Chapter 1 Biological and Machine Vision

In modern mammals, a large proportion of the cerebral cortex (the outer gray matter of the brain) is involved in visual perception. In the late 1950s, David Hubel and Torsten Wiesel (The Johns Hopkins University) carried out pioneering research on how visual information is processed in the mammalian cerebral cortex (more specifically the primary visual cortex - the first part of the cerebral cortex to receive visual input from the eyes).

Simple neurons receive visual input from the eye and are most responsive to simple straight edges. These neurons are highly specific in that a single simple neuron responds optimally to an edge at a specific orientation. Subsequently, large groups of simple neurons are able to represent all 360 degrees of orientation.

Simple neurons pass along information (presence or absence of lines are specific orientations) to a large number of complex neurons. Thus via many hierarchically organized layers of neurons feeding information into increasingly higher-order neurons, gradually more complex visual stimuli can be represented by the brain.

Machine vision is split into two alternative approaches: the deep learning (DL) approach versus the traditional machine learning (TML) approach.

**The Deep Learning Approach**

Neocognitron – an architecture for machine vision proposed by Kunihiko Fukushima in the late 1970s that was an analog to the Hubel – Wiesel model.

LeNet-5 – a model developed by Yann LeCun and Yoshua Bengio in the late 1990s that was a significant development in the DL approach. LeNet-5 benefited from superior data for model training, faster processing power, and utilization of the back-propagation algorithm (backprop). These three components of LeNet-5 rendered it sufficiently reliable to become an early commercial DL application - it was used by the USPS to automate the reading of ZIP codes.

LeNet-5 exposes a fundamental difference between DL and TML approaches. In TML, practitioners invest a bulk of their efforts in feature engineering – applying clever, oftentimes elaborate, algorithms to raw data in order to preprocess the data into input variables that can be readily modeled by traditional statistical techniques. Feature engineering is often critical since these traditional statistical techniques – such as regression, random forest, and support vector machine – are seldom effective on unprocessed raw data. This leads to TML practitioners’ utilizing less time optimizing ML models.

The DL approach flips this development pipeline around. DL practitioners typically spend little to none of their time with feature engineering, instead spending it modeling data with various artificial neural network architectures that process the raw inputs into useful features automatically.

**The Traditional Machine Learning Approach**

Following LeNet-5, research into artificial neural networks, including deep learning, fell out of favor. This was primarily due to the view that automated feature generation was not pragmatic. In the early 2000s, Paul Viola and Michael Jones leveraged engineered features to detect faces reliably. Their efficient algorithm founds its way into Fujifilm cameras, facilitating real-time auto-focus.

**ImageNet and the ILSVRC (ImageNet Large Scale Visual Recognition Challenge)**

LeNet-5 made gains over neocognitron primary by utilizing a high quality set of training data. ImageNet – an immense catalog of training data [14 million images spread across 22,000 categories] consisting of a labeled index of photographs devised by Fei-Fei Li – enabled researchers to make significant progress in machine vision.

In 2010 and 2011 all entrants into the ILSVRC utilized TML. In 2012, all entrants except one were TML. The lone DL entry by Alex Krizhevsky and Ilya Sutskever (working at the lab headed by Geoffrey Hinton – the godfather of deep learning) called AlexNet demolished the prior TML benchmarks as measured by error rate. Since 2012 all of the top entrants into the ILSVRC have been DL based approaches. Starting in 2015, the top DL algorithms surpassed human accuracy.

AlexNet took advantage of three principal factors to launch it into prominence:

1. The ImageNet training data was not only bigger but transformations applied to the base image catalog expanded the scope of the training data.
2. Processing power had grown exponentially since the days of LeNet-5.
3. AlexNet used deeper architectural layers, some incorporating a new type of artificial neuron. This coupled with a nifty trick helped AlexNet generalize in superior ways relative to its predecessors and competing approaches.

AlexNet is widely useful and disruptive across industries and computational applications. Models like AlexNet dramatically reduce the subject-matter expertise required for building highly accurate predictive models. For a rapidly growing list of use cases, one’s ability to apply deep learning techniques outweighs the value of domain-specific proficiency.

**bit.ly/TPplayground** is a deep learning model in the TensorFlow Playground.

**quickdraw.withgoogle.com** is deep learning algorithm designed to guess what you’ve sketched.