湖南大學

HUNAN UNIVERSITY

综述报告

卷积神经网络结构分析

姓 名：丁浩涛

学 号：S191000905

学 院：信息科学与工程学院

专 业：计算机科学与技术

系 名：计算机科学系

任课教师（职称）：吴强教授

课 程：高等计算机体系结构

2019 年 12 月

卷积神经网络结构分析

摘要

深度卷积神经网络（CNN）是一种特殊类型的神经网络，在各种图像识别及分类的竞赛上表现出了相当优异的成果。深度卷积神经网络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。因此，本综述着重于论述的深度 CNN 架构的最新研究进展以及 CNN 的基本结构。此外，本文还涵盖了对 CNN 现存的几大瓶颈和未来的发展展望。

关键词：卷积神经网络、CNN 架构、表征能力

一. 引言

机器学习（ML）算法属于人工智能（AI）的一个特殊领域，该领域无需明确的编程来实现，只需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，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）任务具有的挑战性促使这个领域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]。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向下采样，这有助于总结结果，并使输入对于几何变形不变[12],[13]。

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的复杂表示能力。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著进步。最近的研究表明，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]。

二．卷积神经网络的基本结构

如今，CNN 被认为是使用最广泛的机器学习技术，尤其是在视觉相关的应用中。CNN 最近在各种 ML 应用中显示了最佳的结果。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图像分类。

典型的 CNN 体系结构通常包括卷积层和池化层的交替，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[28]。CNN 组件的排列在设计新的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。神经网络通道图如图 2.1、2.2 所示。

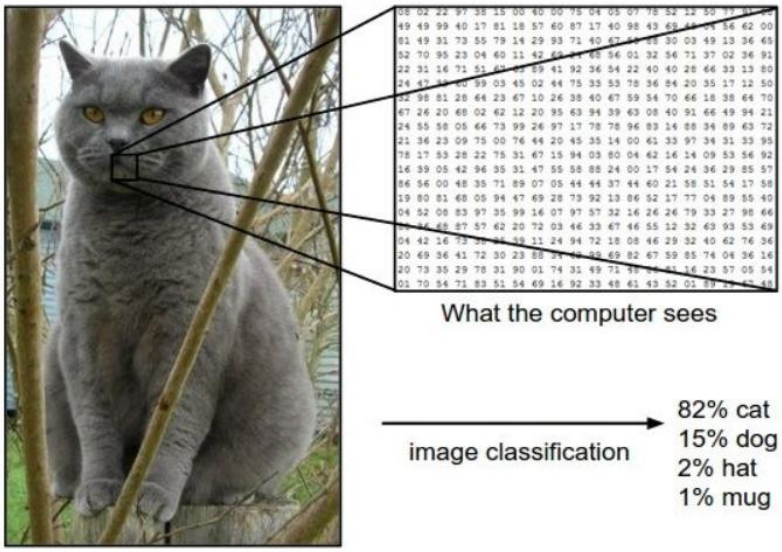

图 2.1 神经网络的基本通道结构

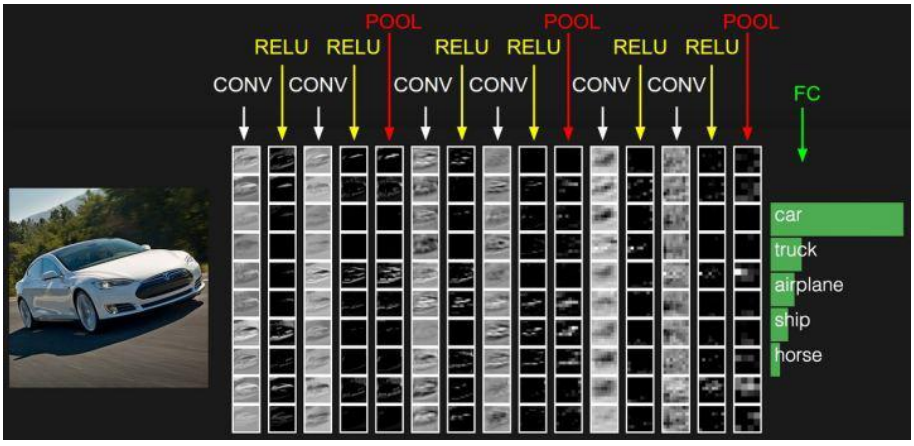

图 2.2 完整神经网络结构

2.1 卷积层

卷积层由一组卷积核（每个神经元充当卷积核）组成。这些卷积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 1 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[29]。另外，如果卷积核是对称的，则卷积运算将变为相关运算[16]。卷积神经网络的卷积层特征提取如图 2.3。

