卷积神经网络简介

Keiron O'Shea[1]和Ryan Nash[2]

[1]阿伯里斯特威斯大学计算机科学系，Ceredigion，SY23 3DB
keo7@aber.ac.uk
[2]兰开斯特大学计算机和通信学院，兰开斯特，LA1 4YW
nashrd@live.lancs.ac.uk

摘要。随着人工神经网络（ANN）的兴起，机器学习领域的发展出现了戏剧性的转折。这些受生物启发的计算模型能够在普通的机器学习任务中远远超过以前的人工智能形式的性能。ANN架构中最令人印象深刻的形式之一是卷积神经网络（CNN）。CNN主要用于解决困难的图像驱动的模式识别任务，其精确而简单的结构，提供了一种简化的ANN入门方法。

本文对CNN进行了简要介绍，讨论了最近发表的论文和新形成的开发这些神奇的图像识别模型的技术。本介绍假定你熟悉ANN和机器学习的基本原理。

关键词模式识别，人工神经网络，机器学习，图像分析

1 简介

人工神经网络（ANNs）是一种计算处理系统，它在很大程度上受到生物神经系统（如人脑）运作方式的启发。ANNs主要由大量相互连接的计算节点（称为神经元）组成，它们以分布式的方式相互缠绕，共同从输入中学习，以优化其最终输出。

ANN的基本结构可以如图1所示进行建模。我们将输入，通常是以多维矢量的形式加载到输入层，输入层将把它分配到隐藏层。然后，隐藏层将从上一层做出决定，并权衡自身的随机变化如何损害或改善最终输出，这被称为学习的过程。多个隐藏层相互叠加，通常被称为深度学习。

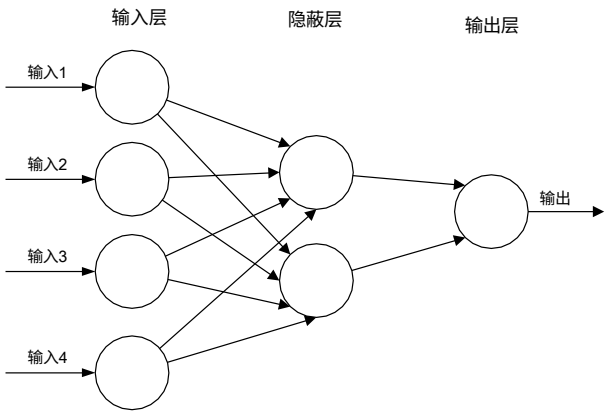

图1：一个简单的三层前馈神经网络（FNN），由一个输入层、一个隐藏层和一个输出层组成。这种结构是许多常见的ANN结构的基础，包括但不限于前馈神经网络（FNN）、受限玻尔兹曼机（RBM）和循环神经网络（RNN）。

图像处理任务中的两个关键学习范式是监督学习和无监督学习。**监督学习**是通过预先标记的输入来学习，这些输入作为目标。对于每个训练实例，都有一组输入值（向量）和一个或多个相关的指定输出值。这种训练形式的目标是通过正确计算训练实例的输出值来减少模型的整体分类误差。

无监督学习的不同之处在于，训练集不包括任何la-bels。成功通常是由网络是否能够减少或增加一个相关的成本函数来决定的。然而，需要注意的是，大多数以图像为中心的模式识别任务通常取决于使用监督学习的分类。

卷积神经网络（CNN）类似于传统的人工神经网络，它们是由通过学习进行自我优化的神经元组成的。每个神经元仍然会接收一个输入并执行一个操作（如标量乘以一个非线性函数）--这是无数个神经网络的基础。从输入的原始图像向量到最终输出的类别分数，整个网络仍将表达一个单一的感知分数函数（权重）。最后一层将包含与类相关的损失函数，所有为传统ANNs开发的常规技巧和窍门仍然适用。

