****

**研究生（ 文献综述与选题 ）报告**

**题 目：卷积神经网络的自适应卷积结构研究**

**学 号 D201880880**

**姓 名 李超**

**专 业 计算机软件与理论**

**指 导 教 师 何琨**

**院（系、所） 计算机科学与技术**

**华中科技大学研究生院制**

**填表注意事项**

一、本表适用于攻读硕士学位研究生选题报告、学术报告，攻读专业硕士学位研究生实践环节报告，攻读博士学位研究生文献综述、选题报告、论文中期进展报告、学术报告等。

二、以上各报告内容及要求由相关院（系、所）做具体要求。

三、以上各报告均须存入研究生个人学籍档案。

四、本表填写要求文句通顺、内容明确、字迹工整。

一、 课题的来源、目的、意义、国内外研究现状

二、 预计需达到的要求、技术指标，预计的技术关键、技术方案和主要实验研究情况

三、 课题研究进展计划

**简介**

深度卷积神经网络（CNN）是一种特殊类型的神经网络，在与计算机视觉和图像处理有关的一些竞赛中表现优异。CNN广泛应用于图像分类和分割，对象检测，视频处理，自然语言处理和语音识别等领域。深度CNN的强大学习能力主要是由于使用了多个特征提取阶段，可以自动从数据中学习表示形式。大量数据的可用性以及硬件技术的改进加速了CNN的研究。研究者们已经探索了一些在CNN中取得进步的启发性想法，例如使用不同的激活和损失函数，卷积层，池化层，参数优化，正则化等创新。尤其是最近涌现出了一批自适应的方法，它们根据输入图像的特征来自动调整网络的某个模块，使得神经网络的学习更加自动化。然而，这些方法往往未能表现出我们期望的性能。我们希望根据训练和测试数据都存在局部对称性这一共性来对网络进行自适应的调整。

**一、 课题的来源、目的、意义、国内外研究现状**

**1. 课题的来源**

图像特征的提取与分类一直是计算机视觉领域的一个基础而重要的研究方向。卷积神经网络 ( Convolutional Neural Network，CNN) 提供了一种端到端的学习模型，模型中的参数可以通过传统的梯度下降方法进行训练， 经过训练的卷积神经网络能够学习到图像中的特征， 并且完成对图像特征的提取和分类。作为神经网络领域的一个重要研究分支， 卷积神经网络的特点在于其每一层的特征都由上一层的局部区域通过共享权值的卷积核激励得到。这一特点使得卷积神经网络相比于其他神经网络方法更适合应用于图像特征的学习与表达。

早期的卷积神经网络结构相对简单，如经典的 LeNet-5 模型[1]，主要应用在手写字符识别、图像分类等一些相对单一的计算机视觉应用领域中。随着研究的不断深入，卷积神经网络的结构不断优化，其应用领域也逐渐得到延伸。近年来，卷积神经网络的结构研究仍然有着很高的热度，一些具有优秀性能的网络结构被提出[2],[3],[4]。并且，随着迁移学习理论[5]在卷积神经网络上的成功应用， 卷积神经网络的应用领域得到了进一步的扩展[6],[7]。卷积神经网络在各个领域不断涌现出来的研究成果，使其成为了当前最受关注的研究热点之一。

**2.课题的目的与意义**

特征表达。图像的特征设计一直是计算机视觉领域的一个基础而重要的课题。在以往的研究中， 一些典型的人工设计特 征 被 证 明 取 得 了 良 好 的 特 征 表 达 效 果， 如 SIFT( Scale-Invariant Feature Transform )[10]、HOG ( Histogram of Oriented Gradient) [11]等。但是， 这些人工设计特征也存在缺乏良好的泛化性能问题。卷积 神 经 网 络 作 为 一 种 深 度 学习[12],[13]模型， 具有分层学习特征的能力[8]。研究[14],[15]表明，通过卷积神经网络学习得到的特征相对于人工设计特征具有更强的判别能力和泛化能力。特征表达作为计算机视觉的研究基础，如何利用卷积神经网络学习、提取、分析信息的特征表达，从而获得判别性能更强， 泛化性能更好的通用特征，将对整个计算机视觉乃至更广泛的领域产生积极的影响。

