深度卷积神经网络是这一波 AI 浪潮背后的大功臣。虽然很多人可能都已经听说过这个名词，但是对于这个领域的相关从业者或者科研学者来说，浅显的了解并不足够。近日，约克大学电气工程与计算机科学系的 Isma Hadji 和 Richard P. Wildes 发表了一篇《我们该如何理解卷积神经网络？》的论文：

- **第一章回顾了理解卷积神经网络的动机；**
- **第二章阐述了几种多层神经网络，并介绍当前计算机视觉领域应用中最成功的卷积结构；**
- **第三章具体介绍了标准卷积神经网络中的各构成组件，并从生物学和理论两个角度分析不同组件的设计方案；**
- **第四章讨论了当前卷积神经网络设计的趋势及可视化理解卷积神经网络的相关研究工作，还重点阐述了当前结构仍存在的一些关键问题。**

通过这篇文章，我们希望帮助大家加深对卷积神经网络的理解，并对这个重要概念有一个全面的认知。

目录

第一章

引言

本文动机

过去几年，计算机视觉研究主要集中在卷积神经网络上（通常简称为 ConvNet 或 CNN），在大量诸如分类和回归任务上已经实现了目前为止最佳的表现。尽管这些方法的历史可以追溯到多年前，但相对而言，对这些方法的理论理解及对结果的解释还比较浅薄。

实际上，计算机视觉领域的很多成果都把 CNN 当作了一种黑箱，这种方式虽然有效的，但对结果的解释却是模糊不清的，这也无法满足科学研究的需求。尤其是当这两个问题是互补关系时：

（1）学习的方面（比如卷积核），它到底学习到的是什么？
（2）模型结构设计方面（比如卷积层数量、卷积核数量、池化策略、非线性函数的选择），为什么某些组合会优于其他组合呢？求解这些问题的答案，不仅有利于我们更好地理解卷积神经网络，而且还能进一步提升它的工程实用性。

此外，当前 CNN 的实现方法都需要大量训练数据，而且模型的设计方案对最终的结果有很大的影响。而更深层的理论理解应该减轻模型对数据的依赖性。尽管大量的研究已经集中在卷积神经网络的实现方式，但目前为止，这些研究结果很大程度上还只局限在对卷积操作内部处理的可视化上，目的是为了理解卷积神经网络中不同层的变化情况。

本文目标

针对以上问题，本文将综述几种当前最优秀的多层卷积结构模型。更重要的是，本文还将通过不同方法来总结标准卷积神经网络的各种组件，并介绍它们所基于的生物学或合理的理论基础。此外，本文还将介绍如何通过可视化方法及实例研究来尝试理解卷积神经网络内部的变化情况。我们的最终目标是向读者详细展示卷积神经网络中所涉及到的每一个卷积层操作，着重强调当前最先进的卷积神经网络模型并说明未来仍需解决的问题。

第二章

多层网络结构

近年来，在深度学习或深层神经网络取得成功前，计算机视觉识别系统最先进的方法主要由两个步骤组成，这两个步骤各自分离但又互补：首先，我们需要通过人工设计操作（如卷积、局部或全局编码方法）将输入数据转换成合适的形式。这种输入的变换形式，通常是为了得到输入数据的一种紧凑或抽象的表征，同时还要根据当前任务的需要手动设计一些不变量。通过这种转换，我们能够将输入数据表征成一种更容易分离或识别的形式，这有助于后续的识别分类。其次，转换后的数据通常作为分类器（如支持向量机）训练的输入信号。通常而言，任何分类器的表现都会受到变换后的数据质量及所使用的变换方法的影响。

多层神经网络结构的出现为解决这一问题带来了新的方式，这种多层结构不仅能够训练目标分类器，还能从输入数据中直接学习所需的变换操作。这种学习方式通常称为表征学习，当将其应用在深度或多层神经网络结构中时，我们称之为深度学习。

多层神经网络定义为是一种从输入数据的层次抽象表征中提取有用信息的计算模型。一般而言，设计多层网络结构的目标是为了在高层凸显输入数据的重要信息，同时能让那些不太不重要的信息变化更具鲁棒性。

近年来，研究者已经提出了很多不同类型的多层架构，而大多数的多层神经网络都是以堆叠的方式，将一些线性和非线性函数模块组合形成多层结构。本章将会覆盖计算机视觉应用中最先进的多层神经网络结构。其中，人工神经网络是我们需要的关注重点，因为这种网络结构的表现非常突出。为了方便起见，在下文我们会直接将这类网络称为神经网络。

神经网络

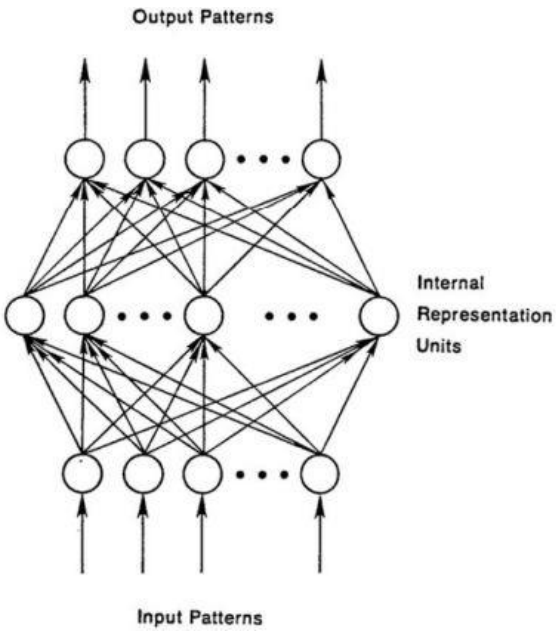
标准的神经网络结构通常由输入层 x，输出层 y 和多个隐藏层 h 堆叠而成，其中每个层还由多个单元组成，如下图所示。通常，每个隐藏单元 hj 接受上一层所有单元的输入，并将其加权组合，其非线性组合的数学形式如下：

$$h_j = F(b_j + \sum_i w_{ij} x_i)$$

wij 是权重值，用于控制输入单位和隐藏单位之间连接的强度，bj 是隐藏单位的偏置，F 是非线性函数，如 Sigmoid 函数。

深度神经网络可以被视为是 Rosenblatt 感知器及多层感知器的实例。 尽管神经网络模型已经存在多年（即自 1960 年代以来），但它们并未被广泛使用。造成这种的原因有很多，最主要的原因是感知器无法模拟像 XOR 这样的简单操作而被外界否定，这也进一步阻碍了研究人员对感知器的研究。