CNN和传统的ANN之间唯一值得注意的区别是，CNN主要用于图像内的模式识别领域。这使得我们可以将图像的具体特征编码到架构中，使网络

更适合于以图像为重点的任务--同时进一步减少建立模型所需的参数。

传统形式的人工神经网络的最大局限性之一是，它们往往在计算图像数据所需的计算复杂性中挣扎。常见的机器学习基准数据集，如MNIST手写数字数据库，适合于大多数形式的ANN，因为它的计算量相对较大。
小图像的维度只有28×28。有了这个数据集，第一个隐藏层中的单个神经元将包含784个权重（28×28×1，其中1是指MNIST被归一化的黑白值），这是可管理的。
用于大多数形式的ANN。

如果你考虑更多的64×64的彩色图像输入，仅第一层的一个神经元上的权重数量就大大增加到12，288。还要考虑到，为了处理这种规模的输入，网络也需要比用于分类颜色规范化的网络大得多。
MNIST数字，那么你就会明白使用这种模型的弊端。

1.1 过度拟合

但这又有什么关系呢？当然，我们可以增加网络中的隐藏层的数量，也许还可以增加其中的神经元数量？这个问题的简单答案是否定的。这有两个原因，一个是没有无限的计算能力和时间来训练这些巨大的ANN的简单问题。

第二个原因是阻止或减少过拟合的影响。**过度拟合**基本上是指一个网络由于一些原因而无法有效地学习。这是大多数（如果不是所有）机器学习算法的一个重要概念，采取一切预防措施以减少其影响是很重要的。如果我们的模型出现了过拟合的迹象，那么我们可能会看到，不仅我们的训练数据集，而且我们的测试集和预测集的概括性特征的能力下降。

这是降低我们的ANN复杂度背后的主要原因。需要训练的参数越少，网络就越不可能过度拟合--当然，也会提高模型的预测性能。

2 CNN架构

如前所述，CNN主要关注的是输入将由图像组成的基础。这使得架构的设置方式最适合处理特定类型的数据的需要。

其中一个关键的区别是，CNN内各层的神经元是由三个维度组成的，即输入的空间维度（**高度**和**宽度**）和**深度**。深度并不是指ANN内的总层数，而是指激活量的第三维。与标准的人工神经网络不同，任何给定层的神经元只与前面一层的一个小区域相连。

在实践中，这意味着对于前面的例子，输入 "体积 "的维度为64×64×3（高度、宽度和深度），导致最终输出层的维度为1×1×*n*（其中*n*代表可能的类的数量），因为我们已经浓缩了。
将全部输入维度转化为较小的深度维度上的类别分数。

2.1 整体架构

CNN是由三种类型的层组成的。它们是卷积层、集合层和**全连接层**。当这些层堆叠在一起时，就形成了一个CNN架构。图2显示了用于MNIST分类的简化CNN架构。

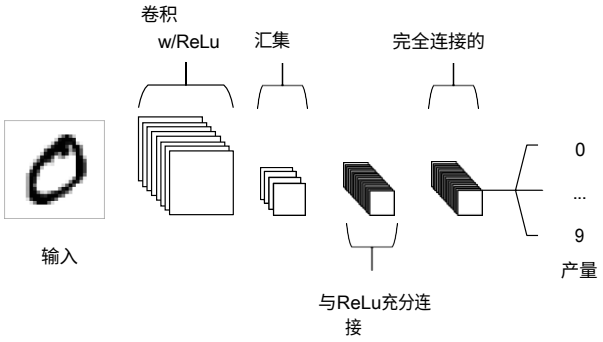

图2：一个简单的CNN架构，仅由五层组成。

上述CNN例子的基本功能可以分解为四个关键领域。

1. 正如在其他形式的ANN中发现的那样，**输入层**将保存图像的像素值。
2. **卷积层**将通过计算神经元的权重和与输入量相连的区域之间的标量乘积，确定与输入的局部区域相连的神经元的输出。**整顿线性单元**（通常简称为ReLu）旨在应用

