

卷积神经网络简介： 什么是机器学习？—— 第一部分

Ole Dreessen，现场应用工程师

摘要

随着人工智能(AI)技术的快速发展，AI可以越来越多地支持以前无法实现或者难以实现的应用。本系列文章基于此解释了卷积神经网络(CNN)及其对人工智能和机器学习的意义。CNN是一种能够从复杂数据中提取特征的强大工具，例如识别音频信号或图像信号中的复杂模式就是其应用之一。本文讨论了CNN相对于经典线性规划的优势，后续文章《训练卷积神经网络：什么是机器学习？——第二部分》将讨论如何训练CNN模型，系列文章的第三部分将讨论一个特定用例，并使用专门的AI微控制器对模型进行测试。

什么是卷积神经网络？

神经网络是一种由神经元组成的系统或结构，它使AI能够更好地理解数据，进而解决复杂问题。虽然神经网络有许多种类型，但本系列文章将只关注卷积神经网络(CNN)，其主要应用领域是对输入数据的模式识别和对象分类。CNN是一种用于深度学习的人工神经网络。这种网络由输入层、若干卷积层和输出层组成。卷积层是最重要的部分，它们使用一组独特的权重和滤波器，使得网络可以从输入数据中提取特征。数据可以是许多不同的形式，如图像、音频和文本。这种提取特征的过程使CNN能够识别数据中的模式从而让工程师能够创建更有效和高效的应用。为了更好地理解CNN，我们首先将讨论经典的线性规划。

经典控制技术中的线性规划

控制技术的任务是借助传感器读取数据并进行处理，然后根据规则做出响应，最后显示或发送结果。例如，温度调节器每秒测量一次温度，通过微控制器单元(MCU)读取温度传感器的数据。该数值用于闭环控制系统的输入，并与设定的温度进行比较。这就是一个借助MCU执行线性规划的例子，这种技术通过比较预编程值和实际值来给出明确的结论。相比之下，AI系统通常依据概率论来发挥作用。

复杂模式和信号处理

许多应用所使用的输入数据必须首先由模式识别系统加以判别。模式识别可以应用于不同的数据结构。本文讨论的例子限定为一维或二维的数据结构，比如音频信号、心电图(ECG)、光电容积脉搏波(PPG)、一维的振动数据或波形、热图像、二维的瀑布图数据。

在上述模式识别中，将应用通过MCU的代码来实现是极其困难的。一个例子是识别图像中的具体对象（例如猫）：这种情况下无法区分要分析的图像是很早摄录的，还是刚刚由相机读取的。分析软件基于一些特定的规则来判断图片中是否有猫，比如说猫必须有典型的尖耳朵、三角形的鼻子和胡须。如果可以在图像中识别出这些特征，软件便可以报告在图像中发现了猫。但是这存在一些问题：如果图像只显示了猫的背面，模式

识别系统会怎么办？如果猫没有胡须或者在事故中失去了腿，会发生什么？尽管这些异常情况不太可能出现，但模式识别的代码将不得不考虑所有可能的异常情况，从而增加大量额外的规则。即使在这个简单的例子中，软件设置的规则也会变得非常复杂。

机器学习如何取代经典规则

AI背后的核心思想是在小范围内模仿人类进行学习。它不依赖于制定大量的if-then规则，而是建立一个通用的模式识别的机器模型。这两种方法的关键区别在于，与一套复杂的规则相比，AI不会提供明确的结果。AI不会明确报告“我在图像中识别出了一只猫”，而是提供类似这样的结论：“图像中有一只猫的概率为97.5%，它也可能是豹子(2.1%)或老虎(0.4%)。”这意味着在模式识别的过程结束时，应用的开发人员必须通过决策阈值做出决定。

另一个区别是AI并不依赖固定的规则，而是要经过训练。训练过程需要将大量猫的图像展示给神经网络以供其学习。最终，神经网络将能够独立识别图像中是否有猫。关键的一点是，未来AI可以不局限于已知的训练图像开展识别。该神经网络需要映射到MCU中。

AI的模式识别内部到底是什么？

AI的神经网络类似于人脑的生物神经网络。一个神经元有多个输入，但只有一个输出。基本上，这些神经元都是输入的线性变换——将输入乘以数字（权重w）并加上一个常数（偏置b），然后通过一个固定的非线性函数产生输出，该函数也被称为激活函数¹。作为网络中唯一的非线性部分，激活函数用于定义人工神经元值的激活范围。神经元的功能在数学上可以描述为

$$Out = f(w*x+b) \quad (1)$$

其中，f为激活函数，w为权重，x为输入数据，b为偏置。数据可以是单独的标量、向量或矩阵。图1显示了一个神经元，它拥有三个输入和一个激活函数ReLU²。网络中的神经元总是分层排列的。

