

深度学习笔记 | 第4讲：写一篇人人都能看得懂的卷积神经网络

louwill 狗熊会 2018-08-27



新朋友点蓝色字免费订阅



大家好！又到了每周一狗熊会的深度学习时间了。在前面三期的分享中，小编和大家介绍了神经网络的基本原理、优化方法和训练算法等基础性的知识，主题基本是围绕着深度神经网络的结构来进行讲述的。从本期开始，小编将为大家介绍另一种著名的神经网络结构——卷积神经网络，也就是平常大伙儿眼熟的 CNN，CNN 在图像识别、目标检测和语义分割等多个计算机视觉领域有着广泛的应用。本期小编就和大家介绍卷积神经网络的基本原理。



——1——

卷积神经网络发展简史

早在上个世纪 60 年代的时候，生物神经学领域的相关研究就表明，生物视觉信息从视网膜传递到大脑是由多个层次的感受野逐层激发完成的。到了 80 年代，出现了相应的早期感受野的理论模型。这一阶段是

最早期的具备了部分朴素的涉及到卷积网络的理论时期。

到了 1985 年, Rumelhart 和 Hinton 等人提出了 BP 神经网络, 也就是著名的反向传播算法来训练神经网络模型。这基本奠定了神经网络的理论基础, 如今大家在谷歌学术上提出 BP 算法的这篇文章的引用次数是 25292 ... 所以如果你的论文里提到了 BP 算法, 这个数字还得继续往上涨。

Learning internal representations by error propagation

DE Rumelhart, [GE Hinton](#), RJ Williams - 1985 - dtic.mil

This paper presents a generalization of the perception learning procedure for learning the correct sets of connections for arbitrary networks. The rule, called the generalized delta rule, is a simple scheme for implementing a gradient descent method for finding weights that minimize the sum squared error of the system's performance. The major theoretical contribution of the work is the procedure called error propagation, whereby the gradient can be determined by individual units of the network based only on locally available information ...

☆ 被引用次数: 25292 相关文章 所有 35 个版本

有了此前的理论积累, 在 BP 算法提出三年之后, 如今深度学习三巨头之一的 Yann LeCun (小编喜欢叫他燕乐村) 发现可以用 BP 算法来训练一种构造出来的多层卷积网络结构, 并用其训练出来的卷积网络识别手写数字。在 (*Backpropagation applied to handwritten zip code recognition, 1989*) 一文中, LeCun 正式提出了卷积神经网络 (Convolutional Neural Network) 的概念。因而, 现在我们提到 Yann LeCun 这个人物时, 除了深度学习三巨头之外, 还有个名号就是深度学习之父。够霸气吧? 下图最左边这位就是燕乐村巨佬。



深度学习三巨头和吴恩达

至于图中其他人，也一并介绍了吧。除了第一位卷积神经网络之父的燕乐村之外，第二位就是前面提到的发明反向传播算法之一的 Geoffrey Hinton，第三位则是‘幺叔’Yoshua Bengio，这三位便是前面小编说的深度学习三巨头，是他们支撑起了深度学习的发展。那么第四位大家可能就很熟了，是我们的 Adrew Ng 吴恩达。大家若想更多的了解以上各位大神的故事，可参看 NG 对三位大神的采访，看他们对大家学习机器学习和深度学习是怎样建议的。

咱们继续说卷积神经网络的发展历程。LeCun 正式提出卷积神经网络之后，经过一些年的酝酿，在 1998 年提出了卷积神经网络的开山之作——LeNet-5 网络。提出 LeNet-5 的这篇论文 ([Gradient-based learning applied to document recognition , 1998](#)) 引用次数已达 13637 次：

Gradient-based learning applied to document recognition

[Y LeCun, L Bottou, Y Bengio...](#) - Proceedings of the ..., 1998 - [ieeexplore.ieee.org](#)

Multilayer neural networks trained with the back-propagation algorithm constitute the best example of a successful gradient based learning technique. Given an appropriate network architecture, gradient-based learning algorithms can be used to synthesize a complex decision surface that can classify high-dimensional patterns, such as handwritten characters, with minimal preprocessing. This paper reviews various methods applied to handwritten character recognition and compares them on a standard handwritten digit ...

