对卷积神经网络的理解

Saad ALBAWI , Tareq Abed MOHAMMED
计算机工程系 工程和建筑学院
伊斯坦布尔Kemerburgaz大学 伊斯坦
布尔，土耳其

Saad AL-ZAWI
电子工程系 工程学院
迪亚拉大学 迪
亚拉 , 伊拉克

***摘要*-- 深度学习或深度神经网络一词是指具有多层的人工神经网络（ANN）。在过去的几十年里，它被认为是最强大的工具之一，并在文献中变得非常流行，因为它能够处理大量的数据。最近，人们对拥有更深的隐藏层的兴趣已经开始超越不同领域的经典方法性能；特别是在模式识别方面。最受欢迎的深度神经网络之一是卷积神经网络（CNN）。它的名字来源于矩阵之间的数学线性操作，称为卷积。CNN有多个层次；包括卷积层、非线性层、池化层和全连接层。卷积层和全连接层有参数，但池化层和非线性层没有参数。CNN在机器学习问题上有出色的表现。特别是处理图像数据的应用，如最大的图像分类数据集（Image Net），计算机视觉，以及自然语言处理（NLP），所取得的结果非常惊人。在本文中，我们将解释和定义与CNN有关的所有元素和重要问题，以及这些元素如何工作。此外，我们还将说明影响CNN效率的参数。本文假设读者对机器学习和人工神经网络都有足够的了解。**

关键词-机器学习，人工神经网络，深度学习，卷积神经网络，计算机视觉，图像识别。

I. 简介

在过去十年中，卷积神经网络在与模式识别相关的各种领域中取得了突破性的成果；从图像处理到语音识别。CNN最有利的方面是减少了ANN中的参数数量。这一成就促使研究人员和开发人员接近更大的模型，以解决复杂的任务，这在经典的ANNs中是不可能的； 。关于由CNN解决的问题，最重要的假设是不应该有空间依赖性的特征。换句话说，例如，在人脸检测应用中，我们不需要注意人脸在图像中的位置。我们唯一关心的是，无论他们在给定图像中的位置如何，都要检测他们。CNN的另一个重要方面是，当输入向深层传播时，获得抽象的特征。例如，在图像分类中，边缘可能在第一层被检测到，然后在第二层检测到更简单的形状，然后是更高层次的特征

*如图*1所示，如[1,3,5,15]所示，在下一层中的面。

II. 卷积神经网络元素

为了很好地掌握CNN，我们从其基本要素开始。

A. 卷积

让我们假设我们的神经网络的输入具有*图*2中呈现的形状，它可以是一个图像（例如CIFAR-10数据集的彩色图像，宽度和高度为32×32像素，深度为3，其中RGB通道）或一个视频（灰度视频，其高度和宽度为分辨率，深度为帧）甚至是一个实验视频，其宽度和高度为（L×L）传感器值，深度与不同时间帧相关，如[2,10,15]。

为什么是卷积？.让我们假设网络接收原始像素作为输入。因此，以CIFAR-10数据集为例，为了将输入层只连接到一个神经元（例如在多层感知器的隐藏层），应该有32×32×3的权重连接。

