深度卷积神经网络最新进展综述

摘要：深度卷积神经网络（**CNN**）是一种特殊类型的神经网络，在各种竞赛基准上表现出了当前最优结果。深度 **CNN** 的超强学习能力主要是通过使用多个非线性特征提取阶段实现的，这些阶段能够从数据中自动学习分层表征。大量可用的数据和硬件处理单元的改进加速了 **CNN** 的研究，最近也报道了非常有趣的深度 **CNN** 架构。近来，深度 **CNN** 架构研究比赛表明，创新的架构理念以及参数优化可以提高 **CNN** 在各种视觉相关任务上的性能。鉴于此，关于 **CNN** 设计的不同想法被提出，如使用不同的激活函数和损失函数、参数优化、正则化以及处理单元的重构。然而，在表征能力方面的主要改进是通过重构处理单元来实现的。尤其是，使用块而不是层作为结构单元的想法获得了极大关注。因此，本综述着重于最近报道的深度**CNN**架构中存在的内在分类法，将**CNN**架构中的最新创新分为七个不同类别。这七个类别分别基于空间利用、深度、多路径、宽度、特征图利用、通道提升和注意力。此外，本文还涵盖了对 **CNN** 组成部分的基本理解，并揭示了 **CNN** 目前面临的挑战及其应用。

关键词：深度学习，卷积神经网络，结构，表征能力，残差学习，**CNN**通道提升

1. 引言

机器学习（**ML**）算法属于人工智能（**AI**）的一个特殊领域，该领域无需明确编程，通过学习数据之间的潜在关系并做出决策，从而将智能赋予计算机。自**1990**年代末以来，已经开发出了不同的**ML**算法来模拟人类的感官反应言，如言语和视觉等，但是它们通常无法达到人类水准的满意度[1]-[6]。 机器视觉（**MV**）任务具有挑战性促使产生了一类特殊的神经网络（**NN**），即卷积神经网络（**CNN**）[7]。

CNN被认为是学习图像内容的最佳技术之一，并且在图像识别、分割、检测和检索相关任务方面显示了最佳的成果[8]，[9]。 **CNN**的成功所引起的关注已超出学术界。在行业中，诸如**Google**，**Microsoft**，**AT＆T**，**NEC**和**Facebook**之类的公司已经建立了活跃的研究小组，以探索**CNN**的新架构[10]。目前，大多数图像处理竞赛的领跑者都采用基于深度**CNN**的模型。

CNN拓扑分为多个学习阶段，由卷积层、非线性处理单元和下采样层的组合组成[11]。每层使用一组卷积核（过滤器）[12]执行多次转换。卷积运算通过将图像分成小片（类似于人眼的视网膜）来提取局部相关的特征，从而使其能够学习合适的特征。卷积核的输出被分配给非线性处理单元，这不仅有助于学习抽象表示，而且还将非线性嵌入到特征空间中。这种非线性为不同的响应生成了不同的激活模式，因此有助于学习图像中的语义差异。非线性函数的输出通常经过下采样，这有助于总结结果，并使输入对于几何变形不变[12]，[13]。

CNN的结构设计灵感来自Hubel和Wiesel的工作，因此在很大程度上遵循了灵长类动物视觉皮层的基本结构[14]，[15]。CNN最早在1989年通过LeCuN的工作而备受关注，该技术用于处理网格状拓扑数据（图像和时间序列数据）[7]，[16]。CNN的普及很大程度上是由于其分层特征提取能力。CNN的分层组织模拟人脑中新皮质深层和分层学习过程，该过程会自动从基础数据中提取特征[17]。CNN的学习过程分期与灵长类动物的视觉皮层腹侧通路（V1-V2-V4-IT/VTC）非常相似[18]。灵长类动物的视觉皮层首先从视网膜位区域接收输入，在该区域通过外侧膝状核执行多尺度高通滤波和对比度归一化。此后，通过分类为V1，V2，V3和V4的视觉皮层的不同区域执行检测。实际上，视觉皮层的V1和V2部分类似于卷积层和下采样层，而颞下区类似于CNN的较高层，这可以推断图像[19]。在训练期间，CNN通过反向传播算法根据输入调整权重变化来学习。CNN使用反向传播算法最小化损失函数类似于基于响应的人脑学习。CNN具有提取低、中和高层特征的能力。高级特征（更抽象特征）是低级和中级特征的组合。借助自动特征提取功能，CNN减少了合成单独的特征提取器的需要[20]。因此，CNN可以通过少量处理从原始像素中学习良好的内部表示。

人们发现，通过增加CNN的深度可以增强CNN的表达能力，随后到来的是使用CNN进行图像分类和分割的热潮[21]。当处理复杂的学习问题时，深层架构比浅层架构具有优势。以分层的方式堆叠多个线性和非线性处理单元，可为深度网络提供学习不同抽象级别上复杂表示的能力。此外，硬件的进步及其带来的高计算资源也是深度CNN近期成功的主要原因之一。较深的CNN架构显示出比基于浅层和传统视觉模型性能的显着进步。除了在监督学习中使用外，深层CNN还有从大量未标记的数据中学习有用表示的潜力。CNN使用多个映射功能使它能够改进不变表示的提取，因此使其能够处理数百个类别的识别任务。最近，研究表明，可以利用转移学习（TL）的概念将不同层特征（包括低级和高级）转移到通用识别任务中[22]–[24]。CNN的重要属性是分层学习，自动特征提取，多任务处理和权重共享[25]–[27]。

CNN学习策略和结构经过了多种改进，以使CNN可扩展到大而复杂的问题。这些创新可以归类为参数优化，正则化，结构重构等。但是，据观察，在AlexNet在ImageNet数据集上获得优异性能之后，基于CNN的程序变得更为流行[21]。因此，CNN的重大创新主要在2012年以来提出，主要包括处理单元的重组和新区块的设计。类似地，Zeiler和Fergus[28]引入了特征的逐层可视化的概念，这将趋势转向了在诸如VGG[29]的深层结构中以低空间分辨率提取特征的趋势。如今，大多数新架构都是基于VGG引入的简单同质拓扑原理构建的。另一方面，Google小组提出了一个有趣的想法，即分割、变换和合并，并且相应的块称为inception块。inception块首次给出了在层内进行分支的概念，该概念允许在不同的空间尺度上提取特征[30]。2015年，ResNet[31]引入的用于深层CNN训练的跳跃连接概念广为人知，随后，此概念被大多数后续的Nets使用，例如Inception-ResNet，WideResNet，ResNext等[32]–[34]。

为了提高CNN的学习能力，不同的结构设计，例如WideResNet，Pyramidal Net，Xception等，从附加基数和增加宽度的角度探讨了多尺度转换的效果[32]，34]，[35]。 因此，研究重点从参数优化和连接重新调整转向改进网络的架构设计（层结构）。这种转变带来了许多新的体系结构思想，例如通道提升，空间和通道智能开发以及基于注意力的信息处理等[36]–[38]。

在过去的几年中，研究人员对深层CNN进行了各种有趣的研究，详细阐述了CNN的基本组成部分及其替代方案。[39]的综述回顾了2012-2015年的著名架构及其组成部分。同样，在文献中，有一些著名的综述讨论了CNN的不同算法，并专注于CNN的应用[20]，[26]，[27]，[40]，[41]。同样，[42]中的综述讨论了基于加速技术的CNN分类。另一方面，在本综述中，我们讨论了近期和著名的CNN体系结构的内在分类。本综述中讨论的各种CNN架构可以大致分为以下七个主要类别：空间利用，深度，多路径，宽度，特征图利用，通道提升和基于注意的CNN。本文的其余部分按以下顺序组织（如图1所示）：第1节总结了CNN的基础知识，其与灵长类动物的视觉皮层的相似性以及对机器视觉的贡献。第2节概述了基本CNN组件，第3节讨论了深度CNN的体系结构演变。第4节讨论了CNN结构的最新创新，并将CNN分为七个大类。第5节和第6节阐明了CNN的应用和当前的挑战，而第7节讨论了未来的工作，最后一节得出了结论。

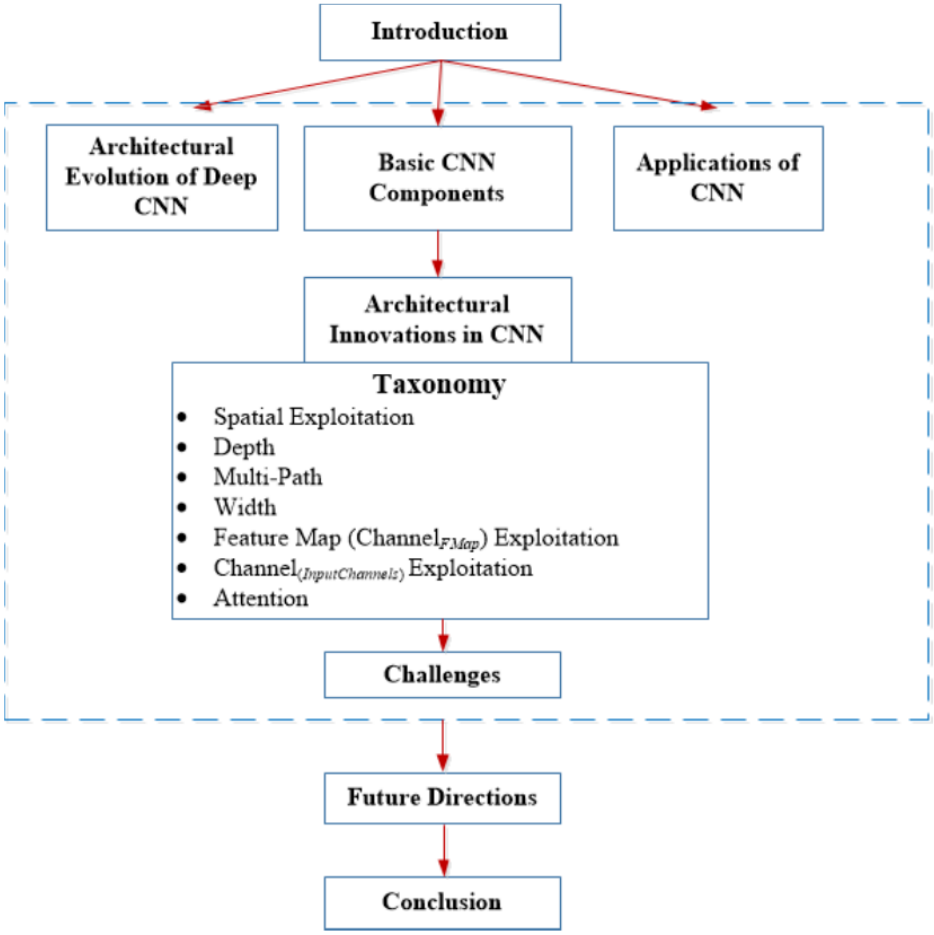

图1 本综述结构

2、CNN基本组件

如今，CNN被认为是使用最广泛的ML技术，尤其是在视觉相关应用中。CNN最近在各种ML应用中显示了最佳的结果。ML系统的典型框图如图2所示。由于CNN既具有良好的特征提取能力，又具有较强的辨别能力，因此在ML系统中，它主要用于特征提取和分类。

典型的CNN体系结构通常包括卷积和池化层的交替，最后是一个或多个全连接层。在某些情况下，全连接层替换为全局平均池化层。除了学习的各个阶段外，还结合了不同的正则化单元，例如批次归一化和dropout，以优化CNN性能[43]。CNN组件的排列在设计新体系结构和获得增强性能方面起着基本作用。本节简要讨论了这些组件在CNN体系结构中的作用。

2.1 卷积层

卷积层由一组卷积核（每个神经元充当核）组成。这些核与图像的一小部分区域相关，称为感受野。它通过将图像划分成小块（感受野）并将其与一组特定的权重（滤波器的元素与相应的感受野元素相乘）进行卷积来工作[43]。 卷积运算可以表示如下：

$$F_l^k = (I_{x,y} * K_l^k) \tag{1}$$

其中，输入图像用$I_{x,y}$表示，x,y表示具体位置，K_l^k表示第k层的第l个卷积核。将图像分成小块有助于提取局部相关的像素值。这种局部汇总的信息也称为特征图。通过使用相同的权重卷积核在整个图像上滑动来提取图像中的不同特征集。与全连接网络相比，卷积运算的这种权重共享功能使CNN参数更有效。根据滤波器的类型和大小，填充的类型以及卷积的方向，可以进一步将卷积操作分为不同的类型[44]。另外，如果核是对称的，则卷积运算将变为相关运算[16]。