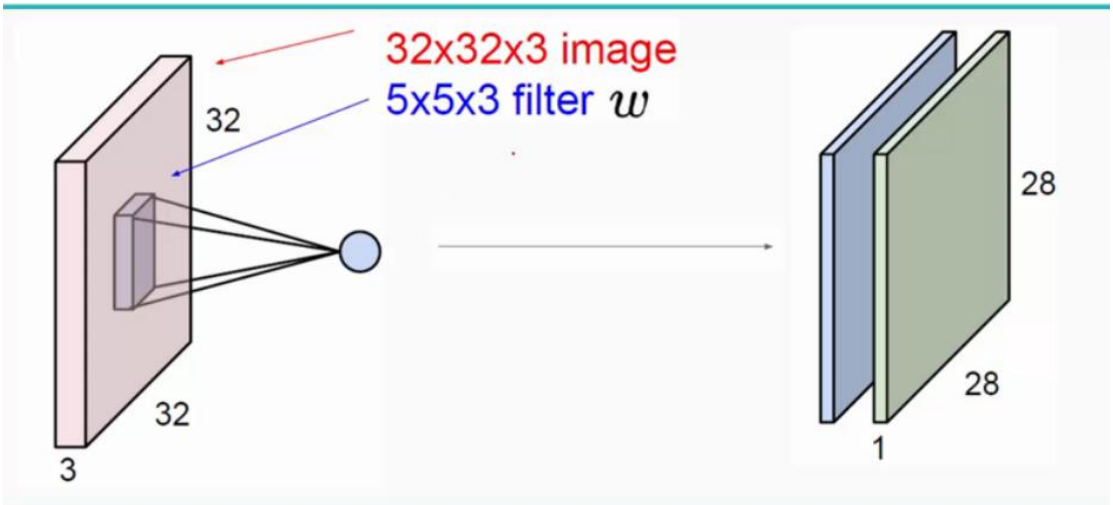

图 2.3 卷积神经网络的卷积层特征提取

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向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[30]。

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

公式（2）表示池化操作，其中Z_l表示第 1 个输出特征图，$F_{x,y}^{\ l}$表示第 1 个输入特征图，而$f_p(.)$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], [31]。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[32]-[34]。

池化层操作如图 2.4。

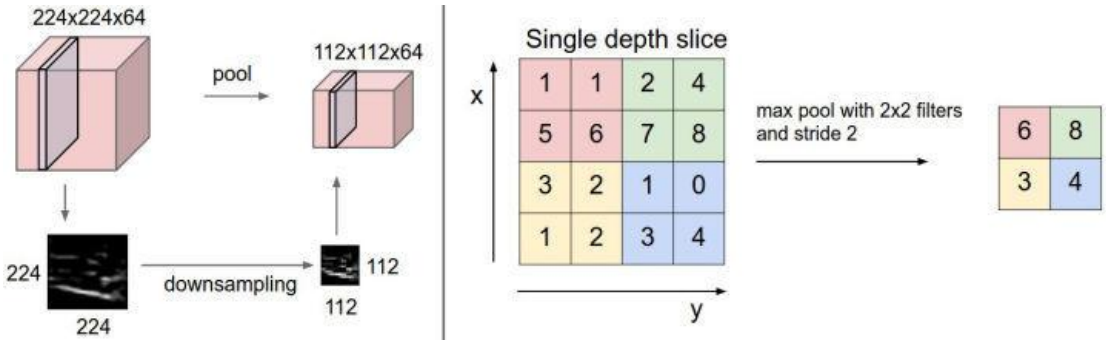

图 2.4 池化层操作

2.3 激活函数

激活函数能起决策功能，有助于学习复杂图片的模式。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。

$$T_1^k = f_A(F_l^k) \tag{3}$$

在上式中，F_1^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_1^k。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，因为它有助于解决梯度消失问题。

常用的激活单元如图 2.5 所示。

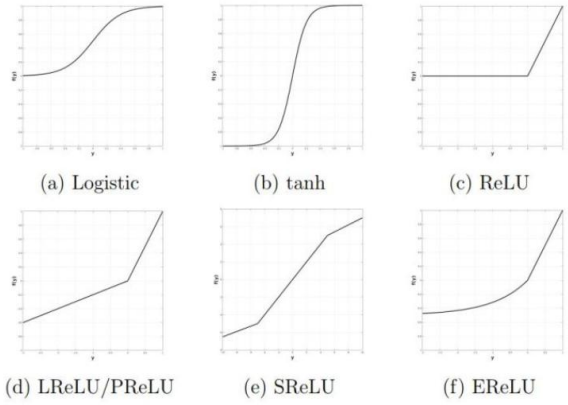

图 2.5 常用的激活单元

2.4 批次归一化

批次归一化用于解决和特征图内部协方差偏移有关的问题。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最小值），并对参数初始化要求高。等式（4）中显示出了变换后的特征图的批次归一化。

$$N_1^k = \frac{F_l^k - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \quad (4)$$