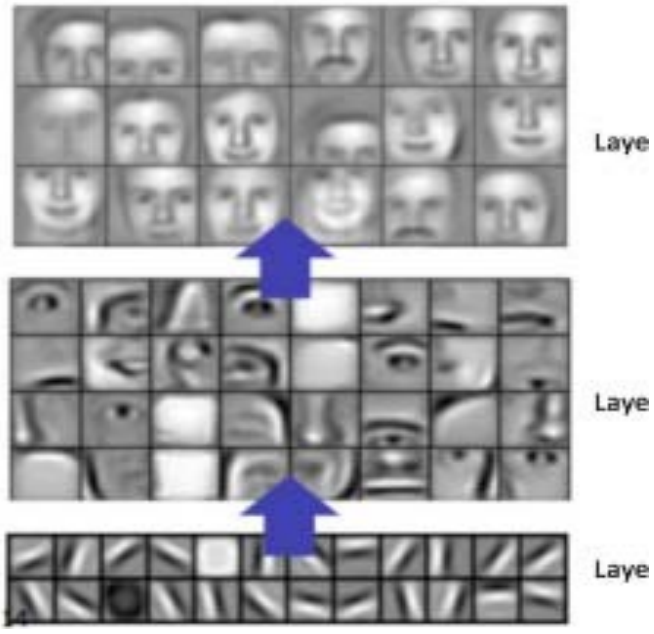


图1.从卷积神经网络学到的特征

ICET2017, 安塔利亚, 土耳

978-1-5386-1949-0/17/$31.00 ©2017 IEE

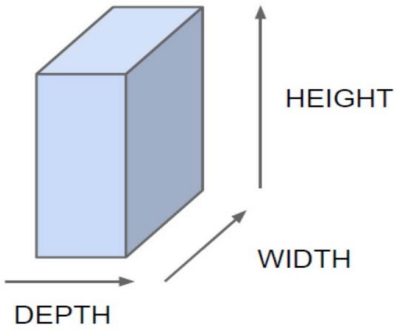


图2.CNN的三维输入表示

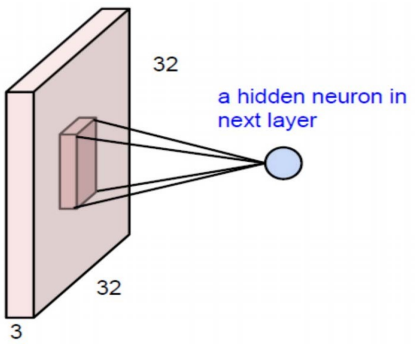


图3.卷积作为全连接网络的替代品。

如果我们在隐藏层中再增加一个神经元，那么我们将需要另一个32×32×3的权重连接，这将成为总共32×32×3×2的参数。为了更清楚地说明问题，6000多个权重参数被用来连接输入到仅有的两个节点。人们可能会认为，对于图像分类应用来说，两个神经元可能不足以进行任何有用的处理。为了使其更有效率，我们可以用完全相同的高度和宽度值将输入图像连接到下一层的神经元。可以认为这个网络适用于图像中的边缘等类型的处理。然而，上述网络需要32×32×3×32的权重连接，即（3,145,728），如[4,5,14]。

因此，为了寻找一种更有效的方法，人们发现，与其说是全面连接，不如说是在图片中寻找局部区域，而不是在整个图片中寻找局部区域，这是一个好主意。*图*3，显示了下一层的区域连接。换句话说，下一层的隐藏神经元只从上一层的相应部分获得输入。例如，它只能与5×5的神经元连接。因此，如果我们想在下一层有32×32个神经元，那么我们将有5×5×3乘以32×32的连接，这就是76,800个连接（而全连接则是3,145,728），如[1,9,13,14,17]。

虽然连接的大小急剧下降，但它仍然留下许多参数需要解决。另一个简化的假设是，保持下一层的整个神经元的局部连接权重固定。这将使下一层的邻居神经元以完全相同的权重连接到上一层的局部区域。因此，它再次放弃了许多额外的参数，并将权重的数量减少到只有5×5×3=75，以便将32×32×3的神经元连接到下一层的32×32[5,8,11]。

这些简单的假设有很多好处。首先，连接的数量从大约3.5万个减少到1.5万个。 3
在所提出的例子中，只有75个连接是百万级的。其次，一个更有趣的概念是，固定局部连接的权重类似于在输入神经元中滑动一个5×5×3的窗口，并将生成的输出映射到相应的位置。它提供了一个检测和识别特征的机会，无论它们在图像中的位置如何。这就是为什么它们被称为卷积的原因[6,7,16]。

为了显示卷积矩阵的惊人效果，*图*4描述了如果我们在一个3×3的窗口中手动挑选连接权重会发生什么。

我们可以看到图4，矩阵可以被设置为检测图像中的边缘。这些矩阵也被称为过滤器，因为它们的作用就像图像处理中的经典过滤器。然而，在卷积神经网络中，这些过滤器被初始化，随后的训练过程中形成过滤器，这更适合于给定的任务。

为了使这种方法更有利，可以在输入层之后添加更多的层。每一层都可以与不同的过滤器相关联。因此，我们可以从给定的图像中提取不同的特征。*图*5，显示了它们与不同层的连接方式。每一层都有自己的过滤器，因此从输入中提取不同的特征。*图*5中所示的神经元使用不同的过滤器，但看的是输入图像的同一部分。[6,8,15,17]