2.2 池化层

作为卷积运算输出而产生的特征图可能出现在图像的不同位置。一旦提取特征后，只要保留相对于其他特征的近似位置，其精确位置就不再重要。像卷积一样进行池化或下采样是一个有趣的局部操作。它汇总了感受野附近的相似信息，并在该局部区域内输出主要响应[45]。

$$Z_l = f_p(F_{x,y}^l) \tag{2}$$

公式（2）表示池化操作，其中Z_l表示第l个输出特征图，$F_{x,y}^l$表示第l个输入特征图，而$\mathrm{f_p}$（.）定义了池化操作的类型。 合并操作的使用有助于提取特征的组合，这些特征对于平移和轻微变形是不变的[13]，[46]。 将特征图的大小减小到不变的特征集不仅可以调节网络的复杂性，而且可以通过减少过度拟合来帮助提高通用性。CNN中使用了不同类型的池化公式，例如最大值，平均值，L2，重叠，空间金字塔合并等[47]-[49]。

2.3 激活函数

激活功能起决策功能，有助于学习复杂的模式。选择适当的激活功能可以加快学习过程。等式（3）定义了卷积特征图的激活函数。

$$T_l^k = f_A(F_l^k) \tag{3}$$

在上式中，F_l^k是卷积运算的输出，分配给激活函数；$f_A(.)$会添加非线性并返回第k层的转换输出T_l^k。在文献中，不同的激活函数,例如sigmoid，tanh，maxout，ReLU和ReLU的变体,例如leaky ReLU，ELU和PReLU[39]，[48]，[50]，[51]用于引入特征的非线性组合。 然而，ReLU及其变体优于其他激活函数，因为它有助于克服梯度消失问题[52]，[53]。

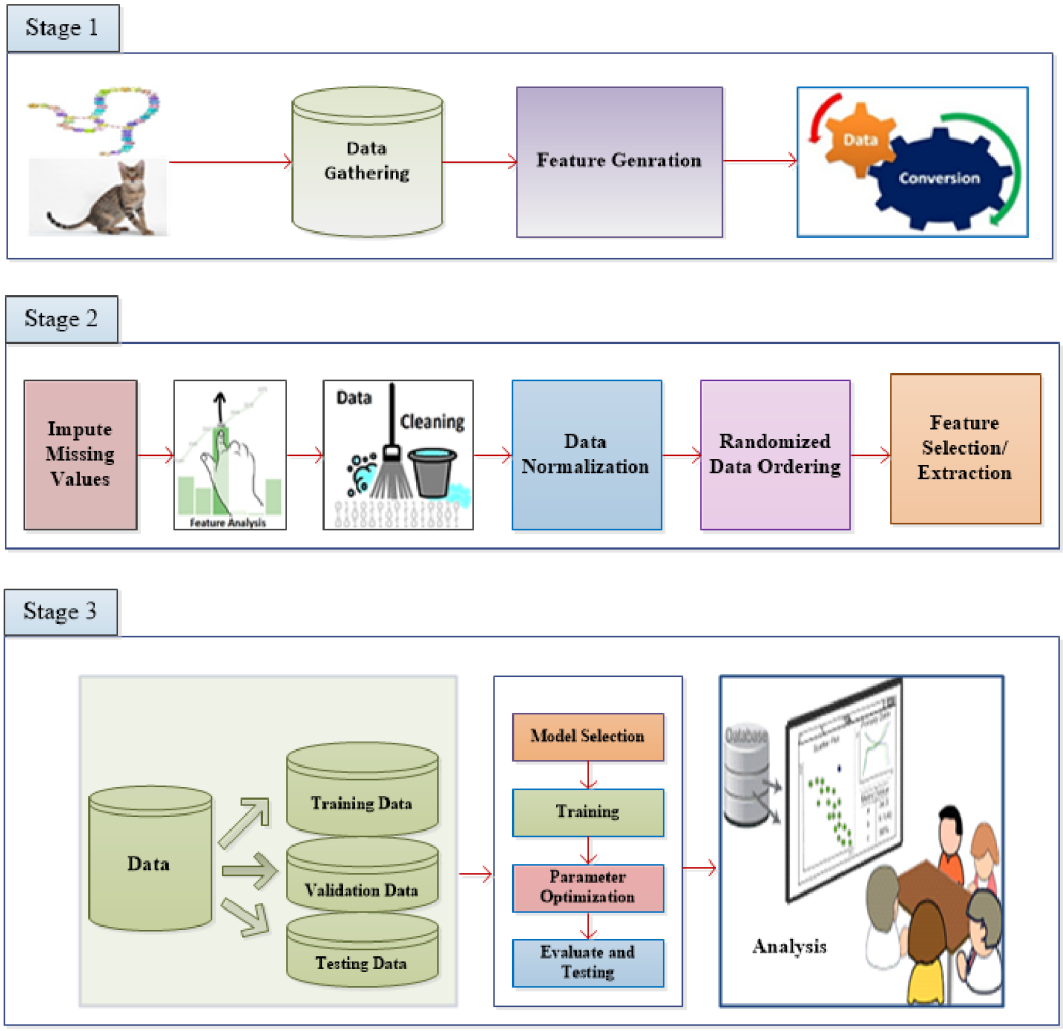

图2 典型ML系统的基本布局。 在与ML相关的任务中，首先对数据进行预处理，然后将其分配给分类系统。典型的ML问题包括以下三个步骤：阶段1与数据收集和生成相关，阶段2执行预处理和特征选择，而阶段3基于模型选择，参数调整和分析。CNN具有良好的特征提取能力和较强的判别能力，因此在ML系统中，它可以用于特征提取和分类。

2.4 批次归一化

批次归一化用于解决与特征图中内部协方差平移有关的问题。内部协方差偏移量随隐藏单位值分布变化，这会降低收敛速度（通过将学习率强制为小值），并对参数初始化要求高。等式（4）中示出了对变换后的特征图T_l^k的批次归一化。

$$N_l^k = \frac{F_l^k - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \tag{4}$$

在等式（4）中，N_l^k表示归一化特征图，F_l^k是输入特征图，μ_B和σ_B^2分别表示小批次特征图的均值和方差。批次归一化通过将特征图值设为零均值和单位方差来统一其分布[54]。此外，它可以平滑梯度流并充当调节因素，从而有助于改善网络的泛化。

2.5 Dropout

Dropout引入了网络内的正则化，最终通过以一定概率随机跳过某些单元或连接来最终提高泛化性。在神经网络中，有时学习某个非线性关系的多个连接会相互适应，这会导致过拟合[55]。某些连接或单元的这种随机丢弃会产生几种稀疏的网络体系结构，最后选择一个权重较小的代表性网络。 然后，将这种选择的架构视为所有提议网络的近似[56]。

2.6 全连接层

全连接层通常在网络末端用于分类任务。与池化和卷积不同，它是全局操作。它从前一层获取输入，并全局分析所有前一层的输出[57]。这将选定特征进行非线性组合，用于数据分类 [58]。

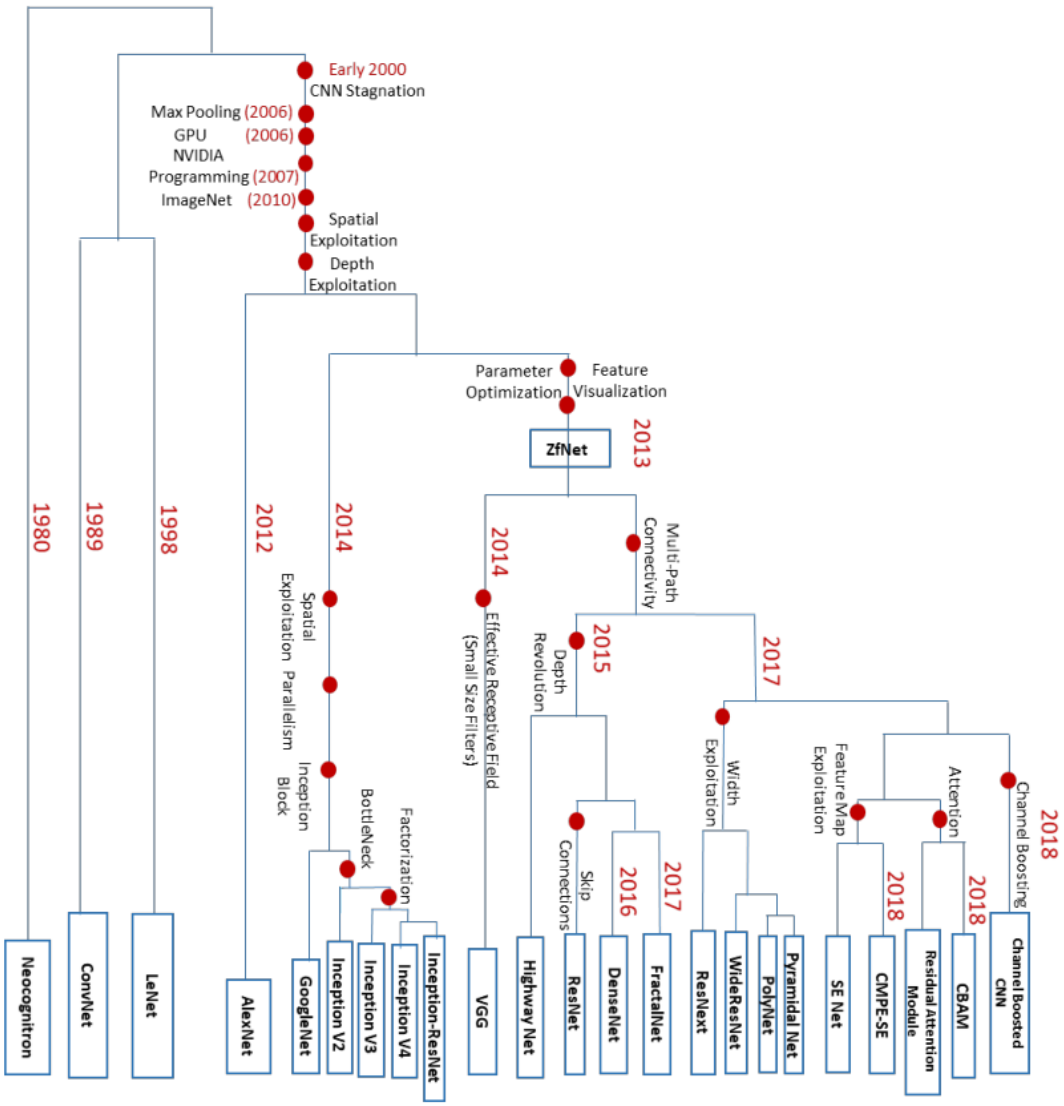

图3 深度CNN演化史

3、深度CNN结构演化史

如今，CNN被认为是受到生物学启发的AI技术中使用最广泛的算法。 CNN的历史始于Hubel和Wiesel（1959，1962）进行的神经生物学实验[14]，[59]。 他们的工作为许多认知模型提供了平台，后来几乎所有这些模型都被CNN取代。 几十年来，人们为提高CNN的性能做出了不同的努力。 图3中用图形表示了这一历史。这些改进可以分为五个不同的时代，下面将进行讨论。

3.1 1980年代末至1999年：CNN的起源

自1980年代后期以来，CNN已应用于视觉任务。 1989年，LeCuN等人提出了第一个名为ConvNet的多层CNN，其起源于福岛的Neocognitron [60]，[61]。 LeCuN提出了ConvNet的监督训练，与它的前身Neocognitron所采用的无监督强化学习方案相比，使用了反向传播算法[7]，[62]。因此，LeCuN的工作为现代2D CNN奠定了基础。监督训练 使CNN具有从原始输入

中自动学习特征的能力，而无需设计传统ML方法使用的手工特征。这个ConvNet成功解决了手写数字和邮政编码识别相关问题[63]。 1998年， LeCuN改进了ConvNet，并用于文档识别程序中的字符分类[64]。修改后的架构被命名为LeNet-5，它是对最初的CNN的改进，因为它可以从原始像素中以分层的方式提取特征表示[65]。LeNet-5依赖更少参数，同时考虑了图像空间拓扑，使CNN能够识别图像的旋转变体[65]。由于CNN在光学字符识别方面的良好性能，其分别于1993年和1996年开始在ATM和银行中商业化使用。尽管LeNet-5实现了许多里程碑式成功，但是与之相关的主要问题是，其识别能力并未扩展到除手识别之外的其他分类任务。

