

图像数据通常表示为像素值数组。我们将从嵌入式摄像头模块中获取图像数据，这些模块会以像素值数组的形式将数据提供给我们。同时，我们的模型本就期望它的输入为像素值数组。因此，在将数据输入模型之前，我们不需要进行太多的预处理。

考虑到我们无须进行太多的预处理，我们的应用程序将会非常简单。它从摄像头模块中获取数据快照，将数据输入模型，并确定检测到哪个输出类别。然后，程序以某种简单的方式显示结果。

在继续之前，让我们进一步了解将要使用的模型。

9.2.1 模型介绍

在第 7 章中，我们了解到卷积神经网络是用于处理多维张量的神经网络，这些多维张量的隐藏信息包含在相邻值之间的关系中。卷积神经网络特别适合处理图像数据。

我们的行人检测模型是在 Visual Wake Words 数据集上训练的卷积神经网络。这个数据集包含 115 000 张图像，每张图像都被标记了是否包含行人。

我们的模型大小为 250KB，比我们的语音模型大得多。除了占用更多内存外，它还意味着运行单次推断将会花费更长的时间。

我们的模型接收 96 像素 × 96 像素的灰度图像作为输入。每个图像都是一个形状为 (96, 96, 1) 的三维张量，其中最终维包含一个表示单个像素的 8 位数值。它表示像素的灰度值，范围从 0（全黑）到 255（全白）。

我们的摄像头模块可以返回各种分辨率的图像，因此我们需要确保将它的尺寸调整为 96 像素 × 96 像素。我们还需要将全彩色的图像转换为灰度图，以便它们可以与模型一起使用。

你可能会认为 96 像素 × 96 像素听起来是很小的分辨率，但是这已经足够让我们在每个图像中检测是否有行人。与图像相关的模型通常都接收小得惊人的小分辨率图像。增加模型的输入的大小会导致收益递减，同时，网络的复杂度也随着输入输出的增多而大大提高。因此，即使是最先进的图像分类模型，通常最多也只能处理 320 像素 × 320 像素的图像。

我们的模型会输出两个概率：一个表示输入中存在一个人的概率，另一个表示输入中不存在行人的概率。概率的范围为 0~255。

我们的行人检测模型使用广为人知、久经考验的 MobileNet 架构，它被设计用于在移动电话等设备上进行图像分类。在第 10 章中，你将学习如何调整此模型以使其适合微