应用价值。卷积神经网络经过多年的发展， 从最初较为简单的手写字符识别[1]应用， 逐渐扩展到一些更加复杂的领域，如: 行人检测[16]、行为识别[9],[17]、人体姿势识别[18]， 等。

近期，卷积神经网络的应用进一步向更深层次的人工智能发展， 如: 自 然 语 言 处 理[19],[20]、语 音 识 别[21]， 等。最 近， 由Google 开发的人工智能围棋程序 Alphago[22]成功利用了卷积神经网络分析围棋盘面信息， 并且在挑战赛中接连战胜了围棋欧洲冠军和世界冠军， 引起了广泛的关注。从当前的研究趋势来看，卷积神经网络的应用前景充满了可能性，但同时也面临着一些研究难题， 比如: 如何改进卷积神经网络的结构，以提高网络对于特征的学习能力; 如何将卷积神经网络以合理的形式融入新的应用模型中。

**3.** **深度CNN的架构演变**

CNN被认为是受生物学启发的人工智慧（AI）技术中使用最广泛的算法。 CNN的历史始于Hubel和Wiesel[23][24]进行的神经生物学实验。 他们的工作为许多认知模型提供了平台，而CNN几乎取代了所有这些模型。 几十年来，人们为提高CNN的性能做出了各种努力。 深度CNN架构的演进历史如图3所示。CNN架构的改进可以归纳为以下讨论的五个不同时代。

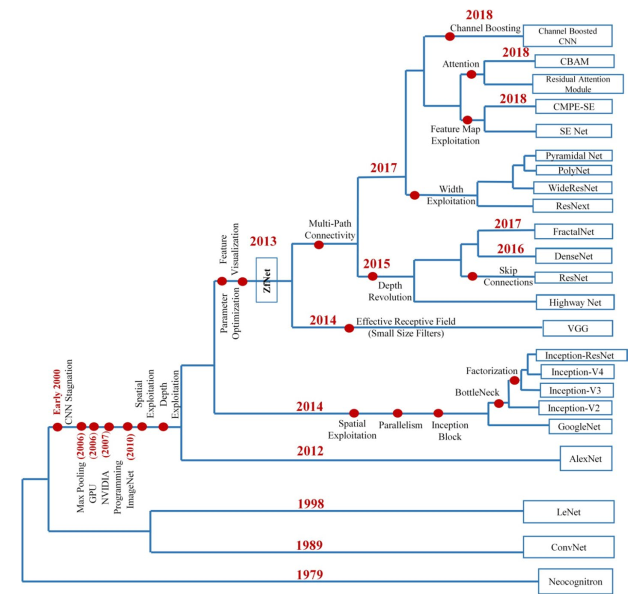


图1. 深度卷积神经网络的发展历史

**3.1 CNN的起源：1980年代末至1999年**

自1980年代后期以来，CNN已应用于视觉任务。 1989年，LeCuN等人。提出了名为ConvNet的第一个多层CNN，其起源于Fukushima的Neocognitron[25],[26]。 与前身Neocognitron所使用的无监督强化学习方案相比，LeCuN提出了使用反向传播算法进行ConvNet的有监督训练[27],[28]。因此，LeCuN的工作为现代2D CNN奠定了基础。该ConvNet展示了成功解决手写数字和邮政编码识别相关问题的结果。 1998年，LeCuN提出了ConvNet的改进版本，即著名的LeNet-5，它开始在文档识别相关应用程序中使用CNN对字符进行分类[29]。由于CNN在光学字符和指纹识别方面的良好性能，其分别于1993年和1996年开始在ATM和Banks中投入商业使用。在这个时代，LeNet-5在光学字符识别任务上取得了许多成功的里程碑，但是在其他图像识别问题上却表现不佳。

**3.2 CNN停滞：2000年初**

在1990年代末和2000年初，研究人员对CNN的内部运作了解甚少，因此被视为黑匣子。复杂的架构设计和繁重的处理使其难以训练CNN。在2000年初，人们普遍认为，用于训练CNN的反向传播算法不能有效地收敛到误差面的整体最小值。因此，与手工特征相比，CNN被认为是效果较差的特征提取器。而且，那时还没有可用的各种图像类别的综合数据集。因此，由于以高计算时间为代价的CNN性能的微不足道的改善，因此很少关注研究其在不同应用中的作用，例如物体检测，视频监视等。当时，其他统计方法尤其是SVM由于其相对较高的性能而变得比CNN更受欢迎。

