Math Matters

Apply It.

The math behind... Image Recognition

Technical terms used:

Thresholding, wavelet transform, feature extraction, artificial neural networks, machine learning, invariance

Uses and applications:

Image recognition has a wide array of applications in security, web search, medical imaging, artificial intelligence, robotics, and more. Technology companies such as Google and Facebook have utilized image recognition to build highly intelligent software to learn and describe entire scenes in pictures and videos.

How it works:

The complex problem of image recognition—finding and identifying objects of interest in images and videos—draws from techniques in computer science, information theory, and statistics. Most approaches typically involve an initial processing stage known as image segmentation, where an image is partitioned into meaningful groups of pixels to better identify objects. One technique for segmentation is thresholding, or converting a grayscale or color grid of pixels into a binary (black and white) image. There are different criteria for finding an ideal threshold value to separate black and white pixels, such as Otsu's method, which searches for a threshold value that minimizes the variance of pixel values assigned to either of the two classes. Using a Gaussian noise filter (which performs a convolution over pixels in the image), edge detection can also be used to extract important pixels.

After this segmentation phase, a resulting set of visual descriptors for scene objects are created. The challenge is then to select low-dimensional, identifiable features from this set of pixels. This can be accomplished using statistical techniques such as principal component analysis, which finds a principal set of dimensions (or pixel neighborhoods, in this case) that are linearly uncorrelated and can best separate clusters in a dataset. Bounding boxes can also be used to reduce pixel groups into size and coordinate vectors. After reducing an image's objects to a set of essential representations, they can be labeled and classified using machine learning. This involves training or calibrating a classification algorithm using prelabeled image objects, so it can then predict the labels of unseen objects. A popular computer vision classifier is the artificial neural network, a multilayer topology of nodes with adaptive weights during "learning" to estimate a function between given input (preprocessed image data) and output (image / object label).

Perhaps the biggest challenge in the image recognition task is recognizing patterns and objects in images that are invariant of scale, orientation, rotation, image quality, and other "nonsemantic" factors. Techniques to address this involve image reconstruction, orthogonal projections and affine transformations of pixel segments, and careful feature selection.

Interesting fact:

Many techniques in the field of computer vision draw on understandings of human biological systems. For instance, the neural network classifier was inspired by the way the human brain propagates signals in parallel. Invariant feature selection similarly draws on human cognitive models of learning. Although many theoretical and processing frameworks were developed more than 20 years ago, with the advent of high-performance and distributed computing, powerful image recognition tasks have become more computationally feasible. Researchers at Google have developed software that can give descriptive sentence captions to entire images.

Reference:

Szeliski, Richard. Computer vision: algorithms and applications. Springer Science & Business Media, 2010.

Submitted by Soubhik Barari, Tufts University, USA, First place, Math Matters, Apply it! contest February 2016

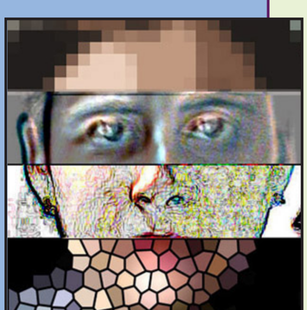

Image courtesy of MIT Open Courseware

Society for Industrial and Applied Mathematics
www.siam.org
www.siam.org/careers/matters.php