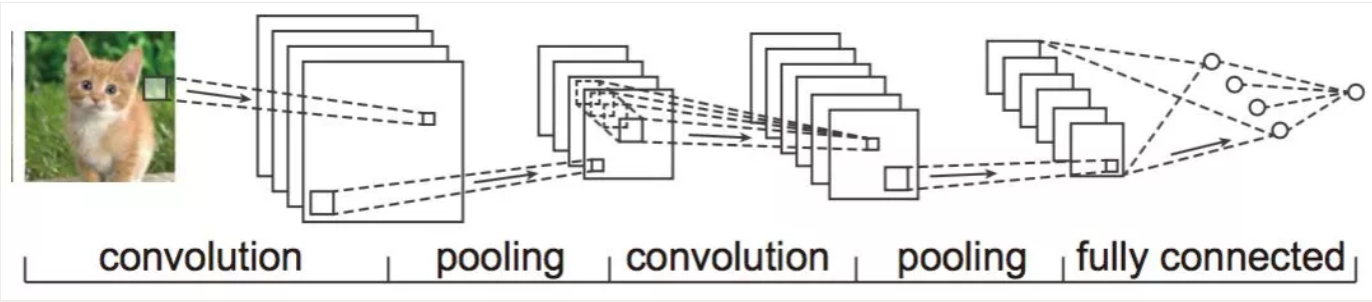
☆ 被引用次数: 13637 相关文章 所有 55 个版本

此后进入新世纪，因为计算能力和可解释性等多方面的原因，神经网络的发展经历了短暂的低谷，直到 2012 年 ILSVRC 大赛上 AlexNet 一举夺魁，此后大数据兴起，以卷积神经网络为代表的深度学习方法逐渐成为计算机视觉、语音识别和自然语言处理等领域的主流方法，卷积神经网络才真正实现开宗立派。

2

卷积

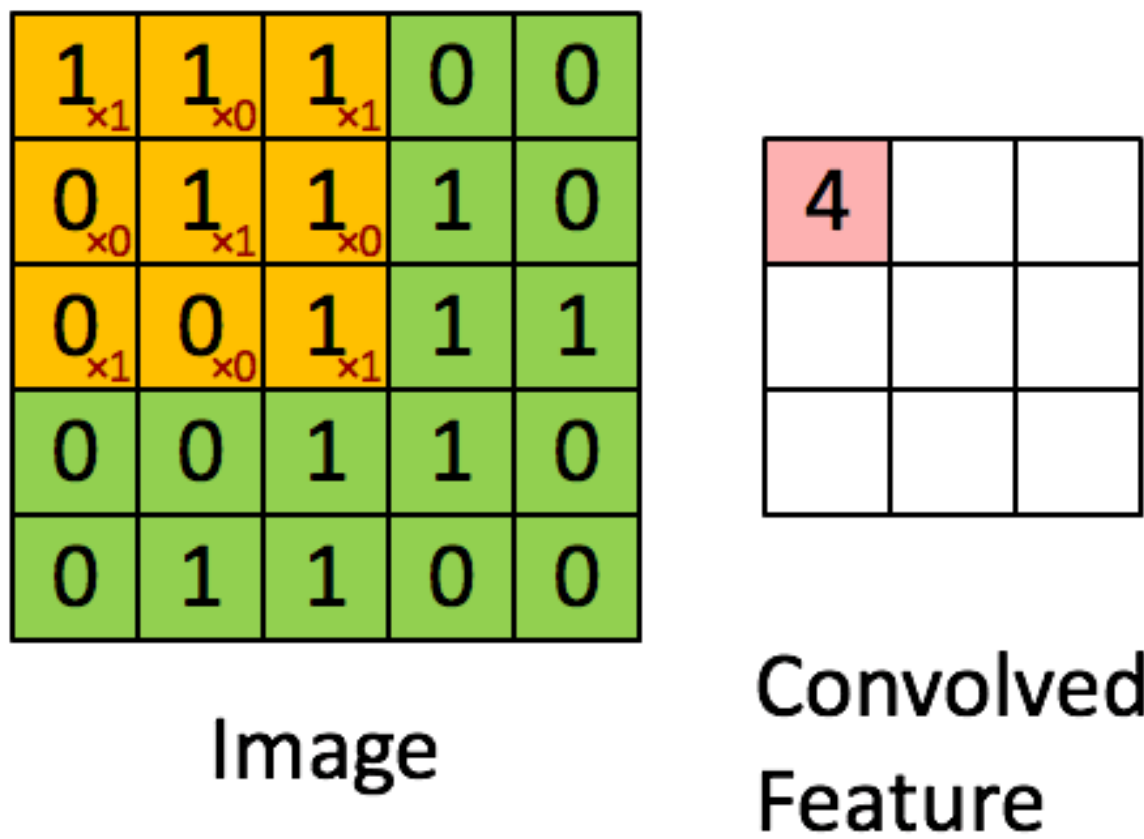
从前面的学习中，我们了解了深度神经网络的一般结构、前向传播和反向传播机制，而卷积神经网络相较于深度神经网络，其主要区别就在于卷积层，卷积层的存在使得神经网络具备更强的学习和特征提取能力。除了卷积层之外，池化层（Pooling layer）的存在也使得卷积神经网络的稳健性更好，最后则是 DNN 中常见的全连接层（Fully Connected layer）。一个典型的卷积神经网络通常包括这三层。



