

我们检测到你可能使用了 Adblock 或 Adblock Plus，它的部分策略可能会影响到正常功能的使用（如关注）。

你可以设定特殊规则或将知乎加入白名单，以便我们更好地提供服务。（为什么？）

机器学习 神经网络 卷积神经网络 (CNN)

关注者
2168

被浏览
100345

卷积神经网络工作原理直观的解释？

我目前大四，在用CNN做手写识别毕业设计，已经接触机器学习4个月了。但CNN是目前最让我困惑的，其简直就像黑匣子，只有输入输出，然后看看准确率，完全不...显示全部

2 条评论 分享 邀请回答

关注问题

写回答

查看全部 29 个回答



下载知乎客户端

卷积神经网络工作原理直观的解释？



张觉非

7 人赞同了该回答

我感觉 CNN 的卷积层就是一个基于滤波器的图像处理器。图像处理领域使用一些人们预先定义好的滤波器起到模糊图像、识别边缘等操作。CNN 就是把滤波器也纳入学习过程，把它们当作可训练的参数。这样在 CNN 的靠近输入图像的那部分（卷积层）就可以认为训练出来了一些滤波器，这些滤波器滤出来的图像再送往网络后端——一个类似传统神经网络的分类器。

至于那些训练出来的滤波器是起什么作用的，那就有点黑盒的意思了。常见的滤波器：高通、低通、高斯模糊、SOBEL 查找边缘 ... 这些，是白盒。是人有目的设计出来的。CNN 卷积层的那些滤

关于作者



张觉非

回答
12

文章
5

关注者
709

关注他

发私信

波器则是根据训练样本拟合出来的。用这些训练得到的滤波器去滤一滤图像看看，也许能看出其中一些滤波器的“目的”。

发布于 2017-02-06

▲ 7

▼

添加评论

分享

★ 收藏

❤ 感谢

更多回答



雨宫夏一

XMU（厦门大学）硕士在读，移动视觉搜索

705 人赞同了该回答

开始正式的答题，学历尚浅，多有偏颇之处，还望指正

来源资料：PRML，cs231n，以及部分论文

其实我们在做线性回归也好，分类（逻辑斯蒂回归）也好，本质上来讲，就是把数据进行映射，要

被收藏 5 次

cnn

0 人关注

风萧萧 创建

cnn

0 人关注

李渊 创建

可以

0 人关注

Sakerjin 创建

深度学习

0 人关注

破碎世界 创建

能力

0 人关注

卷积神经网络工作原理直观的解释？

说，对于一个二分类问题，特别是高维度复杂化之后，数据不一定是线性可分的，这个时候，我们的basis function隆重登场，我们可以把数据进行一定的映射，转变，非线性的线性的，转变之后，就可以进行分类，最明显的例子在andrew NG在讲SVM里面的例子就很好的说明了，但是这个时候问题来了，对于一个很复杂，高维度的数据，我们如何才能找到最好的basis function呢？这个时候，神经网络隆重登场，我们把我们的basis function打开来，我们把误差转递到basis function的里面，通过这样的方式，来得到最好的basis function，同理，我们可以无限打开basis function，一直打开，对应的也就是一层神经网络（具体出自于prml关于神经网络的章节最开始简介的部分），但是问题来了，对于图片怎么办？我们知道，对于图片而言，图片是一个二维度的数据，我们怎样才能通过学习图片正确的模式来对于一张图片有正确的对于图片分类呢？这个时候，有人就提出了一个观点，我们可以这样，对于所有的像素，全部都连接上一个权值，我们也分很多层，然后最后进

相关问题

机器学习有很多关于核函数的说法，核函数的定义和作用是什么？ 45 个回答

BRETT 机器人算有「学习」的能力么？ 16 个回答

如何简单形象又有趣地讲解神经网络是什么？ 42 个回答

在自然科学领域，复杂的模型(如神经网络)在逐渐淘汰掉简单的模型吗？ 29 个回答

行分类，这样也可以，但是对于一张图片来说，像素点太多，参数太多了。然后就有人提出来，我们只看一部分怎么样，就是对于一张图片来说，我们只看一个小窗口就可以了，对于其他的地方，我们也提供类似的小窗口，我们知道，当我们对图片进行卷积的时候，我们可以对图片进行很多操作，比如说图片整体模糊，或者是边缘的提取，卷积操作对于图片来说可以很好的提取到特征，而且通过BP误差的传播，我们可以根据不同任务，得到对于这个任务最好的一个参数，学习出相对于这个任务的最好的卷积核，之所以权值共享的逻辑是：如果说一个卷积核在图片的一小块儿区域可以得到很好的特征，那么在别的地方，也可以得到很好的特征。

