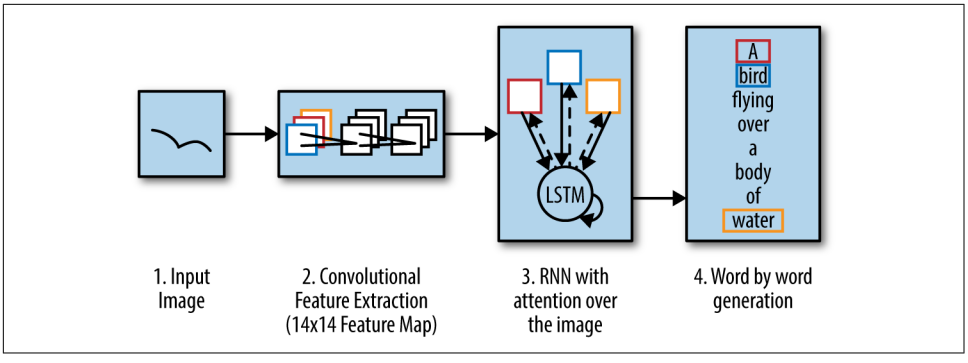

Figure 1-8. A neural captioning architecture. Relevant input features are extracted from the input image using a convolutional network. Then a recurrent network is used to generate a descriptive sentence.

Google Neural Machine Translation

Google's neural machine translation (Google-NMT) system uses the paradigm of end-to-end training to build a production translation system, which takes sentences from the source language directly to the target language. The Google-NMT system depends on the fundamental building block of the LSTM, which it stacks over a dozen times and trains on an extremely large dataset of translated sentences. The final architecture provided for a breakthrough advance in machine-translation by cutting the gap between human and machine translations by up to 60%. The Google-NMT architecture is illustrated in Figure 1-9.

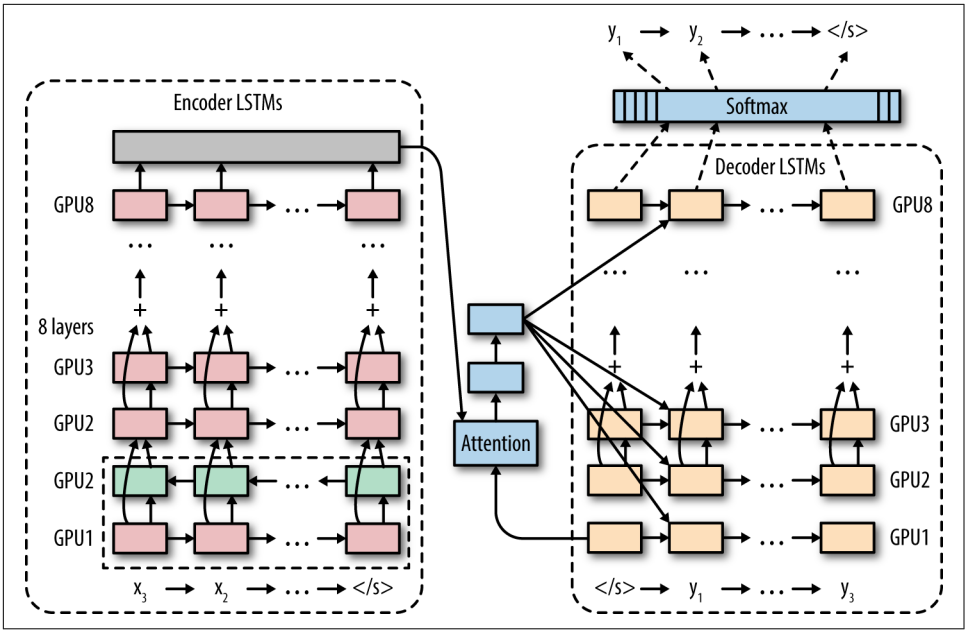

Figure 1-9. The Google neural machine translation system uses a deep recurrent architecture to process the input sentence and a second deep recurrent architecture to generate the translated output sentence.

One-Shot Models

One-shot learning is perhaps the most interesting new idea in machine/deep learning. Most deep learning techniques typically require very large amounts of data to learn meaningful behavior. The AlexNet architecture, for example, made use of the large ILSVRC dataset to learn a visual object detector. However, much work in cognitive science has indicated that humans can learn complex concepts from just a few examples. Take the example of baby learning about giraffes for the first time. A baby shown a single giraffe at the zoo might be capable of learning to recognize all giraffes she sees from then on.

Recent progress in deep learning has started to invent architectures capable of similar learning feats. Given only a few examples of a concept (but given ample sources of side information), such systems can learn to make meaningful predictions with very few datapoints. One recent paper (by an author of this book) used this idea to demonstrate that one-shot architectures can learn even in contexts babies can't, such as in medical drug discovery. A one-shot architecture for drug discovery is illustrated in Figure 1-10.