卷积神经网络的基本结构

说了半天的卷积神经网络，想必大家也等不及要问：到底什么是卷积？

从数学来说，卷积可以理解为一类似于加权运算一样的操作。在图像处理中，针对图像的像素矩阵，卷积操作就是用一个卷积核来逐行逐列的扫描像素矩阵，并与像素矩阵做元素相乘，以此得到新的像素矩阵。这个过程是为卷积。其中卷积核也叫过滤器或者滤波器，滤波器在输入像素矩阵上扫过的面积称之为感受野。可能你还有点晕，让我来更详细的解释下。



卷积过程（来自deeplearningai课程）

且看上面的动图（这里感谢一下 NG 大大给我们提供这么好的教学资料），我们用一个 3x3 的滤波器去扫描一个 5x5 的像素矩阵，用滤波器中每一个元素与像素矩阵中感受野内的元素进行乘积运算，可得到了一个 3x3 的输出像素矩阵，这个输出的 3x3 像素矩阵能够较大程度的提取原始像素矩阵的图像特征，这也是卷积神经网络之所以有效的原因。为防止有同学不清楚卷积是如何计算的，笔者以输出像素矩阵中第一个元素 4 为例，演示一下计算过程：

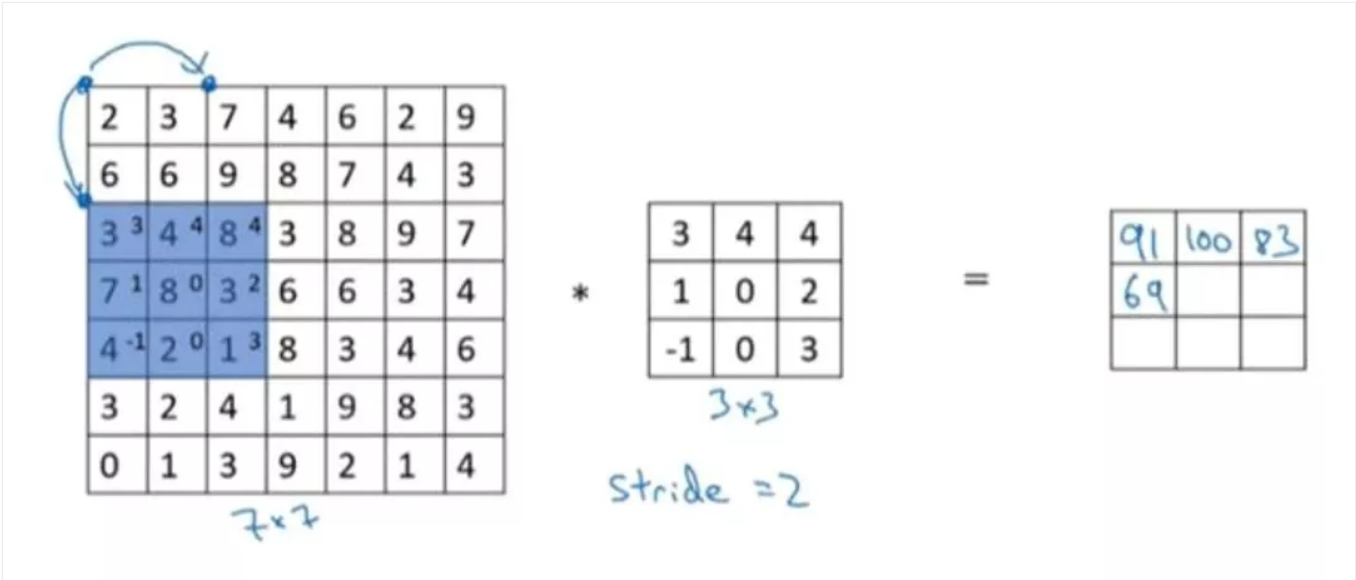
$$1 \times 1 + 1 \times 0 + 1 \times 1 + 0 \times 0 + 1 \times 1 + 1 \times 0 + 0 \times 1 + 0 \times 0 + 1 \times 1 = 4$$