这就有了alex net的提出，通过对图片进行五层（不知道有没有记忆错误）的卷积，然后后面三层的全连接，我们可以得到一个很好的结果，特别的相对于更大的数据集而言，最好参数越多越好，也就是网络最好更加深，更加的宽。

但是神经网络到底是什么？对于一批数据我们有很多的问题，为什么设置五层最好，batchsize多少比较好，每一层多少个卷积核（这个到现在我依旧没有一个更好的解释，每一个应该多少卷积核），宽度多少？要不要LRN？每一层都代表了什么？

这些的解释，就要好好看看今年CVPR的文章Visualizing and Understanding Convolutional Networks这篇文章写的很棒，而且2015 CVPR出了很多对于卷积神经网络理解的文章，这篇文章提出了一个反卷积的方法(De-convolution)的方法，这样我们就可以好好看看每一层卷积神经网络到底做了什么

答

机器学习，神经网络在控制科学中的前景和应用大吗？为什么？ 14 个回答

相关 Live 推荐



IBM 机器学习 CTO：解密机器学习核心商业价值



机器学习入门之特征工程



和吴军、刘长明在 RSA 谈创业



深度学习入门误区



如何快速攻克传统算法和数据结构？

卷积神经网络工作原理直观的解释？

联系我们 © 2017 知乎

我们看到，第一个卷积层只是表达了简单的图片的边缘而已，我们来看第二层：



卷积神经网络工作原理直观的解释？

第二层稍稍复杂了一点点，可以包含的不仅仅是一个边缘，可以是几个边缘的组合

第三层：

第四层：



卷积神经网络工作原理直观的解释？



卷积神经网络工作原理直观的解释？

第五层：



卷积神经网络工作原理直观的解释？

我们看到，每一层都是对于一张图片从最基础的边缘，不断到最复杂的图片自己本身。

同时在进行反卷积的时候M.D. Zeiler and R. Fergus也发现，对于第一层的alexnet，会得到频度很高的像素（也就是颜色很深），所以他们也提出了应该要减小窗口，这样可以得到频度比较适中的像素：

当图片卷积完之后，会把一个图片对于这一类本身最独特的部分凸显出来，然后来进行判断，这一类到底是什么？有下面的实验截图：



卷积神经网络工作原理直观的解释？

如何判断的，我们看到卷积到最后（左二），比较凸显出来的是狗的头部，左二和右二的意思是，当我们遮住不同的区域，判断是狗的几率，红色区域代表概率很高，蓝色区域代表概率很低，我们发现，当我们遮挡住狗的头的地方的时候，我们得到这个物体时狗的概率最低，这个侧面证明了，**所谓卷积神经网络，就是会自动的对于一张图片学习出最好的卷积核以及这些卷积核的组合方式，也就是对于一张图片的任务来说，求出最好的图片对于本任务的特征的表达，然后来进行判断**还有一篇文章也助于理解，

Understanding Deep Image Representations by Inverting Them

这篇对于卷积每一层都不断的还原到最原始的图片：

越是到后面，图片越模糊，但是它自己独特的部分，却凸显了出来。（也就是这个猩猩还是狒狒的部分）

还望指正，多谢

反正我写的也没什么人看，随意干啥都行，标明作者就好了。

编辑于 2016-01-18

▲ 705



● 45 条评论

➦ 分享

★ 收藏

♥ 感谢



张旭

薛定谔的猫

108 人赞同了该回答

谈点个人理解，占个坑。

卷积神经网络工作原理直观的解释？

如果学过数字图像处理，对于卷积核的作用应该不陌生，比如你做一个最简单的方向滤波器，那就是一个二维卷积核，这个核其实就是一个模板，利用这个模板再通过卷积计算的定义就可以计算出一幅新的图像，新的图像会把这个卷积核所体现的特征突出显示出来。比如这个卷积核可以侦测水平纹理，那卷积出来的图就是原图水平纹理的图像。