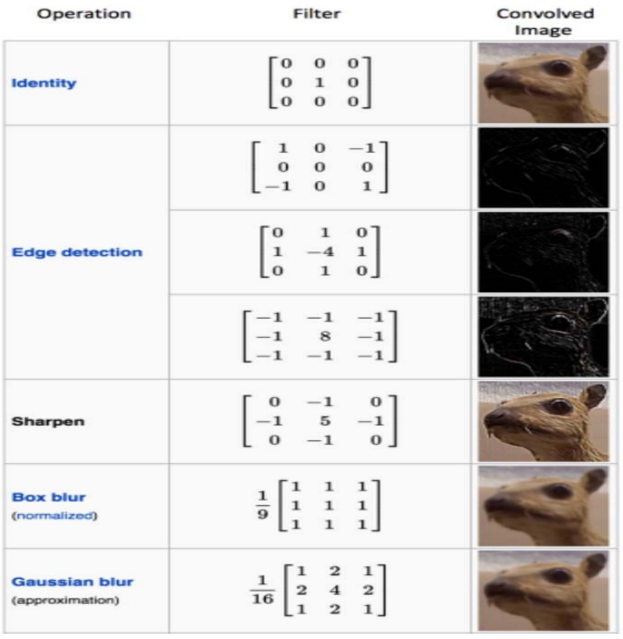


图4.不同卷积矩阵的影响。

ICET2017, 安塔利亚, 土耳

978-1-5386-1949-0/17/$31.00 ©2017 IEE

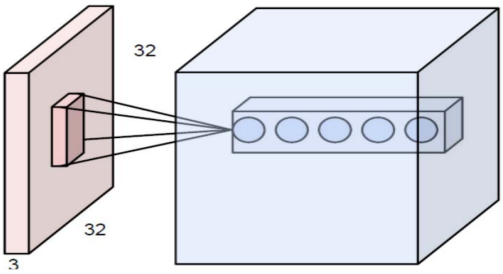


图5.多个图层，每个图层对应不同的过滤器，但在给定的图像中看的是同一个区域。

B. 踔厉风行

事实上，CNN有更多的选项，提供了很多机会，甚至可以越来越多地减少参数，同时减少一些副作用。其中一个选项是跨度。在上面提到的例子中，通过观察区域，我们简单地假设下一层的节点与它们的邻居有很多重叠的地方。我们可以通过控制stride来操纵重叠。*图*6，显示了一个给定的7×7图像。如果我们每次移动过滤器的一个节点，我们就可以只有一个5x5的输出。请注意，*图*6中左边的三个矩阵的输出有重叠（中间的三个矩阵也有重叠，右边的三个矩阵也有重叠）。然而，如果我们移动并使每一个步长为2，那么输出将是3x3。简单地说，不仅是重叠，而且输出的大小也会减少。[5,12,16].

公式(1)对此进行了形式化，给定图像的尺寸为$N\times N$和滤波器的尺寸为$F\times F$，输出尺寸为O，*如图*7所示。

$$O = 1 + \frac{N-F}{S} \quad (1)$$

其中N为输入大小，F为滤波器大小，S为跨度大小。

C. 填充物

卷积步骤的缺点之一是可能存在于图像边界的信息的损失。因为它们只有在滤波器滑动时才会被捕捉到，所以它们永远没有机会被看到。解决这个问题的一个非常简单而又有效的方法是使用零填充。零填充的另一个好处是可以管理输出的大小。例如，在*图*6中，N=7，F=3，步长1，输出将是*5*×5（从7×7*的*输入中缩减）。

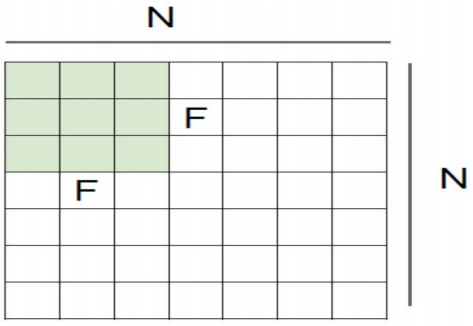


图7.输出中跨度的影响。

然而，通过添加一个零填充，输出将是7×7，这与原始输入完全相同（现在实际的N变成了9，使用公式（1）。包括零填充的修改后的公式为公式（2）。

$$O = 1 + \frac{N+2P-F}{S} \quad (2)$$

其中P是零填充的层数（例如*图*8中的P=1），这种填充的想法有助于我们防止网络输出大小随深度的增加而缩小。因此，可以有任意数量的深度卷积网络。[2,12].