直到最近，一些研究人员将简单感知器扩展到多层神经网络模型。 此外，缺乏适当的训练算法也会延缓感知度的训练进度，而反向传播算法的提出也使得神经网络模型得以普及。更重要的是，多层神经网络结构依赖于大量的参数，这就意味着我们需要大量的训练数据和计算资源来支持模型训练及学习参数过程。



标准神经网络结构示意图

受限波尔茨曼机（RBM）的提出是深层神经网络领域的一大重要贡献。受限玻耳兹曼机可以看作是两层的神经网络，只允许网络以前馈连接的方式堆叠。而神经网络可以看作是使用受限波尔茨曼机进行分层无监督预训练的一种模型，在图像识别任务中，这种无监督学习方法主要包括三个步骤：首先，对于图像中的每个像素，对 xi 及初始化的 wij、偏置 bj、隐藏层状态 hj，其概率可以被定义为：

$$p_j = \sigma(b_j + \sum_i x_i w_{ij})$$

其中，σ（y）= 1 /（1 + exp（-y））。

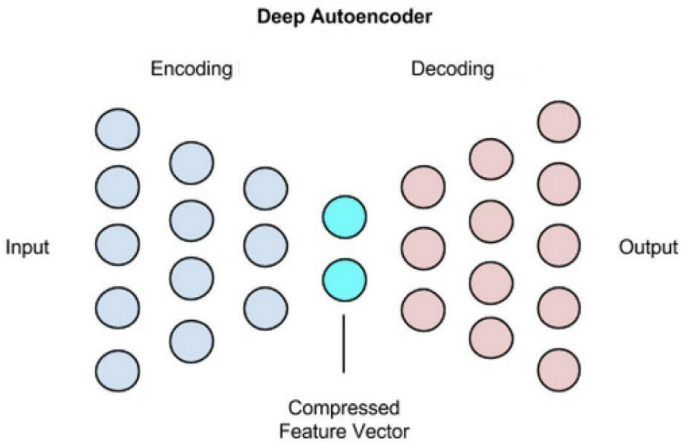
其次，如上式所示，一旦所有的隐藏状态都被随机设定，我们可以根据概率 $p_i = \sigma(b_i + \sum_j h_j w_{ij})$ 将每个像素设定为 1，并以此重建图像。

然后，隐藏单元将通过重建的权重和偏差来更新校正单位的误差：

$$\Delta w_{ij} = \alpha(\langle x_i h_j \rangle_{input} - \langle x_i h_j \rangle_{reconstruction})$$

其中，α 是学习率，（xihj）表示隐藏单元 hj 中像素 xi 出现的次数。整个训练过程将重复 N 次或直到误差下降到预设的阈值 τ。训练完一层后，使用它的输出作为下一层的输入，然后接着重复上述过程训练下一层。通常，网络中的所有层经过预训练后，它们还将通过梯度下降的方式，反向传播误差来进一步微调标记数据。使用这种分层无监督预训练的方式可以不需大量标记数据的情况下，训练深层神经网络结构。因为利用受限波尔茨曼机进行无监督预训练，能够为模型参数的初始化提供了一种有效途径。受限波尔茨曼机的第一个成功应用案例是用于人脸识别的降维，它们被当作是一种自编码器。

自动编码器主要是通过引入不同的正则化方法来防止模型学习一些无关紧要的数据特征。目前一些比较优秀的编码器包括稀疏自编码器、去噪自编码器（DAE）和压缩自编码器（CAE）等。稀疏自编码器允许中间编码表示的大小（即由输入生成编码器）大于输入的大小，同时通过稀疏表示来正则化负相的输出。相反，去噪自编码器改变了编码重建本身的目标，试图重建一个干净、不带噪声的输入版本，得到一个更加强大的表示。类似地，压缩自编码器是通过惩罚噪声中最敏感的单位来实现类似去噪自编码器的过程。



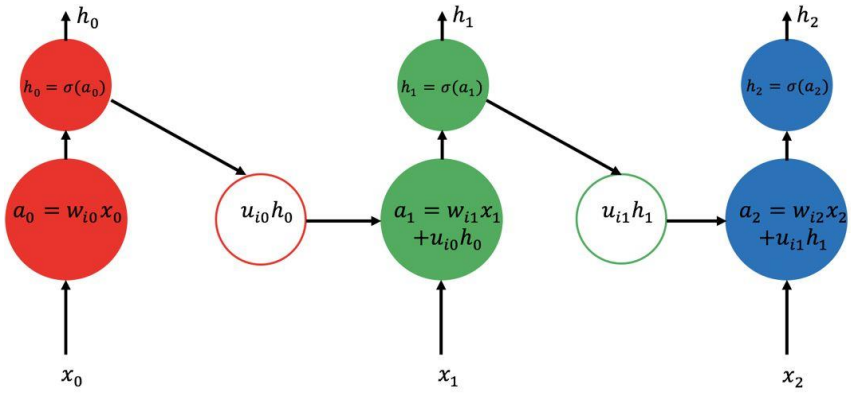
标准的自编码器结构

循环神经网络

循环神经网络是处理序列数据相关任务最成功的多层神经网络模型（RNN）。RNN，其结构示意图如下图所示，它可以看作是神经网络的一种特殊类型，隐藏单元的输入由当前时间步所观察到的数据中获取输入以及它在前一个时间步的状态组合而成。循环神经网络的输出定义如下：

$$h_t = \sigma(w_i x_t + u_i h_{t-1})$$

其中 σ 表示一些非线性函数，wi 和 ui 是网络参数，用于控制当前和过去信息的相对重要性。



标准的循环神经网络结构示意图

每个循环单元的输入将由当前时刻的输入 xt 及上一时刻 ht-1 组成，新的输出表示可通过上式计算得到，并传递给循环神经网络中的其他层。

虽然循环神经网络是一类强大的多层神经网络模型，但其的主要问题是模型对时间的长期依赖性，由于梯度爆炸或梯度消失，这种限制将导致模型训练过程在网络回传过程中误差的不平稳变化。为了纠正这个困难，引入了长短期记忆网络（LSTM）。

