vision, many pretrained models (usually trained on the ImageNet dataset) are now publicly available for download and can be used to bootstrap powerful vision models out of very little data. This is one of the greatest strengths of deep learning: feature reuse. You'll explore this in the next section.

Let's start by getting our hands on the data.

8.2.2 Downloading the data

The Dogs vs. Cats dataset that we will use isn't packaged with Keras. It was made available by Kaggle as part of a computer-vision competition in late 2013, back when convnets weren't mainstream. You can download the original dataset from `www.kaggle.com/c/dogs-vs-cats/data` (you'll need to create a Kaggle account if you don't already have one—don't worry, the process is painless). You can also use the Kaggle API to download the dataset in Colab (see "Downloading a Kaggle dataset in Google Colaboratory" below).

SIDEBAR

Downloading a Kaggle dataset in Google Colaboratory

Kaggle makes available an easy-to-use API to programatically download Kaggle-hosted datasets. You can use it to download the Dogs vs Cat dataset to a Colab notebook, for instance. This API is available as the `kaggle` package, which is pre-installed on Colab. Downloading our dataset is as easy as running the following command in a Colab cell:

```
!kaggle competitions download -c dogs-vs-cats
```

However, access to the API is restricted to Kaggle users, so in order to run the command above, you first need to authenticate yourself. The `kaggle` package will look for your login credentials in a JSON file located at `~/.kaggle/kaggle.json`. Let's create this file.

First, you need to create a Kaggle API key and download it to your local machine. Just navigate to the Kaggle website in a web browser, log in, and go to the "My Account" page. In your account settings, you'll find an "API" section. Clicking the button "create new API token" will generate a `kaggle.json` key file and download it to your machine.

Second, go to your Colab notebook, and upload the API key JSON file to your Colab session by running the following code in a notebook cell:

```
from google.colab import files
files.upload()
```

When you run this cell, you will see a button "Choose files" appear. Click it and select the `kaggle.json` file you just downloaded. This uploads the file to the local Colab runtime.

©Manning Publications Co. To comment go to liveBook
https://livebook.manning.com/book/deep-learning-with-python-second-edition/discussion
Licensed to Noriko Tomuro <tomuro@cs.depaul.edu>

Finally, create a `~/.kaggle` folder (`mkdir ~/.kaggle`), copy the key file to it (`cp kaggle.json ~/.kaggle/`). As a security best practice, you should also make sure that the file is only readable by the current user, yourself (`chmod 600`).

```
!mkdir ~/.kaggle
!cp kaggle.json ~/.kaggle/
!chmod 600 ~/.kaggle/kaggle.json
```

You're now able to download the data we're about to use:

```
!kaggle competitions download -c dogs-vs-cats
```

The first time you try to download the data, you may get a "403 - Forbidden" error. That's because you need to accept the terms associated with the dataset before you download it—you'll have to go to `www.kaggle.com/c/dogs-vs-cats/rules` (while logged into your Kaggle account) and click the button "I Understand and Accept". You only need to do this once.

Finally: the training data is a compressed file named `train.zip`. Make sure to uncompress it (`unzip`) silently (`-qq`):

```
!unzip -qq train.zip
```

The pictures in our dataset are medium-resolution color JPEGs. Figure 8.8 shows some examples.

©Manning Publications Co. To comment go to liveBook
https://livebook.manning.com/book/deep-learning-with-python/discussion
Licensed to Noriko Tomuro <tomuro@cs.depaul.edu>