3.2 2000年初：CNN停滞不前

在1990年代末和2000年代初，人们对NN的兴趣减少，很少探索CNN在不同应用中的作用，例如物体检测，视频监视等。由于在性能上的微不足道的改进（以高计算时间为代价），CNN在ML相关任务中变得沉寂。当时，其他统计方法，尤其是SVM比CNN更为流行[66]-[68]，由于其相对较高的性能。在2000年初，普遍认为用于CNN训练的反向传播算法无法有效收敛到最佳点，因此与手工制作的特征相比，无法以监督的方式学习有用的特征[69]。同时，任有研究人员继续研究CNN，试图优化其性能。Simard等人在2003年改进了CNN架构，并在手写体基准数据集MNIST[64]，[68]上显示了与SVM相比更好的结果[70]-[72]。通过将其在光学字符识别（OCR）中的应用扩展到其他的字符识别[72]-[74]，如部署在视频会议中用于面部检测的图像传感器中以及对街头犯罪的管制等，这种性能的改进加快CNN的研究速度。同样，基于CNN的系统已在超市跟踪客户[75]-[77]方面实现了工业化。此外，研究人员还探索了CNN在医学图像分割、异常检测和机器人视觉等其他应用中的潜力[78]-[80]。

3.3 2006-2011年：CNN的复兴

深度神经网络通常具有复杂的体系结构和时间密集型训练阶段，有时跨越数周甚至数月。在2000年初，只有很少的技术可以训练深度网络。此外，人们认为CNN无法解决复杂的问题。这些挑战使人们停止在ML相关任务中使用CNN。

为了解决这些问题，2006年出现了许多有趣的方法来克服在训练深度CNN和学习不变特征时遇到的困难。Hinton在2006年针对深度架构提出了贪婪逐层预训练方法，从而复兴并恢复了深度学习的重要性[81]，[82]。深度学习的复兴[83] [84]是使深度CNN成为众人关注的因素之一。 黄等（2006）使用最大池化而不是下采样，通过学习不变特征显示出良好的结果[46]，[85]。

在2006年末，研究人员开始使用图形处理单元（GPU）[86]，[87]来加速深度NN和CNN体系结构的训练[88]，[89]。NVIDIA在2007年推出了CUDA编程平台[90]，[91]，该平台可以更大程度地利用GPU的并行处理功能[92]。从本质上讲，使用GPU进行神经网络训练[88]，[93]和其他硬件改进是CNN研究得以复兴的主要因素。2010年，李飞飞在斯坦福大学的小组建立了一个名为ImageNet的大型图像数据库，其中包含数百万个带有标签的图像[94]。基于该数据库举办年度ImageNet大规模视觉识别挑战赛（ILSVRC），对各种模型的性能进行了评估和评分[95]。ILSVRC和NIPS在加强研究和增加CNN的使用方面一直非常活跃，从而使其流行起来。这是改善CNN性能和增加其使用的转折点。

3.4 2012-2014年：CNN的崛起

可获得的大规模训练数据，硬件改进和计算资源有助于CNN算法的改进。在此期间，可以发现CNN在目标检测，图像分类和分割相关任务中的复兴[9]，[96]。然而，CNN在图像分

类任务中的成功不仅归因于上述因素的结果，而且很大程度上归功于架构修改、参数优化、管理单元的合并以及网络内连接的重新制定和调整[39]，[42]，[97]。

CNN性能的主要突破体现在AlexNet[21]。AlexNet赢得了2012-ILSVRC竞赛，这是图像检测和分类中最困难的挑战之一。 AlexNet通过利用深度（合并了多个转换层级）提高了性能，并在CNN中引入了正则化。与2012-ILSVRC中传统ML技术相比，AlexNet[21]的表现堪称典范（AlexNet将错误率从25.8降低至16.4），这表明2006年前CNN性能饱和的主要原因是由于没有足够的训练数据和计算资源。综上所述，在2006年之前，这些资源不足使得在不降低性能的情况下很难训练高容量的CNN [98]。

随着CNN在计算机视觉（CV）领域中越来越普遍，人们进行了许多尝试以降低计算成本来提高CNN的性能。因此，每种新架构都试图克服先前提出的架构与新结构重新组合的缺点。在2013年和2014年，研究人员主要集中在参数优化上，以在计算复杂性略有增加的情况下，在各种应用中加速CNN性能。2013年，Zeiler和Fergus[28]定义了一种机制，可以可视化每个CNN层学习的过滤器。可视化方法用于通过减小过滤器的尺寸来改进特征提取阶段。同样，牛津大学小组提出的VGG架构[29]在2014年ILSVRC竞赛中获得亚军，与AlexNet相比，其接感受野要小得多，但体积却增加了。在VGG中，特征图体积在每一层加倍，深度从9层增加到16层。同年，赢得2014-ILSVRC竞赛的GoogleNet[99]不仅致力于通过更改层设计来降低计算成本，而且根据深度扩展了宽度，以改善CNN性能。 GoogleNet引入了分割、变换和合并块的概念，其中合并了多尺度和多层转换信息以获取局部和全局信息[33]，[99]，[100]。使用多层转换信息有助于CNN处理不同层级的图像细节。在2012-14年度，CNN学习能力的提高主要是通过增加CNN的深度和参数优化策略实现。这表明CNN的深度有助于改善分类器的性能。

3.5 2015年至今：CNN的结构创新和应用迅速增长

通常来说，CNN的性能重大改善出现在2015-2019年期间。CNN的研究仍在进行中，并且有很大的改进潜力。CNN的表示能力取决于其深度，从某种意义上说，它可以通过定义从简单到复杂的各种级别特征来帮助学习复杂的问题。通过将复杂的问题分成较小的模块，多层转换使学习变得容易。但是，深度架构所面临的主要挑战是负面学习的问题，这是由于网络较低层的梯度减小而发生的。为了解决这个问题，不同的研究小组致力于层连接的重新调整和新模块的设计。2015年初，Srivastava等人使用跨通道连接和信息门控机制的概念来解决梯度消失问题并提高网络表示能力[101]-[103]。这个想法在2015年末成名，并提出了类似的概念：残差块或跳跃连接[31]。残差块是跨通道连接的一种变体，它通过规范跨块的信息流来平滑学习[104]-[106]。这个想法在ResNet体系结构中用于训练150层深度网络[31]。跨通道连接的思想被Deluge，DenseNet等进一步扩展到了多层连接，以改善表示性[107]，[108]。

在2016年，研究人员还结合深度探索了网络的宽度，以改进特征学习[34]，[35]。除此之外，没有新的突出的体系结构修改，而是使用已经提出的体系结构的混合来提高深层CNN性能[33]，[104]-[106]，[109]，[110]。这一事实使人感觉到，对有效调节CNN性能，相比适当组装网络单元，可能还有其他更为重要因素。对此，Hu等（2017）确定网络表示在深度CNN的学习中发挥作用[111]。Hu等人介绍了特征图开发的思想，并指出少量信息和领域无关的特征可能会在更大程度上影响网络的性能。他利用了上述想法，并提出了名为“挤压和激发网络（SE-Network）[111]”的新架构。它通过设计专门的SE块来利用特征图（在文献中通常称为通道）信息。该块根据每个特征图在类识别中的作用为每个特征图分配权重。不同的研究

人员对该想法进行了进一步的研究，他们通过利用空间和特征图（通道）信息将注意力转移到重要区域[37]，[38]，[112]。在2018年，Khan等人[36]引入了一种新的通道提升思路。用通道提升表示进行网络训练的动机是使用丰富的表示。通过学习各种特征以及通过TL概念利用已经学习的特征，该想法有效地提高了CNN的性能。

从2012年至今，已经出现许多CNN架构的改进。关于CNN的架构进步，最近的研究重点是设计新的块，这些块可以通过利用特征图和空间信息或通过添加人工通道来增强网络表示。

4、CNN中的结构创新

从1989年至今，CNN架构已进行了不同的改进。这些改进可以归类为参数优化、正则化、结构重构等。但是，可以观察到，CNN性能改进的主要动力来自处理单元的重组和新模块的设计。 CNN架构中的大多数创新都与深度和空间利用有关。 根据架构修改的类型，CNN可以大致分为以下七个类别：空间利用，深度，多路径，宽度，特征图利用，通道提升和基于注意力的CNN。图4所示的Deep CNN的分类法显示了七个不同的类，而它们的摘要在表1中。