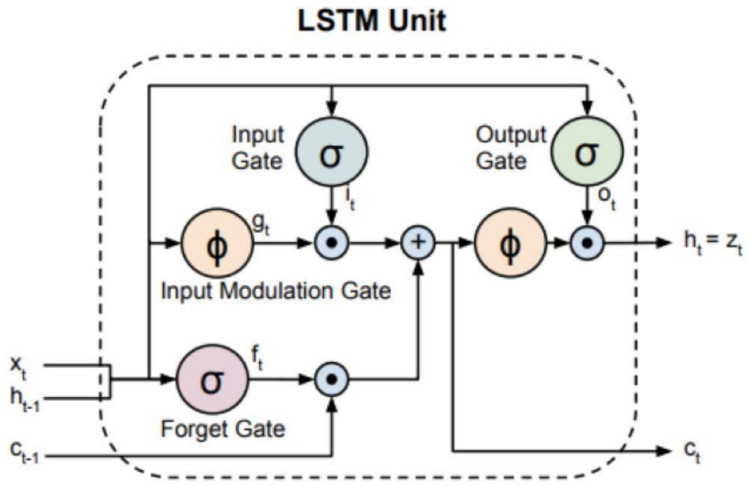
长短期记忆网络（LSTM）的结构示意图下图所示，拥有存储单元或记忆单元，随着时间的推移存储记忆信息。LSTM 的存储单元是通过门控机制从中读取信息或写入信息。 值得注意的是，LSTM 还包含遗忘门，即网络能够删除一些不必要的信息。总的来说， LSTM 的结构主要包含有：三个控制不同的门（输入门、遗忘门及输出门），以及存储单元状态。 输入门由当前输入 xt 和前一个状态 ht-1 控制，它的定义如下：

$$i_t = \sigma(w_i x_t + u_i h_{t-1} + b_i)$$

其中，wi,ui,bi 表示权重和偏差项，用于控制与输入门相关的权重，σ 通常是一个 Sigmoid 函数。类似地，遗忘门定义如下：

$$f_t = \sigma(w_f x_t + u_f h_{t-1} + b_f)$$

相应地，权重和偏差项由 wf,uf,bf 控制。 可以说，LSTM 最重要的一点是它可以应对梯度消失或梯度爆炸时网络中误差传播不平稳的挑战。这种能力的实现是通过遗忘门和输入门的状态进行加法结合来确定存储单元的状态。



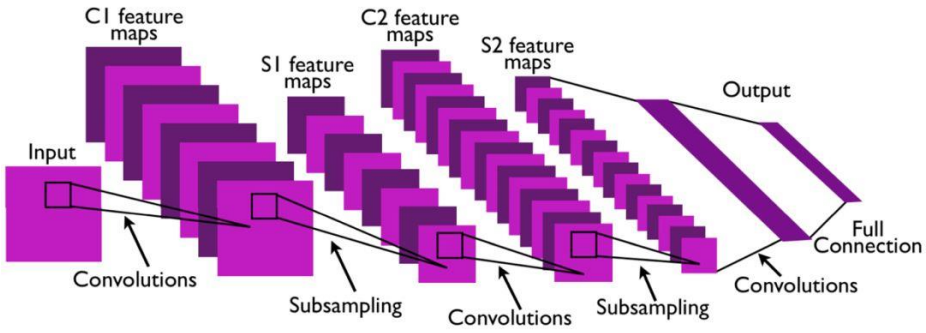
标标准的长短期记忆网络结构示意图

每个循环单元的输入将由当前时刻的输入 xt 及上一时刻 ht-1 组成，网络的返回值将馈送到下一时刻 ht。LSTM 最终的输出由输入门 it，遗忘门 ft 及输出门 ot 和记忆单元状态 ct 共同决定。

卷积神经网络

卷积网络（ConvNets）是一种特殊的神经网络类型，其特别适合计算机视觉应用，因为它们对于局部操作有很强的抽象表征能力。推动卷积神经网络结构在计算机视觉中成功应用的两个关键性的因素：

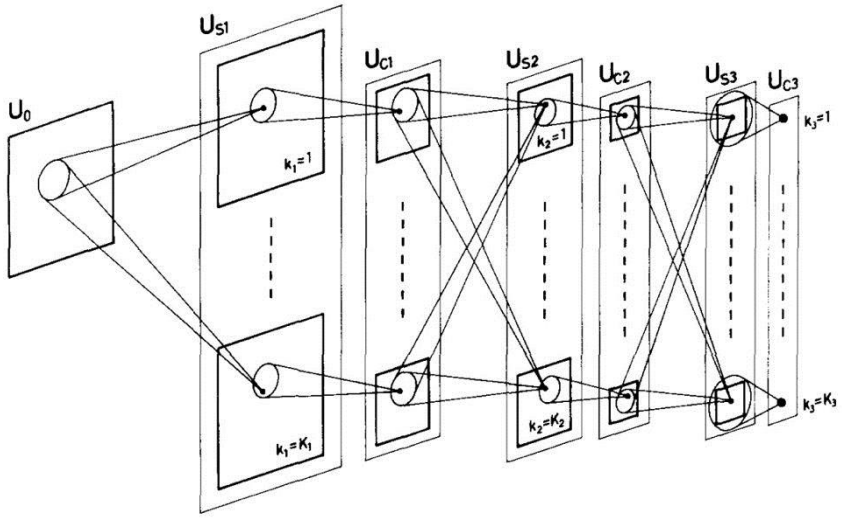
第一，卷积神经网络能够利用图像的 2D 结构和图像相邻像素之间的高度相关性，从而避免在所有像素单元之间使用一对一连接（即如同大多数全连接的神经网络），这有利于使用分组的局部连接。此外，卷积神经网络结构依赖于特征共享原则，正如下图所示，每个通道的输出（或输出的特征映射）都是通过所有位置的相同滤波器的卷积生成。相比于标准的神经网络结构，卷积神经网络的这个重要特性依赖于很少的模型参数。



标准的卷积神经网络结构示意图

第二，卷积神经网络还引入一个池化步骤，在一定程度上保证了图像的平移不变性，这使得模型不受位置变化的影响。还值得注意的是，池化操作使得网络拥有更大的感受野，从而能够接受更大的输入。感受野的增大，将允许网络在更深层学习到更加抽象的特征表征。例如，对于目标识别任务，卷积网络中的浅层将学习到图像的一些边、角特征，而在更深层能够学习到整个目标的特征。

卷积神经网络的结构最早是受生物视觉机制启发而设计的，正如 Hube 在其开创性的研究中所描述的人类视觉皮层的工作原理。随后，Fukushima 提出的神经感知器（Neocognitron）是卷积神经网络的前身，它依赖局部连接的方式，由 K 层神经网络层级联而成，每层神经网络由 S-cell 单元，U sl 及复杂的单元相间分布而成，这种交替分布的形式是模仿生物简单细胞中的处理机制而设计的，其结构示意图如下图所示。