当然，这里你可能会问：如何确定经过卷积后的输出矩阵的维度？我们是有计算公式的。假设原始输入像素矩阵的 shape 为 `nxn`，滤波器的 shape 为 `fxf`，那么输出像素矩阵的 shape 为 `(n-f+1)x(n-f+1)`。比如说 3x3 的滤波器去扫描一个 5x5 的输入图像，按照公式计算的话输出就是 `(5-3+1)x(5-3+1)=3x3`。在训练卷积网络时，我们需要初始化滤波器中的卷积参数，在训练中不断迭代得到最好的滤波器参数。

大体上卷积操作就是这么个过程，是不是非常简单。但这里我们也需要注意两个问题：第一个就是滤波器移动的步幅问题，上面的例子中我们的滤波器的移动步长为 1，即在像素矩阵上一格一格平移。但如果滤波器是以两个单位或者更多单位平移呢？这里就涉及到卷积过程中的 stride 问题。第二个问题涉及到卷积操

作的两个缺点，第一个缺点在于每次做卷积，你的图像就会变小，可能做了几次卷积之后，你的图像就变成 1x1，这就不好办了。第二个缺点在于原始输入像素矩阵的边缘和角落的像素点只能被滤波器扫到一次，而靠近像素中心点的像素点则会被多次扫到进行卷积。这就使得边缘和角落里的像素特征提取不足，这就涉及到卷积过程中的 padding 问题。

针对第一个问题，也就是卷积步长问题，其实也很简单，就是按照正常的卷积过程去操作，只不过每次多走一个像素单位而已。且看卷积步幅为 2 的卷积操作示例：



我们用一个 3x3 的滤波器去对原始像素为 7x7 的图像进行卷积操作，设定卷积步长为 2，可看到输出像素矩阵的第二行第一个元素 69 的计算跨越了两个像素格点，计算过程为：

$$3 \times 3 + 4 \times 4 + 8 \times 4 + 7 \times 1 + 8 \times 0 + 3 \times 2 + 4 \times -1 + 2 \times 0 + 1 \times 3 = 69$$