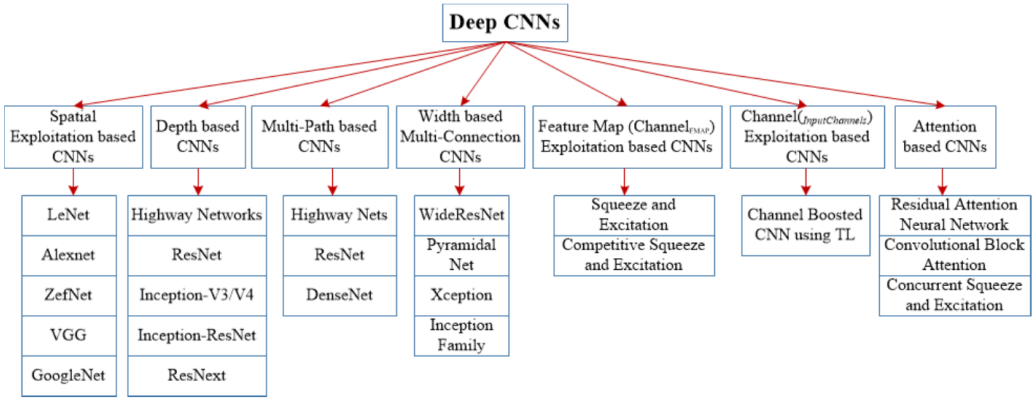

图4 深度CNN结构分类

表1不同类别最新体系结构性能比较，报告所有架构的Top 5个错误率

Architecture Name	Year	Main contribution	Parameters	Error Rate	Depth	Category	Reference
LeNet	1998	- First Popular CNN architecture	0.060 M	[dist]MNIST: 0.8 MNIST: 0.95	7	Spatial Exploitation	[65]
AlexNet	2012	- Deeper and wider than the LeNet - Uses Relu ,Dropout and overlap Pooling - GPUs NVIDIA GTX 580	60 M	ImageNet: 16.4	8	Spatial Exploitation	[21]
ZefNet	2014	- Intermediate layers feature visualization	60 M	ImageNet: 11.7	8	Spatial Exploitation	[28]
VGG	2014	- Homogenous topology - Small kernel size	138 M	ImageNet: 7.3	19	Spatial Exploitation	[29]
GoogLeNet	2015	- Split Transform Merge - Introduces block concept	4 M	ImageNet: 6.7	22	Spatial Exploitation	[99]
Inception-V3	2015	- Handles the problem of a representational bottleneck - Replace large size filters with small filters - Replaces the bigger filter with smaller filters	23.6 M	ImageNet: 3.5 Multi-Crop: 3.58 Single-Crop: 5.6	-	Depth	[100]
Highway Networks	2015	- Multi-Path Idea	2.3 M	CIFAR-10: 7.76	19	Depth + Multi-Path	[101]
Inception-V4	2016	- Split, Transform, Merge Uses asymmetric filter	-	ImageNet: 4.01	-	Depth	[100]
Inception-ResNet	2016	- Split, Transform, Merge and Residual Links	-	ImageNet: 3.52		Depth + Multi-Path	[100]
ResNet	2016	- Residual Learning and Identity mapping based skip connection	6.8 M 1.7 M	ImageNet: 3.6 CIFAR-10: 6.43	152 110	Spatial Exploitation + Depth + Multi-Path	[31]
DelugeNet	2016	- Allow cross layer information inflow in Deep Networks	20.2 M	CIFAR-10: 3.76 CIFAR-100: 19.02	146	Multi-path	[108]
FractalNet	2016	- Different path lengths are interacting with each other without any residual connection	38.6 M	CIFAR-10: 7.27 CIFAR-10+: 4.60 CIFAR-10++: 4.59 CIFAR-100: 28.20 CIFAR-100+: 22.49 CIFAR100++: 21.49	20 40	Multi-Path	[113]
WideResNet	2016	- Width is increased and depth is decreased	36.5 M	CIFAR-10: 3.89 CIFAR-100: 18.85	28 -	Width	[34]
Xception	2017	- Depth wise Convolution followed by point wise convolution	22.8 M	ImageNet: 0.055	36	Width	[114]
Residual Attention Neural Network	2017	- Introduces Attention Mechanism	8.6 M	CIFAR-10: 3.90 CIFAR-100: 20.4 ImageNet: 4.8	452	Attention	[38]
ResNexT	2017	- Cardinality - Homogeneous topology - Grouped convolution	68.1 M	CIFAR-10: 3.58 CIFAR-100: 17.31 ImageNet: 4.4	29 101	Spatial Exploitation	[115]
Squeeze & Excitation Networks	2017	- Models Interdependencies between feature maps	27.5 M	ImageNet: 2.3	152	Feature Map Exploitation	[116]
DenseNet	2017	- Cross-layer information flow	25.6 M 25.6 M 15.3 M 15.3 M	CIFAR-10+: 3.46 CIFAR100+: 17.18 CIFAR-10: 5.19 CIFAR-100: 19.64	190 190 250 250	Multi-Path	[107]
PolyNet	2017	- Experimented structural diversity - Introduces Poly Inception Module - Generalizes residual unit using Polynomial compositions	92 M	ImageNet: Single:4.25 Multi:3.45	- -	Width	[117]
PyramidalNet	2017	- Increases width gradually per unit	116.4 M 27.0 M 27.0 M	ImageNet: 4.7 CIFAR-10: 3.48 CIFAR-100: 17.01	200 164 164	Width	[35]
Convolutional Block Attention Module (ResNeXt101 (32x4d) + CBAM)	2018	- Exploit both spatial and feature map information	48.96 M	ImageNet: 5.59	101	Attention	[37]
Concurrent Squeeze & Channel Excitation Mechanism	2018	- Squeezing spatially followed by exciting channel-wise - Squeezing channel-wise followed by exciting spatially - Performing spatial and channel squeeze & excitation in parallel	-	MALC: 0.12 Visceral: 0.09	-	Attention	[112]
Channel Boosted CNN	2018	- Boost the original channels with extra generated information rich channels	-	-	-	Channel Boosted	[36]
Competitive Squeeze & Excitation Network CMPE-SE-WRN-28	2018	- Residual and identity mappings both are responsible for rescaling the channel	36.92 M 36.90 M	CIFAR-10: 3.58 CIFAR-100: 18.47	28 28	Feature Map Exploitation	[118]

4.1 基于空间利用的CNN

CNN具有大量参数和超参数，例如权重、偏置、处理单元（神经元）数量、层数，滤波器大小、步幅、学习率、激活函数等[119]，[120]。由于卷积运算考虑了输入像素的邻域（局部性），因此可以通过使用不同的滤波器大小来探索不同级别的相关性。因此，在2000年初，研究人员利用空间滤波器来改善这方面的性能，探索了各种大小的过滤器，以评估它们对网络学习的影响。不同大小的过滤器封装了不同级别的粒度；通常，小尺寸滤波器会提取细粒度信息，大尺寸过滤器会提取粗粒度信息。这样，通过调整滤波器大小，CNN可以在粗粒度和细粒度细节上均表现良好。

4.1.1 LeNet

LeNet由LeCuN在1998年提出[65]。它以其历史重要性而闻名，因为它是第一个CNN，显示了手写体识别的最佳性能。它具有对数字进行分类的能力，而不会受到较小的失真，旋

转以及位置和比例变化的影响。 LeNet是一个前馈NN，由五个交替的卷积和池化层组成，然后是两个全连接层。在2000年初，GPU未广泛用于加速训练，甚至CPU也很慢[121]。传统多层全连接神经网络的主要局限性在于，它将每个像素视为一个单独的输入并对其进行转换，这是一个巨大的计算负担，特别是在当时[122]。 LeNet利用了图像的潜在基础，即相邻像素彼此相关并分布在整个图像中。因此，使用可学习的参数进行卷积是一种在很少参数的情况下从多个位置提取相似特征的有效方法。这改变了传统的训练观点，即每个像素被视为与其邻域分离的单独输入特征，而忽略了它们之间的相关性。 LeNet是第一个CNN架构，它不仅减少了参数数量和计算量，而且能够自动学习特征。

4.1.2 AlexNet

LeNet [65]虽然开始了深层CNN的历史，但是在那时，CNN仅限于手数字识别任务，并且不能很好地适用于所有类别的图像。 AlexNet[21]被认为是第一个深度CNN架构，它显示了图像分类和识别任务的开创性成果。 AlexNet由Krizhevesky等人提出，他们通过加深CNN并应用许多参数优化策略来增强CNN的学习能力[21]。 AlexNet的基本体系结构设计如图5所示。在2000年初，硬件限制了深度CNN结构的学习能力，迫使其限制在较小的尺寸。为了利用CNN的表达能力，Alexnet在两个NVIDIA GTX 580 GPU上进行了并行训练以克服硬件的短板。在AlexNet中，特征提取阶段从5（LeNet）扩展到了7，从而使CNN适用于各种类别的图像。尽管事实上通常情况下，深度会提高图像不同分辨率的泛化能力，但是与深度增加相关的主要缺点是过拟合。为了应对这一挑战，Krizhevesky等人（2012）利用了Hinton[56]，[123]的想法，即他们的算法在训练过程中随机跳过了一些变换单元，以强制模型学习更鲁棒的特征。除此之外，ReLU还被用作非饱和激活函数，通过在某种程度上减轻梯度消失的问题来提高收敛速度[53]，[124]。重叠下采样和局部响应归一化也被用于减少过度拟合来提高泛化性。与先前提出的网络相比，其他调整是在初始层使用了大型过滤器（11x11和5x5）。由于AlexNet的高效学习方法，它在新一代CNN中具有重要意义，并开始了CNN体系结构进步研究的新时代。

4.1.3 ZefNet

在2013年之前，CNN的学习机制主要是基于反复试验，而不知道改进背后的确切原因。缺乏了解限制了深层CNN在复杂图像上的性能。2013年，Zeiler和Fergus提出了一种有趣的多层反卷积神经网络（DeconvNet），该网络以ZefNet闻名[28]。开发ZefNet是为了定量可视化网络性能。网络活动可视化的想法是通过解释神经元的激活来监视CNN的性能。在先前的一项研究中，Erhan等人（2009）利用了相同的想法通过可视化隐藏层的特征[125]，优化了深度信念网络（DBN）的性能。Le等人（2011年）以同样的方式通过可视化输出神经元生成的图像类别来评估深度无监督自动编码器（AE）的性能[126]。 DeconvNet的工作方式与前向CNN相同，但颠倒了卷积和池化操作的顺序。这种反向映射将卷积层的输出投影回视觉上可感知的图像模式，从而给出了在每一层学习的内部特征表示的神经元级别的解释[127]，[128]。ZefNet的目标是在训练期间监视学习方案，从而将发现用于诊断与模型相关的潜在问题。这个想法在AlexNet上应用DeconvNet得到了实验验证，结果表明在网络的第一层和第二层中只有少数神经元处于活动状态，而其他神经元则死了（处于非活动状态）。此外，它表明第二层提取的特征表现出混叠伪像（aliasing artifacts，这个是。。。）。基于这些发现，Zeiler和Fergus调整了CNN拓扑并进行了参数优化。Zeiler和Fergus通过减小过滤器尺寸和步幅以在前两个卷积层中保留最大数量的特征，从而最大限度地提高了CNN的学习能力。 CNN拓扑结构的这种重新调整带来了性能提高，这表明特征可视化可用于识别设计缺陷并及时调整参数。

4.1.4 VGG

随着CNN成功用于图像识别，Simonyan等人提出了一种简单有效的CNN架构设计原则。他们的名为VGG的体系结构是模块化的分层模式[29]。与AlexNet和ZefNet相比，VGG的深度为19层，以模拟深度与网络表示能力的关系[21]，[28]。ZefNet是2013年ILSVRC竞赛的一线网络，它建议使用小型滤波器可以提高CNN的性能。基于这些发现，VGG用一堆3x3卷积层代替了11x11和5x5滤波器，并通过实验证明，同时放置3x3滤波器可以达到大尺寸滤波器的效果（感受野同大尺寸滤波器同样有效（5x5和7x7））。小尺寸滤波器的另一个好处是通过减少参数的数量提供了较低的计算复杂性。这些发现为在CNN中使用较小尺寸的滤波器创造了新的研究趋势。 VGG通过在卷积层之间放置1x1卷积来调节网络的复杂性，此外，还可以学习所得特征图的线性组合。为了调整网络，将最大池化层放置在卷积层之后，同时执行填充以保持空间分辨率[46]。VGG在图像分类和定位问题上均显示出良好的效果。虽然VGG未在2014-ILSVRC竞赛中名列前茅，但由于其简单、同质的拓扑结构和增加的深度而闻名。与VGG相关的主要限制是计算成本高。即使使用小尺寸的滤波器，由于使用了约1.4亿个参数，VGG仍承受着很高的计算负担。

4.1.5 GoogleNet

GoogleNet赢得了2014-ILSVRC竞赛的冠军，也被称为Inception-V1。GoogleNet体系结构的主要目标是在降低的计算成本同时实现高精度[99]。它在CNN中引入了inception块的新概念，通过拆分、变换和合并思想整合了多尺度卷积变换。inception块的体系结构如图6所示。该块封装了不同大小的滤波器（1x1、3x3和5x5），以捕获不同尺度（细粒度和粗粒度）的空间信息。在GoogleNet中，传统的卷积层被替换为小块，类似于在网络中网络（NIN）体系结构中提出的用微型NN替换每层的想法[57]。GoogleNet对分割、变换和合并的想法的利用，有助于解决与学习同一图像类别中存在的各种类型的变体有关的问题。除了提高学习能力外，GoogleNet的重点还在于提高CNN参数的效率。在采用大尺寸内核之前，GoogleNet通过使用1x1卷积滤波器添加瓶颈层来调节计算。它使用稀疏连接（并非所有输出特征图都连接到所有输入特征图），从而通过省略不相关的特征图（通道）来克服冗余信息和降低成本的问题。此外，通过在最后一层使用全局平均池来代替连接层，从而降低了连接密度。这些参数调整使参数量从4000万个大大减少到500万个。应用的其他正则因素包括批量标准化和使用RmsProp作为优化器[129]。GoogleNet还引入了辅助学习器的概念以加快收敛速度。但是，GoogleNet的主要缺点是其异构拓扑，需要在模块之间进行自定义。GoogleNet的另一个限制是表示瓶颈，它极大地减少了下一层的特征空间，因此有时可能会导致有用信息的丢失。

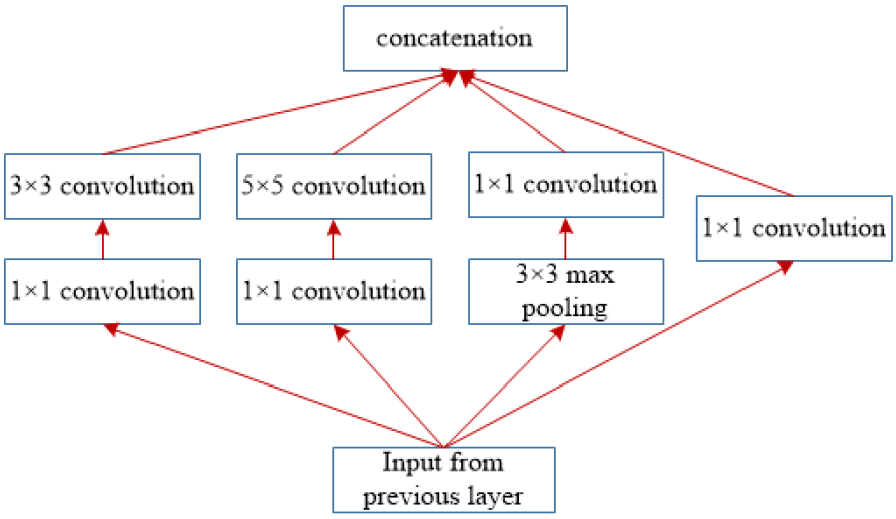

图6 inception块基本结构

4.2 基于深度的CNN

深度CNN架构基于以下假设：随着深度的增加，网络可以通过许多非线性映射和改进的特征表示来更好地近似目标函数[130]。网络深度在监督训练的成功中发挥了重要作用。理论研究表明，与浅层架构相比，深层网络可以更有效地表示某些功能类别[131]。Csáji在2001年提出了通用逼近定理，该定理指出单个隐藏层足以逼近任何函数，但这是以成倍增加许多神经元为代价的，因此经常使它在计算上不可行[132]。在这方面，Bengio和Delalleau [133]提出，更深层的网络有潜力以降低的成本维持网络的表现力[134]。2013年，Bengio等人根据经验表明，对于复杂任务，深度网络在计算上更有效[84]，[135]。在2014年ILSVRC竞赛中表现最好的Inception和VGG，进一步强化了以下观点：深度是调节网络学习能力的重要维度[29]，[33]，[99]，[100]。