神经感知器结构示意图

此外，在卷积操作后都会跟随一个非线性变化单元，常见的非线性函数是修正线形单元 ReLu，其数学表达式如下：

$$\varphi(x) = \begin{cases} x; & if\, x \geq 0 \\ 0; & x < 0 \end{cases}$$

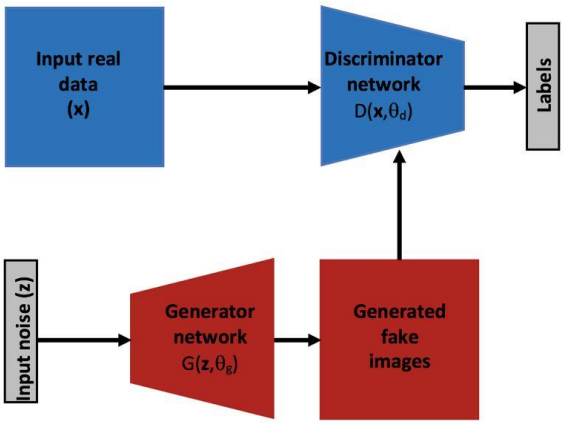
在非线性变换后，通常会引入池化单元。平均池化操作是常用的池化操作之一，通过平均化感受野中的像素值，来综合考虑周围像素的特征。而最大池化则是用来提取相邻像素间最重要的特征信息，避免模型学习到一些无关紧要的特征。经典的卷积网络由四个基本处理层组成： 卷积层、非线性变换层、归一化层及池化层。

近年来，在计算机视觉领域中所应用的卷积神经网络结构，大多是基于 Lecun 在 1998 年提出的用于手写字母识别的 LeNet 卷积模型结构。LeNet 的一个关键是加入反向传播过程来更有效地学习卷积参数。与全连接神经网络相比，虽然卷积神经网络有其独特的优势，但其对于标签数据的严重依赖性，也是其未被广泛使用的主要原因之一。直到 2012 年，随着大型 ImageNet 数据集的发布及计算能力的提高，人们重新恢复对卷积神经网络的研究兴趣。

生成对抗网络

生成对抗网络是 2014 年首次引入的一种新型多层神经网络模型，这种模型结构充分体现了多层网络架构的强大性。虽然生成对抗网络并没有多种不同的网络构建模块，但这种网络结构具有一些特殊性，最关键的是引入了无监督学习方式，使得模型的训练学习不再依赖大量的标记数据。

一个标准的生成对抗模型主要由两部分子网络组成：生成网络 G 和判别网络 D，如下图所示，两个子网络都是预先定义好的多层网络结构（最初提出的模型中二者都是多层全连接网络）。经过交替对抗训练，判别网络的目标是鉴别生成网络的生成数据标签与真实数据标签之间的真伪，而生成网络的目标是生成更加优化的数据，以“欺骗”判别网络，训练的最终结果是使得生成的数据达到以假乱真的目的。



通用的生成对抗网络结构示意图

生成对抗网络自提出以来，因其强大的多层网络结构及独特的无监督学习方式，得到了广泛的关注和研究。GAN 的成功应用包括：文本到图像合成（其中网络的输入是要呈现图像的文字描述）；超分辨率图像的生成，即用较低分辨率的输入生成逼真的高分辨率图像；图像修复，即用 GAN 来生成来自输入图像中的缺失信息；纹理合成，即从输入噪声中生成逼真的纹理特征。

多层网络的训练

如前所述，当前各种多层神经网络结构所取得的成功，在很大程度上取决于网络训练学习过程的进步。通常，神经网络的训练首先需要进行多层无监督预训练，随后，将预训练好的模型进行有监督训练，训练过程都是基于梯度下降的反向传播原则，通过反向传播网络误差，来更正修正模型的参数值，从而优化网络结构及输出结果。

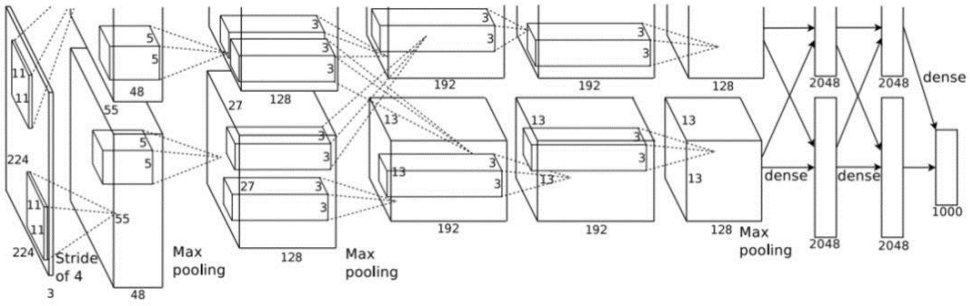
迁移学习

多层神经网络结构的一大益处是在跨数据集甚至跨不同任务中，模型所学得的特征具有通用的适用性。在多层网络结构中，随着层次的增加，所学得的特征表征通常也是

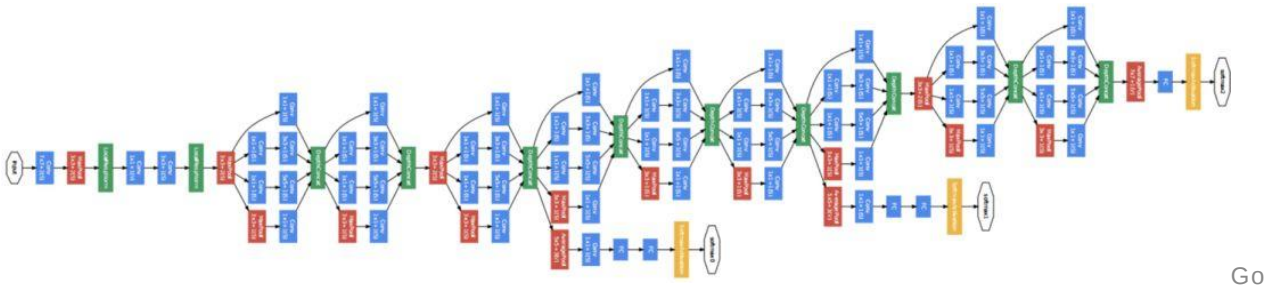
从简单到复杂、从局部到全局发展。因此，在低层次提取的特征往往适用于多种不同任务，这使得多层结构更容易进行迁移学习。