同时，一些研究小组继续研究CNN，并试图优化其性能。LeCun等人在2003年改进了CNN架构并在手位基准数据集MNIST上显示出比SVM更好的结果[28]。性能的提高加速了CNN将其应用范围从光学字符识别扩展到其他脚本的字符识别，部署在图像传感器中以进行视频会议中的面部检测以及对街头犯罪的监管等。同样，基于CNN的系统已在市场上进行了工业化以进行客户跟踪。此外，还有研究者探索了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[29]。

**3.3 CNN的复兴：2006-2011年**

深度CNN通常具有复杂的体系结构和耗时的培训阶段，有时可能长达数周。在2000年初，用于深度网络培训的并行处理技术和硬件资源有限。具有典型激活函数（例如Sigmoid）的深层CNN的训练可能会受到梯度的指数衰减和爆炸的影响。自2006年以来，研究者们为解决CNN优化问题做出了巨大努力。在这方面，一些文章报告了一些有趣的初始化和训练策略，以克服在深层CNN训练和不变特征学习中遇到的困难。Hinton在2006年报告了贪婪分层预训练的概念，这使深度学习的研究得以复兴[32]。实验研究表明，有监督和无监督的预训练都可以以比随机初始化更好的方式初始化网络。 Bengio和其他研究人员提出，sigmoid激活函数不适用于具有随机初始化权重的深度结构的训练。该观察结果开始使用Sigmoid以外的激活函数，例如ReLU，tanh等[33]。深度学习的复兴是使深度CNN成为众人瞩目的因素之一。

Ranzato等人使用最大池化代替二次采样，通过学习不变特征显示了很好的结果[34]。在2006年末，研究人员开始使用图形处理单元（GPU）来加速深度NN和CNN体系结构的训练[35]。 NVIDIA在2007年推出了CUDA编程平台，该平台可以更大程度地利用GPU的并行处理功能。从本质上讲，使用GPU进行NN和CNN训练以及其他硬件改进是促使CNN复苏研究的主要因素[36]。 2010年，斯坦福大学的Fei-Fei Li小组建立了一个名为ImageNet的大型图像数据库，其中包含数百万个包含大量类别的带注释的图像[37]。该数据库与年度ImageNet大规模视觉识别挑战赛（ILSVRC）结合在一起，在其中对各种模型的性能进行了评估和评分。同样，同年，斯坦福大学发布了用于对象检测的PASCAL 2010 VOC数据集。 ILSVRC和神经信息处理系统会议（NIPS）在加强研究和增加CNN的使用并因此使其普及方面起到了主导作用。

**3.4 CNN的兴起：2012-2014年**

广泛的训练数据和硬件的进步是推动CNN研究发展的因素。但是推动研究加速并在图像分类和识别任务中使用CNN的主要驱动力是参数优化策略和新的架构思想[38]。 CNN性能的主要突破是AlexNet带来的，与传统的CV技术相比，该技术在2012-ILSVRC中表现出色（错误率从25.8降低到16.4[39]。在这个时代，人们为提高CNN的性能进行了一些尝试。探索了深度和参数优化策略，从而显着降低了计算成本。类似地，提出了不同的框架设计，从而每种新框架都试图克服先前提出的框架与新结构重新组合的缺点。随着设计深度更深的CNN的趋势，通常很难为每层独立地确定filter dimensions, stride, padding和其他超参数。通过设计具有可重复多次的固定拓扑的卷积层，可以解决此问题。这将趋势从定制层设计转变为模块化和统一层设计。 CNN中的模块化概念使针对不同任务定制网络变得轻松[40]。应当指出的是，在这个时代，研究者使用了两种不同类型的体系结构：深层和窄层以及深层和宽层。

