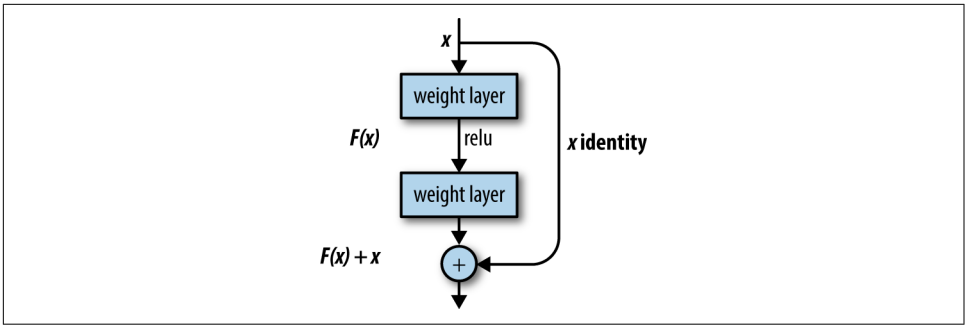

Figure 1-7. The ResNet cell. The identity connection on the righthand side permits an unmodified version of the input to pass through the cell. This modification allows for the effective training of very deep convolutional architectures.

Neural Captioning Model

As practitioners became more comfortable with the use of deep learning primitives, they experimented with mixing and matching primitive modules to create higher-order systems that could perform more complex tasks than basic object detection. Neural captioning systems automatically generate captions for the contents of images. They do so by combining a convolutional network, which extracts information from images, with an LSTM layer that generates a descriptive sentence for the image. The entire system is trained *end-to-end*. That is, the convolutional network and the LSTM network are trained together to achieve the desired goal of generating descriptive sentences for provided images.

This end-to-end training is one of the key innovations powering modern deep learning systems since it lessens the need for complicated preprocessing of inputs. Image captioning models that don't use deep learning would have to use complicated image featurization methods such as SIFT, which can't be trained alongside the caption generator.

A neural captioning model is illustrated in Figure 1-8.

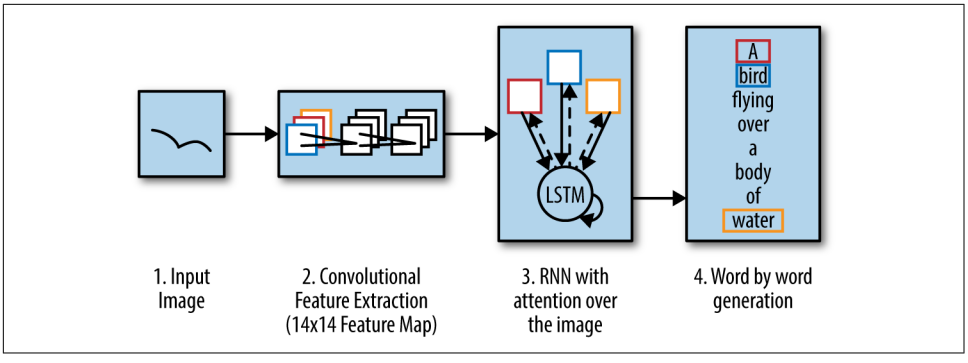

Figure 1-8. A neural captioning architecture. Relevant input features are extracted from the input image using a convolutional network. Then a recurrent network is used to generate a descriptive sentence.

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.