在等式（4）中，N_1^k 表示归一化特征图，F_1^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[35]。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。我们在前向传播的时候，让某个神经元的激活值以一定的概率 p 停止工作，这样可以使模型泛化性更强，因为它不会太依赖某些局部的特征。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[36]。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值[37]。

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，用来把前边提取到的特征综合起来。这将选定特征进行非线性组合，用于数据分类 [38]。

三. 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]。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 2.6 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。

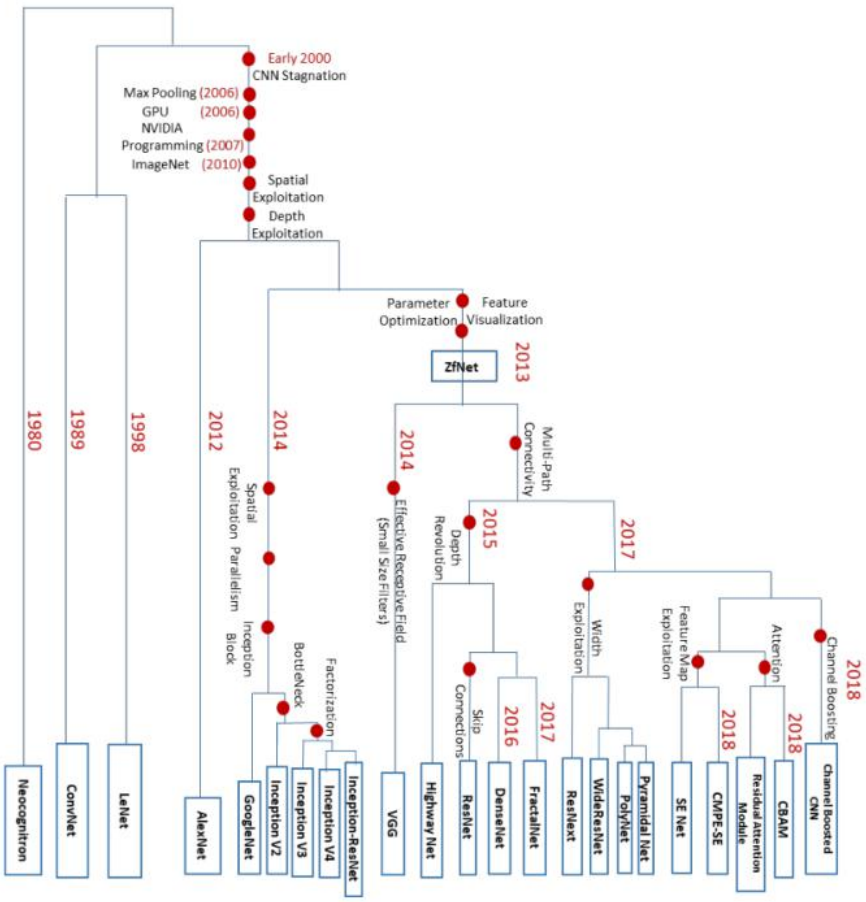

图 2.6 深度 CNN 演化史

四. 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 4.1 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。

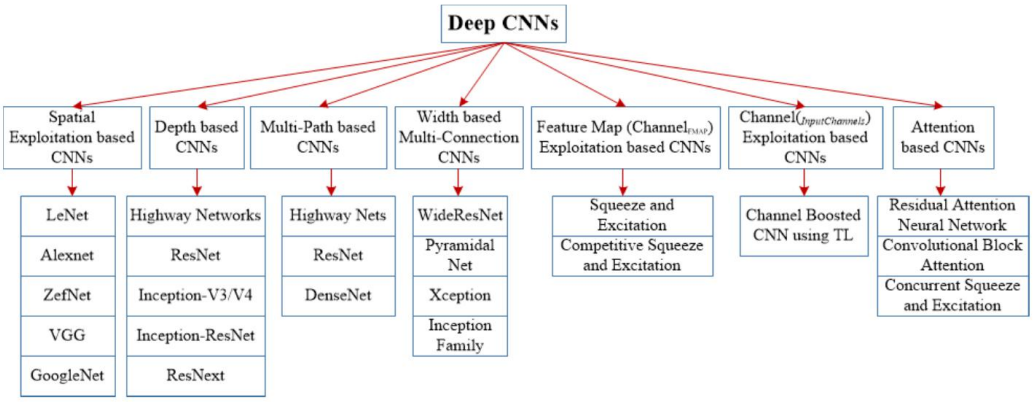

图 4.1 深度 CNN 结构分类

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]

