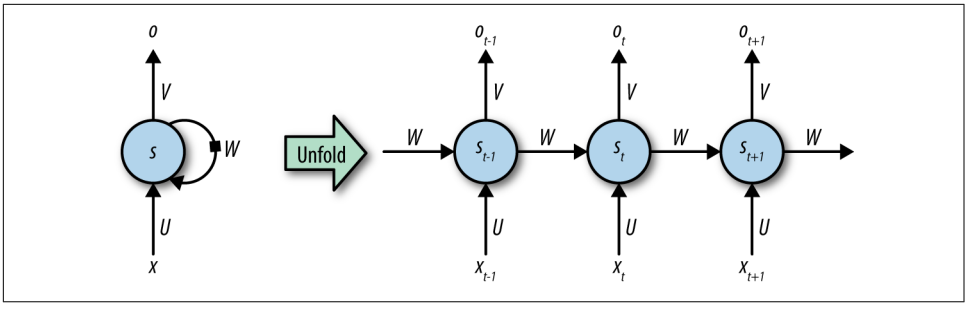

Figure 1-3. A recurrent neural network (RNN). Inputs are fed into the network at the bottom, and outputs extracted at the top. W represents the learned transformation (shared at all timesteps). The network is represented conceptually on the left and is unrolled on the right to demonstrate how inputs from different timesteps are processed.

Long Short-Term Memory Cells

The RNN layers presented in the previous section are capable of learning arbitrary sequence-update rules in theory. In practice, however, such layers are incapable of learning influences from the distant past. Such distant influences are crucial for performing solid language modeling since the meaning of a complex sentence can depend on the relationship between far-away words. The long short-term memory (LSTM) cell is a modification to the RNN layer that allows for signals from deeper in the past to make their way to the present. An LSTM cell is illustrated in Figure 1-4.

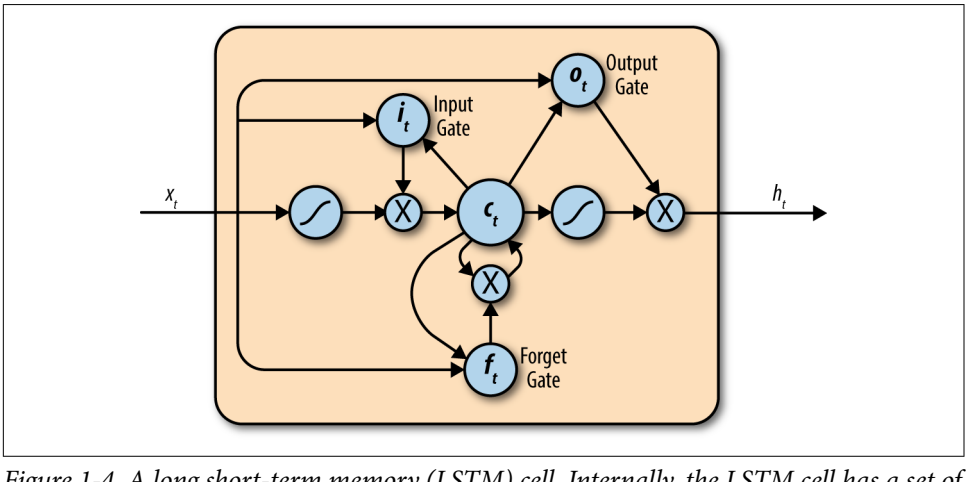

Figure 1-4. A long short-term memory (LSTM) cell. Internally, the LSTM cell has a set of specially designed operations that attain much of the learning power of the vanilla RNN while preserving influences from the past. Note that the illustration depicts one LSTM variant of many.

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.

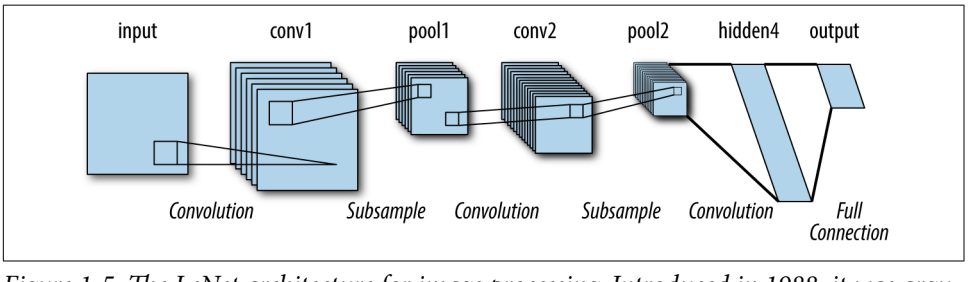

Figure 1-5. The LeNet architecture for image processing. Introduced in 1988, it was arguably the first deep convolutional model for image processing.

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