Deep Learning Architectures

There have been hundreds of different deep learning models that combine the deep learning primitives presented in the previous section. Some of these architectures have been historically important. Others were the first presentations of novel designs that influenced perceptions of what deep learning could do.

In this section, we present a selection of different deep learning architectures that have proven influential for the research community. We want to emphasize that this is an episodic history that makes no attempt to be exhaustive. There are certainly important models in the literature that have not been presented here.

LeNet

The LeNet architecture is arguably the first prominent "deep" convolutional architecture. Introduced in 1988, it was used to perform optical character recoginition (OCR) for documents. Although it performed its task admirably, the computational cost of the LeNet was extreme for the computer hardware available at the time, so the design languished in (relative) obscurity for a few decades after its creation. This architecture is illustrated in Figure 1-5.

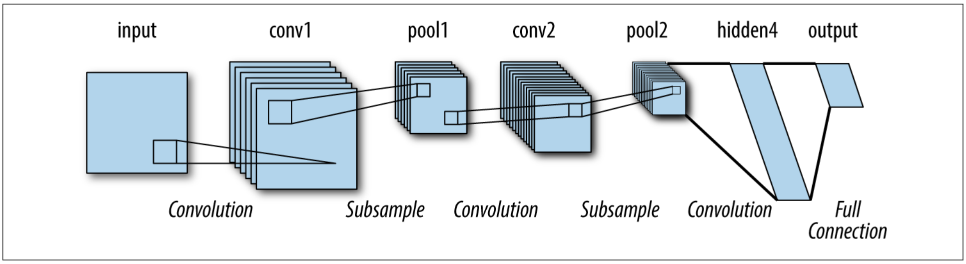

Figure 1-5. The LeNet architecture for image processing. Introduced in 1988, it was arguably the first deep convolutional model for image processing.

AlexNet

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) was first organized in 2010 as a test of the progress made in visual recognition systems. The organizers made use of Amazon Mechanical Turk, an online platform to connect workers to requesters, to catalog a large collection of images with associated lists of objects present in the image. The use of Mechanical Turk permitted the curation of a collection of data significantly larger than those gathered previously.

The first two years the challenge ran, more traditional machine-learned systems that relied on systems like HOG and SIFT features (hand-tuned visual feature extraction methods) triumphed. In 2012, the AlexNet architecture, based on a modification of LeNet run on powerful graphics processing units (GPUs), entered and dominated the challenge with error rates half that of the nearest competitors. This victory dramatically galvanized the (already nascent) trend toward deep learning architectures in computer vision. The AlexNet architecture is illustrated in Figure 1-6.

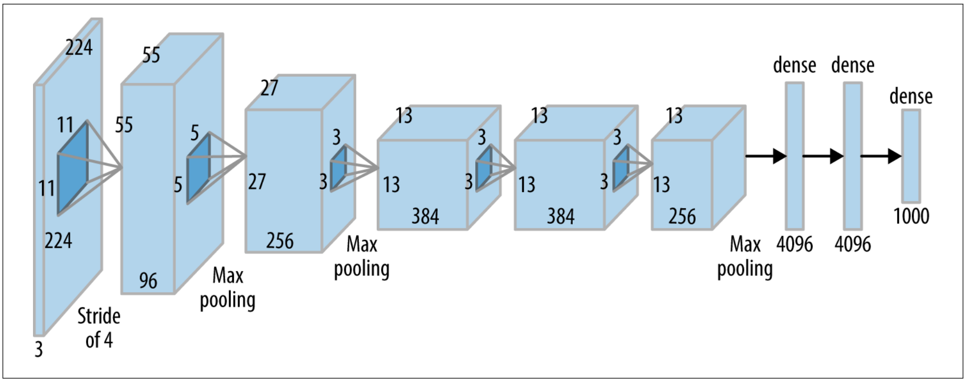

Figure 1-6. The AlexNet architecture for image processing. This architecture was the winning entry in the ILSVRC 2012 challenge and galvanized a resurgence of interest in convolutional architectures.

ResNet

Since 2012, convolutional architectures consistently won the ILSVRC challenge (along with many other computer vision challenges). Each year the contest was held, the winning architecture increased in depth and complexity. The ResNet architecture, winner of the ILSVRC 2015 challenge, was particularly notable; ResNet architectures extended up to 130 layers deep, in contrast to the 8-layer AlexNet architecture.

Very deep networks historically were challenging to learn; when networks grow this deep, they run into the vanishing gradients problem. Signals are attenuated as they progress through the network, leading to diminished learning. This attenuation can be explained mathematically, but the effect is that each additional layer multiplicatively reduces the strength of the signal, leading to caps on the effective depth of networks.

The ResNet introduced an innovation that controlled this attenuation: the bypass connection. These connections allow part of the signal from deeper layers to pass through undiminished, enabling significantly deeper networks to be trained effectively. The ResNet bypass connection is illustrated in Figure 1-7.