五.里程碑研究和最新进展

5.1 AlexNet

LeNet 虽然开始了深层 CNN 的历史，但是在那时，CNN 仅限于手写数字识别任务，并且不能很好地适用于所有类别的图像。 AlexNet[39]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。AlexNet 的基本体系结构设计如图 5.1 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范围内。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。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。

贡献或者创新点：

（1） AlexNet 首次将卷积神经网络应用于计算机视觉领域的海量数据集 ImageNet,揭示了卷积神经网络的强大特征表达能力和学习能力。另一方面，海量数据同时也使卷积神经网络免于过拟合。自此便引发了深度学习，特别是卷积神经网络在计算机视觉中“井喷式”的研究。

（2）使用 GPU 实现网络训练。模型可以借助 GPU 从而将原本需数周甚至数月的网络训练过程大大缩短为几天（目前利用分布式训练仅需要数小时）。这无疑大大缩短了深度网络和大模型开发研究的周期与时间成本。

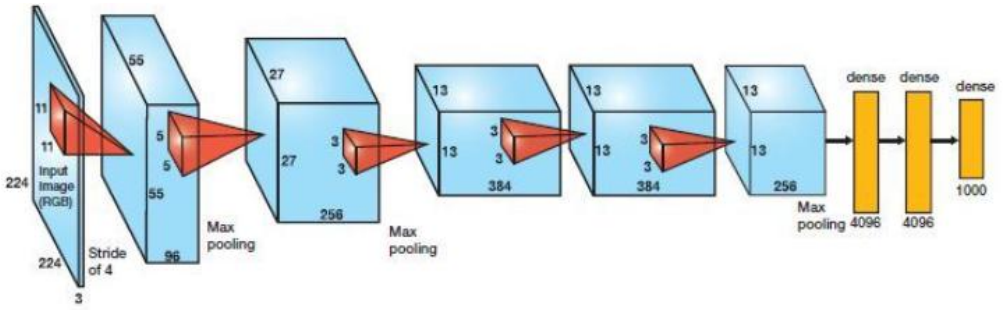

图 5.1 AlexNet 网络结构

5.2 ResNet

ResNet 被认为是 Deep Nets 的延续。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网络。与 Highway Networks 类似，它属于基于多路径的 CNN。ResNet 提出了 152 层深度 CNN，赢得了 2015-ILSVRC 竞赛。ResNet 残差块的 ResNet 体系结构如图 5.2、5.3 所示。分别比AlexNet 和 VGG 深 20 倍和 8 倍的 ResNet 比以前提出的 Nets，表现出更少的计算复杂性。具有 50/101/152 层的 ResNet 在图像分类任务上的错误少于 34 层的 Net。此外,ResNet 在著名的图像识别基准数据集 COCO 上提高了 28%。ResNet 在图像识别和定位任务上的良好的性能表明，深度对于许多视觉识别任务至关重要。ResNet 的贡献及创新点：