**3.5 CNN的结构创新和应用的快速增长：2015年至今**

CNN的研究仍在继续，并且有很大的改进潜力。CNN性能从2015年到2019年发生了显着改善。CNN的表示能力通常取决于其深度，从某种意义上说，从简单到复杂抽象的丰富功能集可以帮助学习复杂问题。 然而，深层架构面临的主要挑战是梯度的减小。最初，研究人员试图通过将中间层与辅助学习模型连接来缓解这一问题[41]。 2015年，新兴的研究领域主要是新连接的开发，以提高深层CNN架构的融合率。在这方面，引入了不同的想法，例如跨多个层的信息门控机制，跳过连接和跨层通道的连通性[42]，[43]。不同的实验研究表明，最先进的深度架构（例如VGG，ResNet，ResNext等）对于挑战识别和定位问题如语义和基于实例的对象分割，场景解析，场景位置等有良好的性能。大多数著名的对象检测和分割架构，例如单镜头多框检测器（SSD），基于区域的CNN（R-CNN），Faster R-CNN，Mask R-CNN和全卷积神经网络（FCN）是基于ResNet，VGG，Inception等构建的。类似地，许多有趣的检测算法（例如功能金字塔网络，Cascade R-CNN，Libra R-CNN等）对前面提到的体系结构进行了改进以提升性能[44]。通过将这些网络与递归神经网络（RNN）结合使用，深层CNN的应用也扩展到了图像字幕。

同样，在2016年，研究者观察到不仅在深度方向而且以并行方式进行的多个变换的堆叠都显示了对复杂问题的良好学习能力[45]。 不同的研究人员使用已经提出的架构的混合体来改善深层CNN性能[43]。 2017年，研究人员的重点主要是设计通用块，可以在CNN架构的任何学习阶段插入通用块，以改善网络表示[46]。新块的设计是CNN研究的增长领域之一，其中通用块用于分配对空间和特征图（通道）信息的关注。 Khan等人在2018年提出了一种新的渠道提升思路，通过学习独特的特征以及通过TL概念利用已经学习的特征来提高CNN的性能。

但是，深入和广泛的体系结构的两个主要问题是较高的计算成本和内存需求。因此，在资源受限的环境中部署最先进的广泛而深入的CNN模型是非常具有挑战性的。传统的卷积运算需要大量的乘法运算，这增加了推理时间，并限制了CNN在低内存和时间约束应用程序中的适用性[47]。许多现实世界的应用程序，例如自动驾驶汽车，机器人技术，医疗保健和移动应用程序，都需要在计算受限的平台上以有限的时间执行。因此，需要修改CNN以使其适合资源受限的环境。著名的方法有知识蒸馏，小型网络训练或压缩预训练网络（例如修剪，量化，哈希，哈夫曼编码等[48]。 GoogleNet利用了小型网络的思想，该思想将传统的卷积替换为逐点分组卷积运算，以提高计算效率。类似地，ShufeNet使用了逐点分组卷积，但是有了信道清洗的新思想，该思想显着减少了操作次数，却不影响准确性。同样，ANTNet提出了一种称为ANT-Block的新型架构模块，该模块以较低的计算成本在基准数据集上取得了良好的性能[49]。

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功能图来增强网络表示，或者通过添加人工通道来操纵输入表示。 此外，与此同时，趋势是朝着轻量级体系结构的设计方向发展，而又不牺牲性能以使CNN适用于资源受限的硬件。