4.2.1 Highway Networks

基于直觉，可以通过增加网络深度来提高学习能力。2015年，Srivastava等人提出了一个名为Highway Networks的深层CNN [101]。与深度网络有关的主要问题是训练慢和收敛慢[136]。Highway Networks通过引入新的跨层连接（在第4.3.1节中讨论），利用深度来学习丰富的特征表示。 因此，Highway Networks也被归类为基于多路径的CNN体系结构。 在ImageNet数据集上，具有50层的Highway Networks的收敛速度要好于薄而深的架构[94]，[95]。 Srivastava等人的实验表明，添加10层以上的隐藏单元后，普通网络的性能会降低[137]。 另一方面，即使深度为900层，Highway Networks的收敛速度也比普通网络快得多。

4.2.2 ResNet

ResNet由He等人提出，被认为是Deep Nets的延续[31]。ResNet通过在CNN中引入残差学习的概念彻底改变了CNN架构竞赛，并设计了一种有效的方法来训练深度Nets。与Highway Networks类似，它属于基于多路径的CNN，因此其学习方法将在4.3.2节中讨论。ResNet提出了152层深度CNN，赢得了2015-ILSVRC竞赛。ResNet残差块的体系结构如图7所示。分别比AlexNet和VGG深20倍和8倍的ResNet比以前提出的Nets[21]，[29]表现出更少的计算复杂性。何等人根据经验表明，具有50/101/152层的ResNet在图像分类任务上的错误少于

34层的纯Net。此外，ResNet在著名的图像识别基准数据集COCO [138]上提高了28%。ResNet在图像识别和定位任务上的良好性能表明，深度对于许多视觉识别任务至关重要。

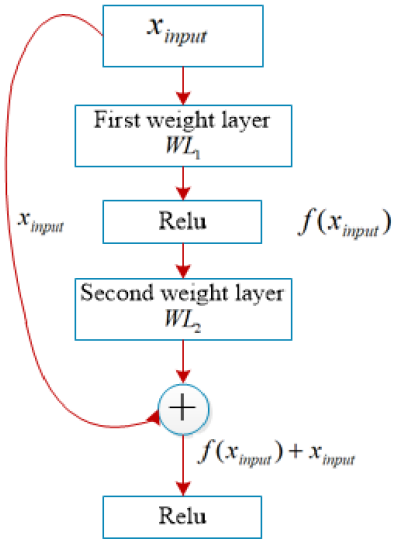

图7 残差块

4.2.3 Inception-V3, V4 and Inception-ResNet

Inception-V3，V4和Inception-ResNet是Inception-V1和V2的改进版本[33]，[99]，[100]。Inception-V3的想法是在不影响泛化的情况下降低更深Nets的计算成本。为此，Szegedy等用小型非对称滤波器（1x7和1x5）替换大型滤波器（5x5和7x7），并在大型过滤器之前使用1x1卷积作为瓶颈[100]。这使得传统的卷积运算更像跨通道相关的。在以前的工作之一，林等充分利用了1x1滤波器在NIN架构中的潜力[57]。Szegedy等 [100]以一种智能的方式使用了相同的概念。在Inception-V3中，使用了1x1卷积运算，该运算将输入数据映射到小于原始输入空间的3或4个独立空间中，然后通过常规3x3或5x5卷积映射这些较小的3D空间中的所有相关性。在Inception-ResNet中，Szegedy等人结合了残差学习和Inception块的作用[31]，[33]。这样做时，滤波器级联被残差连接代替。此外，Szegedy等实验表明，带有残差连接的Inception-V4（Inception-ResNet）具有与普通Inception-V4相同的泛化能力，但深度和宽度增加了。但是，他们观察到Inception-ResNet的收敛速度比Inception-V4更快，这清楚地说明了使用残差连接进行训练会显著加快对Inception网络的训练。

4.2.4 ResNext

ResNext，也称为聚合残差变换网络，是对Inception网络的改进[115]。谢等人通过引入cardinality的概念，以强大而简单的方式利用了分割，变换和合并[99]。cardinality是一个附加维，它是指转换集的大小[139]，[140]。Inception网络不仅提高了传统CNN的学习能力，而且使网络资源有效。但是，由于在转换分支中使用了多种空间嵌入（例如使用3x3、5x5和1x1滤波器），因此需要分别自定义每一层。实际上，ResNext从Inception，VGG和ResNet [29]，[31]，[99]中得出了特征。ResNext通过将split，transform和merge块中的空间分辨率固定为3x3滤波器，利用了VGG的深度同质拓扑和简化的GoogleNet架构。它还使用残差学习。ResNext的构建块如图8所示。ResNext在split，transform和merge块中使用了多个转换，并根据cardinality定义了这些转换。 Xie等人（2017）表明，cardinality的增加显着改善了性能。 ResNext的复杂度是通过在3x3卷积之前应用低嵌入（1x1滤波器）来调节的，优化训练[141] 使用跳跃连接。

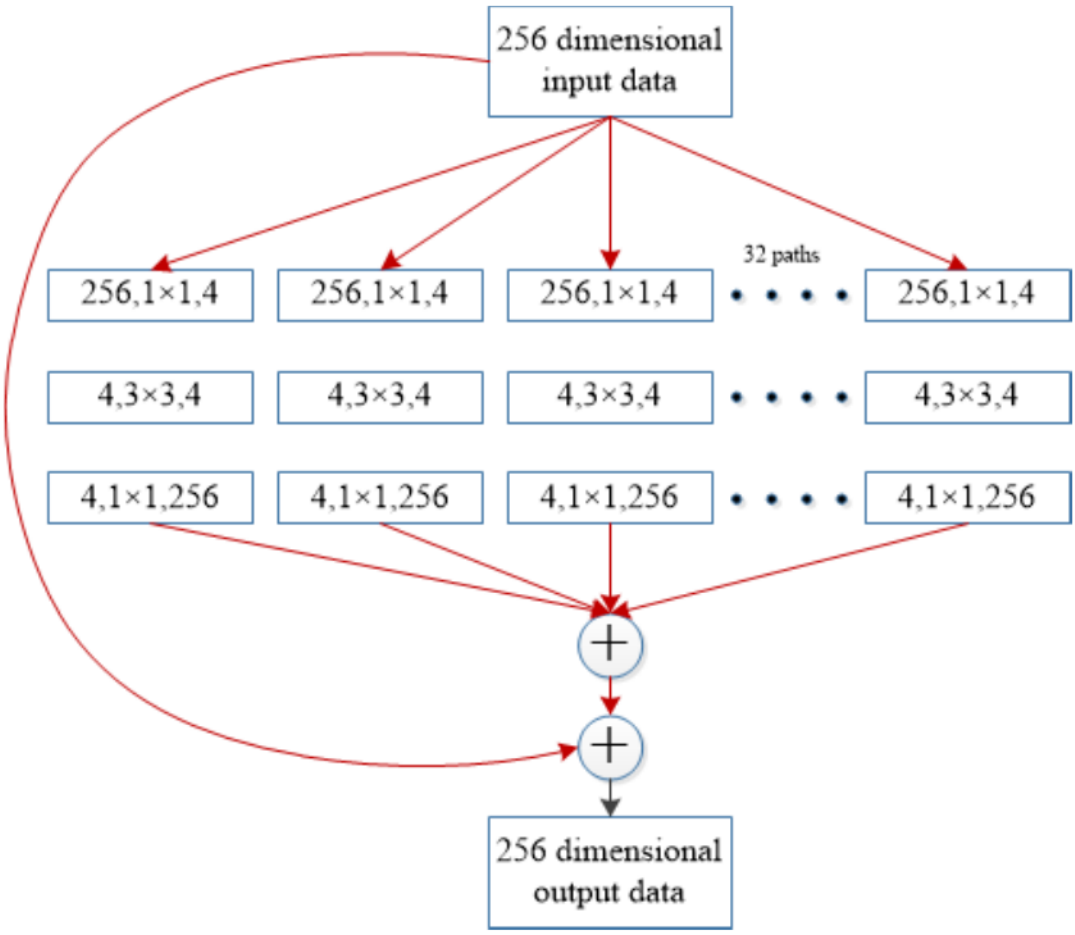

图8 ResNext构建块

4.3 基于多路径的CNN

深度网络的训练是一项艰巨的任务，这已成为最近有关深度网络研究的主题。深度CNN通常在复杂任务上表现良好。然而，更深的网络可能会遭受性能下降，梯度消失或爆炸的问题，这不是由过度拟合引起的，而是由深度的增加引起的[53]，[142]。消失的梯度问题不仅会导致更高的测试误差，而且会导致更高的训练误差[142] –[144]。为了训练更深的网络，研究人员提出了多路径或跨层连接的概念[101]，[107]，[108]，[113]。多个路径或快捷方式连接可以通过跳过某些中间层来系统地将一层与另一层连接，以允许专门信息流的跨层 [145]，[146]。跨层连接将网络分为几个块。这些路径还尝试通过使较低的层可访问梯度来解决梯度消失问题。为此，使用了不同类型的快捷连接，例如零填充，基于投影的，dropout，跳跃连接和1x1连接等。

4.3.1 Highway Networks

网络深度的增加主要是针对复杂问题提高了性能，但同时也使网络训练变得困难。在深网中，由于层数众多，误差的反向传播可能会导致较低层的梯度值较小。为了解决这个问题，Srivastava等人 [101]在2015年，基于跨层连接的想法，提出了一种新的CNN架构，称为Highway Networks。在Highway Networks中，通过在层中分配两个门单元（等式（5）），可以实现跨层的信息畅通无阻。门控机制的思想是从基于长期短期记忆（LSTM）的递归神经网络（RNN）[147]，[148]中得到启发的。通过组合第l层和之前的l-k层信息来聚合信息，产生正则化效果，从而使基于梯度的深度网络训练变得容易。这样就可以

使用随机梯度下降（SGD）算法训练具有100多个层甚至多达900层的网络。Highway Networks网络的跨层连接性在公式（5和6）中定义。

$$\mathrm{y} = H_l\left(x_i,\ W_{H_l}\right).T_g\left(x_i, W_{T_g}\right) + x_i.C_g(x_i, W_{C_g}) \quad (5)$$

$$C_g\left(x_i, W_{C_g}\right) = 1 - T_g\left(x_i, W_{T_g}\right) \quad (6)$$

在公式（5）中，T_g为转换门，表示所产生的结果，而C_g为进位。在网络中，$H_l\left(x_i,\ W_{H_l}\right)$表示隐藏层的作用和残差的实现。而$1 - T_g\left(x_i, W_{T_g}\right)$充当层中的开关，决定信息流的路径。

4.3.2 ResNet

为了解决在训练更深网络时遇到的问题，He等人（2015年）利用了Highway Network中使用的旁路途径提出了ResNet[31]。ResNet的数学公式用公式（7和8）表示。

$$\mathrm{g}\left(x_i\right) = f\left(x_i\right) + x_i \quad (7)$$

$$\mathrm{f}\left(x_i\right) = g\left(x_i\right) - x_i \quad (8)$$

其中，$f\left(x_i\right)$是转换后的信号，而x_i是原始输入。 原始输入x_i通过旁路路径添加到$f\left(x_i\right)$。 本质上，$g\left(x_i\right) - x_i$进行残差学习。ResNet在层内引入了快捷连接以实现跨层连接，但是与Highway Networks相比，这些门是独立于数据且无参数的。 在Highway Networks中，当关闭门控快捷时，这些图层表示非残差功能。但是，在ResNet中，始终传递残差信息，并且永远不会关闭快捷连接。残差链接（快捷连接）加快了深层网络的收敛速度，从而使ResNet能够避免梯度消失问题。152层深度的ResNet（分别比AlexNet和VGG的深度分别高20倍和8倍）赢得了2015-ILSVRC冠军[21]。即使深度增加，ResNet的计算复杂度仍比VGG [29]低。

