Convolutional Neural Networks

Convolutional neural networks allow deep networks to learn functions on structured spatial data such as images, video, and text. Mathematically, convolutional networks provide tools for exploiting the local structure of data effectively. Images satisfy certain natural statistical properties. Let's assume we represent an image as a two-dimensional grid of pixels. Parts of an image that are close to one other in the pixel grid are likely to vary together (for example, all pixels corresponding to a table in the image are probably brown). Convolutional networks learn to exploit this natural covariance structure in order to learn effectively.

Convolutional networks are a relatively old invention. Versions of convolutional networks have been proposed in the literature dating back to the 1980s. While the designs of these older convolutional networks were often quite sound, they required resources that exceeded hardware available at the time. As a result, convolutional networks languished in relative obscurity in the research literature.

This trend reversed dramatically following the 2012 ILSVRC challenge for object detection in images, where the convolutional AlexNet achieved error rates half that of its nearest competitors. AlexNet was able to use GPUs to train old convolutional architectures on dramatically larger datasets. This combination of old architectures with new hardware allowed AlexNet to dramatically outperform the state of the art in image object detection. This trend has only continued, with convolutional neural networks achieving tremendous boosts over other technologies for processing images. It isn't an exaggeration to say that nearly all modern image processing pipelines are now powered by convolutional neural networks.

There has also been a renaissance in convolutional network design that has moved convolutional networks well past the basic models from the 1980s. For one, networks have been getting much deeper with powerful state-of-the-art networks reaching hundreds of layers deep. Another broad trend has been toward generalizing convolu-

tional architectures to work on new datatypes. For example, graph convolutional architectures allow convolutional networks to be applied to molecular data such as the Tox21 dataset we encountered a few chapters ago! Convolutional architectures are also making a mark in genomics and in text processing and even language translation.

In this chapter, we will introduce the basic concepts of convolutional networks. These will include the basic network components that constitute convolutional architectures and an introduction to the design principles that guide how these pieces are joined together. We will also provide an in-depth example that demonstrates how to use TensorFlow to train a convolutional network. The example code for this chapter was adapted from the TensorFlow documentation tutorial on convolutional neural networks. We encourage you to access the original tutorial on the TensorFlow website if you're curious about the changes we've made. As always, we encourage you to work through the scripts for this chapter in the associated GitHub repo for this book.

Introduction to Convolutional Architectures

Most convolutional architectures are made up of a number of basic primitives. These primitives include layers such as convolutional layers and pooling layers. There's also a set of associated vocabulary including local receptive field size, stride size, and number of filters. In this section, we will give you a brief introduction to the basic vocabulary and concepts underlying convolutional networks.

Local Receptive Fields

The local receptive field concept originates in neuroscience, where the receptive field of a neuron is the part of the body's sensory perception that affects the neuron's firing. Neurons have a certain field of "view" as they process sensory input that the brain sees. This field of view is traditionally called the local receptive field. This "field of view" could correspond to a patch of skin or to a segment of a person's visual field. Figure 6-1 illustrates a neuron's local receptive field.