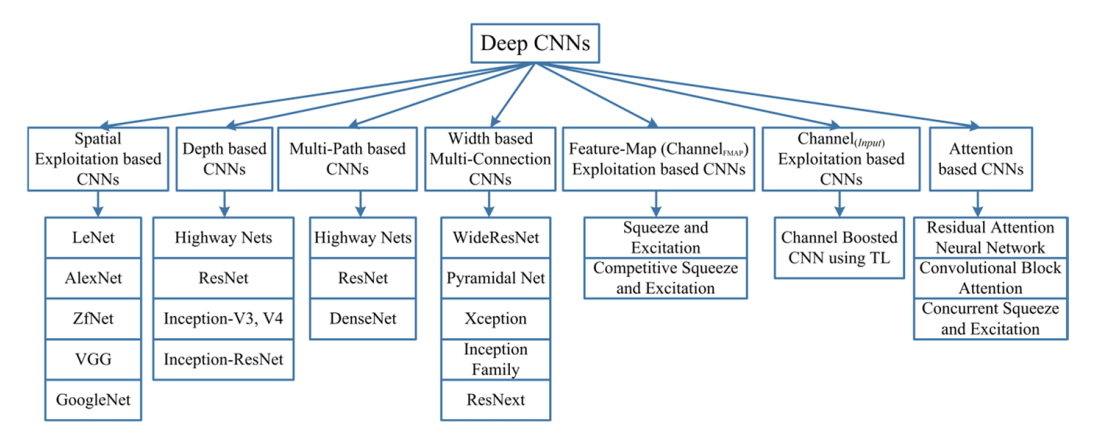


图2. 深度卷积神经网络的七个不同结构类型

**二、预计需达到的要求、技术指标，预计的技术关键、技术方案和主要实验研究情况**

**1. 预计需达到的要求和技术指标**

在本课题中，希望能在充分理解和分析国内外已知算法的基础上，创新性地提出新的卷积神经网络结构自适应研究模型，使其性能可以达到同类算法同等或者更高水平。衡量性能的指标主要包括算法的泛化能力、采样效率、收敛速率和计算代价等。

**2. 预计的技术关键、技术方案**

在本课题的研究中，除了应用国内外已有的相关研究的方法之外，还将通过对可变形卷积网络（Deformable Convolutional Networks[50]）展开深入研究，进行基于原有自适应卷积的优化工作。

传统的CNN只能靠一些简单的方法(比如max pooling)来适应物体的形变，如果形变的太厉害就无能为力了。为了改变这种情况专家们想了很多方法，最常见的有两种：

（1）、使用大量的数据进行训练。比如用ImageNet数据集，再在其基础上做翻转等变化来扩展数据集，通俗地说就是通过穷举的方法使模型能够适应各种形状的物体，这种方法收敛较慢而且要设计复杂的网络结构才能达到理想的结果；

（2）、设计一些特殊的算法来适应形变。比如SIFT，目标检测时用滑动窗口法来适应目标在不同位置上的分类也属于这类。

对第一种方法，如果用训练中没有遇到过的新形状物体(但同属于一类)来做测试，由于新形状没有训练过，会造成测试不准确，而且靠数据集来适应形变的训练过程太耗时，网络结构也必须设计的很复杂；对于第二种方法，如果物体的形状极其复杂，要设计出能适应这种复杂结构的算法就更困难了。

为了解决这个问题，可形变的卷积结构对感受野上的每一个点加一个偏移量，偏移的大小是通过学习得来的，偏移后感受野不再是个正方形，而是和物体的实际形状相匹配。这么做的好处就是无论物体怎么形变，卷积的区域始终覆盖在物体形状的周围。

图3是个示意图，（a）中绿色的点是原始的感受野范围，（b）~（d）中的蓝色点是对感受野加上偏移量后的新的感受野位置，可以看到叠加偏移量的过程可以模拟出目标移动、尺寸缩放、旋转等各种形变。

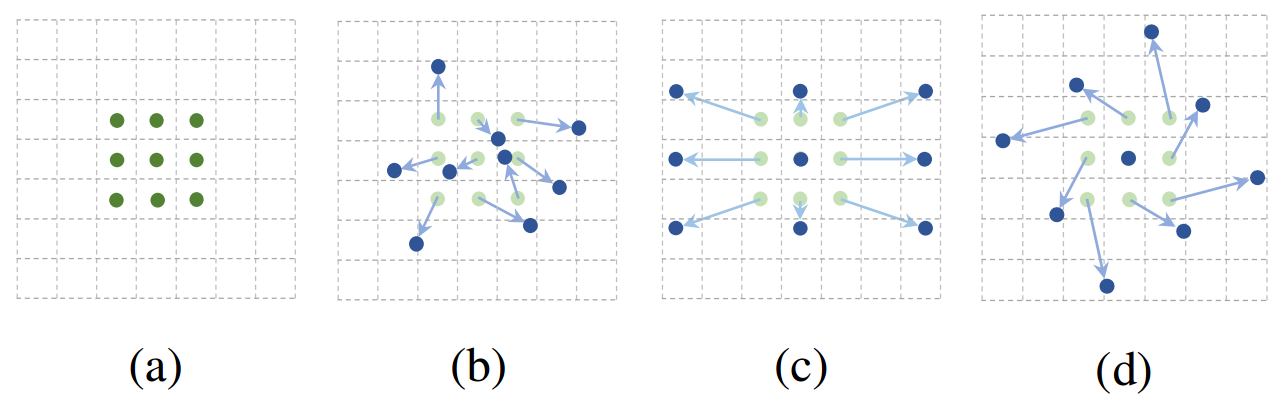
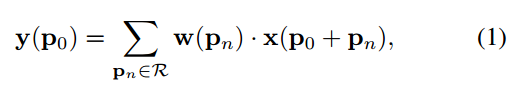
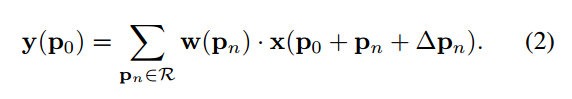