加入步长之后我们的输出像素矩阵的 shape 的计算公式需要更新一下为：

$((n-f)/s+1) \times ((n-f)/s+1)$ 。其中 s 为步长。

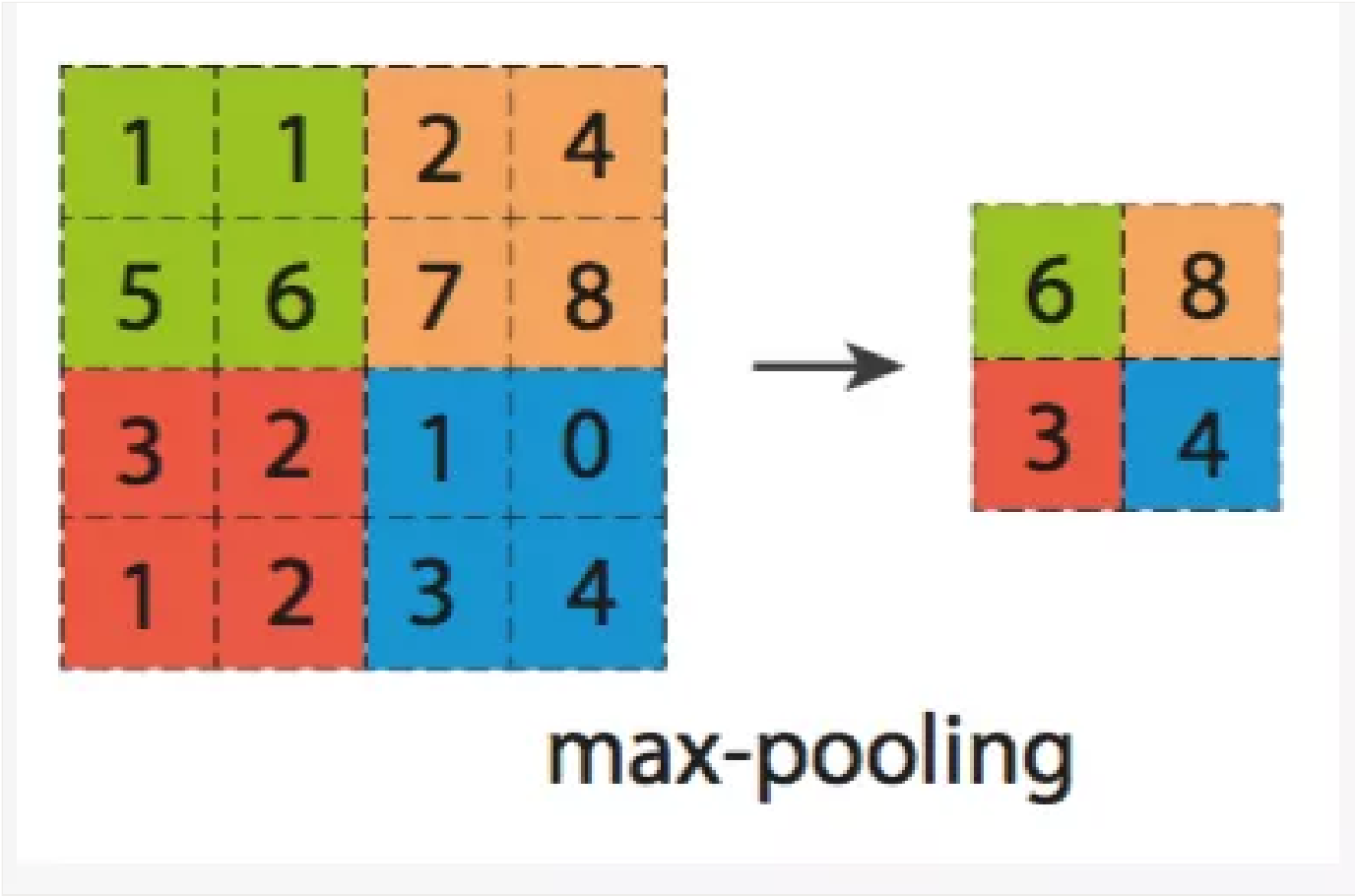
针对第二个问题，卷积神经网络采用一种叫做 padding 的操作，即对原始像素边缘和角落进行零填充，以期能够在卷积过程中充分利用边缘和角落的像素特征。至于填充多少 0 像素值，一般有两个选择，一是 valid 填充，也就是不填充，所以就不用管它了。我们在意的是有填充，就是第二种，same 填充方法。即填充后，输入和输出大小是一致的，对于 nxn 大小的输入像素，如果你用填充了 p 个像素点之后， n 就变成了 $n+2p$ ，最后输出像素的 shape 计算公式就变成了 $((n+2p-f)/s+1) \times ((n+2p-f)/s+1)$ ，要想让 $n+2p-f+1=n$ 的话，输入输出大小相等，则 $p=(f-1)/2$ 。所以，一般而言，滤波器的大小 f 都会选择为奇数个。