D. CNN的特点

权重共享给模型带来了不变性的转换。它有助于过滤学习特征，而不考虑空间属性。通过为过滤器启动随机值，如果能提高性能，它们将学习检测边缘（*如图*4）。重要的是要记住，如果我们需要知道某个东西在给定的输入中的空间重要性，那么使用共享权重是一个极其糟糕的主意。

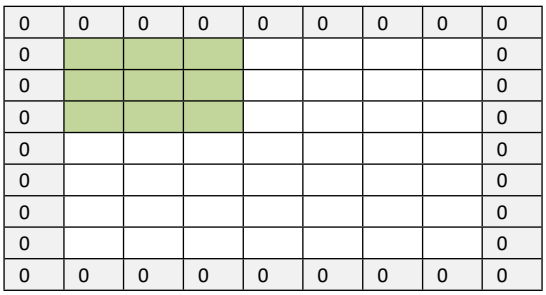


图8.零填充

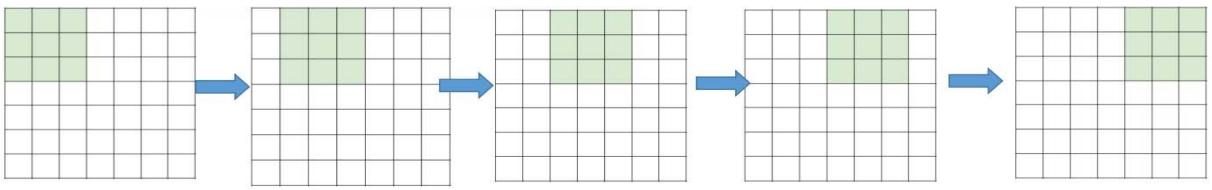
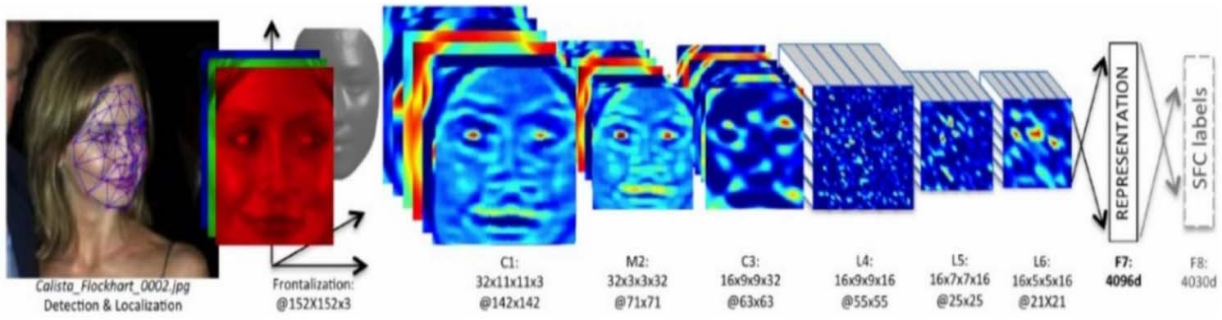


图6.第1步，过滤窗口对每个连接只移动一次

ICET2017, 安塔利亚, 土耳
978-1-5386-1949-0/17/$31.00 ©2017 IEE



图（9）。可视化卷积深度神经网络层

这个概念也可以扩展到不同的维度。例如，如果它是连续的数据，如音频，那么它可以采用一维的音频。如果是图像，如图所示，可以采用二维卷积。而对于视频，或三维图像，可以采用三维卷积。这个简单的想法在2012年ImageNet挑战赛中击败了计算机视觉中所有经典的物体识别方法，*如图9*所示，[5,14,18] 。

E. 卷积公式

下一层的一个像素的卷积是根据公式（3）计算的。

$$\text{[illegible]}(i,j) = (x * w)[i,j] = \sum_{\text{[illegible]}} \sum_{\text{[illegible]}} x[\text{[illegible]}][i-m, j-n] \qquad (3)$$

其中[illegible]是下一层的输出，是输入图像，是内核或过滤矩阵，是卷积操作。*图*10显示了卷积的工作方式。可以看出，输入和核的逐元乘积被聚集起来，然后代表下一层中的相应点。[4,9].