（1）提出了残差模块使得训练几百层的神经网络变为可能。之后提出的升级版 Pre-Activate Resnet 甚至能够训练上千层的网络。

（2）提出了 bottleneck 的 1x1 的卷积层，对输入的数据在深度上进行降维，大大减少了计算量以及减少了过拟合的可能性。

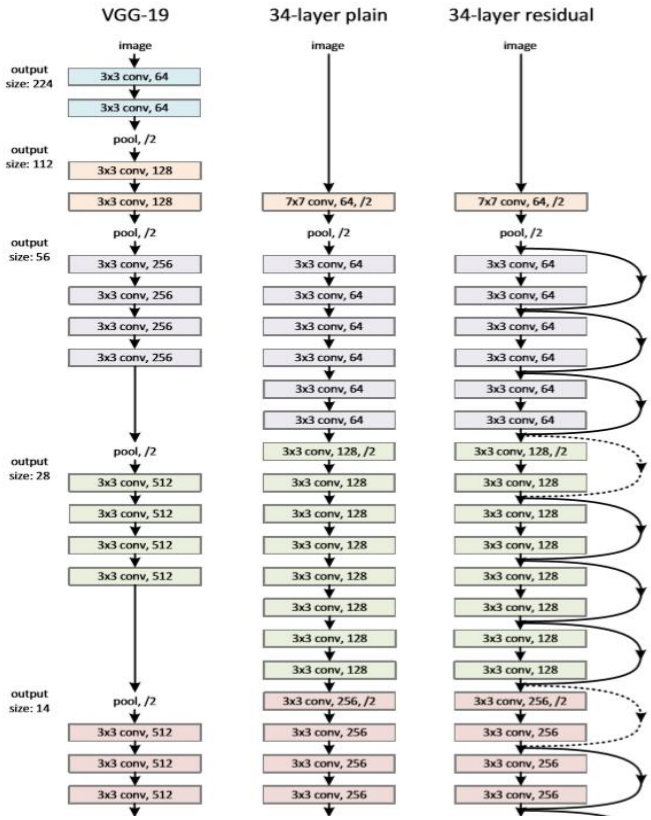

图 5.2 ResNet 网络结构

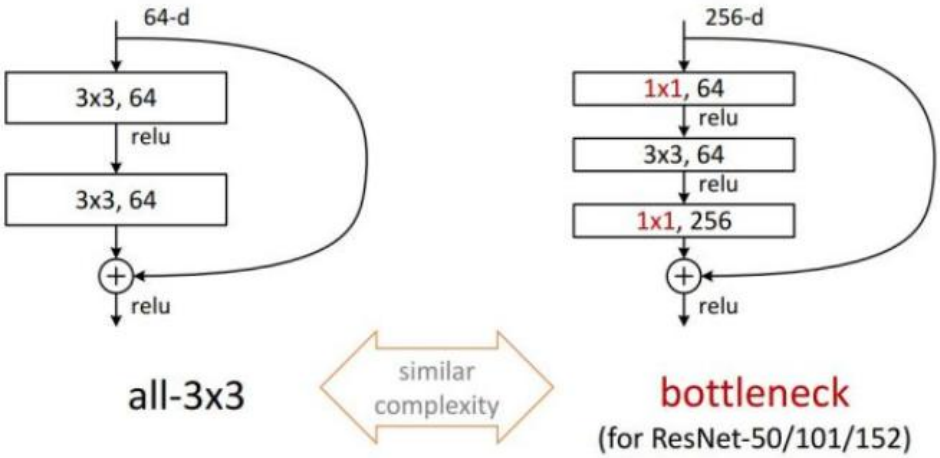

图 5.3 ResNet 具体残差方式

5.3 ResNext

ResNext，也称为聚合残差变换网络，是对 Incepcon网络的改进。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。

cardinality 是一个附加维度，它是指转换集的大小[40], [41]。Incepcon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 Incepcon，VGG 和 ResNet 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 5.4 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。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，优化训练使用跳跃连接。

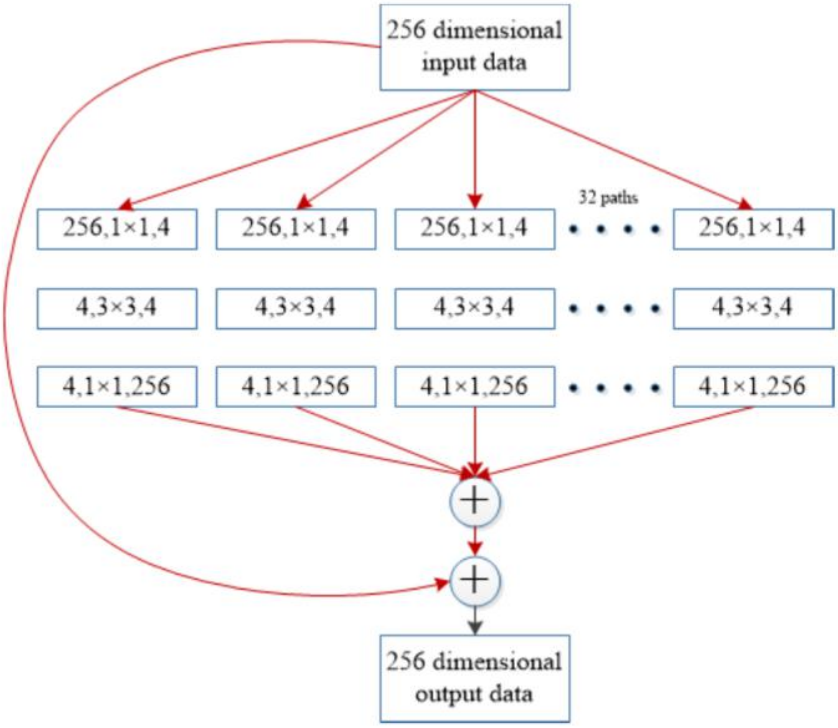

图 5.4 ResNext 构建块

5.4 Xception

Xcepcotion 可以被认为是一种极端的 Incepcon 架构，它利用了 AlexNet [21] 引入的深度可分离卷积的思想。Xcepcon 修改了原始的 incepcon 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。Xcepcon 块的体系结构如图 5.5 所示。Xcepcon 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。Xcepcon 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 Xcepcon 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，Incepcon 使用三个变换段，而在 Xcepcon 中，变换段的数量等于特征图的数量。尽管 Xcepcon 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。

