Image Recognition with Machine Learning on Python
Problem definition

The human brain is making vision easy for most of us, just like the phrase *What you see is what you get* states. It doesn't take any effort for humans to tell apart a dog, a hot dog or a flying saucer. On the other hand, this whole process is quite difficult to be imitated by a computer; it only seems easy because the human brain is incredibly well designed when it comes to recognize patterns of images.

Different varieties of the recognition problem are described in the literature:

Object recognition (also called object classification) – one or several pre-specified or learned objects or object classes can be recognized, usually together with their 2D positions in the image or 3D poses in the scene. See [1] and [4].

Identification – an individual instance of an object is recognized. Examples include identification of a specific person's face or fingerprint, identification of handwritten digits, or identification of a specific vehicle.

Detection – the image data are scanned for a specific condition. Examples include detection of possible abnormal cells or tissues in medical images or detection of a vehicle in an automatic road toll system. Detection based on relatively simple and fast computations is sometimes used for finding smaller regions of interesting image data which can be further analyzed by more computationally demanding techniques to produce a correct interpretation. See [2].

We will use the currently best algorithms, which are based on convolutional neural networks (CNN), to prove how an Image Classifier works. Performance of convolutional neural networks, on the ImageNet[1] tests, is now close to that of humans. See [3]. The best algorithms still struggle with objects that are small or thin, such as a small ant on a stem of a flower or a person holding a quill in their hand. They also have trouble with images that have been distorted with filters (an increasingly common phenomenon with modern digital cameras). By contrast, those kinds of images rarely trouble humans. Humans, however, tend to meet difficulties in another manner; for example, they are not good at classifying objects into fine-grained classes, such as the particular breed of dog or species of bird, whereas convolutional neural networks handle this with ease.

---

[1] An illustration of CNN capabilities is given by the ImageNet Large Scale Visual Recognition Challenge; this is a benchmark in object classification and detection, with millions of images and hundreds of object classes.

We will train our model from scratch and then classify a set of data containing cars and planes. *Train data*: it contains 200 images of each cars and planes, i.e. there are 400 images in the training dataset. *Test data*: it contains 50 images of each cars and planes i.e. there are 100 images in the testing dataset.

The dataset and it's arrangement:

- Train
    - Cars
        - 1.jpg
        - 2.jpg
        - 3.jpg

          .

          .

          .
    - Planes
        - 1.jpg
        - 2.jpg
        - 3.jpg

          .

          .

          .
- Validation
    - Cars
        - 1.jpg
        - 2.jpg
        - 3.jpg

          .

          .

          .
    - Planes
        - 1.jpg
        - 2.jpg
        - 3.jpg

          .

          .

          .

References

[1] Aude, Oliva and Torralba, Antonio , "The role of context in object recognition," *Trends in cognitive sciences,* vol. 11, no. 12, pp. 520-527, 2007.

[2] Girshick, Ross; Donahue, Jeff; Darrell, Trevor and Malik, Jitendra, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 2014.

[3] Krizhevsky, Alex; Sutskever, Ilya and Hinton, Geoffrey, "ImageNet Classification with Deep Convolutional," in *Proceedings of the conference, "Neural Information Processing Systems 2012."*, Toronto, 2012.

[4] Linde, Oskar and Lindeberg, Tony, "Composed Complex-Cue Histograms: An Investigation of the Information Content in Receptive Field Based Image Descriptors for Object Recognition," *Computer Vision and Image Understanding,* vol. 116, no. 4, pp. 538-560, 2012.