III. 非线性

卷积之后的下一层是非线性。非线性可以用来调整或切断生成的输出。这一层的应用是为了使输出饱和或限制生成的输出。

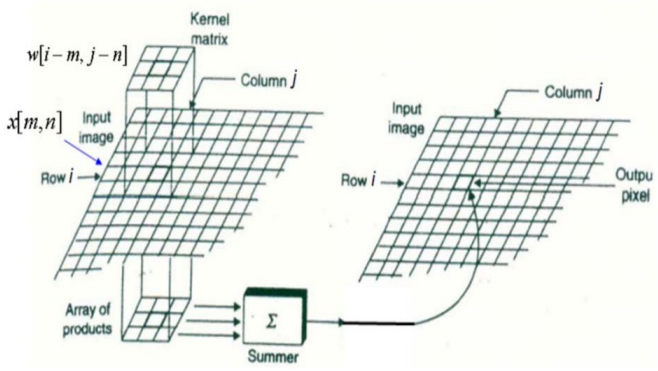


图10.卷积层的细节

许多年来，sigmoid和tanh是最受欢迎的非线性。*图*11，显示了常见的非线性类型。然而，最近，由于以下原因，整流线性单元（ReLU）被更经常地使用。

1. ReLU在函数和梯度方面都有更简单的定义。

$$\text{[illegible]}(x) = \max(0, x) \qquad (4)$$

$$\frac{d}{dx}\text{[illegible]}(x) = \{1 \text{ if } x > 0;\ 0 \text{ otherwise}\} \qquad (5)$$

2. 饱和的函数如sigmoid和tanh在反向传播中会造成问题。随着神经网络设计的深入，梯度信号开始消失，这被称为 "梯度消失"。发生这种情况是因为这些函数的梯度是非常
除了中心以外，几乎所有地方都接近于零。然而，ReLU对于正的输入有一个恒定的梯度。虽然这个函数是不可微调的，但在实际执行中可以忽略它。
3. ReLU创造了一个更稀疏的表示。因为梯度中的零导致获得一个完全的零。然而，sigmoid和tanh总是有来自梯度的非零结果，这可能对训练不利。[2,5,13].

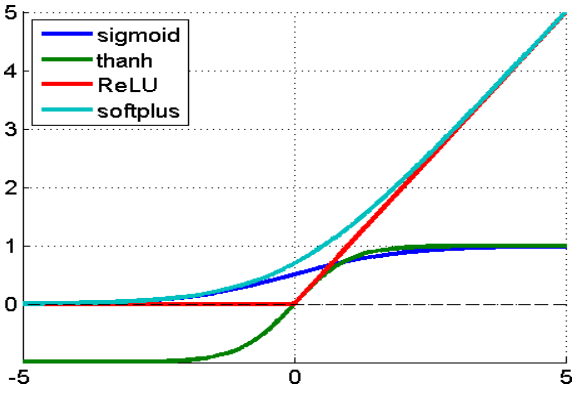


图11.常见的非线性类型。

978-1-5386-1949-0/17/$31.00 ©2017 IEE

IV. 集合

池化的主要思想是向下采样，以减少进一步层的复杂性。在图像处理领域，它可以被视为类似于降低分辨率。池化并不影响过滤器的数量。最大池化是最常见的池化方法之一。它将图像分割成子区域的矩形，并且只返回该子区域内部的最大值。最大集合法中最常用的尺寸之一是2×2。*从图*12中可以看出，当在左上角的2×2块（粉色区域）进行池化时，它移动了2，并将焦点放在右上角的部分。这意味着跨度2被用于池化。为了避免向下取样，跨度
1可以使用，这并不常见。应该考虑的是，下采样并不能保留信息的位置。因此，只有当信息的存在是重要的（而不是空间信息）时，才应该应用它。此外，池化可以与非等效滤波器和步长一起使用以提高效率。例如，一个3x3的最大集合，步长为2，可以保持区域之间的一些重叠。[5,10,16].

