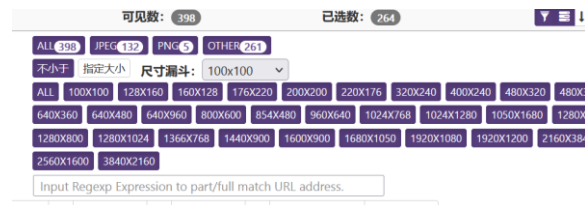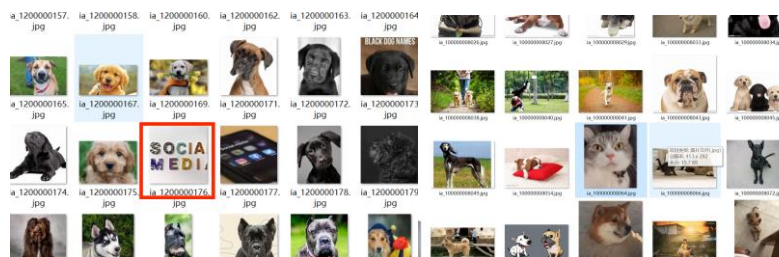# Assignment1 – Make a dataset

I chose the most basic build dataset samples in machine learning, cats and dogs, which are the most familiar and accessible image samples, and being contrasting as domestic pets, simple samples are easier to understand when learning the Ai course.
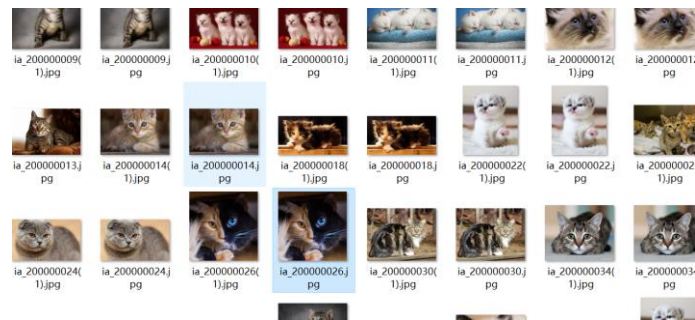
Use the browser to search for photos of cats and dogs, referring to the classic case of algorithms for cats and dogs. Use ImageAssistant plugin to filter out photos below 100px and download images of cats and dogs respectively.

In the total number of samples to reach about 1600 began to filter, after the discovery of duplicate images and irrelevant images, manually deleted to each sample to 550, the total number to 1100.
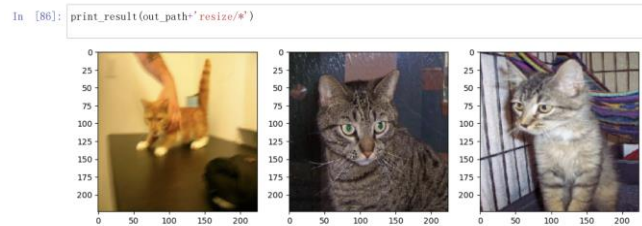

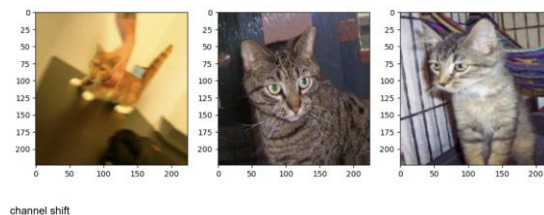*Irrelevant pictures Logos or cats appear*

Duplicate images

Use your browser to search for photos of cats and dogs, referring to the classic case of algorithms for cats and dogs. Use ImageAssistant plugin to filter out photos below 100px and download images of cats and dogs respectively. In the total number of samples to reach about 1600 to start screening, after the discovery of duplicate pictures and irrelevant pictures, manually deleted to each sample to 550, the total number to 1100.
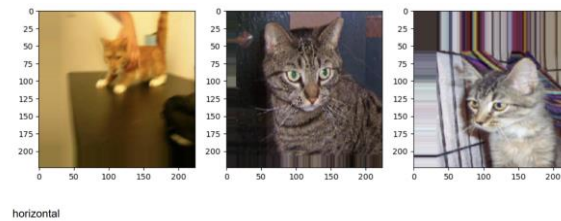
During the deletion process, it was found that the irrelevant images included not only logos and icons, but also screenshots of films and movies related to the words dog and cat, as well as images in which the main subject is a dog or a cat, but there are characters appearing in the images. As well as in the cat search images appear in the picture of the dog and other issues.
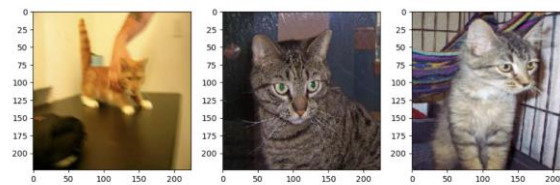

*Same img resize 224*224*
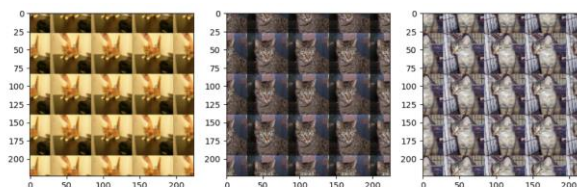

channel shift

rotation_range

horizontal

shift

After the use of ai function, modify the size of the cat, dog pictures, and add more data samples, tilt the picture, picture cropping, increase data. Increasing the number of images is the number, when the sample is not enough when doing deep learning, the dataset of different transformation angles expands the number of samples, and it is also more convenient to improve the accuracy of machine learning.

Horizontal

Fill mode