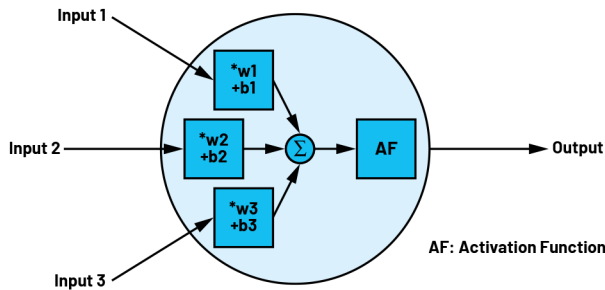


图1. 拥有三个输入和一个输出的神经元

如上所述，CNN用于输入数据的模式识别和对象分类。CNN分为不同的部分：一个输入层、若干隐藏层和一个输出层。图2显示了一个小型网络，它包含一个具有三个输入的输入层、一个具有五个神经元的隐藏层和一个具有四个输出的输出层。所有神经元的输出都连接到下一层的所有输入。图2所示的网络不具有现实意义，这里仅用于演示说明。即使对于这个小型网络，用于描述网络的方程中也具有32个偏置和32个权重。

CIFAR神经网络是一种广泛用于图像识别的CNN。它主要由两种类型的层组成：卷积层和池化层，这两种层分别使用卷积和池化两种方法，在神经网络的训练中非常有效。卷积层使用一种被称为卷积的数学运算来识别像素值数组的模式。卷积发生在隐藏层中，如图3所示。卷积会重复多次直至达到所需的精度水平。如果要比较的两个输入值（本例是输入图像和滤波器）相似，那么卷积运算的输出值总会特别高。滤波器有时也被称为卷积核。然后，结果被传递到池化层提取特征生成一个特征图，表征输入数据的重要特征，称为池化。池化层的运行需要依赖另一个滤波器，称为池化滤波器。训练后，在网络运行的状态下，特征图与输入数据进行比较。由于特征图保留了特定的特征，所以只有当内容相似时，神经元的输出才会被触发。通过组合使用卷积和池化，CIFAR网络可用于高精度地识别和分类图像中的各种对象。

¹ 通常使用 sigmoid、tanh 或 ReLU 函数。

² ReLU：修正线性单元。对于该函数，输入值为负时，输出为零；输入值大于零时，输出值为输入值。

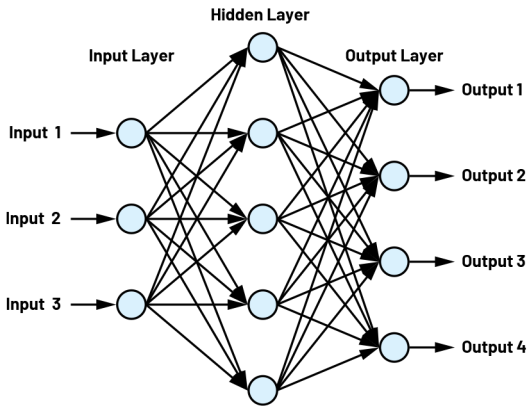


图 2. 一个小型神经网络

CIFAR-10是一个特定数据集，通常用于训练CIFAR神经网络。它由60000幅 32×32 彩色图像组成，分为10个类别。这些图像是从各种来源收集的，例如网页、新闻和个人图像集。每个类别包含6000幅图像，平均分配在训练集、测试集和验证集中，使其成为测试计算机视觉和其他机器学习模型的理想图像集。

卷积神经网络和其他类型网络的主要区别在于处理数据的方式。卷积神经网络通过滤波依次检查输入数据的属性。卷积层的数量越多，可以识别的细节就越精细。在第一次卷积之后，该过程从简单的对象属性（如边或点）开始进行第二次卷积以识别详细的结构，如角、圆、矩形等。在第三次卷积之后，特征就可以表示某些复杂的模式，它们与图像中对象的某些部分相似，并且对于给定对象来说通常是唯一的。在我们最初的例子中，这些特征就是猫的胡须或耳朵。特征图的可视化（如图4所示）对于应用本身而言并不是必需的，但它有助于帮助理解卷积。

即使是像CIFAR这样的小型网络，每层也有数百个神经元，并且有许多串行连接的层。随着网络的复杂度和规模的增加，所需的权重和偏置数量也迅速增长。图3所示的CIFAR-10示例已经有20万个参数，每个参数在训练过程中都需要一组确定的值。特征图可以由池化层进一步处理，以减少需要训练的参数数量并保留重要信息。