4.3.3 DenseNets

在Highway Networks和ResNet的延续中，研究人员提出了DenseNet来解决梯度消失问题[31]，[101]，[107]。ResNet的问题在于它通过附加信息转换显式地保留信息，因此许多层可能贡献很少或根本没有信息。为了解决此问题，DenseNet使用了跨层连接，但是以一种修改的方式。 DenseNet以前馈的方式将每一层连接到其他每一层，将所有先前层的特征图用作所有后续层的输入。与传统CNN中一层与其上一层之间的l个连接相比，这在DenseNet中建立了$\frac{l(l+1)}{2}$个直接连接。它加强了跨层深度卷积的效果。由于DenseNet级联了先前层特征而不是添加它们，因此，网络可以具有显式区分添加到网络的信息和保留的信息的能力。 DenseNet具有窄层结构，但是，随着特征图数量的增加，它在参数上变得昂贵。通过损失函数使每一层直接进入梯度，可以改善整个网络中的信息流。这具有正则化效果，可减少使用较小训练集任务的过拟合。

4.4 基于宽度的多连接CNNs

在2012年至2015年期间，研究重点主要是开发深度以及网络规范化中多通道监管连接的有效性上[31]，[101]。然而，川口等报告说网络的宽度也很重要[149]。多层感知器通过在层中并行使用多个处理单元，获得了比感知器映射复杂功能的优势。这表明，宽度是和深度同样重要的定义学习原则的参数。 Lu（2017年）以及Hanin和Sellke（2017年）等人最近表明，具有ReLU激活功能的NN必须足够宽，以随着深度的增加保持通用逼近性质[150]。此

外，如果网络的最大宽度不大于输入维数，则紧凑集上的一类连续函数不能被任意深度的网络很好地近似[135]，[151]。虽然，多层的堆叠（深度增加）可以学习各种特征表示，但不一定可以提高NN的学习能力。与深层架构相关的一个主要问题是某些层或处理单元可能无法学习有用的功能。为了解决这个问题，研究的重点从深层和狭窄的体系结构转向薄和宽的体系结构。

4.4.1 WideResNet

值得关注的是，深度残差网络相关的主要缺点是特征重用问题，其中某些特征转换或块可能对学习的贡献很小[152]。WideResNet解决了这个问题[34]。Zagoruyko和Komodakis提出，深层残差网络的学习潜力主要是由于残差单元，而深度具有补充作用。 WideResNet通过使ResNet变宽而不是变深来利用残差块的功能[31]。WideResNet通过引入附加因子k，该因子控制网络的宽度。 WideResNet表明，与使残差网络更深相比，拓宽层可能会提供更有效的性能改善方法。尽管深度残差网络提高了表示能力，但是它们具有一些缺点，例如时间密集型训练，许多特征图的失活（特征重用问题）以及梯度消失和爆炸问题。何等人通过将dropout引入残差块以有效地规范网络来解决特征重用问题[31]。同样，黄等人引入了随机深度的概念来解决梯度消失和学习缓慢的问题[105]。目前，即使性能的部分改善也可能需要添加许多新层。一项经验研究表明，WideResNet的参数数量是ResNet的两倍，但可以比深度网络更好地进行训练[34]。更宽的残差网络是基于以下观察结果：与ResNet相比，残差网络之前的几乎所有体系结构（包括最成功的Inception和VGG）都更宽。在WideResNet中，通过在卷积层之间而不是在残差块内部添加dropout来使学习有效。

4.4.2 Pyramidal Net

在早期的深度CNN架构中，例如AlexNet，VGG和ResNet，由于多个卷积层的堆叠，特征图的深度在后续层中增加。但是，空间维数会减小，因为每个卷积层后都有子采样层[21]，[29]，[31]。因此，Han等人认为，在深层的CNN中，丰富的特征表示可以通过减小特征图的大小来弥补[35]。特征图深度的急剧增加同时，空间信息的丢失限制了CNN的学习能力。ResNet在图像分类问题上显示出了非凡的成果。但是，在ResNet中，删除残差块通常会降低性能，在残差块中，空间图和特征图（通道）的尺寸都会发生变化（特征图深度增加，而空间尺寸减小）。在这方面，随机ResNet通过随机丢弃残差单元减少信息损失来提高性能[105]。为了提高ResNet的学习能力，Han等人提出了金字塔网络（Pyramidal Net）[35]。与ResNet随深度的增加而导致的空间宽度的急剧减小相反，金字塔形网络逐渐增加了每个残差单位的宽度。这种策略使金字塔网络能够覆盖所有可能的位置，而不是在每个残差块内保持相同的空间尺寸，直到下采样为止。由于特征图的深度以自上而下的方式逐渐增加，因此被命名为金字塔网。在金字塔网络中，特征图的深度由因子l调节，并使用公式（9）计算。

$$D_l = \begin{cases} 16 & if\ l = 1 \\ \left\lfloor D_{l-1} + \frac{\gamma}{n} \right\rfloor & if\ 2 \le l \le n+1 \end{cases} \tag{9}$$

其中，D_l表示第l个残差单元的维数，n是残差单元的总数，而γ是阶跃因子，并且$\frac{\gamma}{n}$调节深度的增加。深度调节因子试图分配特征图增加的负担。通过使用零填充identity mapping将残差连接插入到层之间。零填充identity mapping的优点是，与基于投影的shortcut连接相比，它需要较少的参数数量，因此可以得到更好正则化[153]。金字塔形网络使用两种不同的方法来扩展网络，包括基于加法和乘法的扩宽。两种类型的拓宽之间的区别在于，加法的的金字塔结构线性增加，乘法的金字塔结构在几何上增加[50]，[54]。然而，金字塔形网的主

要问题在于，随着宽度的增加，空间和时间都发生二次方的增加。

4.4.3 Xception

Xception可以被认为是一种极端的Inception架构，它利用了AlexNet [21]，[114]引入的深度可分离卷积的思想。Xception修改了原始的inception块，使其更宽，并用一个单一的维度（3x3）紧跟1x1替换了不同的空间维度（1x1、5x5、3x3），以调节计算复杂度。Xception块的体系结构如图9所示。Xception通过解耦空间和特征图（通道）相关性来提高网络的计算效率。它先使用1x1卷积将卷积输出映射到低维嵌入，然后将其空间变换k次，其中k为cardinality的宽度，它确定变换的次数。Xception通过在空间轴上分别对每个特征图进行卷积，使计算变得容易，然后进行逐点卷积（1x1卷积）以执行跨通道关联。在Xception中，使用1x1卷积来调节特征图深度。在传统的CNN架构中，传统的卷积运算仅使用一个变换段，Inception使用三个变换段，而在Xception中，变换段的数量等于特征图的数量。尽管Xception采用的转换策略不会减少参数的数量，但是它使学习更加有效并提高了性能。

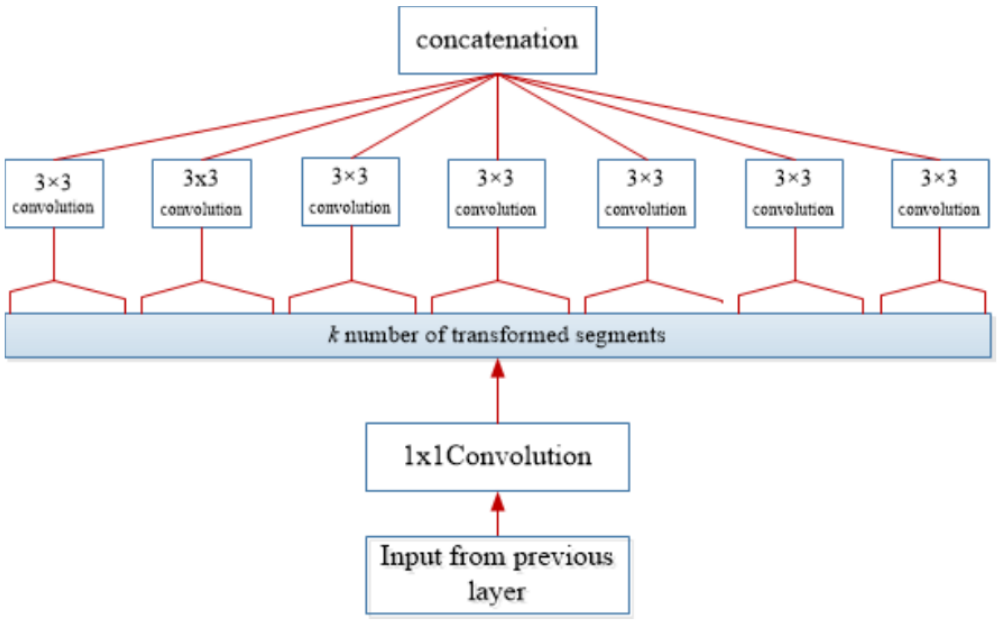

图9 Xception构建块

4.4.4 Inception家族

CNN的Inception家族也属于基于宽度的方法类别[33]，[99]，[100]。在Inception网络中，在一层内，使用了不同大小的滤波器，从而增加了中间层的输出。使用不同大小的滤波器有助于捕获多样的高级特征。 在4.1.4和4.2.3节中讨论了Inception系列的显着特征。

4.5 基于特征图（Channel FMap）利用的CNN

CNN因其分层学习和自动特征提取能力而在MV任务中变得很流行[12]。特征的选择在分类、分割和检测模块的性能中起着重要作用。传统的特征提取技术通常是静态的，并且由于特征类型的限制而限制了分类模块的性能[154]。在CNN中，特征是通过调整与内核（掩码）关联的权重来动态设置的。此外，使用多层特征提取，可以提取各种类型的特征（在CNN中称为特征图或通道）。但是，某些特征图在对象识别中几乎没有作用或没有作用[116]。巨大的特征集可能会产生噪声影响，从而导致网络过拟合。这表明，除了网络工程

之外，特征图的选择在改善网络的泛化方面可以发挥重要作用。在本节中，特征图和通道将可互换使用，因为许多研究人员已将词通道用于特征图。

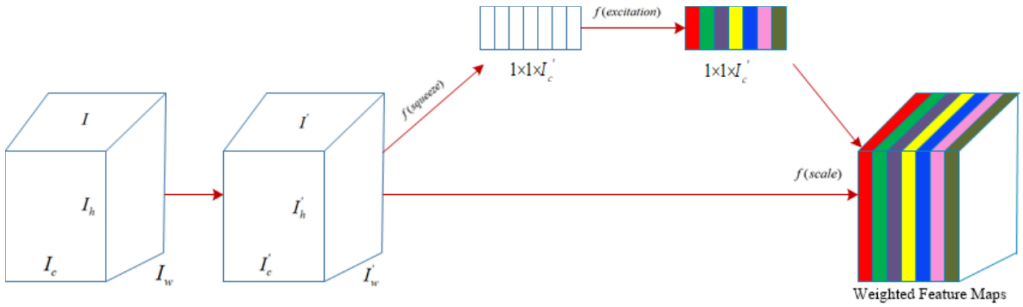

图10 Squeeze和Excitation块

4.5.1 Squeeze和Excitation网络

Hu等人报道了Squeeze和Excitation网络（SE-Network）[116]。他们提出了一个新的块，用于选择与物体识别相关的特征图（通常称为通道）。这个新块被称为SE块（如图10所示），它抑制了不太重要的特征图，但赋予了指定特征图类较高的权重。SE-Network报告了ImageNet数据集错误的减少记录。SE块是一种以通用方式设计的处理单元，因此可以在卷积层之前的任何CNN体系结构中添加。该块的工作包括两个操作：挤压和激发。卷积核捕获局部信息，但是它忽略了该感受野之外特征的上下文关系（相关性）。为了获得特征图的全局视图，压缩块通过抑制卷积输入的空间信息来生成特征图合理统计信息。由于全局平均池化具有有效学习目标对象范围的潜力，因此，挤压操作将其用于使用以下公式生成特征图合理统计信息[57]，[155]：

$$D_M = \frac{1}{m*n}\sum_{i=1}^{m}\sum_{j=1}^{m} x_c(i,j) \tag{10}$$

其中，D_M是特征图描述符，$m*n$是输入的空间维度。 挤压操作输出D_M分配给激励操作，该激励操作通过利用门控机制来建模基于主题的相互依赖性。激励操作使用两层前馈NN将权重分配给特征图，这在数学上用公式（11）表示。

$$V_m = \sigma(\omega_2\delta(\omega_1 D_m)) \tag{11}$$

在等式（11）中，V_m表示每个特征图的权重，其中δ和σ分别表示ReLU和S形函数。在激励操作中，ω_1和ω_2用作调节因子，以限制模型的复杂性并有助于泛化[50]，[51]。压缩块的输出之前是ReLU激活函数，该函数在特征图中增加了非线性。SE块中使用S形激活函数的门控机制，该函数可模拟特征图之间的相互依赖性并根据特征图的相关性分配权重[156]。 SE块很简单，并且通过将卷积输入与主题响应相乘来自适应地重新校准每个图层特征图。