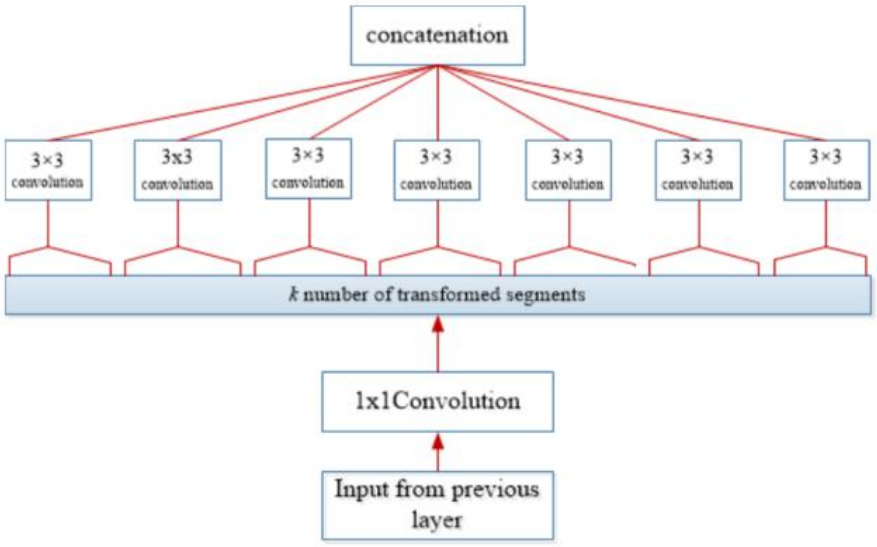

图 5.5 Xcepcon 构建块

5.5 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）[42]。CB-CNN 的框架图如图 5.6 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。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。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，可增强基于深度 CNN 的分类器表示能力。尽管仅评估了在通过开始时插入提升块来提升通道的潜力，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 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 5.7 所示。

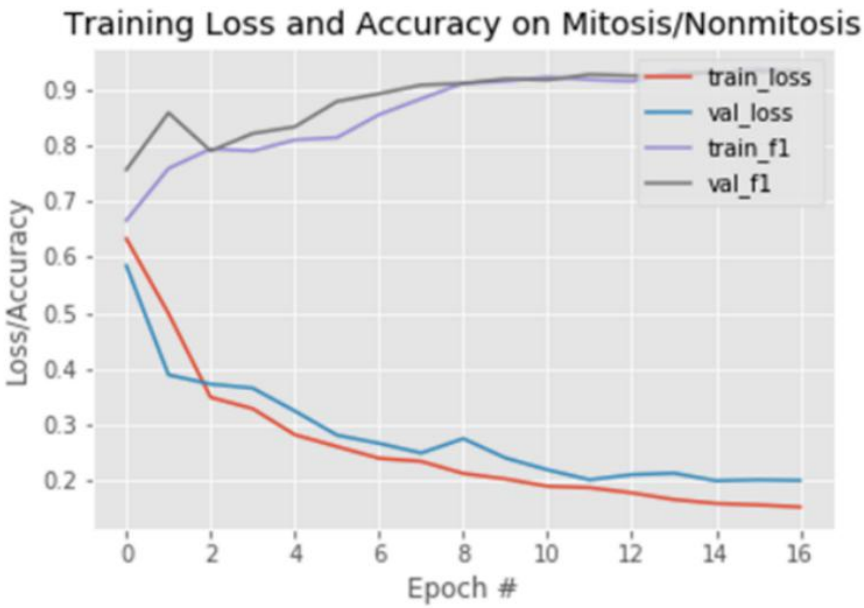

图 5.6 CB-CNN 的框架图

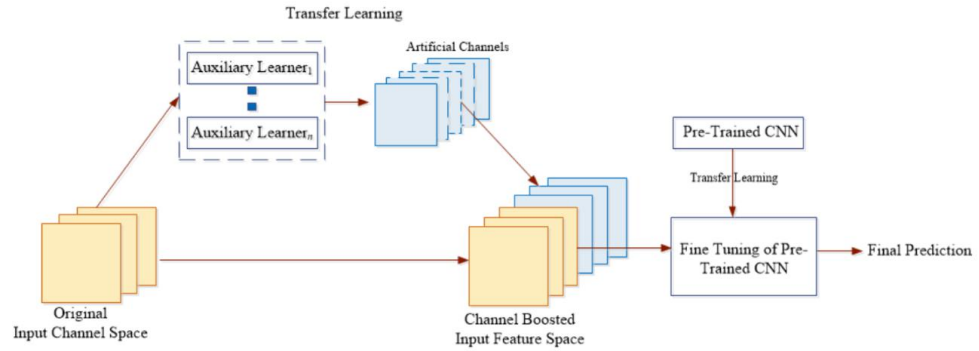

图 5.7 CB-CNN 在有丝分裂数据集上的收敛曲线。损失和精度显示在 y 轴上，而x 轴表示 Epoch。CB-CNN 的训练图表明，该模型在约 14 个 Epoch 后收敛。