空间卷积神经网络

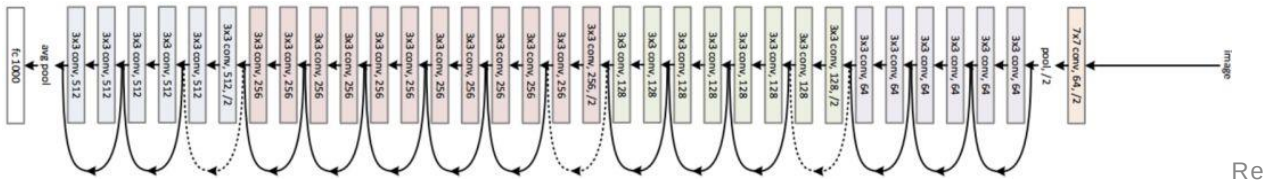
理论上，卷积神经网络可以应用于任意维度的数据，特别适用于二维的图像数据，因此卷积结构在计算机视觉领域受到了相当关注。随着可用的大规模数据集和强大的计算机能力的发展，卷积神经网络在计算机视觉领域的应用也日益增长。本节我们将介绍几种最突出的卷积神经网络结构，包括 AlexNet,VGGNet,GoogleNet,ResNet,DenseNet 等，其结构示意图依次如下，这些体系结构都是基于原始的 LeNet 发展起来的。



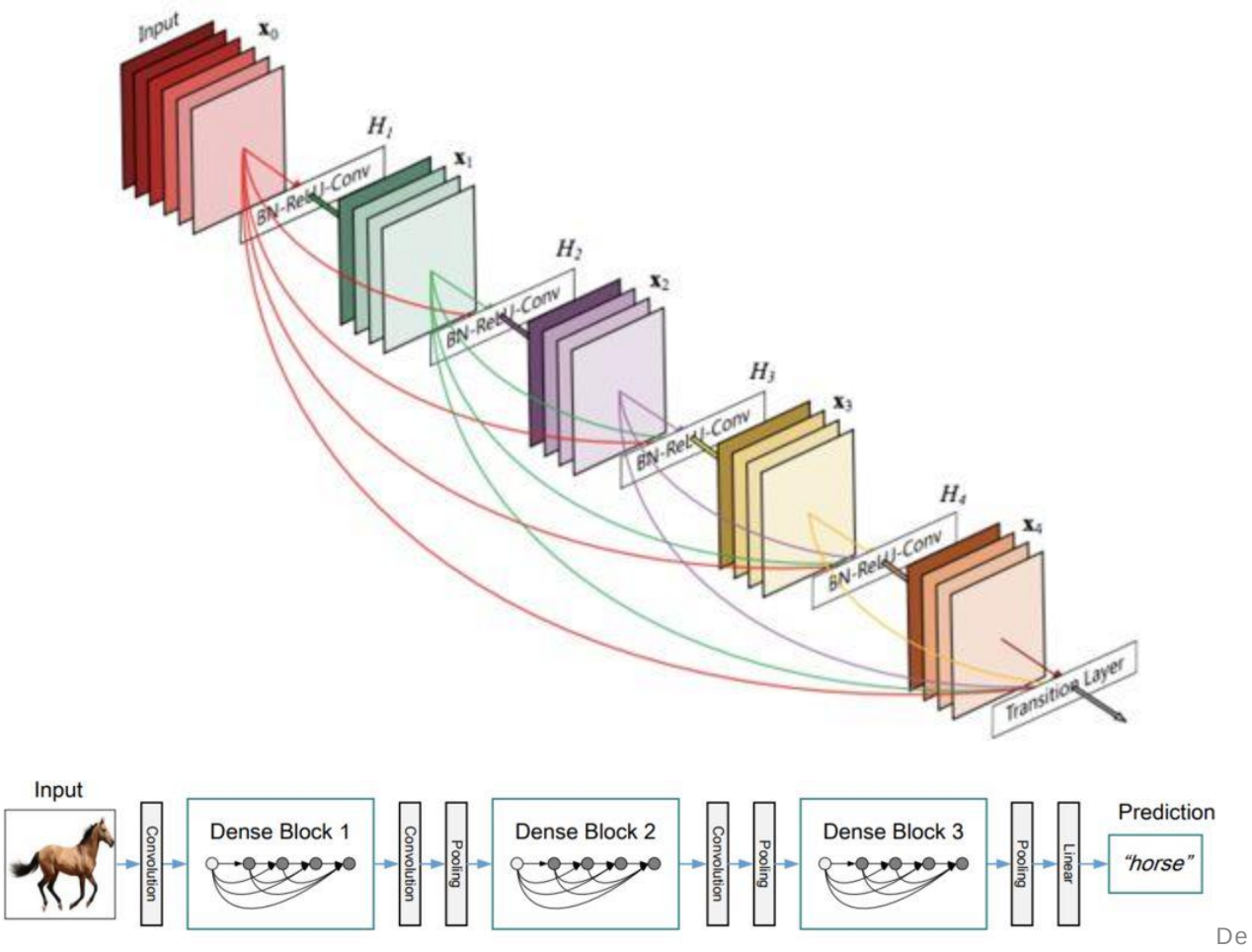
AlexNet 模型结构示意图。值得注意的是，这种结构由两个分支网络构成，分别在两个不同的 GPU 上并行训练。



GoogleNet 模型结构示意图。该模型由多个 Inception 模块构成。



ResNet 模型结构示意图。该模型由多个残差模块构成。



De

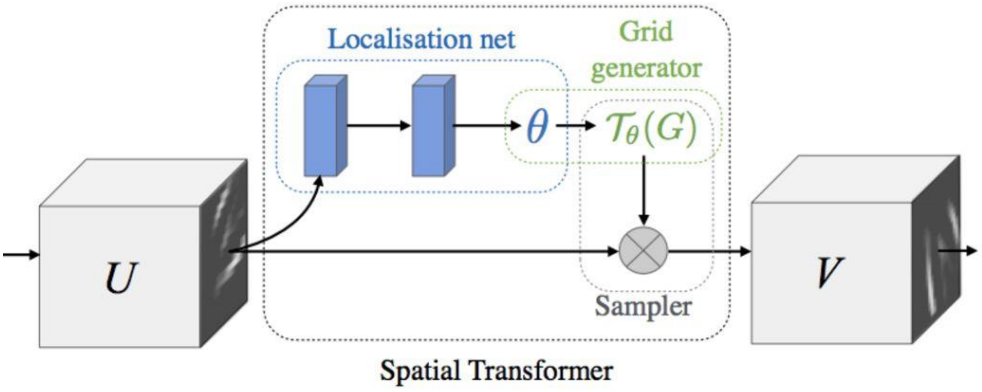
nseNet 模型结构示意图。该模型由多个密集模块堆叠而成。

卷积神经网络的不变形

使用卷积神经网络的一大挑战是需要非常大的数据集来训练并学习模型的所有基本参数。但即便是当前大规模的数据集，如 ImageNet 拥有超过一百万张图像数据的数据集，仍然无法满足深层卷积结构训练的需要。通常，在模型训练前，我们会通过数据增强操作来处理数据集：即通过随机翻转、旋转等操作来改变图像，从而增加数据样本的数量。