一个 "元素 "激活函数，如sigmoid，用于前一层产生的激活输出。

3. 然后，**池化层**将简单地沿着给定输入的空间维度进行降维采样，进一步减少该激活中的参数数量。

4. 然后，**全连接层**将执行标准ANN中的相同职责，并试图从激活中产生类别分数，以用于分类。还建议在这些层之间使用ReLu，以提高性能。

通过这种简单的转换方法，CNN能够使用卷积和下采样技术逐层转换原始输入，产生用于分类和回归的类分数。

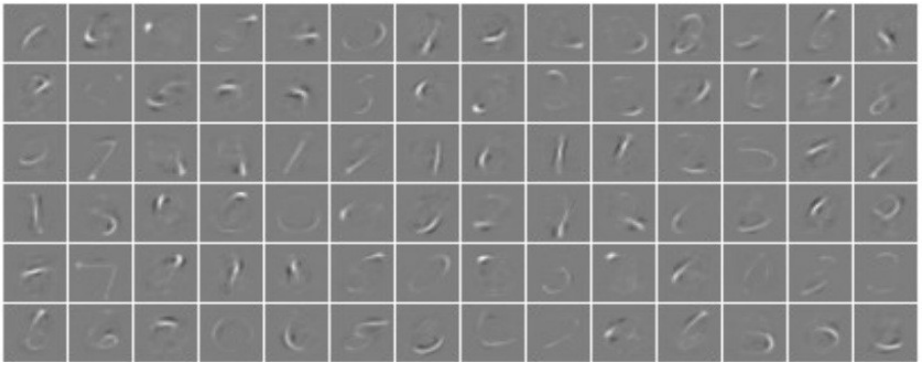

图3：在对MNIST手写数字数据库进行训练后，从一个简单的深度CNN的第一个卷积层提取的激活。如果你仔细观察，你可以看到网络已经成功地捕捉到了特定数字的独特特征。

然而，需要注意的是，仅仅了解CNN架构的整体结构是不够的。这些模型的创建和优化可能需要相当长的时间，而且可能相当令人困惑。我们现在将详细探讨各个层，详细说明它们的超参数和连接性。

2.2 卷积层

顾名思义，卷积层在CNN的运作方式中起着至关重要的作用。该层的参数主要围绕着使用可学习的**内核**。

这些内核的空间维度通常较小，但沿着输入的整个深度扩散。当数据到达卷积层时，该层将每个滤波器在输入的空间维度上进行卷积，产生一个二维激活图。这些激活图可以被可视化，如图3所示。

当我们在输入中滑行时，该核中的每个值都会计算标量积。(图4）由此，网络将学习内核，当它们在输入的特定空间位置看到一个特定的特征时，就会 "开火"。这些通常被称为**激活**。

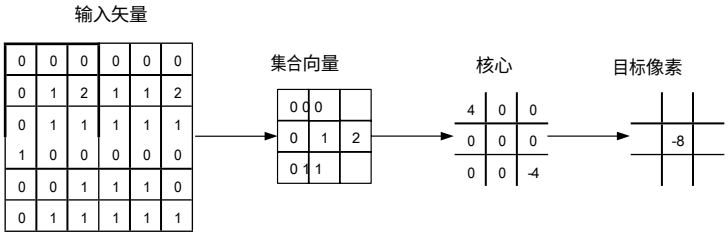

图4：卷积层的直观表示。核的中心元素被放在输入向量上，然后计算并替换为其自身和任何附近像素的加权和。

每个内核都会有一个相应的激活图，这些激活图将沿着深度维度堆叠，形成来自卷积层的全部输出量。

正如我们前面提到的，在图像等输入上训练ANN会导致模型过大而无法有效训练。这归结于标准ANN神经元的全连接方式，所以为了减轻这种情况，卷积层中的每个神经元都只与输入量的小区域连接。这个区域的维度通常被称为神经元的**再感知场大小**。通过深度连接的大小几乎总是等于输入的深度。