以上便是卷积神经网络中卷积的基本过程描述。一个完整的卷积神经网络除了最重要的卷积层之外，还有池化层和全连接层。

3

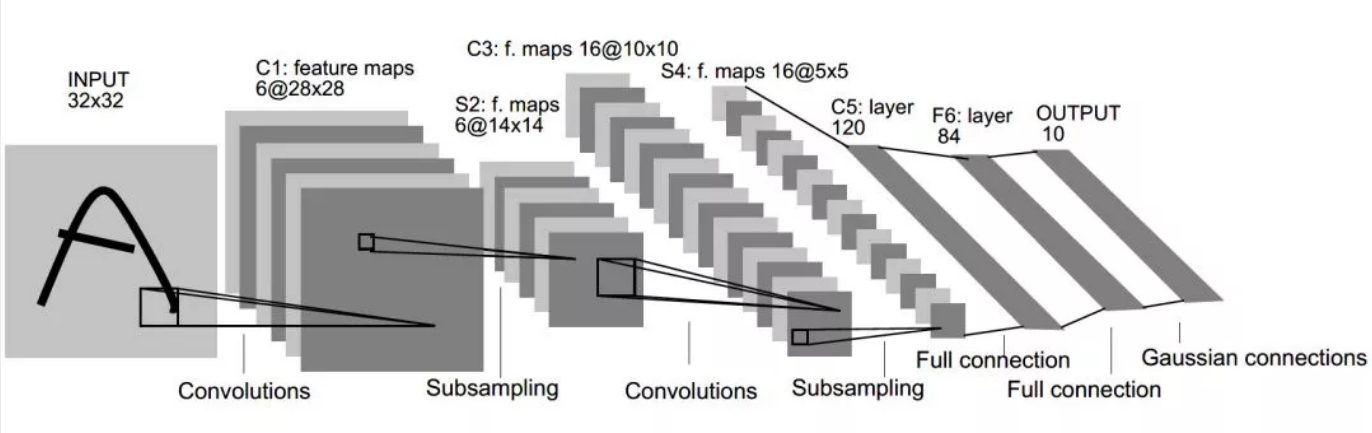
池化和全连接

通常在设计卷积网络结构时，卷积层后会跟着一个池化层。池化（pooling）的操作类似于卷积，只是将滤波器与感受野之间的元素相乘改成了对感受野直接进行最大采样。简单来说，池化层是用来缩减模型大小，提高模型计算速度以及提高所提取特征的鲁棒性。池化操作通常有两种，一种是常用的最大池化（max pooling），另一种是不常用的平均池化（average pooling）。池化操作过程也非常简单，假设池化层的输入为一个 4×4 的图像，我们用最大池化对其进行池化，执行最大池化的卷积核是一个 2×2 的矩阵，执行过程就是将输入矩阵拆分为不同区域，对于 2×2 的输出而言，输出的每个元素都是其对应区域的最大元素值。



最大池化过程就像是应用了一个 2×2 的滤波器以步幅 2 进行区域最大化输出操作。所以简单而言，池化的参数就是滤波器的大小 f 和步幅 s ，池化的效果就相当于对输入图像的高度和宽度进行缩小。值得注意的是，最大池化只是计算神经网络某一层的静态属性，中间并没有什么学习过程。

池化完成之后就是标准神经网络中的全连接层了。全连接层小编在前三讲中有详细介绍，这里就不再赘述。总之，一个典型的卷积层通常包括卷积层-池化层和全连接层。比如说 LeCun 在1998 年提出的 LeNet-5 网络结构就是最初最经典的卷积网络结构。



LeNet-5 网络结构



本篇笔记小编和大家在介绍卷积神经网络发展历程的基础上，对卷积的基本过程进行了详细的描述，之后又介绍了池化和全连接卷积神经网络的另外两大组成部门。总而言之，一个典型的卷积网络结构设计通常包括卷积层、池化层和全连接层三个部分。在下一讲中，小编将和大家继续深入讨论卷积神经网络的结构、训练和其他更加深入的内容。



【参考资料】

<https://www.deeplearning.ai/>

作者简介

鲁伟，狗熊会人才计划一期学员。目前在杭州某软件公司从事数据分析和深度学习相关的研究工作，研究方向为贝叶斯统计、计算机视觉和迁移学习。



识别二维码，查看作者更多精彩文章



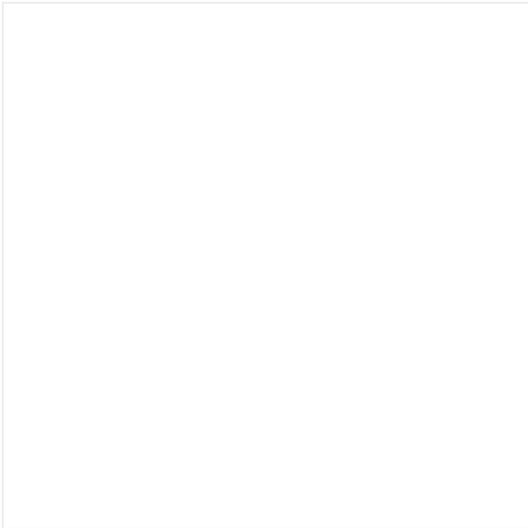
识别下方二维码成为狗熊会会员！

友情提示：

个人会员**不提供数据、代码**，

视频**only**！

个人会员网址：<http://teach.xiong99.com.cn>



点击“[阅读原文](#)”，成为狗熊会会员！

[阅读原文](#)