V. 完全连接的层

全连接层与传统神经网络中神经元的排列方式相似。因此，全连接层中的每个节点都直接连接到上一层和下一层的每个节点，*如图*13所示，从这个图中我们可以注意到，汇集层中最后一帧的每个节点都作为一个向量从全连接层连接到第一层。这些是CNN在这些层内使用最多的参数，并且在训练中需要很长的时间[3,8]。

全连接层的主要缺点是，它包括很多需要在训练实例中进行复杂计算的参数。因此，我们试图消除节点和连接的数量。去除的节点和连接可以通过使用dropout技术来满足。例如，LeNet和AlexNet在保持计算复杂度不变的情况下，设计了一个深而宽的网络[4,6,9]。

CNN网络的本质，也就是卷积，是在引入非线性和池化层的时候。最常见的架构是使用其中的三个作为

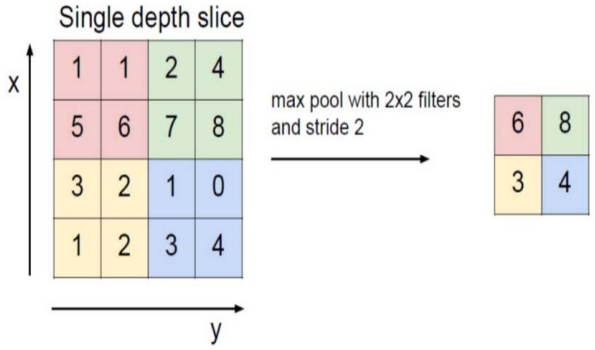


图12.图12展示了最大集合的情况。带有2x2滤波器和跨度2的最大集合导致每个2x2块的下采样被映射为1块（像素）。

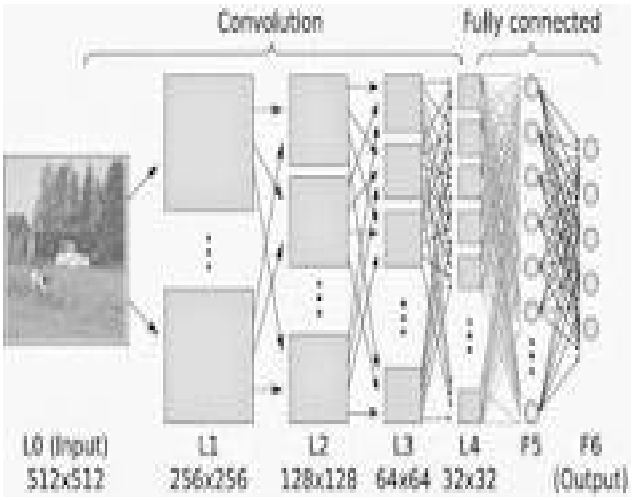


图13.全连接的层

VI. 流行的CNN架构

A. *乐网*

LeNet是由Yan LeCun提出的，用于数字识别，图14，它包括5个卷积层和一个全连接层（像MLP）。[8,19]

B. *淘宝网*

AlexNet包含5个卷积层以及2个全连接层，用于学习特征，图15，它在第一、第二和第五卷积层之后有最大池。它总共有650K个神经元，60M个参数，以及630M个连接。AlexNet是第一个显示深度学习在计算机视觉任务中有效的例子。[2,7]

VII. 融合

在本文中，我们讨论了与卷积神经网络（CNN）相关的重要问题，并解释了每个参数对网络性能的影响。CNN中最重要的一层是卷积层，它占用了网络中大部分的时间。网络的性能也取决于网络中层的数量。但另一方面，随着层数的增加，训练和测试网络所需的时间也在增加。今天，CNN被认为是机器学习中的全能工具，用于很多应用，如人脸检测和图像、视频识别和语音识别。