例如，如果网络的输入是大小为64×64×3的图像（维度为64×64的RGB彩色图像），我们将感受野的大小设置为6×6，我们将在卷积层的每个神经元上总共有108个权重。(6×6×3，其中3是整个体积深度的连通性的大小)为了把这个问题看清楚，在其他形式的ANN中看到的标准神经元将包含12，288个权重。

卷积层也能够通过优化其输出而大大降低模型的复杂性。这些是通过三个超参数进行优化的，即**深度**、**跨度**和设置**零填充**。

卷积层产生的输出量的**深度**可以通过层内神经元的数量来手动设置，以达到输入的同一区域。这可以从其他形式的ANN中看到，隐藏层中的所有神经元都是直接连接到每一个神经元之前的。减少这个超参数可以大大减少网络的神经元总数，但也会大大降低模型的模式识别能力。

我们还能够定义**跨度**，在这个**跨度**中，我们围绕输入的空间维度设置深度，以放置感受野。例如，如果我们将跨度设置为1，那么我们就会有一个严重重叠的感受野，产生极大的激活。另外，将跨度设置为更大的数字将减少重叠量，产生较低空间维度的输出。

零填充是对输入的边界进行填充的简单过程，是一种有效的方法，可以进一步控制输出卷的尺寸。

重要的是要明白，通过使用这些技术，我们将改变卷积层输出的空间维度。为了计算这一点，你可以利用以下公式：

$$\frac{(v - r) + 2z\ s}{+\ 1}$$

其中，*V*代表输入体积大小（高×宽×深），*R*代表感受野大小，*Z*是零填充量的设置，*S*指的是步长。如果从这个方程中计算出的结果不等于一个完整的
整数，那么跨度就被错误地设置了，因为神经元将无法整齐地穿过给定的输入。

尽管到目前为止我们已经做出了最大的努力，但我们仍然会发现，如果我们使用任何*实际*维度的图像输入，我们的模型仍然是巨大的。然而，我们已经开发了一些方法来大大减少卷积层内参数的总体数量。

参数共享的工作原理是，如果一个区域的特征在一个设定的空间区域是有用的计算，那么它很可能在另一个区域也是有用的。如果我们将输出体积内的每个单独的激活图限制在相同的权重和偏置上，那么我们将看到卷积层产生的参数数量大量减少。

因此，随着反向传播阶段的发生，输出的每个神经元将代表整体的梯度，可以在整个深度上进行合计。
- 因此，只更新一组权重，而不是每一组。

2.3 集合层

池化层的目的是逐步降低代表的 维度，从而进一步减少参数的数量和模型的计算复杂性。

池化层对输入的每个激活图进行操作，并使用 "MAX "函数对其维度进行扩展。在大多数CNN中，这些层以**最大集合层**的形式出现，其内核的维度为2×2。
沿着输入的空间维度，跨度为2。这样就可以将
激活图缩小到原始尺寸的25%--同时保持深度
缩小到其标准尺寸。

由于池化层的破坏性，一般只观察到两种最大池化的方法。通常，汇集层的跨度和过滤器都被设置为2×2，这将允许汇集层延伸通过
覆盖了输入的全部空间维度。此外，还可以利用**重叠集合**，其中跨度设置为2，核大小设置为
3.由于池子的破坏性，内核大小超过3将
通常会大大降低模型的性能。

同样重要的是，除了最大池化，CNN架构还可能包含一般池化。**一般**池化层由池化神经元组成，能够执行多种常见操作，包括L1/L2归一化和平均池化。然而，本教程将主要关注最大池化的使用。

2.4 完全连接的层

全连接层包含的神经元直接与相邻两层的神经元相连，而不与其中任何一层相连。这类似于神经元在传统形式的ANN中的排列方式。(图1)

