Cifar10 dataset and record your results, keep these in mind when you train from the bottleneck features. Don't be discouraged if you get results significantly worse than the Traffic Sign dataset.

NEXT