ICET2017, 安塔利亚, 土耳
978-1-5386-1949-0/17/$31.00 ©2017 IEE

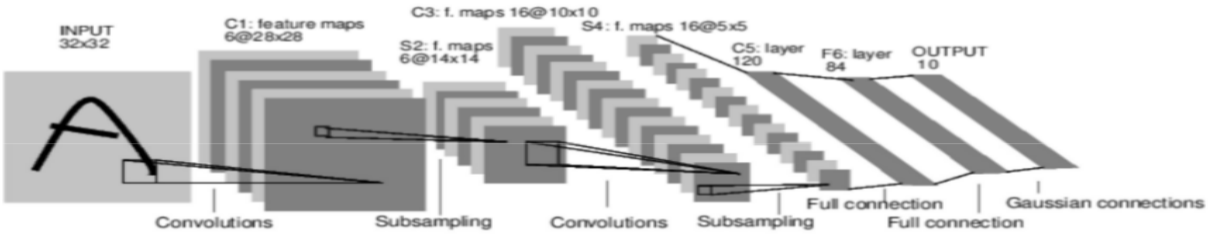


图14.闫乐村介绍的LeNet

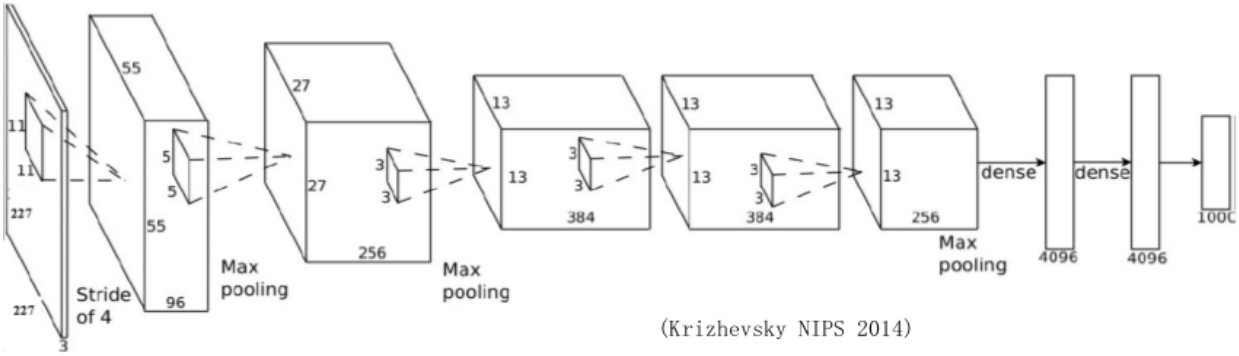


图15.Krizhevsky 2014年介绍的AlexNet

参考文献

[1] O.Abdel-hamid, L. Deng, and D. Yu, "Exploring Convolutional Neural Network Structures and Optimization Techniques for Speech Recognition," no.August, pp. 3366-3370, 2013.

[2] http://www.deeplearningbook.org/contents/convnets.html

[3] V.Dumoulin和F. Visin，"深度学习的卷积算术指南"，第1-28页，2016。

[4] Y.Guo, Y. Liu, A. Oerlemans, S. Wu, and M. S. Lew, "Author ' s Accepted Manuscript Deep learning for visual understanding : A review To appear in： Neurocomputing》，2015年。

[5] https://www.opendatascience.com/blog/an-intuitive-explanation-fconvolutionalneuralnetworks/? utm_source=Open+Data+Science+Newsletter&utm_campaign=f4ea9cc60fEMAIL_CAMPAIGN_2016_12_21& _medium=email&utm_term=0_2ea92bb125-f4ea9cc60f-245860601.

[6] I.Kokkinos, E. C. Paris, and G. Group, "Introduction to Deep Learning Convolutional Networks, Dropout, Maxout 1," pp.1-70。

[7] Krizhevsky, A., Sutskever, I. and Hinton, G.E., 2012.用深度卷积神经网络进行图像网分类。 In Advances inneural information processing systems (pp. 1097-) 1105).

[8] N.Kwak, "卷积神经网络（CNNs）简介)," 2016.

[9] O.Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, C. V Jan, J. Krause, and S. Ma, "ImageNet大规模视觉识别挑战。

[10] D.Stutz和L. Beyer，"理解卷积神经网络"，2014年。

[11] C.Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper with Convolutions, " 2014.

[12] K.Teilo，" 卷积神经网络简介"，No.NOVEMBER 2015，第0-11页，2016。

[13] R.E. Turner, "Lecture 14 : Convolutional neural networks for computer vision," 2014.

[14] J.吴，"卷积神经网络简介"，第1-28页，2016年。

[15] https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

[16] http://www.slideshare.net/hanneshapke/introduction-to-convolutional-神经网络。

[17] 熊伟，杜波，张乐飞，胡瑞敏，陶大成 "Regularizing Deep Convolutional Neural Networks with a Structured Decorrelation Constraint " IEEE 16th International Conference on Data Mining (ICDM) , pp.3366-3370, 2016.

[18] Taigman, Y., Yang, M., Ranzato, M.A. and Wolf, L., 2014.Deepface：缩小人脸验证中与人类水平性能的差距。在IEEE计算机视觉和模式识别会议论文集（第1701-1708页）。

[19] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P., 1998.基于梯度的学习应用于文档识别。IEEE论文集》，86（11），第2278-2324页。

ICET2017, 安塔利亚, 土耳

978-1-5386-1949-0/17/$31.00 ©2017 IEE