3 餐馆菜谱

尽管组成一个CNN所需的层数相对较少，但没有制定CNN架构的固定方法。也就是说，简单地把一些层扔在一起并期望它能发挥作用是愚蠢的。通过阅读相关文献，很明显，与其他形式的人工神经网络一样，CNN倾向于遵循一个共同的架构。这种常见的架构在图2中得到了说明，卷积层被堆叠起来，然后以重复的方式汇集各层，再向前输送到全连接层。

另一个常见的CNN架构是在每个池化层之前堆叠两个卷积层，如图5所示。这是强烈推荐的，因为堆叠多个卷积层可以选择输入矢量的更复杂的特征。

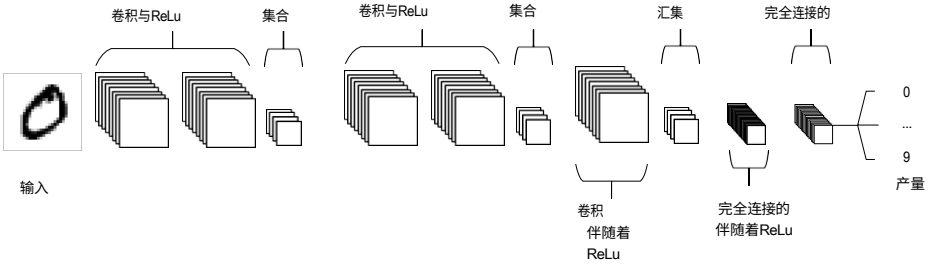

图5：CNN架构的一种常见形式，其中卷积层在ReLus之间连续堆叠，然后再通过pooling层，最后再进入一个或多个全连接的ReLus之间。

我们也建议将大的卷积层分割成许多小的卷积层。这是为了减少特定卷积层内的计算复杂性。例如，如果你要把三个卷积层堆叠在一起，那么你就会发现它的计算复杂度很高。
卷积层相互叠加，感受野为3×3。第一个卷积层的每个神经元将有一个3×3的输入向量视图。第二个卷积层的神经元将有一个5×5的输入向量视图。第三卷积层上的神经元将有一个7×7的输入向量视图。由于这些堆栈具有非线性的特点，这反过来又使我们能够
以较少的参数来表达输入的更强特征。然而，重要的是要明白，这确实伴随着一个明显的内存分配问题--特别是在利用反向传播算法时。

输入层应可递归为二。常见的数字包括32×32、64×64、96×96、128×128和224×224。

在使用小型过滤器时，将跨度设置为1，并使用零填充，以确保卷积层不对输入的任何维度进行重新配置。要使用的零填充量应该通过从接收场大小中取出一个并除以2来计算。

CNN是非常强大的机器学习算法，然而它们可能是可怕的资源密集型。这个问题的一个例子可能是在过滤器
如果输入是227×227（如ImageNet所见），我们用64个内核进行过滤，每个内核的填充值为0，那么结果将是三个227×227×64大小的激活向量--计算起来大约是1000万个激活--或者每幅图像需要70兆字节的巨大内存。在这种情况下，你有两个操作
解决办法。首先，你可以通过以下方式降低输入图像的空间维度

将原始图像的大小调整为不那么重的东西。另外，你也可以违背我们在本文前面所说的一切，选择更大跨度的滤镜尺寸（2，而不是1）。

除了上面概述的几条经验法则外，掌握一些关于通用ANN训练技术的 "技巧 "也很重要。专家们建议阅读Geoffrey Hinton的优秀作品《训练受限玻尔兹曼机的实用指南》。

4 总结

卷积神经网络与其他形式的人工神经网络工作不同，它不是关注整个问题领域，而是利用对特定类型输入的了解。这反过来又允许建立一个更简单的网络结构。

本文概述了卷积神经网络的基本概念，解释了构建卷积神经网络所需的层次，并详细说明了在大多数图像分析任务中如何最好地构建网络。

使用神经网络的图像分析领域的研究在最近一段时间里有些放缓。这部分是由于人们对开始建立这些强大的机器学习算法模型所需的复杂程度和知识的不正确认识。作者希望本文能在某种程度上减少这种混乱，并使初学者更容易进入这个领域。