现在假设要做一个图像的分类问题，比如辨别一个图像里是否有一只猫，我们可以先判断是否有猫的头，猫的尾巴，猫的身子等等，如果这些特征都具备，那么我就判定这应该是一只猫（如果用心的话你就会发现这就是CNN最后的分类层，这一部分是我们传统的神经网络的范畴）。关键在于这些特征是高级的语义特征，这种特征怎么用卷积核提取呢？

原来的卷积核都是人工事先定义好的，是经过算法设计人员精心设计的，他们发现这样或那样的设计卷积核通过卷积运算可以突出一个什么样的特征，于是就高高兴兴的拿去卷积了。但是现在我们所需要的这种特征太高级了，而且随任务的不同而不同，人工设计这样的卷积核非常困难。

于是，利用机器学习的思想，我们可以让他自己去学习出卷积核来！也就是学习出特征！

如前所述，判断是否是一只猫，只有一个特征不够，比如仅仅有猫头是不足的，因此需要多个高级语义特征的组合，所以应该需要多个卷积核，这就是为什么需要学习多个卷积核的原因。

还有一个问题，那就是为什么CNN要设计这么多层呢？首先，应该要明白，猫的头是一个特征，但是对于充斥着像素点的图像来说，用几个卷积核直接判断存在一个猫头的还是太困难，怎么办？简单，把猫头也作为一个识别目标，比如猫头应该具有更底层的一些语义特征，比如应该有猫的眼睛、猫的耳朵、猫的鼻子等等。这些特征有的还是太高级了，没关系，继续向下寻找低级特征，一直到最低级的像素点，这样就构成了多层的神经网络。

最好，CNN最不好理解的就要放大招了。虽然我们之前一直用一些我们人常见的语义特征做例子，但是实际上CNN会学习出猫头、猫尾巴、猫身然后经判定这是猫吗？显然我们的CNN完全不知道什么叫猫头、猫尾巴，也就是说，CNN学习到的特征根本不是我们人类理解的语义特征，而是一种抽



卷积神经网络工作原理直观的解释？

已，但是另一个选取的特征就是有没有猫的毛，猫的爪子，还有的人更加奇怪，他会去通过这张图像里是不是有老鼠去判断，而我们的CNN，则是最奇怪的一个人，他使用了我们完全无法用语言形容的一系列特征去判断。

分割线

最近又看了一些资料，在此纠正和阐明一些问题。

目前CNN的可视化是一个很火的方向了，有些论文中已经提到了中间的卷积层特征其实也是具有现实的语义意义的，但是只是不那么清晰。

CNN称之为深度学习，要义就在这个深字上，对于CNN而言，这个深其实就是意味着层层特征表示。比如浅层的特征，例如点、线、面之类的简单几何形状，都是在底层训练出来的，对于这些底

层的特征继续进行组合表示，就是后面的若干层的任务。最后把从低级特征组合而来的高级特征在进一步变成语义特征，就可以使用全连接层进行分类了。也就是说，最后一次分类并不一定要用神经网络，如果已经拿到了足够好的特征信息，使用其余的分类器也未尝不可。

这就是为什么CNN可以fine-tune的原因，例如你要完成一个分类猫和狗的任务，你需要从头训练一个CNN网络吗？假设你的猫狗图片样本量并不是很大，这并不是一个好主意。好的办法是，拿一个经过大型图像数据集，你如ImageNet，训练过的大规模CNN（比如VGG NET）直接载入训练，这个过程称之为fine-tuning。因为这个CNN底层已经训练到了丰富的细节信息，你所需要训练的其实是上层对这些特征的组合信息，以及最后全连接层的分类信息，所以完全不需要从头再来。这也证明了CNN确实可以有迁移学习的能力。

编辑于 2017-03-14

▲ 108



● 17 条评论

➦ 分享

★ 收藏

♥ 感谢

查看全部 29 个回答



卷积神经网络工作原理直观的解释？