图3. 可变形卷积神经网络的可学习到丰富的采样点

传统的卷积结构可以定义成公式（1），其中pn是卷积输出每一个点相对感受野上的每一个点的偏移量，取整数。



采用全新的可变形卷积后要在公式（1）基础上给每个点再增加一个偏移量Δpn，这个新的偏移量是由另一个卷积得出，所以一般是小数，见公式（2）。



可变形卷积是一种新的自适应学习模型， 其并不采用固定的感受野，而是通过训练数据学习感受野，有较大的研究空间，我们希望在它的基础上根据训练和测试数据都存在局部对称性这一共性来对网络进行自适应的调整。

**3.主要实验研究情况**

目前已对国内外相关研究现状进行了较为充分的调研，同时对关键技术有较为深入的了解，正在对相关技术方案进行设计与改进。

截止到目前，已以第一作者发表计算机视觉顶会论文一篇：

[1]. **Li, Chao**, Yixiao Yang, Kun He, Stephen Lin, and John E. Hopcroft. "Single Image Reflection Removal through Cascaded Refinement." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3565-3574. 2020.

**三、 课题研究进展计划**

（1） 2020.10.16 - 2020.10.31 对国内外研究现状中的算法进一步深入研究，复现其中的关键细节；

（2） 2020.11.1 - 2020.11.30 设计和完善技术方案，并论证其可行性，筛选出可实施的有效技术方案；

（3） 2020.12.1 - 2020.12.31 针对有效技术方案，设计合理的模型并实现，同时进行模型调优，结构优化；

（4） 2021.1.1 - 2021.1.31 与国内外最新研究的算法进行实验对比，分析实验结果，进一步改进设计的模型与机构；

（5） 2021.2.1 – 2021.3.31 撰写并发表相关论文；

（6） 2021.4.1– 2023.1.31 以此论文为基础发表一系列相关论文；

（7） 2023.2.1 – 2023.6.30撰写硕士学位论文， 完成硕士学位论文的后续相关工作。

**参考文献**

1. LECUN Y， BOTTOU L， BENGIO Y， et al． Gradient-based learning applied to document recognition ［J］． Proceedings of the IEEE， 1998， 86( 11) : 2278 － 2324．
2. SIMONYAN K， ZISSERMAN A． Very deep convolutional networks for large-scale image recognition ［EB /OL］． ［2015-11-04 ］． http: / /www． robots． ox． ac． uk: 5000 / ～ vgg /publications /2015 /Simonyan15 /simonyan15． pdf．
3. SZEGEDY C， LIU W， JIA Y， et al． Going deeper with convolutions. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition． Washington， DC: IEEE Computer Society， 2015: 1 － 8．
4. HE K， ZHANG X， REN S， et al． Deep residual learning for image recognition ［EB /OL］． ［2016-01-04 ］． https: / /www． researchgate． net /publication /286512696 \_ Deep \_ Residual \_ Learning \_ for \_Image\_Recognition．
5. PAN S J， YANG Q． A survey on transfer learning ［J］． IEEE Transactions on Knowledge and Data Engineering， 2010， 22( 10 ) : 1345 － 1359．
6. COLLOBERT R， WESTON J， BOTTOU L， et al． Natural language processing ( almost ) from scratch ［J］． Journal of Machine Learning Research， 2011， 12 (1) : 2493 － 2537．
7. OQUAB M， BOTTOU L， LAPTEV I， et al． Learning and transferring mid-level image representations using convolutional neural networks ［C］ / / Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition． Washington， DC: IEEE Computer Society， 2014: 1717 － 1724．
8. ZEILER M D， FERGUS R． Visualizing and understanding convolutional networks ［C］/ / Proceedings of European Conference on Computer Vision， LNCS 8689． Berlin: Springer， 2014: 818 －833．