4.5.2 竞争Squeeze和Excitation网络

Hu等人提出了Competitive Inner-Imaging Squeeze and Excitation for Residual Network（也称为CMPE-SE网络）。在2018年[118]， Hu等人用SE块的思想来改善深度残差网络的学习[116]。SE-Network根据特征图在分类识别中的作用重新校准特征图。但是，SE-Net的主要问题在于，在ResNet中，它仅考虑残差信息来确定每个通道的权重[116]。这样可以最小化SE块的影响，使ResNet信息变得多余。Hu等人通过从基于残差和基于identity mapping的特征生成特征图合理统计信息来解决此问题。在这方面，使用全局平均池化操作来生成特征图的全局表示，而通过基于残差和identity mapping的描述符之间进行竞争来估计特征图的相关性。这种现象称为inner imaging [118]。 CMPE-SE块不仅对残差特征图之间的关系进行建

模，而且还将其与inner imaging图的关系进行映射，并在二者之间进行竞争。 CMPE-SE块的数学表达式使用以下公式表示：

$$y = F_{se}(\mu_r, x_{id}) . F_{res}(x_{id}, \omega_r) + x_{id} \quad (12)$$

其中x_{id}是输入的identity mapping，F_{se}表示应用于残差特征图μ_r和identity mapping特征图x_{id}的挤压操作，F_{res}表示SE块在残差特征图上的实现。挤压操作的输出与SE块输出F_{res}相乘。反向传播算法因此尝试优化identity mapping和残差特征图之间的竞争以及残差块中所有特征图之间的关系。

4.6 基于通道（输入）利用的CNNs

图像表示在确定图像处理算法（包括传统算法和深度学习算法）的性能方面起着重要作用。一种良好图像表示方法可以通过紧凑代码定义图像的显着特征。在文献中，各种类型的常规滤波器被用于为单个类型的图像提取不同级别的信息[157]，[158]。然后将这些不同的表示形式用作模型的输入，以提高性能[159]，[160]。现在，CNN是一个有效的特征学习器，可以根据问题自动提取区分特征[161]。但是，CNN的学习依赖于输入表示。输入中缺乏多样性和类别可辨别信息可能会影响CNN作为判别器的性能。为此，在CNN中引入了使用辅助学习器的通道提升（输入通道维度）的概念，以增强网络的表示[36]。

4.6.1 使用TL的通道提升CNN

在2018年，Khan等人基于增加输入通道数以提高网络的表示能力的想法，提出了一种新的CNN体系结构，称为通道提升CNN（CB-CNN）[36]。CB-CNN的框图如图11所示。通过在深层生成模型人为地创建额外的通道（称为辅助通道），然后通过深层判别模型加以利用，从而进行通道提升。该文认为可以在生成和区分阶段都使用TL的概念。数据表示在确定分类器的性能中起着重要作用，因为不同的表示可能表示信息的不同方面[84]。为了提高数据的代表性，Khan等人利用了TL和深度生成学习器[24]，[162]，[163]。生成型学习器试图在学习阶段表征数据生成分布。在CB-CNN中，自动编码器用作生成学习器，以学习解释数据背后变化的因素。增强以原始通道空间（输入通道）学习到的输入数据分布，归纳TL的概念以新颖的方式用于构建提升输入表示。CB-CNN将通道提升阶段编码为一个通用块，该块插入到深层网络的开头。对于训练，Khan等人使用了预训练的网络以减少计算成本。这项研究的意义在于，将生成学习模型用作辅助学习器的情况下，可以增强基于深度CNN的分类器表示能力。尽管仅评估了通过在开始时插入提升块来提升通道的潜力，但是Khan等人（2003年）建议，这一想法可以拓展到在深度体系结构的任何层提供辅助通道。CB-CNN也已经在医学图像数据集上进行了评估，与以前提出的方法相比，它改进了结果。CB-CNN在有丝分裂数据集上的收敛曲线如图12所示。

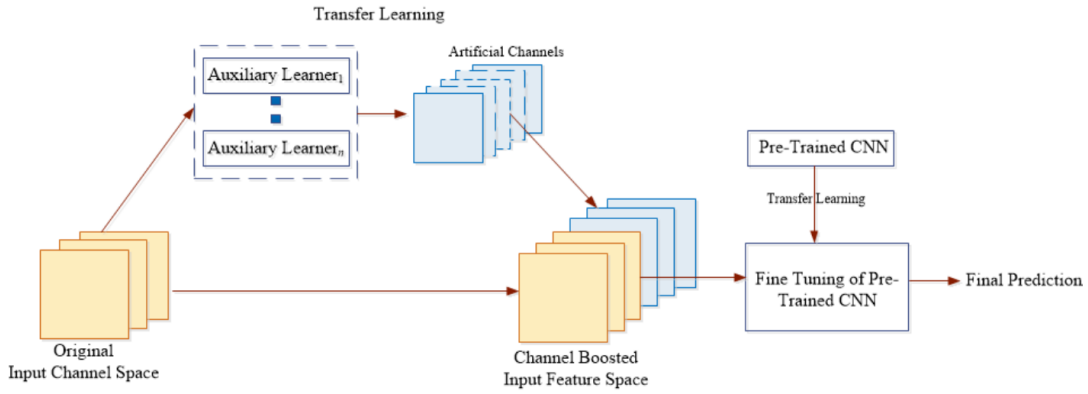

图11 CB-CNN基本结构

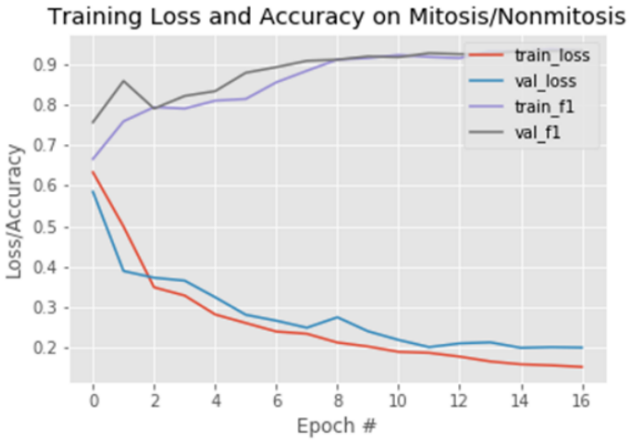

图12 CB-CNN在有丝分裂数据集上的收敛曲线。损失和精度显示在y轴上，而x轴表示Epoch。CB-CNN的训练图表明，该模型在约14个Epoch后收敛。

4.7 基于注意力的CNNs

不同级别的抽象在定义NN的区分能力方面具有重要作用。除了学习不同的抽象级别外，关注与上下文相关的特征在图像定位和识别中也起着重要作用。在人类视觉系统中，这种现象称为注意力。人们瞥见一连串的场景，会注意与上下文相关的部分。此过程不仅用于聚焦选定区域，而且还可以推断该位置处对象的不同解释，从而有助于更好地捕获视觉结构。 RNN和LSTM [147]，[148]或多或少都具有类似的解释性。 RNN和LSTM网络利用注意力模块生成顺序数据，新采样依据先前迭代中的出现分配权重。注意力的概念已被各种研究人员纳入CNN中，以改进表示形式并克服计算限制。这种关注的想法还有助于使CNN足够智能，甚至可以从杂乱的背景和复杂的场景中识别出物体。

4.7.1 残差注意力神经网络

Wang等人提出了一种残差注意力网络（RAN）来改善网络的特征表示[38]。在CNN中纳入注意力的动机是使网络能够学习对象感知特征。RAN是前馈CNN，它是通过将残差块与注意力模块堆叠在一起而构建的。注意力模块采用自下而上、自顶向下学习策略，分为主干和mask分支。将两种不同的学习策略组合到注意力模块中，可以在单个前馈过程中进行快速前馈处理和自上而下的注意力反馈。自下而上的前馈结构产生具有强语义信息的低分辨率特征图。而自顶向下的体系结构会产生密集的特征，以便对每个像素进行推断。在先前提到的研究中，限制波尔兹曼机使用了自上而下、自下而上的学习策略[164]。同样，Goh等在训练的重建阶段，利用自顶向下的注意机制作为深度玻尔兹曼机（DBM）的正则化因子。自上而下的学习策略以在学习过程中逐渐将地输出输入数据特征图的方式全局优化网络[82]，[164]，[165]。RAN中的注意力模块在每一层生成对象感知软掩模$S_{i,FM}(x_c)$[166]。软掩模$S_{i,FM}(x_c)$通过重新校准主干$T_{i,FM}(x_c)$输出，使用等式（13）将注意力分配给对象，因此，对于每个神经元输出，其行为都像控制门。

$$A_{i,FM}(x_c) = S_{i,FM}(x_c) * T_{i,FM}(x_c) \quad (13)$$

在先前的一项研究中，转换网络[167]，[168]也通过将其与卷积块合并来以一种简单的方式利用注意力的概念，但是主要问题是，转换网络中的注意力模块是固定的，无法适应变化的环境。 通过堆叠多个注意模块，使RAN能够有效识别混乱、复杂和嘈杂的图像。RAN的分层结构使其具有基于每个特征图在各层中的相关性，为每个特征图自适应分配权重的能力[38]。残差单元支持了深层次结构的学习。而且，因此，借助捕获不同级别对象感知特征的能力，引入了三种不同级别的注意力：混合注意力、通道注意力和空间注意力[38]。

4.7.2 卷积块注意力模组

注意力机制和特征图利用的重要性已通过RAN和SE-Network验证[38]，[111]。在这方面，Woo等提出了基于注意力的新CNN：卷积块注意模组（CBAM）[37]。 CBAM设计简单，类似于SE-Network。SE-Network仅考虑特征图在图像分类中的作用，但忽略了图像中对象的空间位置。对象的空间位置在对象检测中具有重要作用。CBAM通过先应用特征图（通道）注意力，然后再应用空间注意力来依次查找注意力图，以找到经过改进的特征图。在文献中，通常将1x1卷积和池化操作用于空间注意力。Woo等的结果表明，沿空间轴池化特征会生成有效的特征描述符。CBAM将平均池化与最大池化连接在一起，从而生成强大的空间注意力图。同样，使用最大池化和全局平均池化操作的组合对特征图统计数据进行建模。Woo等表明最大池化可以提供有关独特对象特征的线索，而全局平均池的使用返回特征图注意力的次优推断。利用平均池化和最大池化可提高网络的表示能力。这些精致的特征图不仅专注于重要部分，而且还提高了所选特征图的表示能力。 Woo等的经验表明，通过串行学习过程制定3D注意图有助于减少参数和计算成本。由于CBAM的简单性，它可以轻松地与任何CNN架构集成。

4.7.3 空间和通道并发激励机制

在2018年，Roy等人通过将空间信息的效果与特征图（通道）信息结合起来，使其适用于分割任务[111]，[112]，扩展了胡等人的工作。他们介绍了三个不同的模块：（i）进行spatially和exciting特征图的wise压缩（cSE），（ii）空间上压缩特征图wise和exciting（sSE），以及（iii）同时进行空间和通道压缩与激励（scSE）。在这项工作中，基于自动编码器的卷积神经网络用于分割，而在编码器和解码器层之后插入了建议的模块。在cSE模块中，采用了与SE-block相同的概念。在此模块中，比例因子是基于目标检测中特征图的组合得出的。由于空间信息在分割中起着重要作用，因此在sSE模块中，空间位置比特征图信息更为重要。为此，选择特征图的不同组合并在空间上加以利用以将其用于分割。在最后一个模块中；在SCSE中，通过从空间和通道信息中得出比例因子来分配对每个通道的注意力，从而有选择地突出显示特定对象的特征图[112]。