六. CNN面临的挑战

深度 CNN 在图片识别和图片分类上具备了良好的性能。但是，还存在将深层 CNN 架构用于识别任务的其他挑战。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。

在另项工作中，发现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。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。

深度 CNN 模型训练期间面临的一些挑战如下：

（1）深度 CNN 通常就像一个黑匣子，研究人员没法看到中间的训练过程，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。

七.CNN 未来发展及问题

相较于传统的图像分类方法，卷积神经网络拥有特征自主提取、自主学习的能力，并通过权值共享的方式大大减少了全连接层所需神经元的数量，简化了网络结构使其所需的计算量明显下降。此外，卷积神经网络有着学习迁移的能力，经过训练的网络可以将之前所学到的特征应用于一项新的图像分类任务中，从而有效改善传统图像分类方法通用性差的问题，并且能大大提高图像分类的准确率及效率。

随着基于深度卷积神经网络在各类图像分类系统中的应用越来越广泛，识别效果越来越好，其研究工作一直深受研究者的重视。但是，仍有一些问题还没有较好的解决方案，主要表现在以下几方面：
（1）卷积神经网络的理论研究相对落后，对于图像特征提取、分类的具体机理的理解尚不透彻，导致了网络结构与网络参数的设置需要一定的经验，且随着网络层次的不断加深容易出现网络退化、过拟合等问题。
（2）对于图像分类问题来说，网络的训练需要大量的已标注的数据集来提高其泛化能力，而现有的数据集已经不能满足其发展需求。这是目前制约卷积神经网络发展推广的主要因素。
（3）卷积神经网络在图像分类领域中取得了巨大的成功，其研究仍有广阔的发展前景。目前，进一步理解其工作原理、优化网络结构、发展无监督式学习方法以及借鉴生物视觉系统的机理是其未来发展的主要方向。
（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。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。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。

八.总结

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。

参考文献

[1] O. Chapelle, “Support vector machines for image classification,” Stage deuxième année magistère dinformatique l’École Norm. Supérieur Lyon, vol. 10, no. 5, pp. 1055–1064,1998.

[2] D. G. Lowe, “Object recognition from local scale-invariant features,” Proc. Seventh IEEE Int. Conf. Comput. Vis., pp. 1150–1157 vol.2, 1999.

[3] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-Up Robust Features (SURF),” Comput. Vis. Image Underst., vol. 110, no. 3, pp. 346–359, 2008.

[4] N. Dalal and W. Triggs, “Histograms of Oriented Gradients for Human Detection,” 2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. CVPR05, vol. 1, no. 3, pp. 886–893, 2004.

[5] T. Ojala, M. Pietikäinen, and D. Harwood, “A comparative study of texture measures with classification based on feature distributions,” Pattern Recognit., vol. 29, no. 1, pp. 51–59,1996.

[6] M. Heikkilä, M. Pietikäinen, and C. Schmid, “Description of interest regions with local binary patterns,” Pattern Recognit., vol. 42, no. 3, pp. 425–436, 2009.

[7] Y. LeCun et al., “Backpropagation applied to handwritten zip code recognition,” Neural Comput., vol. 1, no. 4, pp. 541–551, 1989.

[8] X. Liu, Z. Deng, and Y. Yang, “Recent progress in semantic image segmentation,” Artif.Intell. Rev., vol. 52, no. 2, pp. 1089–1106, 2019.

[9] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, “Deep neural networks segment neuronal membranes in electron microscopy images,” in Advances in neural information processing systems, 2012, pp. 2843–2851.

[10] L. Deng, D. Yu, and B. — Delft, “Deep Learning: Methods and Applications Foundations and Trends R in Signal Processing,” Signal Processing, vol. 7, pp. 3–4, 2013.

[11] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun, “What is the best multi-stage architecture for object recognition? BT - Computer Vision, 2009 IEEE 12th International Conference on,” Comput. Vision, 2009 …, pp. 2146–2153, 2009.49

[12] Y. LeCun, K. Kavukcuoglu, and C. Farabet, “Convolutional networks and applications in vision,” in Proceedings of 2010 IEEE International Symposium on Circuits and Systems,2010, pp. 253–256.

[13] D. Scherer, A. Müller, and S. Behnke, “Evaluation of pooling operations in convolutional architectures for object recognition,” in Artificial Neural Networks--ICANN 2010,Springer, 2010, pp. 92–101.