这些数据增强操作的主要优点是使得网络对于各种图像转换更加鲁棒，这项技术也是 AlexNet 取得成功的主要原因之一。因此，除了上述改变网络架构以简化训练的方法之外，其他的研究工作旨在引入新颖的模块结构来更好的训练模型。处理不变性最大化的一种优秀结构是空间变换网络（STN）。具体的说，这种网络结构使用了一个新颖的学习模块，增加了模型对不重要空间变换的不变性，例如， 在物体识别过程中

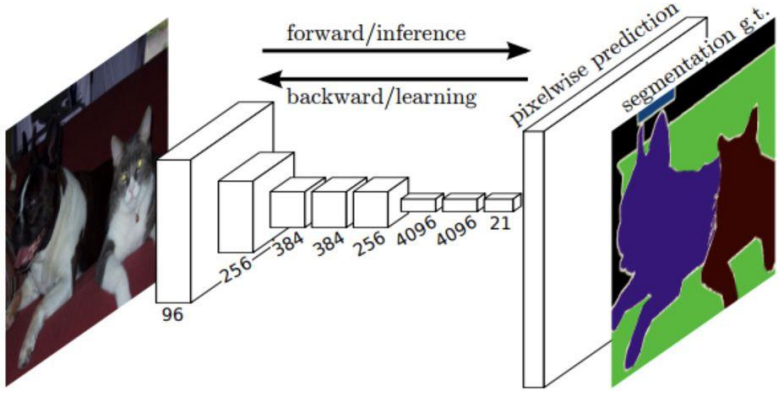
那些由不同视点引起的变换。该模型结构由三个子模块组成：一个定位模块，一个网格生成模块和一个采样模块，如下图所示。



空间变换网络结构示意图

卷积神经网络中的目标定位问题

除了简单的目标识别分类任务，近年来卷积结构在目标精准定位的任务中同样表现出色，如目标检测、语义分割任务等。全卷积网络（FCN）是其中最成功的卷积结构之一，主要用于图像语义分割。顾名思义，FCN 并未使用全连接层，而是将它们转换为卷积层，其感受野范围覆盖整个卷积层的底层特征图。更重要的是，网络通过学习一个上采样或者去卷积滤波器，可以恢复最后一层图像的全分辨率，其结构示意图如下图所示。



全卷积网络结构示意图。经过上采样操作，在模型最后一层得到全分辨率的特征图，适用 softmax 将每个像素分类，并生成最终的分割结果。

在 FCN 中，语义分割问题被转化成一个密集的逐像素分类问题，通过投射来实现。换句话说，每个像素都与 softmax 层关联，通过像素逐类分组来实现图像的语义分

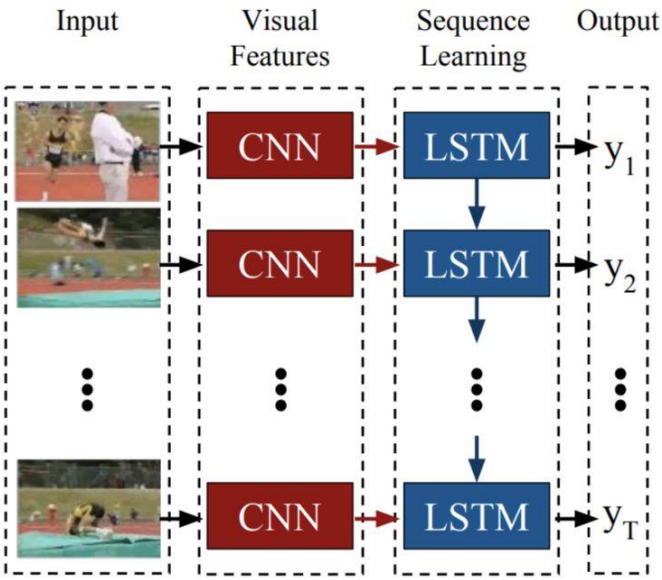
割。更值得注意的是，在这项工作中对较低结构层的特征适用上采样操作，起着至关重要作用。由于较低层特征更倾向于捕捉更精细化的细节，因此上采样操作允许模型进行更精确的分割。此外，反卷积滤波器的一种替代方案是使用扩张卷积，即上采样稀疏滤波器，这有助于在保持参数数量的同时，模型能够学习到更高分辨率的特征图。

R-CNN 是最早用于目标检测任务的卷积结构，这是一种带区域建议的卷积神经网络（RPN），在最初的目标检测任务中取得了最先进的检测结果，特别是使用区域建议的选择性搜索算法来检测可能包含目标的潜在区域，并将这些建议区域做一些变换以便匹配卷积结构的输入大小，经卷积神经网络中特征提取后，最终送入 SVM 中进行分类，并通过非极大值抑制后处理步骤中优化模型的表现。

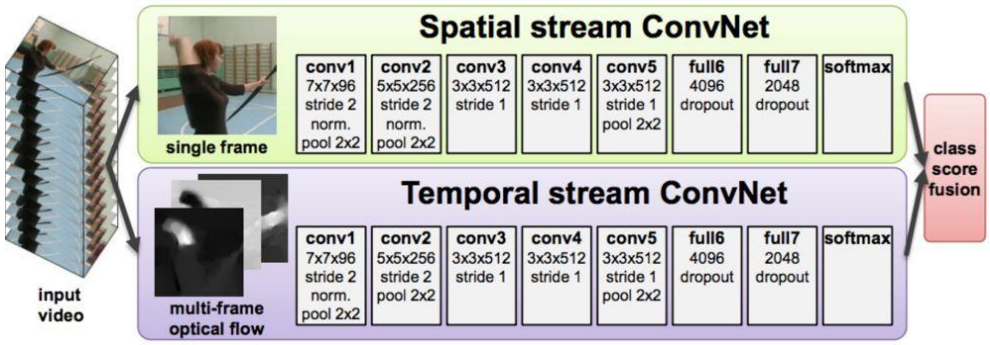
随后，Fast R-CNN,Faster R-CNN,Mask R-CNN 等目标检测模型的提出都是基于最初的 R-CNN 结构。可以说，卷积神经网络在目标检测方面的应用是围绕 R-CNN 结构展开。