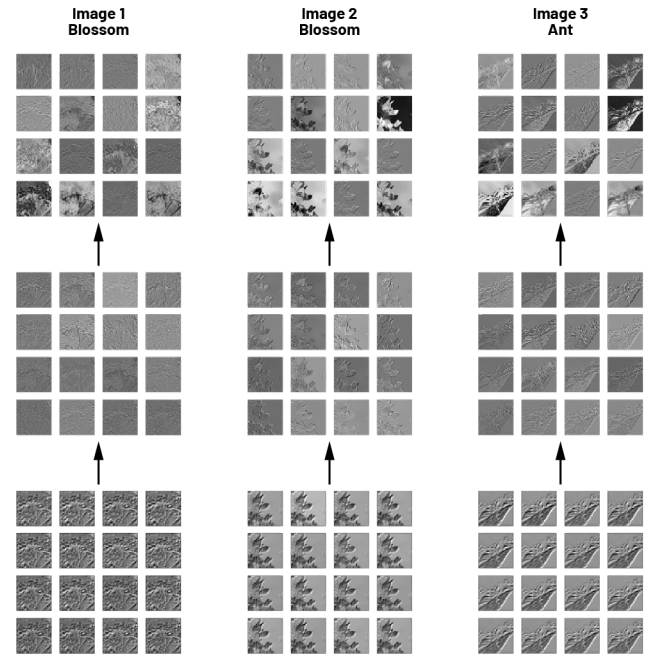


图 4. CNN 的特征图

如上所述，在CNN中的每次卷积之后，通常会发生池化，在一些文献中也常被称为子采样。它有助于减少数据的维度。图4中的特征图里面的很多区域包含很少甚至不含有意义的信息。这是因为对象只是图像的一小部分，并不构成整幅图像。图像

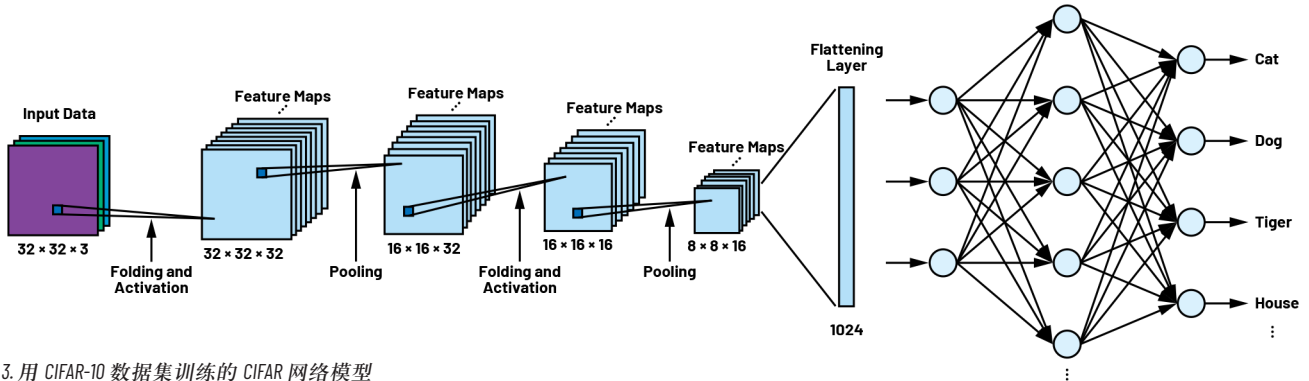


图 3. 用 CIFAR-10 数据集训练的 CIFAR 网络模型

的其余部分未在特征图中使用，因此与分类无关。在池化层中，池化类型（最大值池化或均值池化）和池化窗口矩阵的大小均被指定。在池化过程中，窗口矩阵逐步在输入数据上移动。例如，最大值池化会选取窗口中的最大数据值而丢弃其它所有的值。这样，数据量不断减少，最终形成各个对象类别的唯一属性。

卷积和池化的结果是大量的二维矩阵。为了实现我们真正的目标即分类，我们需要将二维数据转换成一个很长的一维向量。转换是在所谓的压平层中完成的，随后是一个或两个全连接层。全连接层的神经元类似于图2所示的结构。神经网络最后一

层的输出要与需要区分的类别的数量一致。此外，在最后一层中，数据还被归一化以产生一个概率分布（97.5%的猫，2.1%的豹，0.4%的虎，等等）。

这就是神经网络建模的全过程。然而，卷积核与滤波器的权重和内容仍然未知，必须通过网络训练来确定使模型能够工作。这将在后续文章《训练卷积神经网络：什么是机器学习？——第二部分》中说明。第三部分将解释我们上文讨论过的神经网络（例如识别猫）的硬件实现，我们将使用ADI公司开发的带硬件CNN加速器的MAX78000人工智能微控制器来演示。



作者简介

Ole Dreessen是ADI公司的现场应用工程师，于2014年加入ADI公司，此前曾在Avnet Memec和Macnica任职，负责支持通信产品和高性能微处理器。Ole在微控制器和安全方面拥有广泛的专业知识，拥有丰富的会议主讲经验。在业余时间，他是混沌计算机俱乐部的活跃成员，主要研究逆向工程和嵌入式安全等概念。