[14] D. H. Hubel and T. N. Wiesel, “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex,” J. Physiol., vol. 160, no. 1, pp. 106–154, Jan. 1962.

[15] D. H. Hubel and T. N. Wiesel, “Receptive fields and functional architecture of monkey striate cortex,” J. Physiol., vol. 195, no. 1, pp. 215–243, Mar. 1968.

[16] Ian Goodfellow, Y. Bengio, and A. Courville, “Deep learning,” Nat. Methods, vol. 13, no.1, p. 35, 2017.
[17] Y. Bengio, “Learning Deep Architectures for AI,” Found. Trends® Mach. Learn., vol. 2,no. 1, pp. 1–127, 2009.
[18] M. N. U. Laskar, L. G. S. Giraldo, and O. Schwartz, “Correspondence of Deep Neural Networks and the Brain for Visual Textures,” pp. 1–17, 2018.
[19] K. Grill-Spector, K. S. Weiner, J. Gomez, A. Stigliani, and V. S. Natu, “The functional neuroanatomy of face perception: From brain measurements to deep neural networks,” Interface Focus, vol. 8, no. 4, p. 20180013, Aug. 2018.
[20] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E.Muharemagic, “Deep learning applications and challenges in big data analytics,” J. Big Data, vol. 2, no. 1, pp. 1–21, 2015.
[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” Adv. Neural Inf. Process. Syst., pp. 1–9, 2012.
[22] A. S. Qureshi and A. Khan, “Adaptive Transfer Learning in Deep Neural Networks: Wind Power Prediction using Knowledge Transfer from Region to Region and Between Different Task Domains,” arXiv Prepr. arXiv1810.12611, 2018.
[23] A. S. Qureshi, A. Khan, A. Zameer, and A. Usman, “Wind power prediction using deep 50 neural network based meta regression and transfer learning,” Appl. Soft Comput. J., vol.58, pp. 742–755, 2017.
[24] Qiang Yang, S. J. Pan, and Q. Yang, “A Survey on Transfer Learning,” vol. 1, no. 10, pp.1–15, 2008.
[25] Q. Abbas, M. E. A. Ibrahim, and M. A. Jaffar, “A comprehensive review of recent advances on deep vision systems,” Artif. Intell. Rev., vol. 52, no. 1, pp. 39–76, 2019.
[26] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, “Deep learning for visual understanding: A review,” Neurocomputing, vol. 187, pp. 27–48, 2016.
[27] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, “A survey of deep neural network architectures and their applications,”Neurocomputing, vol. 234, no. October 2016, pp. 11–26, 2017.
[28] J. Bouvrie, “1 Introduction Notes on Convolutional Neural Networks,” 2006.
[29] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp.436–444, 2015.
[30]C.-Y. Lee, P. W. Gallagher, and Z. Tu, “Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree,” in Artificial Intelligence and Statistics, 2016, pp.464–472.
[31]F. J. Huang, Y.-L. Boureau, Y. LeCun, and others, “Unsupervised learning of invariant feature hierarchies with applications to object recognition,” in Computer Vision and Pattern Recognition, 2007. CVPR’07. IEEE Conference on, 2007, pp. 1–8.

[32] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 9,52 pp. 1904–1916, 2015.
[33] T. Wang, D. J. D. J. Wu, A. Coates, and A. Y. Ng, “End-to-end text recognition with convolutional neural networks,” ICPR, Int. Conf. Pattern Recognit., no. May, pp. 3304–3308, 2012.
[34] Y. Boureau, “Icml2010B.Pdf,” 2009.
[35] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” 2015.
[36] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors,” pp. 1–18,2012.
[37] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfittin,” J. Mach. Learn. Res., vol. 1,no. 60, p. 11, 2014.
[38] W. Rawat and Z. Wang, “Deep Convolutional Neural Networks for Image Classification:A Comprehensive Review,” vol. 61, no. 5–6, pp. 1120–1132, 2016.
[39] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.
[40] A. Sharma and S. K. Muttoo, “Spatial Image Steganalysis Based on ResNeXt,” 2018 IEEE 18th Int. Conf. Commun. Technol., pp. 1213–1216, 2018.
[41]W. Han, R. Feng, L. Wang, and L. Gao, “Adaptive Spatial-Scale-Aware Deep Convolutional Neural Network for High-Resolution Remote Sensing Imagery Scene Classification,” in IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, 2018, pp. 4736–4739.
[42] X. Liu, Z. Deng, and Y. Yang, “Recent progress in semantic image segmentation,” Artif.Intell. Rev., vol. 52, no. 2, pp. 1089–1106, 2019.