鸣谢

作者要感谢陆川博士和Nicholas Dimonaco的有益讨论和建议。

参考文献

1. Ciresan, D., Meier, U., Schmidhuber, J.: 用于图像年龄分类的多列深度神经网络。In：计算机视觉和模式识别（CVPR），2012年IEEE会议。3642-3649.IEEE (2012)
2. Cires¸an, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J。用深度神经网络对乳腺癌组织学图像进行分析。In： Medical Image Computing and Computer-Assisted Intervention-MICCAI 2013, pp.411-418.Springer (2013)
3. Ciresan, D.C., Meier, U., Masci, J., Maria Gambardella, L., Schmidhuber, J.: 灵活、用于图像分类的高性能卷积神经网络。In： IJCAI Proceedings-International Joint Conference on Artificial Intelligence. vol. 22, p. 1237 (2011)

4. Cires¸an, D.C., Meier, U., Gambardella, L.M., Schmidhuber, J.: 卷积神经网络委员会用于手写字符分类。在：文档分析与识别（ICDAR），2011年国际会议。第1135-1139页。IEEE (2011)
5. Egmont-Petersen, M., de Ridder, D., Handels, H.: Image processing with neural networksa review.模式识别35(10), 2279-2301 (2002)
6. Farabet, C., Martini, B., Akselrod, P., Talay, S., LeCun, Y., Culurciello, E.: 用于合成视觉系统的硬件加速卷积神经网络。在：电路与系统（ISCAS），2010年IEEE国际研讨会论文集。IEEE (2010)
7. Hinton, G.: A practical guide to training restricted boltzmann machines.Momentum 9(1), 926 (2010)
8. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580 (2012)
9. Ji, S., Xu, W., Yang, M., Yu, K.: 人类动作识别的三维卷积神经网络。模式分析与机器智能，IEEE Transactions on 35(1), 221-231 (2013)
10. Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L.: 用卷积神经网络进行大规模视频分类。在：计算机视觉和模式识别（CVPR），2014年IEEE会议，第1725-1732页。IEEE (2014)
11. Krizhevsky, A., Sutskever, I., Hinton, G.E.: 用深度对话神经网络进行图像网分类。In：神经信息处理系统的进展。
pp.1097-1105 (2012)
12. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.：应用于手写邮政编码识别的反向传播法。神经编译1(4), 541-551 (1989)
13. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.：基于梯度的学习应用于文档识别。IEEE 86(11), 2278-2324 (1998)
14. Nebauer, C.：用于视觉识别的卷积神经网络的评估。Neural Networks, IEEE Transactions on 9(4), 685-696 (1998)
15. Simard, P.Y., Steinkraus, D., Platt, J.C.: 应用于视觉文件分析的卷积神经网络工程的最佳实践。In: null. p. 958.IEEE (2003)
16. Srivastava, N.: 改善有辍学现象的神经网络。多伦多大学博士论文(2013)
17. Szarvas, M., Yoshizawa, A., Yamamoto, M., Ogata, J.: 用卷积神经网络检测行人。In：智能车辆研讨会，2005。Proceedings.IEEE。第224-229页。IEEE (2005)
18. Szegedy, C., Toshev, A., Erhan, D.：用于物体检测的深度神经网络。在：2553-2561 (2013)
19. Tivive, F.H.C., Bouzerdoum, A.: 一类新的卷积神经网络（siconnets）及其在人脸检测中的应用。In：神经网络，2003年。第3卷，第2157-2162页。IEEE (2003)
20. Zeiler, M.D., Fergus, R.: 深度卷积神经网络正则化的随机池化. arXiv preprint arXiv:1301.3557 (2013)
21. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks.In：计算机视觉-ECCV 2014，第818-833页。Springer (2014)