时域卷积神经网络

如上所述，卷积神经网络在计算机视觉二维空间的应用中所取得的显着性能，引发了人们对 3D 时空应用的研究。许多文献中提出的时域卷积结构通常只是试图从空间域（x，y）扩展到时间域（x，y，t）的二维卷积结构。而时域神经网络结构有三种不同的形式：基于 LSTM 的时域卷积网络、3D 卷积神经网络和双流卷积神经网络，其模型结构示意图如下图。



基于 LSTM 的时域卷积神经网络。该模型中，由视频流的每帧数据构成模型的输入。



双流卷积神经网络。该模型以 RGB 光流数据作为输入。

总结

相比于手动设计的特征或浅层的特征表示，多层卷积结构是当前计算机视觉领域最先进、最具吸引力的结构之一。总体而言，大多数模型结构都是基于四个共同的构件块，即卷积、非线性单元、归一化和池化操作。虽然这些优秀的卷积模型在大多数计算机视觉任务中取得了最优性能，但它们共同的缺点仍然是对卷积内部操作、特征表征的理解相当有限，依赖于大规模的数据集和模型训练过程，缺乏精确的性能界限和超参数选择的清晰度。这些超参数包括滤波器的大小、非线性函数、池化操作参数以及模型层数的选择。接下来我们将进一步讨论卷积神经网络设计过程中这些超参数的选择。

第三章

理解卷积神经网络的构建模块

考虑到卷积神经网络领域还存在大量未解决的问题，在本章我们将探讨一些典型案例中卷积网络的每一层处理操作的作用及意义，尤其我们将从理论和生物学角度给出合理解释。

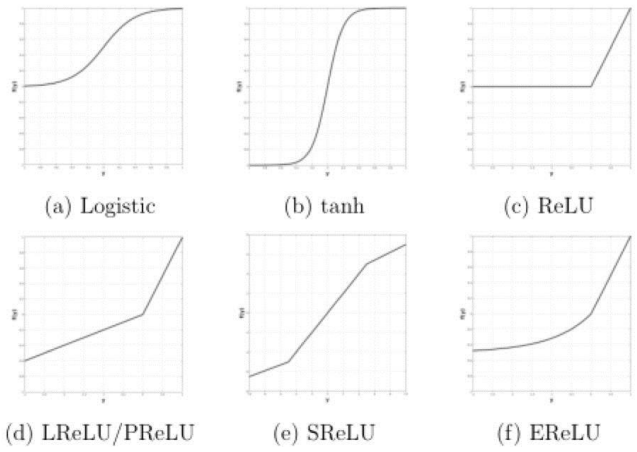
卷积层

卷积神经网络的核心层是卷积层，这是模型最重要的一步。总的来说，卷积是一种线性的、具有平移不变性的运算，它是通过局部加权输入信号来实现的。权重集合是根据点扩散函数（point spread function）来确定的，不同的权重函数能够反映出输入信号的不同性质。

在频率域中，与点扩散函数相关联的是调制函数，这表明了输入的频率组分可以通过缩放和相移来进行调制。因此，选择合适的卷积核，将有助于模型获取输入信号中最显著、最重要的特征信息。

非线性单元

多层神经网络通常是高度的非线性模型，而修正单元（rectification）通常将引入一个非线性函数（也被称为激活函数），即将非线性函数应用到卷积层输出中。引入修正单元的目的，一方面是为了最好的、最合适的模型解释；另一方面是为了让模型能更快和更好地学习。常用的非线性函数主要包括 Logistic 函数、tanh 函数、Sigmoid 函数、ReLU 及其变体 LReLU,SReLU,EReLU 等，其函数图像如下图所示。



多层网络结构中的非线性激活函数

归一化

如上所述，由于这些多层网络中存在级联的非线性运算，因此多层神经网络都是高度的非线性模型。除了上面讨论的修正非线性单元外，归一化（normalization）同样是卷积神经网络结构中重要的非线性处理模块。最广泛使用的归一化形式即所谓的局部响应归一化操作（LRN,Local Response Normalization）。此外，还有诸如批归一化(batch normalization)，分裂归一化（divisive normalization）等。

池化操作

几乎所有的卷积神经网络，都包含池化操作。池化操作是为了提取特征在不同位置和规模上的变化，同时聚合不同特征映射的响应。正如卷积结构中前三个组份，池化操作也是受到生物学启发和理论支持而提出的。 平均池化和最大池化是两个最广泛使用的池化操作，其池化效果依次如下图所示。

经平均池化操作后 Gabor 特征的变化情况

经最大池化操作后 Gabor 特征的变化情况

第四章

当前研究状态

对卷积神经网络结构中各组作用的阐述凸显了卷积模块的重要性，这个模块主要用于捕获最抽象的特征信息。相对而言，我们对卷积模块操作的理解却很模糊，对其中繁琐的计算过程的理解并不透彻。本章我们将尝试理解卷积网络中不同层所学习的内容及不同的可视化方法。同时，我们还将重点展望这些方面仍待解决的问题。

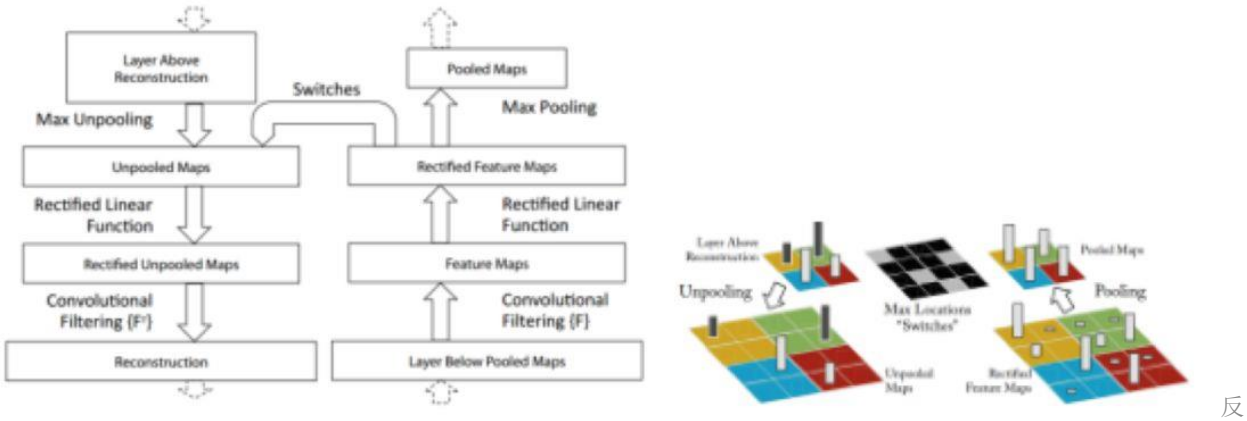
当前趋势

尽管各种优秀的卷积模型在多种计算机视觉应用中取得了最优表现，但在理解这些模型结构的工作方式及探索这些结构的有效性方面的研究进展仍相当缓慢。如今，这个问题已经引起了众多研究者的兴趣，为此很多研究提出用于理解卷积结构的方法。