5、CNN应用

CNN已成功应用于不同的ML相关任务，即对象检测、识别、分类、回归、分割等[169]-[171]。 但是，CNN通常需要大量的数据来学习。 CNN取得了巨大成功的所有上述领域都具有相对丰富的标记数据，例如交通标志识别，医学图像分割以及自然图像中人脸，文字，行人和人的检测。 CNN的一些有趣应用将在下面讨论。

5.1 自然语言处理

自然语言处理（NLP）将语言转换为任何计算机都可以轻松利用的形式。 CNN已被用于基于NLP的应用中，例如语音识别、语言建模和分析等。尤其是，在引入CNN作为一种新的表示学习算法之后，语言建模或语句建模已经发生了变化。执行语句建模以了解语句的语义，从而根据客户要求提供新颖且有吸引力的应用程序。传统的信息检索方法基于单词或特征来分析数据，但忽略了句子的核心。在[172]中，作者在训练过程中使用了动态CNN和动态k-max池化。这种方法无需考虑任何外部来源（如解析器或词汇）就可以找到单词之间的关系。以类似的方式，collobert等[173]提出了基于CNN的架构，该架构可以同时执行各种与MLP相关的任务，例如分块、语言建模，识别名称实体以及与语义相关的角色建模。在另一篇著作中，胡等人提出了一种基于通用CNN的体系结构，该体系结构执行两个句子

之间的匹配，因此可以应用于不同的语言[174]。

5.2计算机视觉相关应用

计算机视觉（CV）致力于开发可以处理包括图像和视频在内的视觉数据并可以有效地理解和提取有用信息的人工系统。 CV包括面部识别、姿势估计、活动识别等多个领域。面部识别是CV中的一项艰巨任务。最近有关面部识别的研究正在努力使原始图像发生很大变化，即使原始图像不存在。这种变化是由照明、姿势变化和不同的面部表情引起的。Farfade等[175]提出了深层CNN，用于检测来自不同姿势的面部并且还能够识别被遮挡的面部。在另一项工作中，Zhang等人 [176]使用新型的多任务级联CNN进行人脸检测。当与最新技术[177]-[179]进行比较时，张的技术显示出良好的效果。由于人体姿势的高度可变性，人体姿势估计是与CV相关的挑战性任务之一。 Li等 [180]提出了一种基于异构深度CNN的姿态估计相关技术。根据李的技术，经验结果表明，隐藏的神经元能够学习身体的局部部位。同样，Bulat等人提出了另一种基于级联的CNN技术[181]。在其级联体系结构中，首先检测热力图，而在第二阶段，对检测到的热力图执行回归。动作识别是活动识别的重要领域之一。开发动作识别系统的困难在于解决属于同一动作类别的不同模式中特征的平移和扭曲。早期的方法包括运动历史图像的构造，隐马尔可夫模型的使用，动作草图的生成等。近来，王等人[182]提出了一种结合LSTM的三维CNN架构，用于识别视频帧中的不同动作。实验结果表明，Wang的技术优于最新的基于动作识别的技术[183]–[187]。同样，Ji等人提出了另一种基于三维CNN的动作识别系统 [188]。在Ji的工作中，三维CNN用于从多个输入帧通道中提取特征。最新动作识别模型是在提取的组合特征空间上开发的。所提的三维CNN模型以有监督的方式进行训练，并且能够在现实世界的应用程序中执行活动识别。

5.3 物体检测

物体检测专注于识别图像中的不同对象。近来，基于区域的CNN（R-CNN）已被广泛用于物体检测。 任等人（2015年）提出了一种改进的R-CNN，称为快速R-CNN，用于对象检测[189]。在他们的工作中，全卷积神经网络用于提取特征空间，可以同时检测位于不同位置对象的边界和得分。同样，戴等人（2016年）提出了使用全连接CNN的基于区域的对象检测[190]。在Dai的工作中，结果通过PASCAL VOC图像数据集测试报告。Gidaris等人提出了另一种物体检测技术 [191]，它基于基于多区域的深度CNN，有助于学习语义感知功能。 使用Gidaris的方法，可以在PASCAL VOC 2007和2012数据集上以高精度检测物体。

5.4 图像分类

CNN已被广泛用于图像分类[192]-[194]。CNN的主要应用之一是医学图像，尤其是使用组织病理学图像诊断癌症的方法[195]。最近，Spanhol等（2016年）使用CNN诊断乳腺癌图像，并将结果与在包含手工描述符[196]，[197]的数据集上训练的网络进行比较。Wahab等人开发了另一种最近提出的基于CNN的乳腺癌诊断技术 [198]。在Wahab的工作中，涉及两个阶段。 在第一阶段，确定了硬非有丝分裂的实例。在第二阶段，执行数据扩充以解决类偏度问题。同样，Ciresan等 [96]使用了与交通标志信号相关的德国基准数据集。他们设计了基于CNN的体系结构，以较高的识别率执行了与交通标志分类相关的任务。

5.5 语音识别

语音被认为是人类之间的交流纽带。在机器学习领域，在硬件资源可用之前，语音识别模型并没有显示出令人满意的结果。随着硬件资源的发展，具有大量训练数据的DNN训练成为可能。深度CNN通常被认为是图像分类的最佳选择，但是，最近的研究表明，它在语音识别任务上也表现良好。哈米德等报道了基于CNN的说话者独立语音识别系统[199]。实

验结果表明，与早期报道的方法相比，错误率降低了10%[200]，[201]。在另一项工作中，探索了基于卷积层中全部或有限数量权重共享的各种CNN架构[202]。此外，还评估了在使用预训练阶段对整个网络进行初始化之后CNN的性能[200]。实验结果表明，几乎所有探索的体系结构在电话和词汇识别相关任务上均具有良好的性能。

6、CNN面临的挑战

深度CNN在具有时间序列性质或遵循诸如网格之类的数据上已取得了良好的性能。但是，还存在将深层CNN架构用于任务的其他挑战。在与视觉相关的任务中，CNN的一个缺点是，当用于估计物体的姿势、方向和位置时，它通常无法显示出良好的性能。在2012年，AlexNet通过引入数据增强的概念在某种程度上解决了这个问题。数据扩充可以帮助CNN学习各种内部表示形式，从而最终提高性能。同样，Hinton报告说，较低的层应仅将其知识移交给下一层的相关神经元。在这方面，Hinton提出了胶囊网络方法[203]，[204]。

在另项工作中，塞格迪等人研究表明，在噪声图像数据上训练CNN体系结构会导致误分类错误的增加[205]。在输入图像中添加少量的随机噪声能够以某种方式欺骗网络，从而使模型可以对原始图像及其受到轻微干扰的版本进行不同的分类。

关于CNN在不同ML任务上的性能，不同的研究人员进行了有趣的讨论。深度CNN模型训练期间面临的一些挑战如下：

①深度NN通常就像一个黑匣子，因此可能缺乏解释性。因此，有时很难对其进行验证，并且在与视觉有关的任务中，CNN可能对噪声和图像的其他更改几乎没有鲁棒性。

② CNN的每一层都会自动尝试提取与任务相关的更好且特定于问题的功能。 但是，对于某些任务，重要的是在分类之前了解深度CNN提取的特征的性质。CNN中特征可视化的想法可以为这个方向提供帮助。

③深度CNN基于监督学习机制，因此，适当的学习需要大量带标注的数据。相反，人类有能力从少量样本中学习和泛化。

④超参数的选择会极大地影响CNN的性能。超参数值的微小变化会影响CNN的整体性能。这就是为什么仔细选择参数是一个主要的设计问题，需要通过一些合适的优化策略来解决。

⑤ CNN的有效训练需要强大的硬件资源，例如GPU。但是，仍然需要探索如何在嵌入式和智能设备中有效地使用CNN。深度学习在嵌入式系统中的一些应用包括受伤度校正，智慧城市中的执法等[206]-[208]。

7、未来方向

CNN结构设计中不同创新思想的使用改变了研究方向，尤其是在MV中。CNN在网格（如拓扑数据）上的良好表现使其成为强大的图像数据表示模型。CNN架构设计是一个有前途的研究领域，在将来，它可能会成为使用最广泛的AI技术之一。

① 集成学习[209]是CNN研究的前瞻性领域之一。多种多样的架构的组合可以通过提取不同级别的语义表示来帮助模型改进各种类别图像的泛化。同样，批次归一化、dropout和新的激活函数等概念也值得一提。

② CNN作为生成学习器的潜力已在图像分割任务中得到了利用，并显示出良好的效果[210]。 在有监督的特征提取阶段（使用反向传播学习过滤器）开发CNN的生成学习能力可以提高模型的表示能力。同样，需要新的范式，通过在CNN的中间阶段结合使用辅助学习器学习信息特征图来增强CNN的学习能力[36]。

③ 在人类视觉系统中，注意力是从图像捕获信息的重要机制之一。 注意机制以这样一种方式运行，它不仅从图像中提取基本信息，而且还存储了它与图像的其他组成部分的上下文关系[211]，[212]。 将来，将在保持对象与后期阶段对象区分特征的空间相关性方向上进行研究。

④通过利用网络的规模来增强CNN的学习能力，这随着硬件处理单元和计算资源的发展而变得可能。但是，深和高容量结构的训练是内存使用和计算资源的重要开销。这需要对硬件进行大量改进，以加速CNN的研究。CNN的主要问题是运行时适用性。此外，由于CNN的计算成本较高，因此在小型硬件中（尤其是在移动设备中）会阻碍CNN的使用。在这方面，需要不同的硬件加速器来减少执行时间和功耗[213]。目前已经提出了一些非常有趣的加速器，例如专用集成电路，Eyriss和Google张量处理单元[214]。此外，通过降低操作数和三值量化的精度，或者减少矩阵乘法运算的数量，已经执行了不同的操作以节省芯片面积和功率方面的硬件资源。现在也该将研究转向面向硬件的近似模型[215]。

⑤深度CNN具有大量超参数，例如激活函数、内核大小、每层神经元数量以及层排列等。在深度学习的背景下，超参数的选择及其评估时间使参数调整变得非常困难。超参数调整是一项繁琐且直观的任务，无法通过明确的表述来定义。在这方面，遗传算法还可用于通过以随机方式执行搜索以及通过利用先前的结果指导搜索来自动优化超参数[216]-[218]。

⑥深度CNN模型的学习能力与模型的大小有很强的相关性。但是，由于硬件资源的限制，深度CNN模型的容量受到限制[219]。为了克服硬件限制，可以利用管道并行概念来扩大深度CNN训练。Google小组提出了一个分布式机器学习库；GPipe[220]使用同步随机梯度下降和管道并行性进行训练。将来，管道的概念可用于加速大型模型的训练并在不调整超参数的情况下扩展性能。

8、结论

CNN取得了显着进步，尤其是在视觉相关任务方面，因此重新唤起了科学家对ANN的兴趣。在这种情况下，已经进行了多项研究工作，以改善CNN在视觉相关任务上的表现。CNN的进步可以通过不同的方式进行分类，包括激活函数、损失函数、优化、正则化、学习算法以及处理单元的重组。本文特别根据处理单元的设计模式回顾了CNN体系结构的进步，从而提出了CNN体系结构的分类法。除了将CNN分为不同的类别外，本文还介绍了CNN的历史，其应用，挑战和未来方向。

多年来，通过深度和其他结构改进，CNN的学习能力得到了显着提高。在最近的文献中观察到，主要通过用块代替常规的层结构已经实现了CNN性能的提高。如今，CNN架构的研究范式之一是开发新型有效的块架构。这些块在网络中起辅助学习作用，它可以通过利用空间或特征图信息或提升输入通道来改善整体性能。这些模块针对问题有意识的学习，在提高CNN性能方面起着重要作用。此外，CNN的基于块的体系结构鼓励以模块化的方式进行学习，从而使体系结构更简单易懂。块作为结构单元的概念将继续存在并进一步提高CNN性能。另外，除了块内的空间信息以外，注意力和利用通道信息的想法有望变得更加重要。

致谢

我们感谢DCIS的模式识别实验室和PIEAS为我们提供了计算设备。

参考文献

略。