9. JI S， XU W， YANG M， et al． 3D convolutional neural networks for human action recognition ［J］． IEEE Transactions on Pattern Analysis and Machine Intelligence， 2013， 35( 1) : 221 － 231．
10. LOWE D G． Distinctive image features from scale-invariant keypoints ［J］． International Journal of Computer Vision， 2004， 60 (2) : 91 －110．
11. DALAL N， TRIGGS B． Histograms of oriented gradients for human detection ［ C ］ / / Proceedings of the 2005 IEEE Conference on Computer Vision and Pattern Recognition． Washington， DC: IEEE Computer Society， 2005: 886 － 893．
12. LECUN Y， BENGIO Y， HINTON G E． Deep learning ［J］． Nature， 2015， 521( 7553) : 436 －444．
13. 孙志军，薛磊，许阳明，等． 深度学习研究综述［J］． 计算机应用研究，2012，29( 8) : 2806 － 2810． ( SUN Z J， XUE L， XU Y M，et al． Overview of deep learning ［J］． Application Research of Computers， 2012， 29 ( 8) : 2806 － 2810)
14. DONAHUE J， JIA Y， VINYALS O， et al． DeCAF: a deep convolutional activation feature for generic visual recognition ［J］． Computer Science， 2013， 50( 1) : 815 － 830．
15. RAZAVIAN A S， AZIZPOUR H， SULLIVAN J， et al． CNN features off-the-shelf: an astounding baseline for recognition ［ EB /OL］． ［2015-11-22 ］． http: / /www． csc． kth． se / ～ azizpour /papers /ha\_cvpr14w． pdf．
16. SERMANET P， KAVUKCUOGLU K， CHINTALA S， et al． Pedestrian detection with unsupervised multi-stage feature learning［C］ / / CVPR '13: Proceedings of the 2013 IEEE Conference on

Computer Vision and Pattern Recognition． Washington， DC: IEEE Computer Society， 2013: 3626 － 3633．

1. KARPATHY A， TODERICI G， SHETTY S， et al． Large-scale video classification with convolutional neural networks ［ C ］ / / CVPR '14: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition． Washington， DC: IEEE Computer Society， 2014: 1725 － 1732．
2. TOSHEV A， SZEGEDY C． DeepPose: human pose estimation via deep neural networks ［C］ / / Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition． Washington，DC: IEEE Computer Society， 2014: 1653 － 1660．
3. KALCHBRENNER N， GREFENSTETTE E， BLUNSOM P． A convolutional neural network for modelling sentences ［EB/OL］． ［2016-01-07］． http: / /anthology． aclweb． org/P/P14 /P14-1062． pdf．
4. KIM Y． Convolutional neural networks for sentence classification［EB /OL ］． ［2016-01-07 ］． http: / /anthology． aclweb． org /D /D14 /D14-1181． pdf．
5. ABDEL-HAMID O， MOHAMMED A， JIANG H， et al． Convolutional neural networks for speech recognition ［ J ］． IEEE /ACM Transactions on Audio， Speech and Language Processing， 2014，