总的来说，这些方法可以分成三个方向：对所学习到的过滤器和提取的特征图进行可视化分析、受生物视觉皮层理解方法所启发的消融学习（ablation study）、以及通过引入主成分分析法设计并分析网络最小化学习过程，我们将简要概述这三种方法。

卷积的可视化分析

卷积可视化的第一种方法是以数据集为中心的方法，因为卷积操作依靠从数据集输入来探测网络在网络中找到最大响应单元。这种方法的第一个应用是反卷积（DeConvNet）。其中可视化是分两步实现：首先，一个卷积结构接收来自数据集 a 的几个图像并记录数据集中输入的特征映射最大响应；其次，这些特征地图使用反卷积结构，通过反转卷积操作模块，将卷积操作中学习到的滤波器特征进行转置来执行“解卷积”操作，从而实现卷积的可视化分析。反卷积操作的示意图如下图所示：



反卷积构建模块

卷积可视化的第二种方法称为以网络为中心的方法，因为它仅使用网络参数而不需要任何用于可视化的数据。这种方法首次应用于深层置信网络的可视化分析中，后来才应用于卷积网络结构中。具体地说，这种卷积可视化是通过合成图像来实现的，该图像将最大化某些神经元（或过滤器）的响应。

卷积的消融学习

另一种流行的可视化方法是使用所谓的网络消融研究。实际上，许多著名的卷积结构都包括模型消融研究实验部分，其目的是隔离卷积结构的不同部分组成网络，来查看删除或添加某些模块如何模拟整体的性能。消融研究能够指导研究者设计出性能更优的网络结构。

卷积结构的控制设计

理解卷积结构的另一种方法是在网络设计时添加先验知识，从而最大限度地减少所需学习的模型参数。例如，一些方法是减少每层卷积层所需学习的过滤器数量，并用转换后的版本在每一层中学习的滤波器来模拟旋转不变性。 其他方法依赖于用基础集合代替过滤器的学习过程，而不是学习过滤器参数，它们的目标是学习如何组合基础集合，以便在每一层形成有效的过滤器。此外，还有一些方法，是通过完全手工设计卷积网络，并针对特定的任务在网络设计阶段加入特定的先验知识，如此设计出可解释的网络。

待解决问题

通过上述内容，我们总结了卷积模型一些关键技术以及如何更好地理解卷积结构的方法。下面，我们将进一步讨论在卷积模型领域仍待解决的一些问题。

基于卷积可视化的研究方法仍待解决的几个关键问题：

- 首先，开发更加客观的可视化评价方法是非常重要的，可以通过引入评价指标来评估所生成的可视化图像质量或含义来实现。
- 此外，尽管看起来以网络为中心的卷积可视化方法更有前景（因为它们在生成可视化结果过程中不依赖模型结构自身），但也缺乏一套标准化的评估流程。一种可能的解决方案是使用一个评估基准来评价同样条件下生成的网络可视化结果。这样的标准化方法反过来也能实现基于指标的评估方法，而不是当前的解释性分析。
- 另一个可视化分析的发展方向是同时可视化网络的多个单元，以更好地理解模型中特征表征的分布情况，甚至还能遵循一种控制方法。

基于 ablation study 的研究方法仍待解决的几个关键问题：

- 使用共同的、系统性组织的数据集。我们不仅要解决计算机视觉领域常见的不同挑战（比如视角和光照变化），还必须要应对复杂度更大的类别问题（如图像纹理、部件和目标复杂度等）。近年来，已经出现了一些这样的数据集。在这样的数据集上，使用 ablation study，辅以混淆矩阵分析，可以确定卷积结构中出错的模块，以便实现更好的理解卷积。

- 此外，分析多个协同的 ablation 对模型表现的影响方式，是一个很受关注的研究方向。这样的研究也能有助于我们理解独立单元的工作方式。

相比于完全基于学习的方法，还有一些受控方法能让我们对这些结构的运算和表征有更深入的理解，因而具有很大的研究前景。这些有趣的研究方向包括：

- 逐层固定网络参数及分析对网络行为的影响。例如，基于当前特定任务的先验知识，一次固定一层的卷积核参数，以分析每一卷积层中卷积核的适用性。这个逐层渐进式的学习方式有助于揭示卷积学习的作用，还可用作最小化训练时间的初始化方法。

- 类似地，可以通过分析输入的特征来研究网络结构的设计（如层的数量或每层中过滤器数量的选择方案），这种方法有助于设计出最适合模型结构。

- 最后，将受控方法应用于网络的同时，可以对卷积神经网络的其它方面的作用进行系统性的研究。通常，我们重点关注的是模型所学习的参数，所以对这方面得到的关注较少。例如，我们可以在固定大多数参数的情况下，研究各种池化策略和残差连接的作用。

作者：Isma Hadji,Richard P. Wildes
论文链接：
https://arxiv.org/pdf/1803.08834.pdf

近期热文

- 干货 | NLP 中的 self-attention【自-注意力】机制
- 这 5 小段代码轻松实现数据可视化（Python+Matplotlib）
- 详解计算机视觉五大技术：图像分类、对象检测、目标跟踪、语义分割和实例分割
- 值得收藏的 27 个机器学习的小抄
- 推荐 | 机器学习中的这 12 条经验，希望对你有所帮助
- 图解机器学习的常见算法

- 利用 Python 实现卷积神经网络的可视化
- 干货丨浅谈强化学习的方法及学习路线

广告、商业合作

更多精彩
请关注《机器学习算法与 python 学习》公众号

请添加微信：guodongwe1991
（备注：商务合作）