22( 10) : 1533 － 1545．

1. SILVER D， HUANG A， MADDISON C J， et al． Mastering the game of Go with deep neural networks and tree search ［J］． Nature， 2016， 529( 7587) : 484 － 489．
2. Hubel DH, Wiesel TN (1959) Receptive fields of single neurons in the cat’s striate cortex. J Physiol. https://doi.org/10.1113/jphysiol.1959.sp006308
3. Hubel DH, Wiesel TN (1962) Receptive felds, binocular interaction and functional architecture in the cat’s visual cortex. J Physiol 160:106–154. https://doi.org/10.1113/jphysiol.1962.sp006837
4. Fukushima K, Miyake S (1982) Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition. In: Competition and cooperation in neural nets. Springer, pp 267–285
5. Fukushima K (1988) Neocognitron: a hierarchical neural network capable of visual pattern recognition. Neural Netw 1:119–130
6. Linnainmaa S (1970) The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors. Master’s Thesis (in Finnish), Univ Helsinki 6–7
7. LeCun Y, Boser B, Denker JS et al (1989) Backpropagation applied to handwritten zip code recognition. Neural Comput 1:541–551
8. Bay H, Ess A, Tuytelaars T, Van Gool L (2008) Speeded-up robust features (SURF). Comput Vis Image Underst 110:346–359. <https://doi.org/10.1016/j.cviu.2007.09.014>
9. LeCun Y, Bottou L, Bengio Y, Hafner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86:2278–2324
10. Chen Y-N, Han C-C, Wang C-T et al (2006) The application of a convolution neural network on face and license plate detection. In: 18th international conference on pattern recognition, 2006. ICPR 2006, pp 552–555
11. Hinton GE, Osindero S, Teh Y-W (2006) A fast learning algorithm for deep belief nets. Neural Comput 18:1527–1554
12. Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics, pp 249–256
13. Ranzato M, Huang FJ, Boureau YL, LeCun Y (2007) Unsupervised learning of invariant feature hierarchies with applications to object recognition. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition. IEEE, pp 1–8
14. Strigl D, Kofer K, Podlipnig S (2010) Performance and scalability of GPU-based convolutional neural networks. In: 2010 18th Euromicro international conference on parallel, distributed and network-based processing (PDP), pp 317–324
15. Cireşan DC, Cireşan DC, Meier U, Schmidhuber J (2018) Multi-column deep neural networks for image

classifcation. In: IEEE conference on computer vision and pattern recognition

1. Russakovsky O, Deng J, Su H et al (2015) imagenet large scale visual recognition challenge. Int J Comput Vis. https://doi.org/10.1007/s11263-015-0816-y
2. Sinha T, Verma B, Haidar A (2018) Optimization of convolutional neural network parameters for image classification. In: 2017 IEEE symposium series on computational intelligence SSCI 2017, pp 1–7. https://doi.org/10.1109/ssci.2017.8285338
3. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. https://doi.org/10.1061/(ASCE)GT.1943-5606.0001284
4. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. ICLR 75:398–406. https://doi.org/10.2146/ajhp170251
5. Szegedy C, Liu W, Jia Y et al (2015) Going deeper with convolutions. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 1–9
6. He K, Zhang X, Ren S, Sun J (2015a) Deep residual learning for image recognition. Multimed Tools Appl 77:10437–10453. https://doi.org/10.1007/s11042-017-4440-4
7. Huang G, Sun Y, Liu Z et al (2016a) Deep networks with stochastic depth. In: European conference on computer vision. Springer, pp 646–661
8. Cai Z, Vasconcelos N (2019) Cascade R-CNN: high quality object detection and instance segmentation. IEEE Trans Pattern Anal Mach Intell. https://doi.org/10.1109/tpami.2019.2956516
9. Zagoruyko S, Komodakis N (2016) Wide residual networks. Proc Br Mach Vis Conf 87(1-87):12. https://doi.org/10.5244/C.30.87
10. Hu J, Shen L, Sun G (2018a) Squeeze-and-excitation networks. In: 2018 IEEE/CVF conference on computer vision and pattern recognition. IEEE, pp 7132–7141
11. Shakeel MF, Bajwa NA, Anwaar AM et al (2019) Detecting driver drowsiness in real time through deep learning based object detection. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)
12. Frosst N, Hinton G (2018) Distilling a neural network into a soft decision tree. In: CEUR workshop proceedings
13. Xiong Y, Kim HJ, Hedau V (2019) ANTNets: mobile convolutional neural networks for resource efficient image classification. arXiv:190403775
14. Dai, Jifeng, et al. "Deformable convolutional networks." Proceedings of the IEEE international conference on computer vision. 2017.

**华中科技大学研究生院**

|  |  |  |
| --- | --- | --- |
| **研究生签字**  **指导教师签字**  **院(系、所)领导签字** |  |  |

**年 月 日**